

特集号
投稿論文

国立病院機構における診療情報分析システムについて

—構築と運用に関する現状と課題—

川島 直美^{†1} 堀口 裕正^{†2} 伏見 清秀^{†3}

^{†1}(独) 国立病院機構 ^{†2} 東京大学 ^{†3} 東京医科歯科大学

国立病院機構本部診療情報分析部は2010年度、診療情報データバンクを整備し、機構144の病院から診療データを収集分析を行っている。ここでは収集データの特性と問題点、分析のためのデータセットの構築に至るまでの各種問題の解決、およびそれらを元に行っている分析を紹介する。

1. はじめに

国立病院機構本部総合研究センター診療情報分析部は2010年4月に発足し、機構144病院から診療データを収集し、分析を行っている。部長(医師)、研究員(いずれも医療職)3名、SE4名(いずれも非医療職)、事務員1名の9名で活動しており、医療を可視化し、比較するためのデータベースの構築と、それによる病院の客観的な評価が部のミッションである。初年度はひとまず41病院個別のレポートの作成を試みた。また、医療の質を定量的に評価するための臨床評価指標事業を機構本部より引き継ぎ、発展させることとした。

医療情報は主に2通りの用途に活用されている。1つはある特定の個人の病気を治癒させるためのもので、カルテやレントゲン等の検査記録がこれに当たる。もう1つは不特定多数の診療記録を収集し、それらを分析することで臨床的・経営的知見を得るものである。本論文では、後者について、収集データの特性と問題点、分析のためのデータセットの構築に至るまでの各種問題の解決、それらを元に行っている分析について述べる。

2. MIAの概要

診療情報分析部では立ち上げにあたり、診療情報分析システム(名称はMedical Information Analysis databank: 以下、MIAとする)の整備を行った。MIAは、テキストファイルをオンラインでアップロードさせる形で収集・蓄積し、リレーショナルデータベース(以下、RDB)を構築するシンプルなシステムである。定形分析画面や帳

票機能等は設けず、診療情報分析部のSEがRDBに対し直接SQLコマンドを発行し集計することで、自由分析を行う。

以下MIAのシステム構成(図1)と処理フロー(図2)を示す。

処理フロー

- 1) 各病院はHospnet(組織内WAN)を利用してデータアップロードサイトからテキストデータを提出する。
- 2) データ処理サーバはテキストデータをRDB化する。DPCデータはそのままテーブル化し、レセプトデータは各レコードにキーとなる情報を追加するなどRDBに取り込める形に変換し、テーブル化している。
- 3) 構築されたDBは適宜、分析用のサーバにコピー

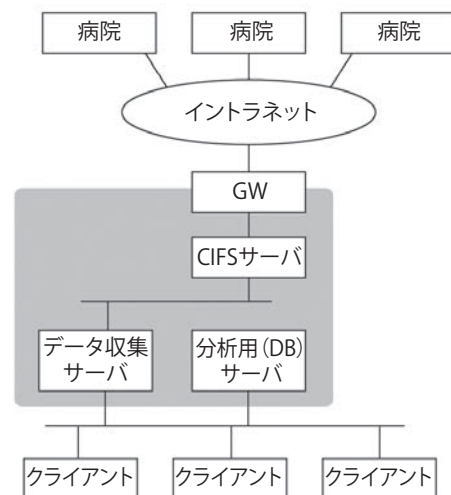


図1 MIA システム構成図

する。

- 4) 研究員は分析したい内容をSEに伝え、SEはクライアントPCからSQL文を発行し分析する。

以上、アップロードサイトおよびRDB化システムを構築し、定型の帳票出力機能や分析画面を設けないことでスタートアップの費用を安く抑え、その分SEを雇用している。新たな分析を始めるたびに、システム改修を行う必要がないため、多様な解析にローコストで柔軟に対応している。

3. 収集データの概要と問題点

機構病院全体では年間約57万人が新たに入院し、延べ約990万人が外来を受診している。それらの患者について横並びに比較、集計するためには正規化されたデータベースの構築が必要不可欠である。個々の病院に個々

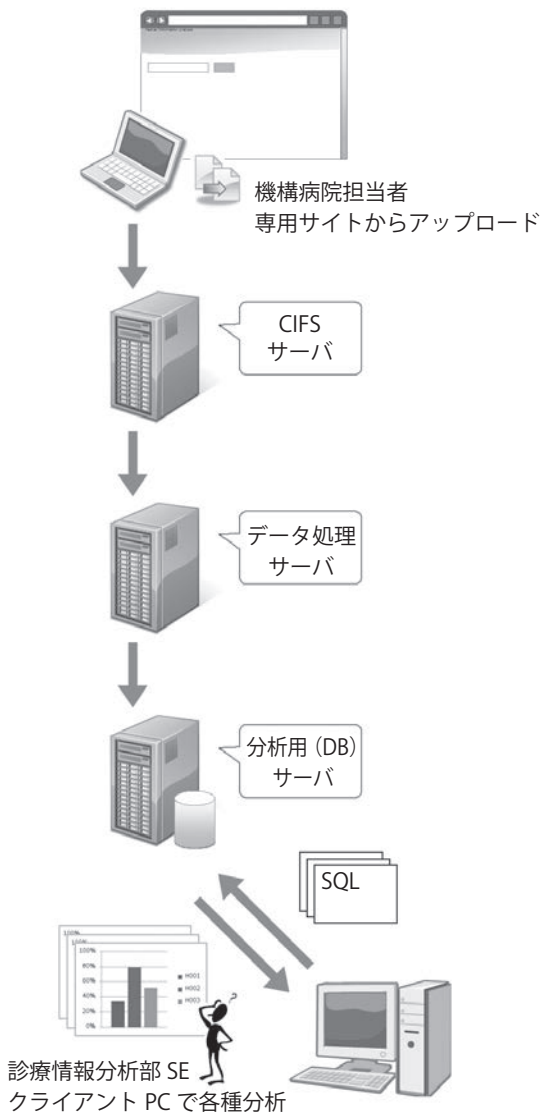


図2 MIA 処理フロー

の患者についての診療情報を定期的に定型フォーマットで入力させることは困難であるため、病院が業務上作成しているデータを2次利用することで追加負担を最低限に抑えることとした。病院は個々に整備したシステムにより独自の形式でさまざまなデータを作成しているが、初年度はまず、全国统一仕様で作成されているDPC導入の影響評価に関する調査データ（以下DPCデータという）、およびレセプトデータの収集とデータベース化を試みた。

3.1 DPC データ

DPC (Diagnosis Procedure Combination: 診断群分類) データ(図3)は医療費の定額支払い制度(DPC/PDPS: Diagnosis Procedure Combination/Per-Diem Payment System)に参加する病院が厚生労働省に対し提出を義務付けられているデータである。DPC/PDPSは患者を入院中に最も医療資源を投入した病名、重症度、手術や処置の種類から2,658種(2010年度改定時点)の診断群分類に分け、その分類ごとに入院1日あたりの費用が決まる制度である。DPCデータは様式1、様式3、様式4、Dファイル、Eファイル、Fファイル、外来Eファイル、外来Fファイルからなり、EXCELファイルである様式3を除けばすべてタブ区切りの正規化されたテキストデータであるため、そのままテーブル化することが可能である。厚生労働省において収集された後、1入院単位で分析を行うことを前提としたデータ形式であるため(厚生労働省は収集したデータを元に、分類や報酬を調整し、次回の改定に活用している)、リレーショナルデータベースとの親和性が高い。

診療情報分析部では日々の医療行為が記録されるEファイル・Fファイルと、退院時に入院期間のサマリが記

様式1ファイル (抜粋)		Eファイル (抜粋)		Fファイル (抜粋)	
病院属性	施設コード	施設コード	施設コード	施設コード	施設コード
患者属性	患者ID	患者ID	患者ID	患者ID	患者ID
	性別	入院年月日	入院年月日	入院年月日	入院年月日
	生年月日	退院年月日	退院年月日	退院年月日	退院年月日
	郵便番号	診療月	診療月	診療月	診療月
入退院情報	入院年月日	データ区分番号	データ区分番号	データ区分番号	データ区分番号
	退院年月日	順序番号	順序番号	順序番号	順序番号
	入院経路	実施年月日	行為明細番号	行為明細番号	行為明細番号
	退院先	診療報酬額	診療行為	診療行為	診療行為
診断情報	医療資源を最も投入した傷病名				
	主病名				
	副傷病				
手術情報	行われた手術				
診療情報	入退院時の意識レベル				
	がんの病期				

図3 DPC データ概要

録される様式1を主に利用している。様式1には体系化された病名コードとテキスト病名のほか、入退院時のJCS（意識障害の深刻度を示す数値）や入退院時のADL（日常生活動作レベルを示す数値）、Burn Index（火傷の程度を示す数値）といった重症度、がんや心筋梗塞など病気の進行度をコードで入力することが義務付けられている。体系化された病名コードは疾病および関連保健問題の国際統計分類（ICD）と呼ばれるもので、約11,000種のコードから医師が選び、記載している。また、診断群分類は010060 x 097 x 01 x のような、14桁のコードで表される。先頭の6桁は疾患コードと呼ばれ、入院中の医療資源が最も投入された病気のICDをグループ化したものである。さらにそのうち先頭の2桁はMDC(Major Diagnostic Category: 主要診断群)と呼ばれ、眼科系疾患や呼吸器疾患といった大まかな分類がなされている。非常に情報量が豊富でカテゴリズしやすいデータであるが、定額支払い制度に参加しているのは比較的大規模の総合病院が中心であるため、収集は機構144病院中53病院にとどまっている。

3.2 レセプトデータ（診療報酬明細書）

3.2.1 データの概要

レセプトデータは病院が健康保険組合等に診療報酬を請求するためのデータで、機構144のすべての病院から収集している。レセプトデータは医科レセプト（社会保険）・医科レセプト（国民健康保険）・DPCレセプト（社会保険）・DPCレセプト（国民健康保険）からなる（DPCレセプトは定額支払い制度に参加し、実際に定額支払いを行っている病院のみの収集）。

現在の医療制度では、患者が病院を受診する際、保険証の提示により医療費は一部負担となり、残額は保険者（健康保険組合などのことをいう）が病院に支払う。病院は月に1度、患者の病名や患者に行った医療行為、およびその金額を記載したもの（これをレセプトという）を審査支払機関に提出する。審査支払機関は医療行為と金額に誤りのないこと、病名と医療行為の間に矛盾のないことなどを審査する。審査に通ったレセプトは患者が加入する保険ごとに仕分けられ、保険者に送られたのち医療費の支払いがなされる（図4）。

これらの業務はすべて紙で運用されていたが、近年電子化が進み2011年度よりレセプトはCSVファイルでの提出が原則義務となった。しかし、患者ごとに支払額とその理由が記載され、審査支払機関において保険者ごとにまとめられるという運用は変わらず、たとえばDPC

データのようにテーブル形式にするなどして、のちのち2次利用して分析を行うことを念頭においたデータ形式とはならなかった。これに対し、基本方針として、レセプトデータをDPCデータの形式を合わせてテーブル化し、様式1に準じた各種状態のラベリングを行い、診断群分類コードに準じたコードを付与することで、分析しやすいデータセットの生成を目指した。しかし、この基本方針で解決しきれない問題点も複数あり、独自の解決方針をとっている。次にこの個々の問題点と解決方針について述べる。

3.2.2 問題点1 レコードごとに列数が異なるCSVデータの成形

レセプトデータは、レコードの先頭にRE（患者情報）、SY（病名情報）等のヘッダを持ち、その種別ごとに各桁の意味が異なる（図5）。まずはすべてのレコードに患者IDと診療月を持たせ、情報種別ごとにテーブル化する。

3.2.3 問題点2 月単位のデータから1入院単位へのデータの統合

未処理のレセプトデータから分かることは、ある月にある患者が患っていた病気、受けた治療とその回数、金額、入院外来の別である。治療は数日の入院で終わるものから、月をまたいで何カ月もかかるものまでさまざまである。医療者は1カ月、1年という単位ではなく、発症から治癒までの一連のエピソード、1つの症例として患者を治療している。抗がん剤治療のように計画的に何

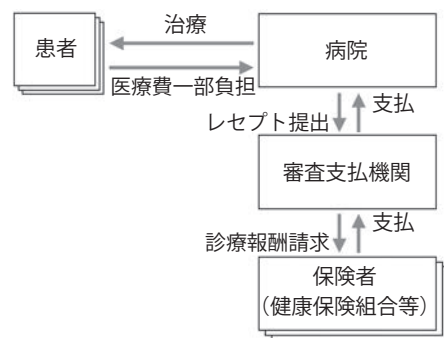


図4 診療報酬の請求と支払いの流れ

```

IR,1,01,1,123456789,,〇〇病院,42504,00,012-345-6789
RE,1,1115,42204, 国病太郎 ,1,3580311,,4220401,18,,00123,,,,,01,,,,,,,,,,,,,
HO,123456,123456,123456,31,96113,,93,66588,,44400,,
SY,3669004,4220401,1,,,01,
SY,0000999,4220401,1,, 糖尿病の疑い ,,
SI,13,1,123456789,,200,1,,,,,1,,,,,,,,,,,,,
IY,21,1,123456789,2,1,,,,,,,,,,,,,
    
```

図5 レセプトデータの例

度か入退院を繰り返して行う治療もあるが、多くの治療は1入院で完結するため、便宜上1入院を1症例とし、DPCデータの形式のように、1入院ごとのデータセットを作成する。

3.2.4 問題点3 病名におけるコードの省略とテキストの表記揺れへの対処

レセプトデータの病名情報レコードには、標準病名コードと呼ばれるコードとテキスト病名を記載する列がそれぞれあるが、標準病名コードは未コード化傷病名と呼ばれる分類外コードの入力が許されている。病名を入力すると関連する標準病名が列挙され、選択すると病名コードと標準病名が入力されるシステムが多いようであるが、テキスト病名のみの記載も散見される。また、テキスト病名は自由入力が入力されているため、漢字、平仮名の違いや、初期、末期等の状態を示す語句等により、入力パターンは無限に存在し得る。このため、たとえばがん患者をカウントしようとするとき、「がん」「ガン」「癌」「悪性腫瘍」「悪性新生物」など、思いつく限りのキーワードで検索する必要があるが、それでも漏れなく把握することは難しい。これらの問題を解決するため、標準病名コードが入力されていない病名については、DPCデータにならない、テキスト病名に対応するICDコードを機械的に付与する。

3.2.5 問題点4 多様な病名修飾語コードへの対処

レセプトデータの病名情報レコードには修飾語と呼ばれる接頭語コード、接尾語コードを記載する列が用意されており、「突発性」や「重症」といった接頭語や、「の疑い」や「の再発」といった接尾語を最大4つまで病名本体に付加できる仕様になっている。「胃」や「大腿骨」のように、付加することによって病名に大きく影響するような修飾語も用意されているが、病名本体との組合せに制限がない。たとえば「B型慢性肝炎」は0703003というコードが用意されているが、接頭語を「慢性」、病名本体を「肝炎」、接尾語を「B型」とすることで、「慢性肝炎B型」という病名が作成できてしまう。この場合病名本体の「肝炎」に付与されるコードは5733005となり、コード化されているからといって0703003で抽出すると漏れてしまう。この事態を避けるため、前述の病名のICDコード化の際、修飾語コードが使用されている病名については標準病名コードは信用せず、テキスト病名をICDコード化する。

3.2.6 問題点5 過剰な病名入力への対処

レセプトデータには請求制度上、実際に治療に至った病名だけでなく罹患が疑われた病名も含めて網羅的に記

載されるため、主に治療した病気は何なのかが把握しにくい。入院する患者の多くは軽重さまざまな病気を抱えている。機構病院のレセプトデータにおいて病名別の集計を行うと、多くの病院で患者数の多い病気は高血圧症、高脂血症、糖尿病、便秘症、不眠症となる。多くの人が持っている持病や、入院中の生活環境の変化によって現れやすい一時的な症状が上位に挙がることになる。しかし、機構病院の入院患者の大多数は高血圧症や高脂血症の治療のために入院しているわけではない。医療者の視点で病気ごとの患者数を集計するためには、主に何の病気の治療のための入院であったのかを把握する必要がある。レセプトデータの仕様上、主病名にはその旨記載する仕様になっているが、請求上問題にならないため記載漏れが多く見受けられる。逆に複数の病名が主病名となっているケースも少なくない。その入院が何のための入院であったのかを判断するため、DPCデータの医療資源を最も投入した傷病に該当する病名を選ぶ。

3.2.7 問題点6 レセプト電算コードからの手術や処置等の把握

前述の通り、レセプトにはその月に行われた検査や手術等の医療行為、投与された薬剤、使用された注射器等の医療材料が記載されている。記載はレセプト電算コードと呼ばれる9桁の数字でなされ、一見しただけでは何が行われたのか判断できない。レセプト電算コードとその内容については、厚生労働省保険局がマスタを公開しているため、それにより判断する。これにより、何のための入院であったのかを判断するための一助とする。

3.2.8 問題点7 重症度の把握

レセプトデータには、DPCデータの様式1のように重症度や病気の進行度を記載する仕様がない。症状詳記というレコードが存在するが、これは診療報酬請求が滞りなく行われるための説明文であり、担当医師や治療エピソードなど、患者個人を特定し得る情報が含まれている可能性があるため、データ収集の際に削除している。よって、行われた手術の術式や投与された薬剤から重症度を推測する。これには診療情報分析部の研究員のみならず、病院で働く現場の医療職や外部の研究者の知見も必要になる。

4. レセプトデータの前処理

前章で述べた数々の問題点を、その解決方針に沿って処理していく。この章では、DPCデータとの統合によるデータベース化の際、レセプトデータに対して具体的

に行った処理について述べる。

4.1 レセプトデータの正規化によるRDBへの格納

MIAは受信したレセプトデータを自動で患者プロフィール情報、病名情報、医療行為情報などに分割、正規化し、RDBに格納している(問題点1の解決)。プロフィール情報はDPCデータ様式1、医療行為情報はFファイルになっている(図6)。

4.2 1入院データの作成

レセプトには入院日コメントと退院日コメントが記載されている(図7)。一連の治療で入退院を繰り返した場合、初回の入院まで遡ってすべての入退院日が記載されており、またすべての月のレセプトに繰り返し記載されているため、一意に抽出しソートすることで入退院日のセットを作成する(問題点2の解決)。

4.3 入院データへのラベリング

前述の入院分析データセットに、各入院において患者はどういう状態であったのか、何が行われたのかをラベリングし、集計しやすくする。

4.3.1 病名の解決

MIAはレセプトデータのテーブル化の際、厚生労働省

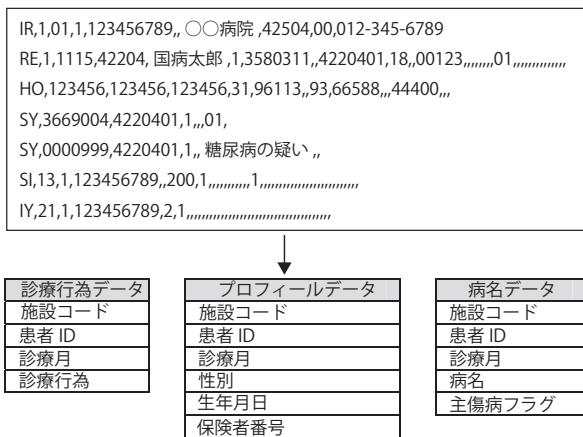


図6 レセプトデータのテーブル化

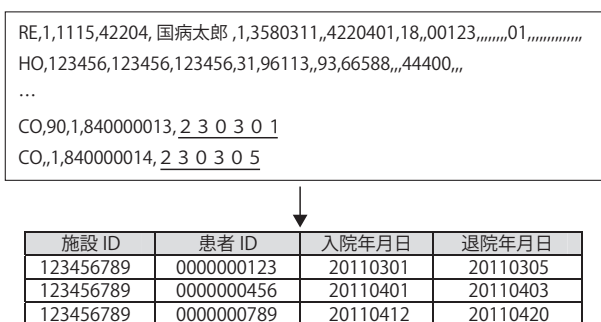


図7 入退院情報の抽出

保険局が公表している傷病名マスタを元に、自動で標準病名コードからICDコードへの変換を行う。標準病名コードが入力されていないレコードについては、ICDコード列を埋める必要がある。また、修飾語コードとの組合せによって本来の標準病名コードが入力されていない恐れのあるレコードについても、いったん選択されたICDコードを破棄し、ふさわしいコードにつけ直す。これらを実現するためには、パターンが無限にあるテキスト病名とICDコードの組合せが必要になる。我々はこれを、DPCデータ様式1の病名列から取得した。様式1には入院中に医療資源を最も投入した病名だけでなく、医療資源を2番目に投入した病名、主病名、入院の契機となった病名、入院時に患っていた病名、入院後に発症した病名について、それぞれICDコードとテキスト病名が入力されている。これらをテーブル化し、辞書としてレセプトデータの病名テーブルにテキスト病名で結合(JOIN)演算することで、病名コードを得ることができる(図8)。様式1に記載されている病名コードは医師が選択したものであり、病名コードとテキスト病名の組を抜き出すということは、どういう病名にどういうコードを選択するかという医師の知見の集合にほかならない。こうして、精度の高いデータを用いて精度の低いデータの精度を高めることに成功した(問題点3、4の解決)。2011年度、レセプトデータのみ提出している病院の分析で用いた病名データは計2,918,280件であったが、このうち237,987件はテキスト病名のみ記載でコード化されていなかった。当該処理を行った結果194,038件(81.5%)についてコード化を行うことができた。残る未コード化病名については、不明として集計している。

4.3.2 主病名の解決

入院期間中のレセプトデータに複数の主病名が記載されている場合、紙イメージでは病名欄の上部に記載される病名がその入院における主病名であろうとの推測の元に、主病名群の中で最も先頭行の病名を主病名とした。主病名が1つもない患者についても同様に、先頭行の病名を主病名としている(問題点5の解決)。2011年度、レセプトデータのみ提出している病院で、1入院に対して主病名が1つ特定できる患者(1つだけ主病名を持っていて、かつ病名がコード化されている)は50.8%にすぎ



図8 未コード化病名のコード化

なかったが、これらの処理により99.0%まで改善した。残るデータについてはごく僅かであるため、分析対象外としている。

4.3.3 入院中に行われた医療行為の解決

レセプトに記載されたレセプト電算コードと診療行為、薬剤、医療材料のマスタを結合することで、入院中に行われた各種治療を把握し、ラベリングする(問題点6の解決)。診療行為のレセプト電算コードはたとえば手術と手術に関する加算だけでも約2,500種あり、薬剤については同じ成分であっても10mgの錠剤と5mgの錠剤、錠剤か点滴薬か、発売元はどこかでレセプト電算コードが異なっている。その数は膨大になるため、集計前にマスタにフラグをつけておくことが望ましい。薬剤マスタには薬価基準収載医薬品コードと呼ばれる12桁のコードが薬剤ごとに記載されており、先頭の3桁は薬効を示している。たとえば解熱鎮痛消炎剤を示すコードは114であり、約500種存在している。

4.3.4 患者状態像の解決

臨床実態の分析のため、前述の主病名や医療行為を組み合わせることで、よりピンポイントな患者集団の抽出を行う。たとえば、気管支喘息の入院患者に対し吸入ステロイドによる治療が行われているかという集計を行ったところ、4割程度という結果になり、現場の医療職の感覚とはかけ離れた数値となった。別の病気で入院し、気管支喘息の症状は落ち着いている患者が多く含まれていたようである。そこで、病名を主病名に限定し、入院期間中にある種の喘息治療薬が処方されている患者に限定すると7割弱となり、おおよそ実態にあった数字となった。単純に病名や手術だけでは抽出できないような患者群については、各学会が発表している治療のガイドラインや、現場の医療職に知見を求め、1つずつ丁寧に抽出ロジックを考えていく必要がある(問題点7の解決)。

4.4 入院データのグループ化

さまざまな視点で患者の構成比を分析するために、患者対条件が多対一となるようにグループ化を行う。厚生労働省が公表している診断群分類(DPC)電子点数表を参考に、各入院レコードに対し診断群分類コードをコーディングした。まず主病名のICDコードから疾患コードを、入院中に行われた手術や処置、投薬から手術コード、処置コードをコーディングした。出生時体重や入院目的等、機械的に分かりかねる部分は不明とした。レセプトデータから作成した入院患者データ約60万件の98%について、診断群分類コードに準じたものを付与することがで

きた。主病名がICDコード化されていなかったなどの理由で診断群分類コードが付与できなかった患者については、診断群分類不明として集計している。

5. データの前処理による効果

前章の手順で作成した分析データセットを用いて、診療機能分析レポートの作成と、臨床評価指標の算出を行った。各種の前処理がどのように役立ったかを事例とともに紹介する。

5.1 入院データの作成による効果

レセプトデータから入院分析用のデータセットを作成したことにより、さまざまな症例を集計し、一般に公表されている退院患者数との比較を行うことが可能になった。診療情報分析部では、部の発足時から診療機能分析レポートを作成している。診療機能分析レポートは、病院を個別に、客観的かつ定量的に評価し、病院運営に役立ててもらうための資料であり、個別病院ごとに1冊の本にまとめてそれぞれの病院に配布している。個々の病院の絶対評価として、入院患者のうち、最も多い病気は何か、何日程度で患者を退院させることができていくのか、手術が行われた患者は何割か、人工呼吸器の装着などが必要な重症患者はどの程度いるのかなどを数値で示している。相対評価としては、厚生労働省中央社会保険医療協議会DPC評価分科会において毎年公開されている全国のDPC病院(2011年4月時点で1,648病院)に関する病院別、疾患別、手術別の退院患者数や、3年に1度、同じく厚生労働省が公開している患者調査にある地域別、疾患別の退院患者数を用いて、病院が存在する地域におけるシェアを計算し、示している。地域の状況を示すことで、病院の果たすべき役割や、今後強化すべき機能、撤退すべき分野などを考えてもらう資料としている。

5.2 入院データへのラベリングによる効果

入院データに対する各種ラベリングにより、さまざまな患者集団の抽出が可能になり、これにより、臨床実態の可視化が実現した。機構本部では2007年度より医療の質を評価する事業(以下、臨床評価指標)に取り組んでいる。診療情報分析部の発足により、その事業を引き継ぎ、発展させることとした。臨床評価指標は、患者に提供される医療の質を定量的に計測し、診療の改善に役立てるものである。各種学会などで標準的とされている

治療が行われているかを測ることで、我流の診療や場当たりの診療が行われていないかを計測できる。現在、機構の臨床評価指標はDPCデータやレセプトデータから集計しているが、診療情報分析部発足前は抽出条件を各病院に伝え、病院職員に該当患者数を報告させていた。

従来、抽出条件の変更は気軽にできるものではなかった。しかし、MIAの整備により、病院職員に負担を強いることなくさまざまな条件での試行集計が行えるようになり、まったく同一の条件で全病院を横並びに比較できるようになった。特に抽出仕様の策定にあたっては、各疾患領域の識者と研究員が検討を重ねながら試行集計を何度も繰り返すため、SQLの変更のみで再集計を行えるようになったことは意義深い。2010年度、厚生労働省の医療の質の評価・公表等推進事業[1]（以下、推進事業）に参加し17個の指標を作成した上で、集計結果をWeb上で公表した[2], [6]。これはDPCデータのみを用いたものであったが、翌年国立病院機構の臨床評価指標[3], [5]として、新たに70指標を整備し（以下、機構指標）、いくつかの指標については前述のレセプトデータから作成したデータセットを用いて算出を行った。機構指標の抽出ロジックについては、DPCデータやレセプトデータにある程度親しんだ者であれば分かるようにまとめ、書籍として出版した[4]。

5.3 入院データのグループ化による効果

レセプトデータから作成した入院分析用データセットに診断群分類コードに準ずるものを付与したことにより、延べ60万人以上の入院患者が重複なく約2,500種に分類された。それにより、たとえば疾患群別に治療に要した期間、リハビリテーションの手厚さ、地域の病院との連携状況、それらの経時変化を示すことが可能になった。また、群集生態学の分野で使われている類似度指数であるC_λ指数を応用し、各診療科における診断群分類ごとの患者構成比から、機構内のすべての病院、すべての診療科の間の類似度を総当たりで計算し、ある病院のある診療科にもっとも似ている他院の診療科を抽出した。自院の各診療科と他院の似ている診療科の治療状況を各病院に示すことで、問題提起を行った。

6. 今後の課題

6.1 精緻な分析を目指す上での課題

レセプトデータでの集計は難点が多いが、DPCデータを提出する病院は半数にも満たないため、分析データ

の質を向上させるためにレセプトデータでの抽出の幅を広げていかなければならない。具体的には、DPCデータの様式1に記載されているような、患者の状態像の評価の実現、現在のところあまり利用できていない外来レセプトデータの積極的利用、機械学習による主病名の選択を考えている。

DPCデータの様式1には、30種以上もの入院中の状態評価や診療状況が体系化されたコードで記載されているのに対し、レセプトデータにはそれがない。それらを医療職の協力を仰ぎながら入院データにラベリングすることで、レセプトデータによる分析の幅を広げることができる。処置や投薬から、重症度の把握をしていくことを考えている。

また、今まではレセプトデータには治療を行った日付の記載義務がなかったため、外来レセプトデータについてはあまり利用してこなかったが、2012年度以降日付の記載が必須となったため、今後は積極的に利用していく。これにより、たとえば外来の通院がメインの症例の分析や、入院前後の通院も合わせた一連の治療エピソードの分析が可能になる。

もう1つ取り組みたいのは、主病名を機械学習によって把握する方法である。現在主病名が特定できないデータについては病名欄の一番上に記載されている病名を主病名としているが、病名データを数件見てみると、患者によっては主病名とは思えない病名が先頭に記載されている場合もある。このような単純な抽出ロジックではなく、蓄積された膨大なデータから、病名と医療行為の組を抽出し、機械学習することで、レセプトに記載された多くの病名の中から、主病名らしい病名を選び出し、患者状態像把握の一助としたい。

6.2 データの管理における課題とフィードバックにより見えてきた課題

2010年度より2年間で、収集したテキストデータの容量は約2.2TBとなっており、様式1テーブルは約86万行、Fファイルは約3億行、レセプトデータの患者プロフィール情報テーブルは約1,500万行となっている。MIA整備時には、DPCデータ、レセプトデータはいずれも公的な機関に提出しているデータであるため、仕様通りの整ったデータが収集できるであろうという予測のもと、細かなチェック機能は備えていなかった。しかし、実際には各病院のレセプト出力システムがベンダによって微妙に異なること、またシステム更新や担当者変更における引き継ぎミスなどにより、ファイル間で患者IDが一

致しないデータやレコード件数の過少過多などが散見され、部内でのチェック、修正、電話やメールによる再提出依頼を行った。これにより、144病院分のデータを整えるのに時間がかかり、診療機能分析レポート、臨床評価指標ともに前年度のデータを用いた集計となってしまった。機構各病院からはもっと鮮度の高い分析結果の提示を求められており、この課題を解決するために、今後はアップロード後のチェック機能やエラー提示画面を強化し、収集したらすぐに使用できるデータの整備を目指したい。また、日々肥大化するデータベースの切り口を見つけるため、ビッグデータのビジュアライゼーションの手法などを取り入れ、データの素早い概要把握に役立てていきたい。

7. おわりに

MIAの整備により国内外でも類を見ない規模である144病院分の診療データを収集することに成功し、すべての病院に対して個別のレポートを提示することができた。また、臨床評価指標として70種の患者の状態を抽出するロジックを整備し、集計結果を提示することができた。2012年度で部の発足から3年目となり、経年比較や長期間のデータの集計を行えるようになった。日々蓄積されるデータをもとに、医療の現場で役に立つ分析が行える基盤として育てていきたいと考えている。

謝辞 本稿の執筆にあたり、日頃DPCデータ、レセプトデータの処理に共に試行錯誤し、取り組んでいるシステム開発専門職の下田俊二氏、中寺昌也氏に謝意を表します。

参考文献

1) 厚生労働省 平成22年度医療の質の評価・公表等推進事業, <http://www.mhlw.go.jp/topics/2010/06/tp0630-6.html>

- 2) (独) 国立病院機構 平成22年度医療の質評価・公表推進事業における臨床評価指標, <http://www.hosp.go.jp/7,9502.html>
- 3) (独) 国立病院機構 臨床評価指標 2010, <http://www.hosp.go.jp/7,10838.html>
- 4) (独) 国立病院機構 臨床評価指標計測マニュアル, <http://www.hosp.go.jp/news/detail.7.11866.html>
- 5) (独) 国立病院機構 臨床評価指標 2011, <http://www.hosp.go.jp/7,11980.html>
- 6) (独) 国立病院機構 平成23年度医療の質評価・公表推進事業における臨床評価指標, <http://www.hosp.go.jp/7,11979.html>
- 7) 秦 温信: ベンチマーク分析によるDPCの評価, 日本医療マネジメント学会雑誌, Vol.11, No.2, pp.80-91 (2010).
- 8) 金子悦代, 藤田貴久子: 院内がん登録と診療科データベースによる共有データベースの構築 大学病院間のベンチマークを目指して, 診療情報管理, Vol.23, No.2, p.172 (2011).
- 9) 藤本俊一郎, 大橋智明: 医療の質の向上・効率化と経営基盤確立に寄与する電子パスの構築 DPC・原価計算・電子パスの3種類のソフトを用いた経済的検討, 日本職業・災害医学会誌, Vol.57, No.6, pp.274-284 (2009).
- 10) 梶島博彰, 山崎信子, 向井知巳: DPC データを用いた院内ベンチマーク分析, 日本病院会雑誌, Vol.55, No.2, pp.163-167 (2008).

川島 直美 (非会員) kawashima-naomi@nho.hosp.go.jp
2005年 会津大学コンピュータ理工学部卒業, 東機エレクトロニクス(株)を経て2011年より(独)国立病院機構本部総合研究センター診療情報分析部システム開発専門職。

堀口 裕正 (非会員) hiromasa-ky@umin.ac.jp
2003年九州大学大学院医学系研究科博士課程環境社会医学専攻修了, (財)日本医療機能評価機構を経て2005年より現在まで東京大学大学院医学系研究科 医療経営政策学講座 寄附講座 教員 (2008年より特任助教)。

伏見 清秀 (正会員) kfushimi.hci@tmd.ac.jp
1985年東京医科歯科大学医学部卒業, 同大医学部第二内科入局, 1995年同助手(腎臓内科), 1997年同大医療情報部助手, 1998~2000年厚生省保険局医療課医療指導監査室特別医療指導監査官, 2000~07年東京医科歯科大学医学部医療情報部助教授, 同大学院医療政策学講座医療情報・システム学分野助教授, 2010年より同講座医療政策情報学分野教授。

投稿受付: 2012年11月5日

採録決定: 2013年5月23日

編集担当: 山川 聡 (日本電気(株))