

200300191A

厚生労働科学研究費補助金 長寿科学総合研究事業

# 国立大学病院に及ぼすDPC導入の影響

— 主要老年病の医療・治療の標準化・充実を目指して —

平成15年度 総括研究報告書

主任研究者 西岡清

(東京医科歯科大学医学部附属病院病院長)

平成16(2004)年4月

厚生労働科学研究費補助金

長寿科学総合研究事業

国立大学病院医療に及ぼすDPC導入の影響  
—主要老年病の医療・治療の標準化・充実を目指して—

平成15年度 総括研究報告書

主任研究者 西岡 清

平成16（2004）年 4月

## 目 次

### I. 総括研究報告

- 国立大学病院医療に及ぼすDPC導入の影響  
ー主要老年病の医療・治療の標準化・充実を目指して …………… 5  
西岡 清

### II. 分担研究報告

1. 高齢者医療の診断群分類のあり方に関する研究 …………… 13  
伏見 清秀
2. 国立大学病院DPCデータ解析システムの構築に関する研究 …………… 17  
石田 達樹

### III. 参考資料

1. DPCデータ分析システムの概要 …………… 23
2. DPCデータマイニングシステムの概要と分析方法の検討 …………… 183

厚生労働科学研究費補助金(長寿科学総合研究事業)  
総括研究報告書

国立大学病院医療に及ぼすDPC導入の影響  
—主要老年病の医療・治療の標準化・充実を目指して—

主任研究者 西岡清 東京医科歯科大学医学部附属病院病院長

研究要旨

平成15年4月より導入されている「特定機能病院の入院医療の包括評価」では全国の特設機能病院等が対象とされており、一般病院入院患者であって傷病名等が診断群分類に該当するものは、「診断群分類ごとの1日当たり点数×医療機関別係数×入院日数」で算定されている。本制度の導入に伴い、特定機能病院の入院医療の診療報酬額は、診断群分類ごとの1日当たりの点数、医療機関別係数、入院日数の包括部分と入院基本料等加算等の出来高の組み合わせにより規定されている。一方国立大学病院は特定機能病院として、また各地域における中核病院として等重要な役割を果たしている。今回の包括評価導入は、DPCの分類、制度の設計等に十分な議論がなされているとは言えない。また医事会計システム、支援システム等の病院情報管理システムやDPCの決定や患者様の説明等院内教育・研修システムも新制度導入の時間的な制約から、不十分なままである。これらのことから高齢者医療をはじめとする特定機能病院、特に大学病院の特色が十分に生かされなくなる可能性が危惧されている。そこで本研究ではDPCのデータ解析を通じて高齢者医療の標準化・充実を図ることを目的とし、国立大学病院におけるDPC制度導入による医療、運営に与えた影響に関して調査研究を行うこととした。

初年度の本研究では、研究目的に沿ったデータ解析を行えるようなデータの収集体制の確立およびデータ分析システムの設計と構築を主題とし、サンプルデータを用いて予備的な解析を行った。データ分析方法としては、年齢の影響、アウトライヤーの分析、施設間格差の検討、診断群分類の問題点と改善案の検討等が行えるよう、非常に詳細な descriptive な解析が行える分析システムと、人工知能の技術を応用して分類ルール等に新たな発見を試みるデータマイニングの手法を併用することとした。予備的分析により、全国立大学附属病院の調査データを解析する機能が確認され、年齢、地域格差等を考慮した分類の改訂案を検討するための方法論等を明かした。

#### 分担研究者

伏見清秀 東京医科歯科大学助教授

石田達樹 東京医科歯科大学専門官

### A. 背景と目的

平成15年4月より「特定機能病院の入院医療の包括評価」が導入されている。全国42の国立大学病院は特定機能病院として、また各地域の中核病院として重要な役割を果たしている。今回のDPCの導入により、高齢者医療をはじめとする医療、経営を含めた運営に関して大学病院は多大な影響を受けることが予想される。これまで、国立大学医学部附属病院長会議常置委員会等において包括評価制度に関する検討を行ってきたが、同様な形態で検討をすすめることは、データの内容、信頼性等からも必要不可欠であると考えられる。したがって本研究では、特定機能病院である大学病院の特色を十分にいかせるよう、国立大学病院におけるDPC制度導入により生じた医療、運営上の変化を分析することで、診断群分類の精緻化、診療内容に見合うコストの算定等包括制度の充実を図ることを目指し、DPC全体の精緻化を行うとともに、高齢者に対する医療費の実態を明らかにし、DPCにおける高齢者医療に対する精緻化を行うことを目的とした。

### B. 方法

全国42大学でのデータを集積、解析して適切な医療費配分のための資料を作成した。また、国立大学病院のみなら

ず公私立大学病院についてもDPCに関するデータの比較分析が可能かどうか検討を行った。全国42国立大学医学部附属病院における医療費データを「出来高払い」、「包括払い」に加えて、国立大学が開発導入している「経営情報分析システム」を用いて「実医療費データ(Costing data)」を集積した。平均在院日数を初めとする入院関連のデータだけでなく、画像検査等を初めとする高齢者の外来診療の与える影響等も併せて検討を行った。また、主要老年病の医療や治療に関する標準化・充実を図ることを目的に大学間の比較検討を行った。さらに、解析した実医療費データを用いて診断群分類の精緻化を検討し、特に高齢者医療におけるDPCの精緻化のための分析を行った。

### C. 結果

初年度は、(1)DPCデータ収集体制の確立、(2)DPCデータ分析システムの設計と構築、(3)DPCデータ解析データマイニングシステムの設計と構築を実施した。

#### (1)DPCデータ収集体制の確立

①各大学病院の協力体制の確認、②個人情報保護に配慮したデータフォーマットとデータ送付方法の検討と決定、③DPCデータ分析のための必要なデータフォーマットの変換、④収集されたデータに含まれるエラーの検出と補正方法の検討を行った。

#### ①各大学病院の協力体制の確認

国立大学の大部分の協力が得られ、データ提供およびデータ解析チームの

編成に関する検討を継続して進めた。平成15年7月から10月にかけて実施された「診断群分類見直しのためのデータ収集」に際して厚生労働省に提出した様式1データ(診療情報に関するデータ)および診療報酬明細データとしてレセコンデータダウンロード方式に準じたフォーマットによる診療報酬請求明細データ(EFファイル)を収集した。EFファイルには、従来の出来高算定に準じた診療行為明細内容とともに、丸め処理等を行う前の日付・回数情報なども含む診療行為実施明細データが含まれているため、出来高請求と包括評価請求の比較とともに、包括評価の下での診療行為明細の日々のデータを解析することが可能である。具体的には、個別の医薬品や医療材料の使用状況、診療行為の日計表分析、診断群分類毎の最適パスの抽出、診療行為の医療機関間差異等の分析のためのデータを用意した。

## ②個人情報保護に配慮したデータフォーマットとデータ送付方法の検討と決定

分析に供用される診療関連情報、診療報酬明細情報からは個人を識別しうる氏名、住所、生年月日等の情報を削除した状態でデータベースに格納し、個人識別情報を除去する前段階で、ダミーコードによって診療関連情報、診療報酬明細情報、包括評価請求情報等相互のリンクを設定した。年齢等の情報は必要に応じて、年齢階級表記に変更した。稀少疾患等で症例数が少なく個人が同定される可能性がある情報は、集計表から除外し、公表されないように配慮した。データフォーマットは各参加医療機関の負担

を極力下げ、また、エラー発生危険性を低めるため、厚生労働省に提出したデータをそのまま研究に使用することとした。厚生労働省への提出データは、エラーチェック等のため訂正、修正後複数回再提出されている部分もあったため、最新版のデータを集約することで統一し、医療機関間のエラー発生状況の不均一性を可能な限り低めることを考慮した。

さらに、将来的なデータの継続的な集約と分析およびデータ解析システムへの各参加医療機関からのアクセスを可能とするために、データ解析システムをファイアーウォールを経てインターネットへ接続する構成とした。情報保護の観点からシステムへのアクセス、データ転送等は全てVPNを介して、暗号化された情報により安全に実施する方法をとった。

## ③DPC データ分析のための必要なデータフォーマットの変換

収集されているデータは反復項目、重複項目等を含む冗長なデータであるためこのまま集計解析を行うには適していないことが明らかとなった。例えば、手術に関する項目は同じものが5回反復されているが、データ解析時は5つの項目を同等に評価する必要がある。入院時併存症、合併症等の項目も同様な問題がある。また、解析に必要な在院日数、外泊日数等のデータの計算導出も必要であることが明らかとなった。これらの問題に関して、データの正規化の方法、必要なデータの導出方法等を検討した。さらに、日毎の診療行為、医療費等を分析するためには、収集されたEFフォーマットのデータでは冗長すぎる事が明らかと

なったため、日計表形式の日毎の推移、変化を解析しやすいフォーマットへの変更を検討した。

#### ④収集されたデータに含まれるエラーの検出と補正方法の検討

「診断群分類見直しのためのデータ収集」によって収集されたデータは、データの作成方法、作成システム等が明確に規定されていないため、各医療機関が独自のシステムを構築しデータを収集、提出している。本研究のために収集されたデータも厚生労働省に提出されたデータと同等であるため、データの作成方法に関連すると考えられる異常データの発生を認めた。収集されるデータのうち診療情報に関連する「様式1」情報は、日付情報、診断情報、診療関連情報を含むが、基本的に文字列データ形式で提出されているため、提出医療機関が文字列フォーマットに変換する際に生じたと考えられるエラーが認められた。文字コードのエラー、全角半角の処理エラー、タイプエラー、漢字変換エラー、不要なスペースの付加、異常な項目区切りキャラクター等がこれらに含まれた。また、日付データ、病名コードデータなどに関しては日付整合性エラー、存在しないコード等を認めた。これらは、各病院が情報システム等から直接データを取得している場合は発生しにくいと考えられるが、何らかのデータ変換操作、または手作業によるデータ入力等の過程でエラーが発生している可能性が考えられた。さらに、EFデータと診療関連情報の不整合エラーが認められ、EFフォーマットの診療報酬関連データは直接データダウン

ロードシステム等で作成していることから信頼性が高い反面、診療関連情報の作成には、医師の認識、医師と診療情報関連部門の情報共有等種々の問題からエラーが発生しやすいと考えられた。

これらのエラーは、データ解析時の異常値、場合によっては異常な解析結果や解釈の誤り等を誘発する危険があるため、注意深く除去する必要性を認めた。コード変換、エラー検出ロジック、多重整合性チェック、異常値の視認等複数の方法を組み合わせてデータのクリーニングを行い、エラー除去を試みた。最終的にデータの正確性が担保出来ないものについては解析から除去することとし、約3～5%のデータが解析に使用されなかった。

#### **(2) DPC データ分析システムの設計と構築**

昨年度までに構築した国立大学病院 DPC 分析評価システムをベースに① DPC データ分析システムの要件の設定とシステム設計、および、②サンプルデータを用いた分析テストラン、等を実施した。

##### ①DPC データ分析システムの要件の設定とシステム設計

本研究に用いられる大量のデータの効率的な解析を実施するためには、収集されたデータを元に、年齢、性別、診断群分類、在院日数、医療費等の視点からインタラクティブな解析を提供できるデータベースと分析システムが必要である。これらの点に関して、分析に適するデータフォーマット、データ統合の手法、分析システムのあり方等を検討した。また、イ

インタラクティブな分析のインターフェース、分析手順等を検討した。高齢者医療を含む診断群分類の妥当性の検証には、散布図、箱ひげ図などの集約的分析とアウトライヤー、インライヤー解析などの個別的な分析を、効率的に平行して進めていく必要がある。そこで、診断群分類の各階層での年齢、在院日数、医療費等のばらつきの程度とその要因の分析、医療費区分毎のばらつきへの影響度の検討、ばらつきの医療機関要因の分析等に散布図、箱ひげ図、変動係数等を活用する方法論を検討した。年齢、診断群分類、在院日数、医療費等の視点からの集約的分析には、入院IDをキーとして「様式1」データ、「様式2」データ、DPCコードを格納するリレーショナルDBを構築し、それを基礎に分析用アプリケーション作成と集計分析を実施した。EFファイルデータはレコードが細分化され、そのままでは集計分析に適していないため、「様式2」相当ファイルに集計して利用した。

データ量、分析手法の複雑性から診断群分類毎の集計データをリアルタイムで供覧するのは実用的ではなかったため、あらかじめバッチ処理によって全ての分析パネルを作成し、分析者のリクエストに応じて必要な集計結果パネルを供覧する方法が有効であった。分析結果はhtml化されていたため、web参照など汎用性、携帯性に優れていた。一方、アウトライヤー分析等で必要な個別的解析には、集約的分析の個別データからリンクされたEFファイルデータ参照機能が効果的であった。箱ひげグラフ上のはず

れ値から直接EFファイルを参照し、アウトライヤー症例の具体的診療内容を精査することにより、アウトライヤー要因の解析が効率的に実施された。

## ②サンプルデータを用いた分析テスト

分析結果パネルは、棒グラフ、散布図、箱ひげ図などの集計グラフと平均、最大、最小、分散、変動係数等の基本統計一覧表を組として作成することにより、正確な検討が可能であった。具体的な項目をいくつか以下に例示する。性別・年齢別の症例数の分布の集計は、疾患の性・年齢特異性を把握する上で必須であった。病院別在院日数の評価は、医療機関別特異性の把握に有益であった。はずれ値の一覧表を付記することにより、個別症例に関するEFファイルの詳細データを参照することが可能であり、医療機関毎のアウトライヤー要因の分析と患者年齢の関与等の解析に重要な役割を果たした。在院日数と1日あたり診療報酬請求額のプロット、在院日数と1入院あたり診療報酬請求額のプロット、および在院日数と診療報酬区分別プロット等は、患者の年齢区分毎に比較が可能であり、高齢者に関する包括評価の妥当性、診断群分類の妥当性、短期入院評価の分析に有用であった。診療報酬区分別の1日あたり診療報酬請求額と1入院あたり診療報酬請求額は、変動要因への患者年齢の影響の分析に必要であった。そのほか病院特異性および地域特異性の把握は、患者年齢分布の相違との関連性を比較する上で有用であった。



### (3) DPC データ解析データマイニングシステムの設計と構築

①データマイニング用データセットの構築、②収集されたデータからのマイニングデータへの変換方法の検討、③データマイニングに不適当なデータのクレンジング作業の検討、④マイニングロジックの検討、等を実施した。

#### ①データマイニング用解析システムの検討

データマイニングは投入された大量のデータから新たな「法則」等を発見する手法であるため、分析にどのようなデータを使用するかを検討が最も重要なステップの一つとなる。データの項目数、データレコード数、データの質、データの粒度等を、必要とされる分析内容、CPU パワー、メモリ量等を含むシステムの能力等を勘案して決定する必要があった。本年度の予備的検討で、システムとして 4 Pentium4-Xeon 3 GHz CPU を搭載する PC サーバーを用意し、DPC 調査項目から様式1、EF データの全項目を使用して全病院からのデータの解析が可能であるかを検討した。マイニングシステムは多くの企業で採用されマイニング機能の実績が高く、大容量のデータの解析にも適している IBM DB2 Intelligent Miner を使用することとした。サンプルデータとして用意された1病院分のデータの解析状況から、予定の解析は可能であることが確認された。

#### ②収集されたデータからのマイニングデータへの変換方法の検討とデータセットの構築

様式1データから在院日数、年齢等に

関するデータを導出し、必要に応じて外泊日数の補正、年齢階級別解析への対応を行った。年齢階級としては、高齢者の状況を詳細に解析出来るよう、65才から74才の前期高齢者、75才以上の後期高齢者、5才以下の乳幼児、14才以下の小児と分割することとした。当初の解析用のデータセットとして、現行の診断群分類の決定ロジックに使用されているデータを全て投入することとした。また、診療報酬の解析のため、1日あたり診療報酬を計算し主要区分毎の診療報酬額も分析に投入することとした。

#### ③データマイニングに不適当なデータのクレンジング作業の検討

データマイニングにあたっては、異常値の存在が解析結果に与える影響が非常に大きいため、データのクレンジング作業が重要であることが明らかとなった。在院日数や1日あたりの診療報酬額の異常値が含まれているとマイニング結果に大きな影響を与える危険があるため、粗データの基礎統計の解析、データ分布状況の把握、はずれ値データの除去方法の検討を行った。上記 DPC データ分析システムの解析進行と併せて、異常値を取り除きながらマイニングを進めていくこととした。

#### ④マイニングロジックの検討

データマイニングには、Decision Tree などのクラス判別、予測、クラスター解析、相関関係分析、時系列順パターン分析、類似時系列分析などがあるが、本研究では、DPC 分類ツリー図の解析に Decision Tree 解析が、DPC 分類基準等を含む定義表の解析に予測解析およ

びクラスター分割解析が適していると考えられた。特に予測解析は、「教師つき」クラスター解析とも捉えられ、ある特定の指数、例えば、在院日数や1日あたりの医療費、1入院あたりの医療費等を指標にグループ分けを試みる解析に適していると考えられた。この方法により、年齢を含めた患者の諸因子が診断群分類にどのような影響を与えているかを検討することが可能となると考えられた。年齢に関しては、年齢階級を解析に含める方法と、年齢階級毎にデータを分割してそれぞれ独立した解析を行う2方法を併用することとした。同様に、患者年齢構成と医療機関間差異、地域格差等の影響を検討することとした。

#### D. 考察

DPC データ収集体制については、各大学病院の協力体制が確認され、個人情報保護に配慮したデータフォーマットとデータ送付方法、DPC データ分析のための必要なデータフォーマットの変換方法等については、実用上問題のないことが確認された。また、収集されたデータに含まれるエラーの検出と補正方法については、マンパワーを必要とする部分が大きく、次年度以降、収集するデータの質の向上とより効率的なエラー修正方法の検討を続ける必要があると考えられた。

DPC データ分析システムの設計と構築については、昨年度までに構築した国立大学病院 DPC 分析評価システムをベースに実施した DPC データ分析システムの要件の設定とシステム設計、およ

び、サンプルデータを用いた分析テストラン等により、構築した分析システムの実用性が高いことが確認され、大量の実データを用いた本稼働に問題がないことが確認された。次年度以降、引き続き、DPC データの解析を進めることが可能であると考えられた。

DPC データ解析データマイニングシステムの設計と構築については、データマイニング用データセットの構築、収集されたデータのマイニングデータへの変換、データマイニングに不適当なデータのクレンジング作業は順調に完了され、データマイニングシステムの稼働環境は十分に整えられた。マイニングロジックは基礎的な検討は完了したので、次年度以降、実効性の高いマイニング結果を求めていく方向で検討を進める必要があると考えられた。

これらの検討結果は、診療実態にみあう診断群分類の精緻化、投入医療コストに見合う定額設定等の基礎データが提供可能となるだけではなく、わが国の医療水準、医療費等の国際比較、年齢・重症度別のデータ比較、高度医療の医療保険への取り入れ、高齢者医療の実態、医師の適正な技術料算定等に視するデータの提供も可能となり、主要老年病の医療や治療法の標準化・充実に資することが期待された。

#### E. 結論

DPC 全体の精緻化、高齢者に対する医療費の実態の明確化、DPC における高齢者医療に対する評価方法のあり方等を明らかにするための、DPC データ収

集体制、データ分析システム、データマイニングシステムがほぼ完成し、次年度以降の研究の基礎的環境が整えられた。

**F.健康危険情報**

該当なし

**G. 研究発表**

該当なし

**H. 知的財産権の出願・登録状況**

該当なし

**I. 特許取得**

該当なし

厚生労働科学研究費補助金(長寿科学総合研究事業)  
分担研究報告書

高齢者医療の診断群分類のあり方に関する研究

分担研究者 伏見清秀 東京医科歯科大学大学院医療情報・システム学分野

研究要旨

平成15年4月より導入された特定機能病院の入院医療の包括評価では高齢者医療に対する評価が十分検討されているとはいえない。そこで本研究ではDPCのデータ解析により、高齢者医療のためのDPCの精緻化の方法論を検討した。初年度は、サンプルデータを用いて予備的な解析を行い、年齢の影響、アウトライヤーの分析、施設間格差の検討、診断群分類の問題点と改善案の検討等を行った。これらの予備的分析により、国立大学附属病院の調査データを解析する分析システムの実効性が確認され、年齢、地域格差等を考慮した分類の改訂案を検討するための方法論等を明かした。

A. 背景と目的

平成15年4月より「特定機能病院の入院医療の包括評価」が導入されている。ここで用いられている日本版DRGであるDPC (Diagnosis Procedure Combination) においては、小児等に対する独立した分類は多くあるが、高齢者に対する評価は存在しない。一般に、高齢者は若年者と疾病構造が異なり、また、併存症の存在、術後併発症の高リスク、リハビリテーションの長期化等、医療資源をより多く必要とする可能性が示唆されてきている。したがって本研究では、高齢者に対する医療費の実態を明らかにし、DPCにおける高齢者医療に対する精緻化を行うことを目的とした。

B. 方法

全国42大学でのデータを集積し、DPCデータ分析システムを構築して高齢者医療におけるDPCの精緻化のための分析を行った。また、DPCデータ解析データマイニングシステムを構築し、主要老年病の医療や治療に関する検討を実施した。

C. 結果

(1) DPCデータ分析システムによる分析

昨年度までに構築した国立大学病院DPC分析評価システムをベースにサンプルデータを用いた分析テストランを実施した。

分析結果パネルは、棒グラフ、散布図、箱ひげ図などの集計グラフと平均、最大、最小、分散、変動係数等の基本統計一覧表を組として作成することにより、正確

な検討が可能であった。具体的な項目をいくつか以下に例示する。性別・年齢別の症例数の分布の集計は、疾患の性・年齢特異性を把握する上で必須であった。病院別在院日数の評価は、医療機関別特異性の把握に有益であった。はずれ値の一覧表を付記することにより、個別症例に関する EF ファイルの詳細データを参照することが可能であり、医療機関毎のアウトライヤー要因の分析と患者年齢の関与等の解析に重要な役割を果たした。在院日数と1日あたり診療報酬請求額のプロット、在院日数と1入院あたり診療報酬請求額のプロット、および在院日数と診療報酬区分別プロット等は、患者の年齢区分毎に比較が可能であり、高齢者に関する包括評価の妥当性、診断群分類の妥当性、短期入院評価の分析に有用であった。診療報酬区分別の1日あたり診療報酬請求額と1入院あたり診療報酬請求額は、変動要因への患者年齢の影響の分析に必要であった。そのほか病院特異性および地域特異性の把握は、患者年齢分布の相違との関連性を比較する上で有用であった。

## (2) DPC データ解析データマイニングシステムによる検討

設計構築されたデータマイニングシステムを利用してマイニングロジックに関する予備的検討を実施した。

データマイニングには、Decision Tree などのクラス判別、予測、クラスター解析、相関関係分析、時系列順パターン分析、類似時系列分析などがあるが、本研究では、DPC 分類ツリー図の解析

に Decision Tree 解析が、DPC 分類基準等を含む定義表の解析に予測解析およびクラスター分割解析が適していると考えられた。特に予測解析は、「教師つき」クラスター解析とも捉えられ、ある特定の指数、例えば、在院日数や1日あたりの医療費、1入院あたりの医療費等を指標にグループ分けを試みる解析に適していると考えられた。この方法により、年齢を含めた患者の諸因子が診断群分類にどのような影響を与えているかを検討することが可能となると考えられた。年齢に関しては、年齢階級を解析に含める方法と、年齢階級毎にデータを分割してそれぞれ独立した解析を行う2方法を併用することとした。同様に、患者年齢構成と医療機関間差異、地域格差等の影響を検討することとした。

## D. 考察

DPC データ分析システムについては、サンプルデータを用いた分析テストラン等により、構築した分析システムの実用性が高いことが確認され、大量の実データを用いた本稼働に問題がないことが確認された。次年度以降、引き続き、DPC データの解析を進めることが可能であると考えられた。

DPC データ解析データマイニングシステムについてはマイニングロジックの基礎的な検討は完了したので、次年度以降、実効性の高いマイニング結果を求めていく方向で検討を進める必要があると考えられた。

## E. 結論

DPC 全体の精緻化、高齢者に対する医療費の実態の明確化、DPC における高齢者医療に対する評価方法のあり方を明らかにするための、データ分析システム、データマイニングシステムがほぼ完成し、次年度以降の研究の基礎的環境が整えられた。

#### F.研究発表

該当なし

#### G.知的所有権の取得状況

該当なし

厚生労働科学研究費補助金(長寿科学総合研究事業)  
分担研究報告書

国立大学病院DPCデータ解析システムの構築に関する研究

分担研究者 石田達樹 東京医科歯科大学医歯学教育システム研究センター

研究要旨

平成15年4月より導入されている特定機能病院の入院医療の包括評価ではDPCの分類、制度の設計等に十分な議論がなされているとは言えない。また医事会計システム、支援システム等の病院情報管理システムやDPCの決定や患者様の説明等院内教育・研修システムも新制度導入の時間的な制約から、不十分なままである。そこで本研究ではDPCのデータ解析を通じて高齢者医療の標準化・充実を図ることを目的とし、国立大学病院におけるDPCデータの解析のための分析システムの検討を行った。初年度は、研究目的に沿ったデータ解析を行えるようなデータの収集体制の確立およびデータ分析システムの設計と構築を実施した。データ分析方法としては、年齢の影響、アウトライヤーの分析、施設間格差の検討、診断群分類の問題点と改善案の検討等が行えるよう、非常に詳細な descriptive な解析が行える分析システムと、人工知能の技術を応用して分類ルール等に新たな発見を試みるデータマイニングの手法を併用することとした。

A. 背景と目的

平成15年4月より「特定機能病院の入院医療の包括評価」が導入されている。全国42の国立大学病院は特定機能病院として、また各地域の中核病院として重要な役割を果たしている。今回のDPCの導入により、高齢者医療をはじめとする医療、経営を含めた運営に関して大学病院は多大な影響を受けることが予想される。したがって本研究では、特定機能病院である大学病院の特色を十分にいかせるよう、国立大学病院におけるDPC制度導入により生じた医療、運営

上の変化を分析することで、診断群分類の精緻化、診療内容に見合うコストの算定等包括制度の充実を図ることを目指し、DPC全体の精緻化を行うとともに、高齢者に対する医療費の実態を明らかにし、DPCにおける高齢者医療に対する精緻化を行うことを目的とした。

B. 方法

全国42国立大学医学部附属病院における医療費データを「出来高払い」、「包括払い」に加えて、国立大学が開発導入している「経営情報分析システム」を

用いて「実医療費データ(Costing data)」を集積した。平均在院日数を初めとする入院関連のデータだけでなく、画像検査等を初めとする高齢者の外来診療の与える影響等も併せて検討を行った。また、主要老年病の医療や治療に関する標準化・充実を図ることを目的に大学間の比較検討を行った。

## C. 結果

### (1) DPC データ収集体制の確立

①各大学病院の協力体制の確認、②個人情報保護に配慮したデータフォーマットとデータ送付方法の検討と決定、③DPC データ分析のための必要なデータフォーマットの変換、④収集されたデータに含まれるエラーの検出と補正方法の検討を行った。

#### ①各大学病院の協力体制の確認

国立大学の大部分の協力が得られ、データ提供およびデータ解析チームの編成に関する検討を継続して進めた。平成15年7月から10月にかけて実施された「診断群分類見直しのためのデータ収集」に際して厚生労働省に提出した様式1データ(診療情報に関するデータ)および診療報酬明細データとしてレセコンデータダウンロード方式に準じたフォーマットによる診療報酬請求明細データ(EFファイル)を収集した。EFファイルには、従来の出来高算定に準じた診療行為明細内容とともに、丸め処理等を行う前の日付・回数情報なども含む診療行為実施明細データが含まれているため、出来高請求と包括評価請求の比較とともに、包括評価の下での診療行為明細の日々の

データを解析することが可能である。具体的には、個別の医薬品や医療材料の使用状況、診療行為の日計表分析、診断群分類毎の最適パスの抽出、診療行為の医療機関間差異等の分析のためのデータを用意した。

#### ②個人情報保護に配慮したデータフォーマットとデータ送付方法の検討と決定

分析に供用される診療関連情報、診療報酬明細情報からは個人を識別しうる氏名、住所、生年月日等の情報を削除した状態でデータベースに格納し、個人識別情報を除去する前段階で、ダミーコードによって診療関連情報、診療報酬明細情報、包括評価請求情報等相互のリンクを設定した。年齢等の情報は必要に応じて、年齢階級表記に変更した。稀少疾患等で症例数が少なく個人が同定される可能性がある情報は、集計表から除外し、公表されないように配慮した。データフォーマットは各参加医療機関の負担を極力下げ、また、エラー発生の危険性を低めるため、厚生労働省に提出したデータをそのまま研究に使用することとした。厚生労働省への提出データは、エラーチェック等のため訂正、修正後複数回再提出されている部分もあったため、最新版のデータを集約することで統一し、医療機関間のエラー発生状況の不均一性を可能な限り低めることを考慮した。

さらに、将来的なデータの継続的な集約と分析およびデータ解析システムへの各参加医療機関からのアクセスを可能とするために、データ解析システムをファイアーウォールを経てインターネットへ接続する構成とした。情報保護の観点から



システムへのアクセス、データ転送等は全て VPN を介して、暗号化された情報により安全に実施する方法をとった。

### ③DPC データ分析のための必要なデータフォーマットの変換

収集されているデータは反復項目、重複項目等を含む冗長なデータであるためこのまま集計解析を行うには適していないことが明らかとなった。例えば、手術に関する項目は同じものが5回反復されているが、データ解析時は5つの項目を同等に評価する必要がある。入院時併存症、合併症等の項目も同様な問題がある。また、解析に必要な在院日数、外泊日数等のデータの計算導出も必要であることが明らかとなった。これらの問題に関して、データの正規化の方法、必要なデータの導出方法等を検討した。さらに、日毎の診療行為、医療費等を分析するためには、収集されたEFフォーマットのデータでは冗長すぎるということが明らかとなったため、日計表形式の日毎の推移、変化を解析しやすいフォーマットへの変更を検討した。

### ④収集されたデータに含まれるエラーの検出と補正方法の検討

「診断群分類見直しのためのデータ収集」によって収集されたデータは、データの作成方法、作成システム等が明確に規定されていないため、各医療機関が独自のシステムを構築しデータを収集、提出している。本研究のために収集されたデータも厚生労働省に提出されたデータと同等であるため、データの作成方法に関連すると考えられる異常データの発生を認めた。収集されるデータのうち

診療情報に関連する「様式1」情報は、日付情報、診断情報、診療関連情報を含むが、基本的に文字列データ形式で提出されているため、提出医療機関が文字列フォーマットに変換する際に生じたと考えられるエラーが認められた。文字コードのエラー、全角半角の処理エラー、タイプエラー、漢字変換エラー、不要なスペースの付加、異常な項目区切りキャラクター等がこれらに含まれた。また、日付データ、病名コードデータなどに関しては日付整合性エラー、存在しないコード等を認めた。これらは、各病院が情報システム等から直接データを取得している場合は発生しにくいと考えられるが、何らかのデータ変換操作、または手作業によるデータ入力等の過程でエラーが発生している可能性が考えられた。さらに、EFデータと診療関連情報の不整合エラーが認められ、EF フォーマットの診療報酬関連データは直接データダウンロードシステム等で作成していることから信頼性が高い反面、診療関連情報の作成には、医師の認識、医師と診療情報関連部門の情報共有等種々の問題からエラーが発生しやすいと考えられた。

これらのエラーは、データ解析時の異常値、場合によっては異常な解析結果や解釈の誤り等を誘発する危険があるため、注意深く除去する必要性を認めた。コード変換、エラー検出ロジック、多重整合性チェック、異常値の視認等複数の方法を組み合わせてデータのクリーニングを行い、エラー除去を試みた。最終的にデータの正確性が担保出来ないものについては解析から除去することとし、約3

～5%のデータが解析に使用されなかった。

## (2) DPC データ分析システムの設計と構築

昨年度までに構築した国立大学病院 DPC 分析評価システムをベースに① DPC データ分析システムの要件の設定とシステム設計を実施した。

### ① DPC データ分析システムの要件の設定とシステム設計

本研究に用いられる大量のデータの効率的な解析を実施するためには、収集されたデータを元に、年齢、性別、診断群分類、在院日数、医療費等の視点からインタラクティブな解析を提供できるデータベースと分析システムが必要である。これらの点に関して、分析に適するデータフォーマット、データ統合の手法、分析システムのあり方等を検討した。また、インタラクティブな分析のインターフェース、分析手順等を検討した。高齢者医療を含む診断群分類の妥当性の検証には、散布図、箱ひげ図などの集約的分析とアウトライヤー、インライヤー解析などの個別的分析を、効率的に平行して進めていく必要がある。そこで、診断群分類の各階層での年齢、在院日数、医療費等のばらつきの程度とその要因の分析、医療費区分毎のばらつきへの影響度の検討、ばらつきの医療機関要因の分析等に散布図、箱ひげ図、変動係数等を活用する方法論を検討した。年齢、診断群分類、在院日数、医療費等の視点からの集約的分析には、入院 ID をキーとして「様式1」データ、「様式2」データ、DPC コードを格納するリレーショナル DB

を構築し、それを基礎に分析用アプリケーション作成と集計分析を実施した。EF ファイルデータはレコードが細分化され、そのままでは集計分析に適していないため、「様式2」相当ファイルに集計して利用した。

データ量、分析手法の複雑性から診断群分類毎の集計データをリアルタイムで供覧するのは実用的ではなかったため、あらかじめバッチ処理によって全ての分析パネルを作成し、分析者のリクエストに応じて必要な集計結果パネルを供覧する方法が有効であった。分析結果は html 化されていたため、web 参照など汎用性、携帯性に優れていた。一方、アウトライヤー分析等で必要な個別的分析には、集約的分析の個別データからリンクされた EF ファイルデータ参照機能が効果的であった。箱ひげグラフ上のはずれ値から直接 EF ファイルを参照し、アウトライヤー症例の具体的診療内容を精査することにより、アウトライヤー要因の解析が効率的に実施された。

### (3) DPC データ解析データマイニングシステムの設計と構築

①データマイニング用データセットの構築、②収集されたデータからのマイニングデータへの変換方法の検討、③データマイニングに不適当なデータのクレンジング作業の検討を実施した。

#### ①データマイニング用解析システムの検討

データマイニングは投入された大量のデータから新たな「法則」等を発見する手法であるため、分析にどのようなデータを使用するかを検討が最も重要なステ

ップの一つとなる。データの項目数、データレコード数、データの質、データの粒度等を、必要とされる分析内容、CPUパワー、メモリ量等を含むシステムの能力等を勘案して決定する必要がある。本年度の予備的検討で、システムとして4 Pentium4-Xeon 3 GHz CPUを搭載するPCサーバーを用意し、DPC調査項目から様式1、EFデータの全項目を使用して全病院からのデータの解析が可能であるかを検討した。マイニングシステムは多くの企業で採用されマイニング機能の実績が高く、大容量のデータの解析にも適しているIBM DB2 Intelligent Minerを使用することとした。サンプルデータとして用意された1病院分のデータの解析状況から、予定の解析は可能であることが確認された。

#### ②収集されたデータからのマイニングデータへの変換方法の検討とデータセットの構築

様式1データから在院日数、年齢等に関するデータを導出し、必要に応じて外泊日数の補正、年齢階級別解析への対応を行った。年齢階級としては、高齢者の状況を詳細に解析出来るよう、65才から74才の前期高齢者、75才以上の後期高齢者、5才以下の乳幼児、14才以下の小児と分割することとした。当初の解析用のデータセットとして、現行の診断群分類の決定ロジックに使用されているデータを全て投入することとした。また、診療報酬の解析のため、1日あたり診療報酬を計算し主要区分毎の診療報酬額も分析に投入することとした。

#### ③データマイニングに不適当なデータの

#### クレンジング作業の検討

データマイニングにあたっては、異常値の存在が解析結果に与える影響が非常に大きいため、データのクレンジング作業が重要であることが明らかとなった。在院日数や1日あたりの診療報酬額の異常値が含まれているとマイニング結果に大きな影響を与える危険があるため、粗データの基礎統計の解析、データ分布状況の把握、はずれ値データの除去方法の検討を行った。上記DPCデータ分析システムの解析進行と併せて、異常値を取り除きながらマイニングを進めていくこととした。

#### D. 考察

DPCデータ収集体制については、各大学病院の協力体制が確認され、個人情報保護に配慮したデータフォーマットとデータ送付方法、DPCデータ分析のための必要なデータフォーマットの変換方法等については、実用上問題のないことが確認された。また、収集されたデータに含まれるエラーの検出と補正方法については、マンパワーを必要とする部分が多く、次年度以降、収集するデータの質の向上とより効率的なエラー修正方法の検討を続ける必要があると考えられた。

DPCデータ解析データマイニングシステムの設計と構築については、データマイニング用データセットの構築、収集されたデータのマイニングデータへの変換、データマイニングに不適当なデータのクレンジング作業は順調に完了され、データマイニングシステムの稼働環境は十分

に整えられた。

これらの検討結果は、診療実態にみ  
あう診断群分類の精緻化、投入医療コス  
トに見合う定額設定等の基礎データが  
提供可能となるだけでなく、わが国の  
医療水準、医療費等の国際比較、年齢・  
重症度別のデータ比較、高度医療の医  
療保険への取り入れ、高齢者医療の実  
態、医師の適正な技術料算定等に視す  
るデータの提供も可能となり、主要老年  
病の医療や治療法の標準化・充実に資  
することが期待された。

#### E. 結論

DPC 全体の精緻化、高齢者に対する  
医療費の実態の明確化、DPC における  
高齢者医療に対する評価方法のあり方  
等を明らかにするための、DPC データ取  
集体制、データ分析システム、データマ  
イニングシステムがほぼ完成し、次年度  
以降の研究の基礎的環境が整えられ  
た。

#### F. 研究発表

該当なし

#### G. 知的所有権の取得状況

該当なし