

医療費の包括支払制度が医療の質に及ぼす影響 ーマッチング推定による検証ー

岡崎康平、杉原茂、田倉智之、村舘靖之

March 2023



内閣府経済社会総合研究所

Economic and Social Research Institute

Cabinet Office

Tokyo, Japan

ESRI Research Note は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

ESRI リサーチ・ノート・シリーズは、内閣府経済社会総合研究所内の議論の一端を公開するために取りまとめられた資料であり、学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

資料は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

The views expressed in “ESRI Research Note” are those of the authors and not those of the Economic and Social Research Institute, the Cabinet Office, or the Government of Japan.

医療費の包括支払制度が医療の質に及ぼす影響 —マッチング推定による検証—

岡崎康平、杉原茂、田倉智之、村舘靖之¹

要約

医療費の包括払制度は、医療サービス提供の効率化を促進すると期待される一方で、医療の質が劣化するのではないかと懸念が存在している。本研究では、レセプトの大規模データベースを利用して、診断群分類包括評価

(DPC)が医療の質に対して及ぼした影響を検証する。その際に、DPCに参加する病院が元々医療の質が高い可能性があるというバイアスを調整するために、病院の詳細な属性に基づいたマッチング推定を行う。傾向スコアの推定が観測される要因のみに基づくという制約を緩和するために、DPC 選択関数の説明変数の選択に当たっては、病院についての豊富な情報を持つデータベースを利用した。分析対象としては、新生物、呼吸器系、消化器系、循環器系、及びその合計とした。

4疾患合計についての結果は、DPC への参加は死亡率を1%程度低下させることが確認された。これは、大きな改善効果と言える。また、バイアスを調整しない場合にはこの倍程度と過大評価になっており、マッチング推定の有効性が示された。個別の疾患については、DPC 参加の効果が大きい疾患と小さい疾患のばらつきがあったが、やはりすべての疾患でDPC 参加は医療の質を向上させることが確認された。さらに、DPC 選択が観測されない要因によって影響を受ける可能性を考慮した頑健性のチェックを行ったが、マッチング推定には内生性の影響は残るものの、それがマッチング推定の妥当性を失わせるようなものではないことが確認された。

¹ 岡崎康平：野村證券、杉原茂：日本大学、田倉智之：東京大学、村舘靖之：内閣府経済社会総合研究所。本稿を執筆するにあたっては、原湖楠研究員（アリゾナ大学）より有益なアドバイスを頂いた。あらためて感謝を申し上げる。なお、本稿中の見解、分析等はあくまで筆者個人のものであり、経済社会総合研究所のものではない。また、データや分析における誤り等の責任はすべて筆者に帰するものである。

目次

1. はじめに.....	4
2. 分析の枠組み.....	7
3. データの記述統計的分析.....	13
4. 推定結果.....	16
5. 終わりに.....	31
参考文献.....	32

1. はじめに

医療提供の効率化の手段として包括化が検討・実施されて久しい。その一方で、医療の質が劣化するのではないかという懸念が存在している。診断群分類包括評価（Diagnostic-Procedure Combination：DPC）が2003年以降段階的に導入されてかなりの期間が経った。データも相当に蓄積されてきており、その医療の質に対する効果を検証することは時宜に適ったことと言える。

DPCの医療の質に対する効果の検証は、当然ながら、DPC参加病院と非参加病院の比較が基本である。しかし、医療の質が高い病院はDPCに参加する確率が高いという自己選択に起因するバイアスが存在する可能性がある。例えば、病院経営がしっかりしている病院は、医療の質が高い一方で、効率化の効果を持つDPCに積極的に参加することがあり得る。このように、DPCの影響を計測する際に、異質の病院を比較しているため、正確な計測ができないことになる。

そこで、同質の病院を比較するために、こうした自己選択バイアスを調整した上で医療の質を比較する必要がある。本論文では、傾向スコアを使ったマッチング法を利用して、自己選択バイアスの調整を試みる。すなわち、傾向スコアを使って等質なサンプルを構成し、そうしたサンプルを使って死亡率関数を推定することによってDPCの影響を正確に計測することとする。傾向スコアとは、ある政策（本論文の場合にはDPC）を適用される確率であるが、この傾向スコアすなわちDPCに参加する確率が近い病院を比較可能なものとみなして、傾向スコアが同じ病院のうち、DPCに参加している病院の死亡率と非参加の病院の死亡率とを比較する。傾向スコア（DPCに参加する確率）は観測できる変数を使って推定するが、自己選択バイアスを適切に調整するには、変数が外生的でかつ参加の有無を良く説明するものでなければならない。本論文では、傾向スコアの

推定に当たって、参加決定についての情報を豊富に含むデータを使って精度の高い調整を行うことにする。具体的には、詳細な病院属性についてのデータベースを利用して、病院規模などの DPC 参加の前提条件に加え、経営姿勢、診療科の技術、各種診療科の有無など詳細な病院属性を使って病院の DPC 参加の意思決定をモデル化し傾向スコアを推定する。

傾向スコアによるマッチングは、観測できる変数によるコントロールであるため、DPC 参加の意思決定が観測できる変数のみによってモデル化できる場合にバイアスは適切に除去できる。上記の傾向スコアによる分析は、精緻な DPC 選択関数の推定によりこの条件を充たそうとするものであるが、観測できない固有効果などが DPC 参加に影響を与えていて傾向スコアではバイアスを除去できないのではないかという懸念を抱く向きもあろう。そこで、頑健性を検討するため、観測できない固有効果を死亡率関数に取り込んで推定も行う。

ここで、包括支払制度と医療の質との関係についての既存研究を概観しておこう。包括支払制度が医療の質やコスト、効率性に与える影響については、アメリカにおいて 1983 年に導入された DRG/PPS (Diagnostic-Related Group/Prospective Payment System) の影響が RAND 研究所の研究者達の一連の研究によって詳細に分析された。Rogers, et al. (1990) の総括的な評価によると、次のような結論が得られたとまとめられている。(1) 死亡率は PPS の導入によって影響を受けず、また、導入前から始まっていた治療プロセスの改善は導入後も継続した。(2) PPS 導入後には、患者が不安定な状態で退院する確率が高まった。

さらに詳しく個々の研究を見てみると、Kahn, et al. (1990a) によると、採り上げた 5 つの疾病について、在院日数は PPS 導入前に比べて 24%減少し、院内死亡率は導入前の 16.1%から 12.6%へ低下した（入院時の重症度を調整した 30 日死亡率は 1%ポイントの低下、180 日死亡率はほぼ変化なし）。しかし、Kosecoff, et al. (1990) によると、自宅へ退院した人のうち不安定な状態だった人の割合は、PPS 導入前には 10%だったのが、導入後には 15%に上昇した。Kosecoff, et al. (1990) は不安定な状態での退院は退院後の死亡率の上昇につながるとしているが（不安定な状態で退院した人の 16%が退院後 90 日までに死亡するが、安定的な状態で退院した人については 11%）、Cutler (1995) の分析においては、PPS により退院後の死亡率が上昇する効果は検出されなかった。なお、アメリカの DRG/PPS では 1 入院に対して包括支払いであるため、入院日数を削減するインセンティブが強い構造となっていることが、在院日数の削減に大きな影響を与えている。一方、日本の DPC では一定の入院期間ごとに 1 日当たりの包括支払い額が設定されており、日にち当たりの支払いが減るところ（例、入院期間 I から入院期間 II への切り替わり点）までは入院を引き伸ばすインセンティブもあり得るため、不安定な状態の患者を早期退院させることは少ないと考えられる。

なお、医療の質を計測する手法として、RAND 研究所の研究者達は、明示的プロセス基準 (Explicit Process Criteria) と陰伏的レビュー (Implicit Review) という手法を編み出した。明示的プロセス基準とは、医学文献に基づいて適切と考えられる治療プロセスを患者に適用しているかどうかを医療の質の指標とするもので、Kahn, et al. (1990b) によると、PPS 導入後の治療プロセスの改善により、30 日死亡率が 1%ポイント低下した。また、陰伏的レビューとは、患者に対する治療内容が医学文献に基づいて適切と考えられるかどうかをカルテから医師が判断するもので、Rubenstein, et al. (1990) によると、PPS 導入後は導入前に比べて医療の質は改善した (劣悪ないし非常に劣悪な治療を受けた患者の割合が、導入前には 25%だったものが、導入後には 12%へ減少した)。

Newhouse (2004) は、DRG/PPS 導入についての研究を総括して、在院日数の減少によりコストが削減される一方で、健康への悪影響は軽微なものであり、医療提供の効率性が向上したとした。同様に、日本における DPC の導入が効率性を改善させたかどうかについて、野口等 (2010) は、DPC 導入により病院のマネジメントが経年的に改善する効果を DPC 病院としての経験年数 (DPC に参加した年数) として把握し、疾病や患者の重症度 (Charlson Index 等) などをコントロールした上で、在院日数や医療費などの医療資源消費量が DPC 経験年数によって低下するかどうかを検証した。結果は、DPC 経験年数が長いほど医療資源の消費量が少ない傾向があるということが示された。すなわち、DPC 病院として活動することにより効率性が向上することが示された。なお、分析対象とした患者は腹腔鏡下での手術であるが、こうした術式の選択にはセレクション・バイアスが存在する。このバイアスを調整するために、患者が入院した曜日を操作変数とした推定を行っている。野口等 (2010) は、同論文の最後の考察のところで自身で述べているように、DPC 参加以降の経験による効率性の向上を対象としており、DPC に参加すること自体により効率化が促進される効果は含まれていない (DPC 参加病院だけのデータではそれは無理である)。我々の論文は、後者の効果を含めて包括支払制度が医療の効率性に与える影響を計測するものであり、その際に、DPC に参加するかどうかを病院が選択することで生じてしまうセレクション・バイアスを傾向スコアを使って調整して、厳密な形で DPC の効果を計測しようとするものである。

本論文の構成は以下のとおりである。第 2 節において分析の枠組みを整理し、第 3 節において使用したデータの説明と記述統計量を示す。第 4 節で推定結果を提示し、第 5 節で結論をまとめる。

2. 分析の枠組み

傾向スコアによる分析は、同じものを比べるということである (Imbens and Rubin, 2015)。例えば、病床数の多い病院ほど死亡率が低く、かつ、DPC へも参加しやすい場合を考えてみよう。DPC が医療の質に与える影響を調べるには、DPC に参加している病院の死亡率と参加していない病院の死亡率を比較することが基本であることに変わりはない。しかし、DPC 参加病院には死亡率が低い病院が多く集まることになり、DPC 参加病院と非参加病院の死亡率を単純に比べることは、DPC 参加の効果を実際に捉えることにはならない。何故なら、もともと死亡率が低い傾向がある DPC 参加病院と死亡率が高い傾向にある非参加病院を比べることになるからである。そこで、例えば病床規模が大きい病院ほど死亡率が低く DPC に参加する傾向も強い場合、病床規模が同じ病院について、そうした病院の中で DPC に参加している病院の死亡率と参加していない病院の死亡率を比較するというのがリーズナブルである。DPC 選択の意思決定に多数の要因が影響を与える場合、多くの要因のすべてについて同じ病院を採ってくることは困難である。そのため、様々な要因により DPC 参加する確率が決まるが、その参加確率が同じ病院について比較する。これが傾向スコア分析である。

本論文における傾向スコアの推定の特徴の一つは、病院について詳細な情報を収集し、DPC 参加の意思決定に影響する様々な側面についての多くの要因を使って DPC の選択関数を推定することである。これにより、DPC への参加病院と非参加病院の死亡率を同じとみなせる病院について比較することが可能となる。傾向スコアが同じ病院をサンプルとして死亡率関数を推定し、その説明変数として DPC 参加の有無を表わすダミー変数を含めることにより、DPC が病院の医療の質へ及ぼす影響を計測することができる。このように病院属性の影響によるバイアスの調整は、単に調整変数を多く説明変数とした多重回帰分析によっても行うことはできる。しかし、傾向スコア分析を利用すると、**common support condition** を強制することになり、例えば、一方で、明らかに DPC に参加するような超大規模病院・大学病院等を、他方で、明らかに DPC に参加しない超小規模病院等を分析から除外することとなり、明らかに比較の対象とならない病院を分析から除外してより同質性が担保されるようにすることができる。

医療の質としては、院内死亡率を採る。DPC に参加する確率が同じ程度の病院について、DPC 参加病院と非参加病院との死亡率を比較することになる。これは当然、推計期間中に DPC 参加病院と非参加病院で院内死亡率に異なる影響を与えるような動きはなかったという仮定に基づく比較であるが、傾向スコアにより同質性を担保しているのこうした条件は充されていると考えられる。また、医療の質としては、院内死亡率よりは 30 日死亡率を採る方が医学研究ではスタンダードであろうが、退院後に 30 日以内に死亡したかどうかをフォロー

することは、脱落が生じやすく、本論文のようなビッグ・データではデータ上の制約が大きい。例えば、データベースに含まれていない病院・医院等で死亡したかどうかを個別にフォローしたり死亡診断書のデータを使って補完したりすることは困難である。このように本論文の死亡率の把握には限界があるが、概念的に適切な30日死亡率を追求することが実際にはバイアスを拡大してしまうことになりかねないことを考慮しての選択である。

さて、やや回り道であるが、セレクションによるバイアスの問題の所在と対応をより分かり易くするために、通常の計量経済学的な形に書いてみよう。死亡率の決定とDPC参加の決定の2つの意思決定が行われている。これを次のような連立方程式で表わすことができる。一つ目の式は医療の質方程式で、本論文では医療の質を死亡率によって測ることとするので要は死亡率関数である。これは、患者が死亡した時は1、生存の時は0の値をとるダミー変数を被説明変数とするロジスティック回帰により推定する。もう一つの式はDPCに参加するかどうかを選択する方程式で、参加した場合には1、非参加の場合には0の値をとるダミー変数を被説明変数とするロジスティック回帰により推定する。推定式の具体的な形は、次のようになる。

$$y_{ijt} = 1[\alpha + \sum_{k=1}^K \beta_k \cdot x_{ijkt} + \gamma \cdot r_{ijt} + u_{ijt}] \quad (1)$$

$$r_{ijt} = 1[\kappa + \sum_{k=1}^K \varphi_k \cdot x_{ijkt} + \delta \cdot z_{jt} + v_{ijt}] \quad (2)$$

ここで、 $1[Z]$ は定義関数 (Indicator function) と呼ばれるもので、 $Z > 0$ の時に1、 $Z < 0$ の時に0という値をとる関数である。 y_{ijt} は個人*i*が病院*j*において*t*年に治療を受けた時の転帰で、患者が死亡した時は1、生存の時は0の値をとるダミー変数である。 r_{ijt} は個人*i*が治療を受けた病院*j*が*t*年においてDPCに参加しているかどうかを表わす変数で、当該病院がDPCに参加していれば1、参加していなければ0の値をとるダミー変数である。 x_{ijkt} は患者のリスク・ファクター、 u_{ijt} 及び v_{ijt} はかく乱項である。選択方程式(2)の右辺には z_{jt} というリスク・ファクター以外の選択に関係する変数が含まれる。

(2)式を(1)式に代入してみると容易に分かるように、もしかく乱項 u_{ijt} 及び v_{ijt} が相関を持てば、(1)式の説明変数 r_{ijt} とかく乱項 u_{ijt} が相関を持つため、(1)式を単独で推定するとバイアスが発生し、DPC参加が医療の質に与える影響を適切に推定することができない。死亡率及びDPC参加の選択には、共通する観測されない要因が影響を与えていると考えられる。代表的なものは、病院のステーク・ホルダーの意向やマネジメントの整備状況などが考えられる。医療の質や効率性に対する意識が高い医師や看護師あるいは経営陣などは、質を高めようと努力するとともに、DPCへの参加に積極的であることは自然な動きである。

また、医療提供のマネジメントがしっかりしている病院ほど、質が高いとともに DPC 参加へのハードルは低い。これらの要因は観測されない固有効果としてかく乱項 u_{ijt} 及び v_{ijt} に含まれ、したがってこの 2 つのかく乱項は相関を持つと想定される。この場合、先に指摘したように、単純に死亡率関数を推定するだけでは DPC の効果が正確に計測できない。

計量経済学において上記のような問題に対処するための手法としては、操作変数法、パネル分析あるいは差分の差分法、マッチング法、回帰不連続分析など様々なものがある²。本論文では、傾向スコアを使って等質なサンプルを構成し、そうしたサンプルで死亡率関数を推定することによって DPC の影響を正確に計測することとする。この意味付けについては、欠落変数 (Missing Data) の文献とのアナロジーを考えるとより明瞭になると思われる。変数の欠落とはデータがサンプルから脱落することであるが、脱落のメカニズムとして大きく 3 つのカテゴリーに分類されている。一つは **Missing completely at random (MCAR)** で、これは欠落するかしないかが結果変数に依存しない場合を指す。二つ目は **Missing at random (MAR)** で、これは欠落するかしないかが観測された結果変数に依存するが欠落した結果変数に依存しない場合を指す。三つ目は **Missing not at random (MNAR)** で、これは欠落するかしないかが欠落した結果変数に依存する場合である。Daniels and Hogan (2008) はセレクション・モデルの文脈でこの 3 つを分かり易く定式化している。セレクション・モデルは、結果変数 y と欠落の指示関数 (Missing data indicator) r との同時確率分布を以下のように分解するものである (x は説明変数、 w は分布の形のパラメーター、 θ 及び ψ も w に依存して分布をインデックスするパラメーター)。

$$p(y, r|x, w) = p(y|x, \theta(w))p(r|y, x, \psi(w)) \quad (3)$$

欠落メカニズムは(3)式の右辺の第 2 項であるが、これは次のように一般的には観測される結果変数 y_{obs} と欠落した結果変数 y_{mis} の両方に依存している。

$$p(r|y, x, w) = p(r|y_{obs}, y_{mis}, x, w) \quad (4)$$

MCAR、MAR 及び MNAR は、それぞれ、次の(5)から(7)式に対応している。

$$\text{MCAR :} \quad p(r|y, x, \psi) = p(r|x, \psi) \quad (5)$$

² 計量経済学の領域における教科書的な解説としては、Angrist and Pischke (2015)や Imbens and Rubin (2015)、田中(2015)などを参照。

$$\text{MAR : } \quad p(r|y, x, \psi) = p(r|y_{obs}, x, \psi) \quad (6)$$

$$\text{MNAR : } \quad p(r|y_{obs}, y_{mis}, x, \psi) \neq p(r|y_{obs}, y'_{mis}, x, \psi) \quad (7)$$

Missing completely at random (MCAR) は欠落するかどうかの結果と全く関係なく完全にランダムであるため、欠落メカニズムを考慮することなく結果方程式 ((3)式の右辺第1項) を推定することができる。Missing at random (MAR) は欠落メカニズムが観測された結果変数に依存するが、観測されたデータだけを使って妥当な推定を行うことが可能である。欠落変数の文献における用語を用いれば、MAR は *Ingorability* と呼ばれる条件の一要素であり (Little and Rubin, 2019)³、*Ingorability* の下ではパラメーター θ に関する推論は観測されたデータに基づいて適切に行うことができる(欠落メカニズムは無視することができる)⁴。

基礎的な説明が長くなってしまったが、欠落変数の文献を上記の計量経済学的モデル(1)(2)に引き直して言えば、DPC 選択関数が欠落した結果変数(死亡したか否か)に依存しないという MAR の条件は、選択関数と死亡率関数のかく乱項同士が相関を持たないということになる。本論文において詳細な病院情報を使って傾向スコアを推定するという趣旨は、選択関数推定において十分な説明変数を使うことにより、DPC 選択が説明し尽くされ、選択が「at random」に行われることである。もう少し説明しよう。かく乱項だけを取り出して、以下のように分解してみる。 c_j^y は死亡率に関する病院 j の固有效果、 c_j^r は DPC 参加選択に関する病院 j の固有效果、 u'_{ijt} 及び v'_{ijt} は相互に相関を持たない2つのかく乱項である。

$$u_{ijt} = c_j^y + u'_{ijt} \quad (8)$$

$$v_{ijt} = c_j^r + v'_{ijt} \quad (9)$$

元々のかく乱項 u_{ijt} と v_{ijt} が相関を持ったのは、固有效果 c_j^y と c_j^r が相関を持ったためである。元々のかく乱項 u_{ijt} と v_{ijt} が相関を持たなくなる一つのケースは、固有效果 c_j^r が説明変数 z_{jt} により説明し尽くされる場合 (Wooldridge, 2002)。ここで「説

³ MAR に次の2つの条件が加わった時に *Ignorable* と呼ばれる (x は落としてある) : (i) w は、フルデータ・モデル $p(y|\theta)$ をインデックスするパラメーター θ と $p(r|y, \psi)$ をインデックスするパラメーター ψ とに分解される : $w = (\theta, \psi)$ 。(ii) θ と ψ は独立 : $p(\theta, \psi) = p(\theta)p(\psi)$ 。

⁴ 少し細かく言えば、パラメーター θ に関する推論は、観測された結果モデル : $p(y_{obs}|w) = \int p(y_{obs}, y_{mis}|\theta(w))dy_{mis}$ に比例する尤度関数に基づいて行うことができる (Little and Rubin, 2019 及び Daniels and Hogan, 2008)。なお、Molenberger and Kenward (2007) はこうした *ignorability* の想定に基づく最尤推定を前面に押し出している。

明し尽くされる」というのは、外生変数 z_{jt} （患者のリスク・ファクターを含んでいても良い）に条件付けた時、 c_j^r の残差 v''_{ijt} が医療の質方程式の残差と相関を持たないという趣旨である。

$$c_j^r = \lambda + \mu \cdot z_{jt} + v''_{ijt} \quad (10)$$

ここで、 v''_{ijt} は u'_{ijt} とも c_j^y とも相関を持たないかく乱項である。固有効果 c_j^r がこのようにモデル化されれば、(1)式の説明変数 r_{ijt} とかく乱項 u_{ijt} が相関を持たず、 r_{ijt} の因果関係としての効果が正確に計測できる。まとめると、DPC 選択の要因を詳細に把握することにより、MAR を実現しようとしており、この場合、死亡率関数(1)を単独に推定しても正確な推定ができる。

ただし、欠落データの文献のアナロジーを続けると、MNAR (Missing Not At Random) という可能性は先験的には排除しにくい。詳細な説明変数を使っても選択方程式のかく乱項に固有効果 c_j^r が残ってしまう可能性はある。この場合、傾向スコアによるマッチングは機能しない。傾向スコア分析の頑健性をチェックするために、選択方程式(2)の残差 \widehat{r}_{ijt} を死亡率関数(1)の説明変数に加える形で死亡率関数を推定する。このような操作を行うことの正当性は、選択方程式(2)が線型である場合には、非常に簡明に理解できる。仮に選択方程式(2)が次のような形をしているとする。

$$r_{ijt} = \kappa + \sum_{k=1}^K \varphi_k \cdot x_{ijkt} + \delta \cdot z_{jt} + c_j^r + v'_{ijt} \quad (2')$$

この時、残差 \widehat{r}_{ijt} は、推定値をハット (^) で表わすと、

$$\begin{aligned} \widehat{r}_{ijt} &= r_{ijt} - (\hat{\kappa} + \sum_{k=1}^K \widehat{\varphi}_k \cdot x_{ijkt} + \hat{\delta} \cdot z_{jt}) \\ &= (\kappa - \hat{\kappa}) + \sum_{k=1}^K (\varphi_k - \widehat{\varphi}_k) \cdot x_{ijkt} + (\delta - \hat{\delta}) \cdot z_{jt} + c_j^r + v'_{ijt} \end{aligned} \quad (11)$$

この式を死亡率関数(1)に代入すると、(1)式の説明変数に病院の固有効果 c_j^r がデータとして含まれることになり、(1)式のかく乱項には固有効果 c_j^r が含まれないことになる。したがって、(1)式のかく乱項と説明変数 r_{ijt} との間に相関がなくなり、DPC の効果 γ が正確に計測されることになる。

これを選択方程式(2)が非線型の場合に拡張する際に、Wooldridge (2005)の方法を参考としよう。Wooldridge は、結果方程式（本論文の場合には死亡率関数）が被説明変数 y_{it} のラグ $y_{i,t-1}$ を説明変数として含むダイナミック・パネル・データ・モデルを考えた。この場合、説明変数（ラグ付き被説明変数）と被説明変数のかく乱項はともに観測されない異質性 c_i を含むが、このために説明変数と被説明変

数のかく乱項の間に相関が存在してバイアスが生じる。この問題に対処するために Wooldridge (2005)は、観測されない異質性を被説明変数の初期値 y_{i0} 及び他の説明変数を各経済主体 i について時間を通じて平均した $z_i = \frac{1}{n_i} \sum_{t=1}^{n_i} z_{it}$ (n_i は i についてのサンプル数) の関数として定式化することを提案した。

$$y_{it} = 1[z_{it} \cdot \beta + \rho \cdot y_{i,t-1} + c_i + u_{it}] \quad (12)$$

$$c_i = \alpha_0 + \alpha_1 \cdot y_{i0} + z_i \cdot \alpha_2 + v_{it} \quad (13)$$

(13)式を(12)式に代入すると、

$$y_{it} = 1[z_{it} \cdot \beta + \rho \cdot y_{i,t-1} + \alpha_0 + \alpha_1 \cdot y_{i0} + z_i \cdot \alpha_2 + u_{it} + v_{it}] \quad (14)$$

(14)式のかく乱項からは観測されない異質性 c_i が取り除かれて自己ラグ $y_{i,t-1}$ の係数はバイアスなく推定される。

Wooldridge (2005)の推定手続きは、直接的には観測されない異質性 c_i を観測できる初期値 y_{i0} で置き換えるということであるが、本質的な部分は、観測されない異質性を含む変数 y_{i0} を結果方程式(7)に代入することにより、観測されない異質性をデータとして結果方程式に含め、それによってかく乱項と説明変数との間の相関を除去しようとするものである。これは意表を突くほど簡便なやり方であるが、上記の選択関数が線型の場合と同様に(2)式の残差を結果方程式(本論文の場合には死亡率関数)に加えるという方法と軌を一にしたものである。

ただ、Wooldridge (2005)は我々の場合にはそのまま適用することができない。何故ならば、死亡率関数のサンプルとなるためには最初は生存していなければならないので、死亡率関数の被説明変数の初期値は必ず0であるためである。それではアナロジーとして選択関数の初期値を加えるということを考えてみても、サンプルに含まれるためには選択関数の初期値は必ず1である。そこで、Wooldridge (2005)の方法の本質的部分、すなわち観測されない異質性を含む変数を死亡率方程式に代入するということを実現するために、選択関数の残差 \widehat{r}_{ijt} を死亡率関数の説明変数に加えることを考える。選択関数の残差 \widehat{r}_{ijt} は、推定値をハット (^) で表わすと、

$$\begin{aligned} \widehat{r}_{ijt} &= r_{ijt} - 1[\hat{\kappa} + \sum_{k=1}^K \widehat{\varphi}_k \cdot x_{ijkt} + \hat{\delta} \cdot z_{jt}] \\ &= 1[\kappa + \sum_{k=1}^K \varphi_k \cdot x_{ijkt} + \delta \cdot z_{jt} + c_j^r + v'_{ijt}] \\ &\quad - 1[\hat{\kappa} + \sum_{k=1}^K \widehat{\varphi}_k \cdot x_{ijkt} + \hat{\delta} \cdot z_{jt}] \end{aligned} \quad (15)$$

選択方程式は非線型であり(11)式のような固有効果が線型で入る簡明な形にならないが⁵、観測されない異質性 c_i を非線型の形ではあっても含むことになり、Wooldridge (2005)の方法の本質的部分である観測されない異質性を含む変数を死亡率方程式に代入するというを体現する変数として、死亡率関数に含めて推定することとする。

3. データの記述統計的分析

データは、東京大学医学系研究科田倉研究室がメンテナンスしている東京大学医療経済ビッグ・データを利用した。データベースに含まれるレセプトの期間は2012年4月～2018年6月であるが、後述するように病院属性データとの接合した後のサンプル数の問題などから、実際の分析に使用したのは2013年度から2017年度までのデータとした。

傾向スコアの推定に利用する病院属性データは、医事日報社『病院情報』の各年版から採った。同資料は地域別に刊行されているが、地域によっては刊行されていない年もある。そうした場合、直近の年のデータで補完した。

レセプト・データにおける医療機関と病院属性データの医療機関との接続に当たっては、付与されている医療機関コードが田倉研究室データベースと『病院情報』では異なるため⁶、同コードによる接続はできなかった。そこで、接続のキーとしては電話番号を利用し、電話番号で接続できなかった場合には病院の名称・住所・統廃合情報から手作業で接続するなどした。

レセプト・データと病院属性を接合する場合、レセプトの発行時点と、接続相手の病院データの時点をどのように対応させるかが問題となる。いくつかの方法が考えられるが、簡便に行うのであれば (i) そのレセプトが発行された年を用い、その年の病院情報シリーズのデータを対応させることが考えられる。この方法は、データハンドリングの観点からは簡便であるが、正確性は一步劣らざるを得ない。特に、『病院情報』は各年において元データが存在しない地域が存在しているため、当該地域に関してはデータの対応がより不正確になりうる。そこで、(ii) レセプト発行月から最も近い時点での病院情報シリーズの情報を紐づけ

⁵ 結果変数が1という値をとる確率を p として、ロジット方程式を次のように書く：

$$\log\left(\frac{p}{1-p}\right) = X\beta + c. \text{ これを } p \text{ について解くと、 } p = \frac{\exp(X\beta+c)}{1+\exp(X\beta+c)}.$$

$$\text{残差は、 } p - \hat{p} = \frac{\exp(X\beta+c)}{1+\exp(X\beta+c)} - \frac{\exp(X\hat{\beta})}{1+\exp(X\hat{\beta})} = \frac{\exp(X\beta+c) - \exp(X\hat{\beta})}{\{1+\exp(X\beta+c)\}\{1+\exp(X\hat{\beta})\}}.$$

⁶ 田倉データベースは厚生労省の医療機関コード、『病院情報』は独自の医療機関コードである。

ることが考えられる。この場合には、レセプト発行時点からの時間的距離が単純に最も短い病院情報シリーズを利用するか、レセプト発行前（あるいは後）で最も時間的に近接する病院情報シリーズのデータを用いるか、という点を考える余地がある。本論文では、（前後を問わず）時間的に最も近接した時点の病院関連データを利用することとする。

分析対象とした疾患は、新生物、呼吸器系、消化器系、循環器系であり、分析はこれら疾患をまとめた場合と個別の疾患ごとの場合の二通り行った。また、分析対象は入院レセプトのみとした。次節で説明する傾向スコアの分析の結果傾向スコアが 0.6~0.8 の医療機関のみをサンプルに含めた。すべての疾患（4疾患以外の疾患を含む）を合わせたサンプル数（レセプトの数）は、106万3126枚である（第1表）。年度別のレセプト枚数は、2013年度15万5752枚、2014年度19万8308枚、2015年度23万6623枚、2016年度24万3759枚、2017年度22万8684枚である。4疾患については、全年度の合計で、新生物が17万3162枚、呼吸器系が9万2692枚、消化器系が11万2888枚、循環器系が7万8513枚である。

第1表 レセプト数（枚数）

	全年度	2013	2014	2015	2016	2017
全疾病	1,063,126	155,752	198,308	236,623	243,759	228,684
（うち）DPC	671,398	99,374	123,032	146,328	156,153	146,511
新生物	173,162	25,285	32,217	38,257	39,883	37,520
（うち）DPC	134,685	20,211	24,630	28,564	31,616	29,664
呼吸器	92,692	12,817	16,839	21,396	21,675	19,965
（うち）DPC	75,442	10,042	13,510	17,217	18,008	16,665
消化器	112,888	16,962	21,589	25,198	25,422	23,717
（うち）DPC	61,481	10,721	11,269	13,136	13,664	12,691
循環器	78,513	10,828	14,516	17,244	18,511	17,414
（うち）DPC	53,641	7,320	9,826	11,552	12,825	12,118

出所）東京大学医療経済ビッグデータより作成

第2表はすべての疾患（4疾患以外の疾患を含む）について患者属性をまとめたものである。全年度のサンプルについてみると、年齢別の構成比は、19歳まで、及び、20歳から39歳がそれぞれ4分の1、40歳から59歳までが37%、60歳以上がそれ以外である。女性の割合は53%、個別の疾患に罹患している割合は、新生物、呼吸器系、消化器系、循環器系とも10%前後である。重症度と関係する指標を見ると、cci（チャールソン併存疾患指数）有りの割合が18%、再入院有りの割合が17%である。また、レセプト種別を見ると、公費併用の割合が14%、低所得の割合が2%である。各年度ごとに見ても、以上の患者属性はほとんど同じである。

第3表は、分析の対象となった病院の属性である。特定機能病院を取得した割合が13%、DPCに参加対象病院の割合が68%である。DPC参加基準に関連した指標を見ると、ICUを有している病院の割合が58%、CCUは33%、HCUは26%である。

第2表 患者属性の記述統計量*

患者属性	全年度	2013	2014	2015	2016	2017
年齢_19歳まで	0.24	0.23	0.23	0.24	0.24	0.24
年齢_20歳から39歳	0.25	0.25	0.25	0.25	0.26	0.25
年齢_40歳から59歳	0.37	0.40	0.38	0.37	0.36	0.36
性別(女性)	0.53	0.53	0.53	0.53	0.53	0.54
picking(以下の疾病を抽出)						
C(新生物)	0.11	0.11	0.11	0.11	0.11	0.11
D	0.05	0.05	0.05	0.05	0.05	0.05
J(呼吸器系)	0.09	0.08	0.09	0.09	0.09	0.09
K(消化器系)	0.11	0.11	0.11	0.11	0.10	0.10
I(循環器系)	0.08	0.07	0.07	0.07	0.08	0.08
cci(チャールソン併存疾患指数)有り	0.18	0.17	0.18	0.17	0.18	0.18
再入院有り	0.17	0.13	0.16	0.17	0.18	0.19
レセプト種別						
公費併用	0.14	0.12	0.12	0.13	0.15	0.16
低所得	0.02	0.02	0.03	0.02	0.02	0.02

*これらの項目は割合である
出所) 東京大学医療経済ビッグデータより作成

第3表 病院属性の記述統計量*

病院属性	全年度	2013	2014	2015	2016	2017
開設主体(公営/私立)						
特定機能病院の取得	0.13	0.13	0.13	0.13	0.13	0.13
DPCに参加対象病院の割合	0.68	0.64	0.67	0.68	0.70	0.70
DPC参加基準						
ICU有無	0.58	0.59	0.58	0.58	0.57	0.32
CCU有無	0.33	0.34	0.34	0.33	0.32	0.32
HCU有無	0.26	0.26	0.26	0.26	0.26	0.26

*これらの項目は割合である
出所) 東京大学医療経済ビッグデータより作成

第4表は死亡率をまとめたものである。全疾患(4疾患以外の疾患を含む)の全年度の死亡率は0.74%であり、うち、DPC参加病院では0.54%、DPC非参加病院では1.09%である。DPC参加病院とDPC非参加病院の死亡率の差は0.56%

ポイントである。年度の動きを見ると、全病院の死亡率は 2013 年の 0.8% から 2017 年の 0.68% へ緩やかに低下している。このような緩やかな低下は DPC 参加病院と DPC 非参加病院に共通に見られるが、DPC 参加病院と DPC 非参加病院の死亡率の差は 2013 年の 0.58%ポイントから 2017 年には 0.54%ポイントへ若干縮小したが、ほぼ一定である。

個別の疾患の死亡率を見ると、新生物と循環器系で高く、呼吸器系と消化器系で低い。消化器系を除いて、年とともに死亡率が緩やかに低下しているのは全疾患と同じである。DPC 参加病院と DPC 非参加病院の死亡率の差は、年とともに、新生物と循環器系ではほとんど変わらず、呼吸器系では差が縮小、消化器系では逆転している。

第 4 表 死亡率(%): DPC 参加・非参加

	全年度	2013	2014	2015	2016	2017
全疾病						
全病院	0.74	0.80	0.81	0.74	0.70	0.68
DPC参加病院	0.54	0.60	0.58	0.53	0.53	0.48
DPC非参加病院	1.09	1.17	1.20	1.10	1.02	1.02
新生物						
全病院	2.20	2.32	2.42	2.21	2.10	2.04
DPC参加病院	1.64	1.76	1.80	1.61	1.63	1.46
DPC非参加病院	4.13	4.48	4.39	3.93	3.83	4.20
呼吸器						
全病院	0.42	0.52	0.48	0.39	0.40	0.34
DPC参加病院	0.24	0.24	0.27	0.24	0.25	0.21
DPC非参加病院	1.17	1.53	1.33	0.97	1.13	1.02
消化器						
全病院	0.29	0.29	0.34	0.29	0.27	0.28
DPC参加病院	0.26	0.24	0.29	0.21	0.28	0.31
DPC非参加病院	0.31	0.37	0.40	0.30	0.25	0.25
循環器						
全病院	2.06	2.16	2.18	2.17	1.96	1.89
DPC参加病院	0.87	0.97	0.86	0.93	0.81	0.80
DPC非参加病院	4.53	4.54	4.83	4.58	4.47	4.28

出所) 東京大学医療経済ビッグデータより作成

4. 推定結果

まず傾向スコアを計算する。選択関数は DPC 参加しているかどうかを被説明変数としてロジスティック回帰を行った。推定式は次のとおりである。

$$\ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (16)$$

ここで、 Y は DPC 参加しているかどうかを表すであり、 $p(Y = 1)$ は DPC 参加病院である確率となる。

変数は基本となる変数 5 つを固定し、その他は選択確率の推定精度が高まるような変数を選択した。こうした変数選択は、変数の有意性を検証する場合には不適切であるが、結果 (DPC に参加するか否か) を予測するためには、正当な手続きである。実際、Chernozhukov, et al. (2018) などにおいて、第一段階の推定は機械学習で行い、興味のあるパラメーターについては第二段階の推定で因果推論的な検証をするというフレームワークが近年盛んに議論されている (本論文に引き写せば、第一段階は傾向スコアをランダムフォレストで推定し、第二段階で包括支払制度の死亡率に与える影響を因果関係として推定することに対応する)。Chernozhukov, et al. (2018) の Example 1.1 を例として見ると、政策 D 及びその他のコントロール変数 X が結果 Y に及ぼす影響を推定するために、第一段階で auxiliary sample を使って政策 D の X を与えた時の平均関数を計算し、それを使って政策 D から X の影響を partial out する orthogonalization を行い、第二段階で main sample を使って orthogonalization に基づく debiased ML estimator を構成し、さらに、overfitting による効率性の欠如を補填するための cross-fitting を提唱している。本論文の Propensity score の推定 + マッチングのケースは Chernozhukov 等のフレームワークにそのまま該当するものではないが (したがって、本論文では orthogonalization/debiased estimator は採用しない)、そこでの基本的考え方は第一段階のランダムフォレスト + 第二段階の政策効果の因果関係の検証という手続きの妥当性を裏付けるものと言える。

傾向スコアを推定する際の基本となる変数として、病床数、医師一人当たりベッド数、ICU 有無、CCU 有無、HCU 有無を採用した。これらは DPC に参加するかどうかを決定する基本的な属性と考えられるためである。その理由は以下のとおり。DPC 対象病院となるためには、急性期一般入院基本料や特定機能病院等の 7 対 1・10 対 1 入院基本料の届出をしていることが必要である。特に、特定機能病院についてみると、特定機能病院に指定されるためには、400 床以上の病床を有すること、人員配置として医師は最低でも通常の 2 倍程度配置されていること、構造設備として集中治療室などを備えていることなどが必要である。上記の 5 変数は、こうした DPC 承認の要件を代表したものである。

その他の変数選択のプロセスは次のようにした。(i) 『病院情報』に含まれる病院属性のうち 195 個の logical 変数 (0 か 1 を取るタイプの変数) を説明変数の候補とする。(ii) 先に挙げた基本 5 変数を先験的に説明変数として固定した上で、195 個の変数から 20 個をランダムに抜き出し、合計 25 個の説明変数 (+ 定数項) で DCP 対象病院か否かを予測する選択関数を構成する。その際、サンプルは全体の 5% とする。(iii) この(ii)の推定式を使ってステップワイズ法で変数

の選択を実施する。前方／後方の両方で変数選択。(iv) 以上の(ii)と(iii)の工程を500回繰り返して、500回分の変数選択結果を前方ステップワイズ／後方ステップワイズ法の両方について得る。選ばれた回数の多かった変数を、説明変数の候補とする。(v) 最後に、(iv)で得た変数（と固定された5変数）を選択関数の変数候補として、再度ステップワイズ法で変数選択を行う。以上の変数選択においてはランダムフォレストを使用した。変数重要度は「その変数をモデルから除いた場合にどれだけ推計精度が落ちるか」という尺度で定義されている。

第5表 選択関数の推定結果

	推定値	標準誤差	p値
切片	2.02	0.03	0.000
要素_a1	0.51	0.01	0.000
要素_b1	0.09	0.01	0.000
要素_c1	0.01	0.01	0.296
医師当たり病床数	0.00	0.00	0.000
病床数	0.00	0.00	0.000
地域周産期母子医療センター_真	0.44	0.01	0.000
日本医療機能評価機構認定病院_偽	-1.25	0.01	0.000
腹腔鏡_偽	-0.20	0.01	0.000
PET（ポジトロンCT）_偽	-1.14	0.02	0.000
IABP駆動装置_偽	-0.41	0.01	0.000
婦人科_偽	-0.31	0.01	0.000
整形外科_偽	-1.15	0.01	0.000
DMAT_真	0.86	0.01	0.000
ガンマナイフ_真	0.39	0.03	0.000
健康管理センター_真	0.33	0.01	0.000
イマトロン（電子走査型CT）_真	-2.26	0.02	0.000
放射線治療室_偽	-0.51	0.01	0.000
3D-CT_真	0.16	0.01	0.000
PET/CT_偽	-0.15	0.01	0.000
腎臓科_偽	-0.31	0.01	0.000
前立腺肥大治療装置（レーザー）_偽	-0.57	0.01	0.000
開放型病院_真	0.43	0.01	0.000
NMR-CT（MRI）_真	0.65	0.01	0.000
人工腎臓（透析）装置_偽	-0.30	0.01	0.000
救急告示病院_真	1.00	0.01	0.000
通所介護_x_真	-0.93	0.03	0.000
検査室_偽	0.59	0.01	0.000
胸腔鏡_偽	-0.25	0.01	0.000
神経科デイ・ナイトケアサービス_真	-1.20	0.05	0.000

*真 or 偽 ダミー変数の真偽値

出所) 東京大学医療経済ビッグデータより作成

傾向スコアの計算のための選択関数の推定は、全疾患（4疾患以外の疾患を含む）、4疾患を合わせたもの、及び、疾患別に行った。4疾患についての推定結果は第5表のとおり（個別疾患の結果は表示せず）。選択された変数をみると、大きく5つのカテゴリーに分類できるとみられる。(i)DPC参加のための施設基準。これはある意味当然ではあるが、施設基準を充たしてなければDPCに参加はできない一方、DPC施設基準を充たしているとDPCに参加する確率が高い。(ii)経営姿勢を反映する変数。地域連携や介護支援など新しい潮流への積極的対応や医療機能評価を受けるなど質の高い医療への取り組み、健診や日帰り手術など予防や効率的医療への関与などが見られる病院⁷はDPCに参加する確率が高い。(iii)当該診療科の技術基盤。固定費用的な設備⁸が存在するとDPCに参加する確率が高い。これは参入コストや sunk・コストを表すと言えよう。例えば、リニアックを新たに導入するのはコストがかかるので新生物治療に参入しにくく、逆に、既にリニアックを保有していれば、新生物治療から退出しにくい、とかいったことがあると推測される。(iv)他診療科の技術基盤。おそらく技術の代替性や補完性ということから、ある科の患者のデータにおいて、その病院がDPCに参加する確率は他の診療科における治療方法が影響をすることがある⁹。例えば、循環器治療は細かな手技を必要とするので、マイクロサージャリーに優位を持つ病院は（コスト的及び医療の質的に）優位に立てるとかいった具合である。

DPC 選択関数を基に傾向スコアを計算した。その結果を度数分布にしたものが第1図に示されている。濃い紫色の棒グラフが実際にDPCに参加した病院についての傾向スコア、薄いピンク色のかかった棒グラフがDPCには参加しなかった病院についての傾向スコアである。傾向スコアが高い病院はかなりの割合で実際に参加している。また、実際には参加していない病院でも、傾向スコアはかなり高い場合も多い。したがって、この傾向スコアを用いてマッチングをすることができる。以下の分析では、傾向スコアが0.6~0.8の病院をサンプルに残すこととする。これは、heterogenous treatment effectの文脈を考慮すると、傾向スコアが0.6~0.8というDPCを採用するかどうかのボーダーラインにいるような病院について、DPCを採用した時とそうではない時で医療の質がどのように違う

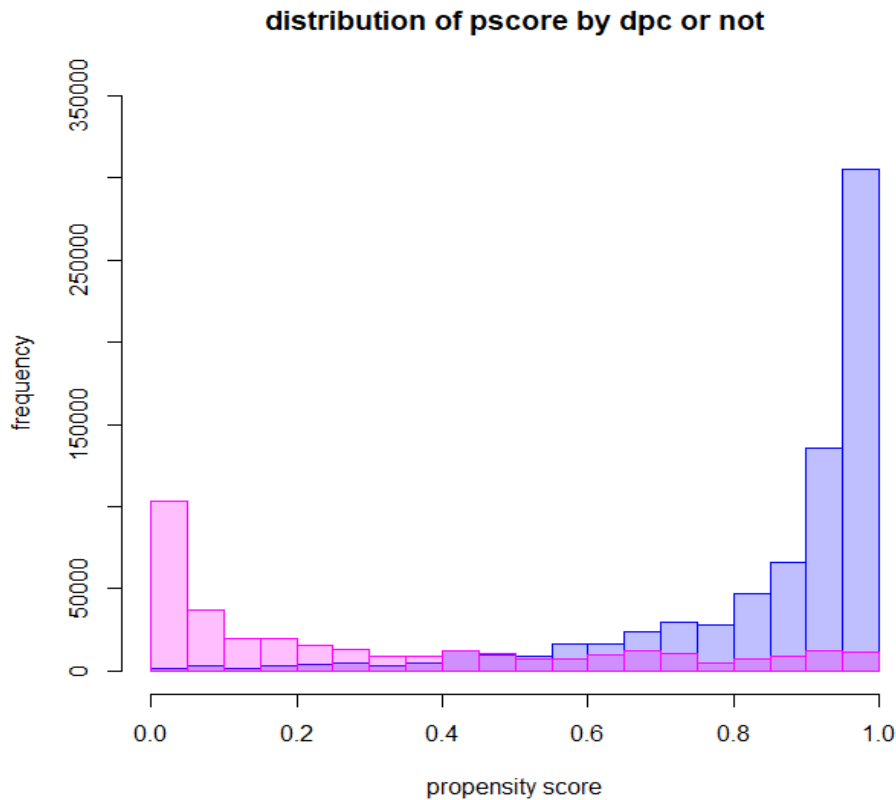
⁷ 例を挙げると、地域医療支援・地域医療連携室や介護支援センター、健康管理センターの存在、医療機能評価の受験、受託健診や日帰り手術の実施などである。

⁸ 例えば、がんにおけるリニアック、呼吸器における胸腔鏡、がんや消化器における高圧酸素法など。

⁹ 例えば、がんに対するシネアンギオ、消化器に対する電子走査型CT、呼吸器に対する体外衝撃波結石粉碎装置、循環器に対するマイクロサージャリー装置、がんに対する精神科作業療法など。

かを比較することになり、明らかに DPC に参加するような病院やその逆のような病院における効果は分析の対象外とすることになる。本論文のセッティングにおいて **heterogenous treatment effect** があるのかどうかは議論の余地があるが、このような効果を計測しているということに留意することは重要である。なお、以上の傾向スコアマッチングに利用したパッケージは、統計ソフト R の “Matching” パッケージである。

第 1 図 傾向スコアの度数分布



出所) 東京大学医療経済ビッグデータより作成

以上のようにマッチングによって構成した等質なサンプルを使って死亡率関数を推定する。死亡率関数は次のように定式化した。

$$\ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (17)$$

ここで、Y は死亡の時に 1、生存の時に 0 をとる変数であり、 $p(Y = 1)$ は死亡率を表わす。説明変数 x_i に含まれるのは、DPC レセプトか否かを示すダミー、レセプト種別で公費併用を示すダミー、同じく低所得を示すダミー、性別 (女性

か否か)、再入院有り、cci (チャールソン併存疾患指数) 有り、年齢について3つのダミー (40歳から59歳、20歳から39歳、19歳以下)、FY2014年度ダミー*DPCレセプト、FY2015年度ダミー*DPCレセプト、FY2016年度ダミー*DPCレセプト、FY2017年度ダミー*DPCレセプトである。

推定結果が第6表に示されている。年度ダミー*DPCレセプト・ダミー以外は有意な結果となっている。DPCレセプト・ダミーの係数がDPCが死亡率に与える影響を表わすことになる。この係数が有意で、かつ、マイナスであるということは、DPC参加により死亡率に影響は低下するということである。年度ダミー*DPCレセプト・ダミーが有意にならないということは、驚くに当たらない。第4表で見たように、生の死亡率では、DPC参加病院と非参加病院の死亡率の差は年度に関係なくほぼ一定であるためである。

第6表 死亡率関数の推定結果

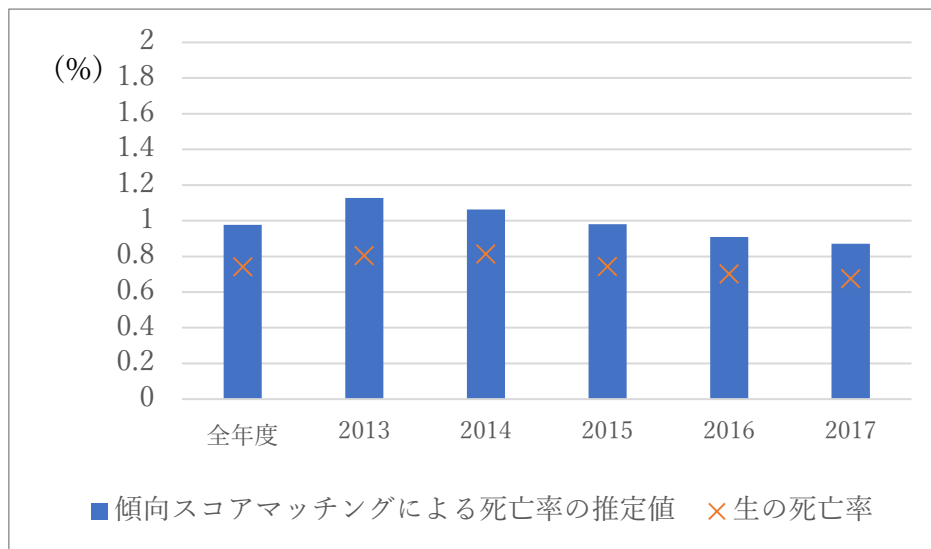
切片	-3.96	0.05	0.000
DPCレセプト	-0.97	0.06	0.000
レセプト種別_公費併用	0.11	0.04	0.009
レセプト種別_低所得	0.32	0.05	0.000
性別_女	-0.20	0.02	0.000
再入院有り	0.76	0.02	0.000
cci有り	1.03	0.02	0.000
年齢_40歳から59歳	-0.40	0.03	0.000
年齢_20歳から39歳	-1.73	0.05	0.000
年齢_19歳まで	-1.79	0.05	0.000
FY2014年度	-0.08	0.05	0.131
FY2015年度	-0.17	0.05	0.001
FY2016年度	-0.25	0.05	0.000
FY2017年度	-0.26	0.05	0.000
DPCレセプト:FY2014年度	0.00	0.08	0.976
DPCレセプト:FY2015年度	0.01	0.07	0.902
DPCレセプト:FY2016年度	0.07	0.08	0.367
DPCレセプト:FY2017年度	-0.01	0.08	0.939

出所) 東京大学医療経済ビッグデータより作成

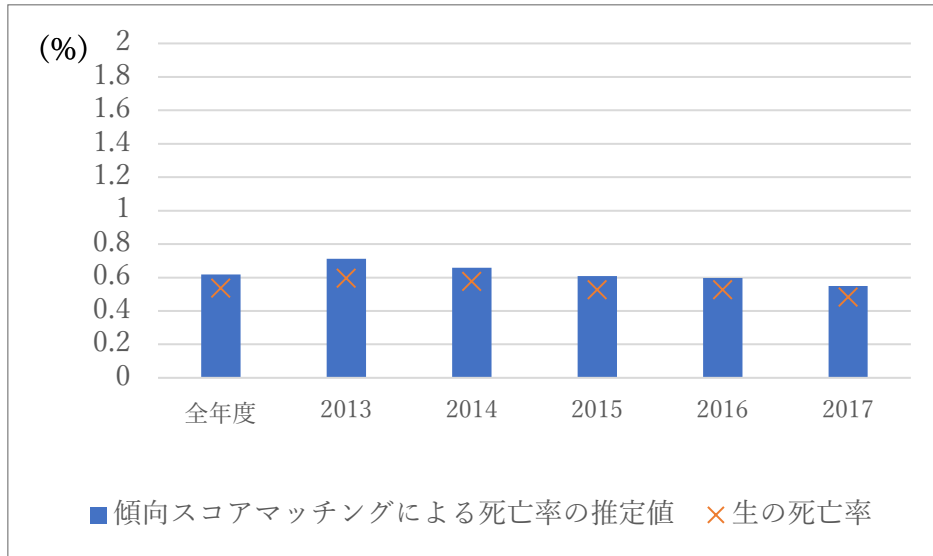
本論文の興味を中心は、サンプル・セレクションの調整である。すなわち、DPC参加が死亡率を低下させるとして、DPC参加が内生的に決まっている以上、単純に回帰分析を行ってもDPCの効果을正確に計測することはできない。先に見た傾向スコアによるマッチングによりこうしたセレクション・バイアスを調整した時に、DPCの効果はどれほどのものとなるかを検証しようということである。

まず、死亡率の水準がどのように推定されているかを調整しない生の死亡率と比較しながら見てみよう。第2図は全疾患（4疾患以外の疾患を含む）の死亡率の水準について、マッチングによる推定値を棒グラフ、調整しない死亡率を×印で描いたものである。最初に2番目のパネルのDPC参加病院について見ると、全年度も各年度も調整後の死亡率は調整前の死亡率より高くなっている。これは、質の高い（死亡率が低い）病院ほどDPCに参加しやすいことによりバイアスが生じる可能性があるという出発点の想定に沿った結果である。下方バイアスを含んだ生の死亡率に対して、マッチングにより調整を施せば死亡率の値は上昇することになる。ただし、3番目のパネルの非DPC病院について見ると、やはり調整後の死亡率が調整前より高く、しかも、かなり高くなっている。これは正直パズルである。最初のパネルに全病院についての結果が示されているが、調整に影響を受けないであろう全病院についても調整後の死亡率が調整前より高くなっており、推定誤差がかなりあることが示唆されているのかもしれない。いずれにしろ、第2-2図と第2-3図を比較すれば明らかなように、セレクション・バイアスを調整した後も、DPC参加病院の方が非参加病院よりも死亡率は低くなっており、DPCが医療の質を向上させる効果があったことが分かる。

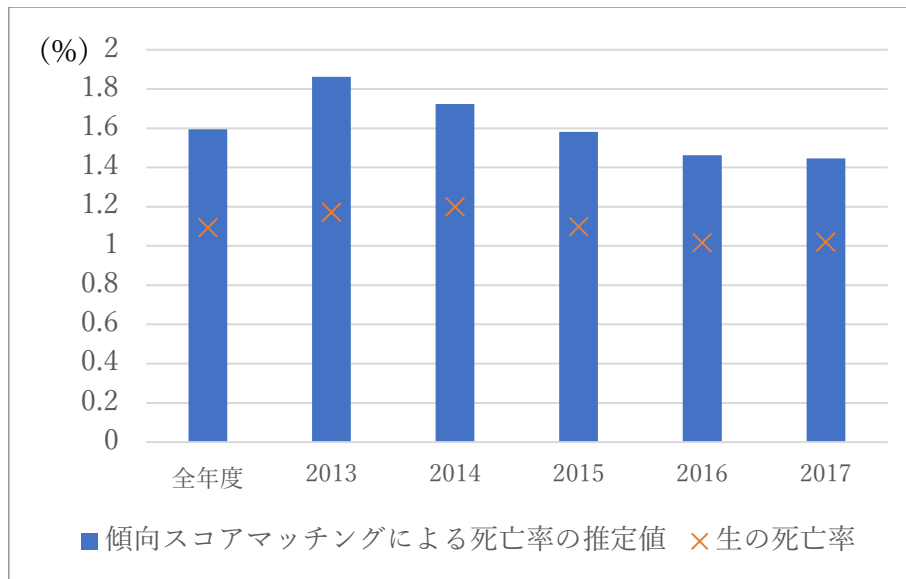
第2-1図 死亡率



第 2-2 図 DPC の死亡率



第 2-3 図 非 DPC の死亡率



出所) 東京大学医療経済ビッグデータより作成

次に、新生物、呼吸器系、消化器系、循環器系の4疾患について、興味を中心であるDPCの効果、すなわち、DPC参加病院と非参加病院の死亡率の差がセレクションの調整によりどのように変化するかを見よう。第3図は、全年度について、4疾患を合わせたもの及び疾患別の調整後の死亡率の差を棒グラフ、調整前の差を点により描いたものである。なお、全年度合計の値は、次の第4図で推定した各年度の数値をサンプル数で加重平均した。一番左の4疾患合計を見ると、

死亡率の差は、マッチングによる調整により調整前よりも縮小している（後者の1.8%ポイントから前者の1.1%ポイントへ縮小）。これは質の高い病院がDPCへ参加しやすいというセレクション・バイアスの影響があったことを示している。したがって、バイアスを調整しないとDPCが死亡率に与える影響を過大に見積もってしまうことになる。しかし、個別の疾患については、調整後の差の方が調整前の差よりも大きくなっている。この場合、調整するとDPC参加の効果はより大きくなるということになる。これは第2図と同様のパズルであるが、その時に見た様に、DPC非参加病院の推定値の動きにややイレギュラーなものがあることが原因であろう。

第3図 マッチング推定による死亡率の差：全年度



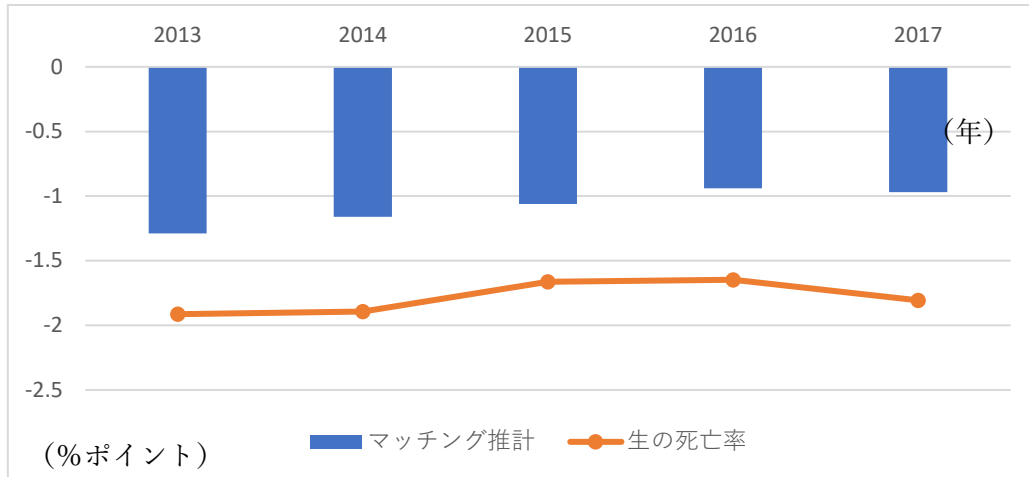
出所) 東京大学医療経済ビッグデータより作成

しかし、いずれにしろ、DPC参加病院と非参加病院の死亡率の差はマイナスになっており、セレクション・バイアスを調整した後も、DPCは医療の質を向上させたことが確認される。その程度は、1%程度死亡率を低下させたというものである。個別の疾患における死亡率低下効果は、消化器系が0.3%ポイント、呼吸器系が2.7%ポイント、新生物が3.5%ポイント、循環器系が15.0%ポイントと、各疾病についての効果はばらばらであるが、やはりDPCは医療の質を向上させたことが確認された。

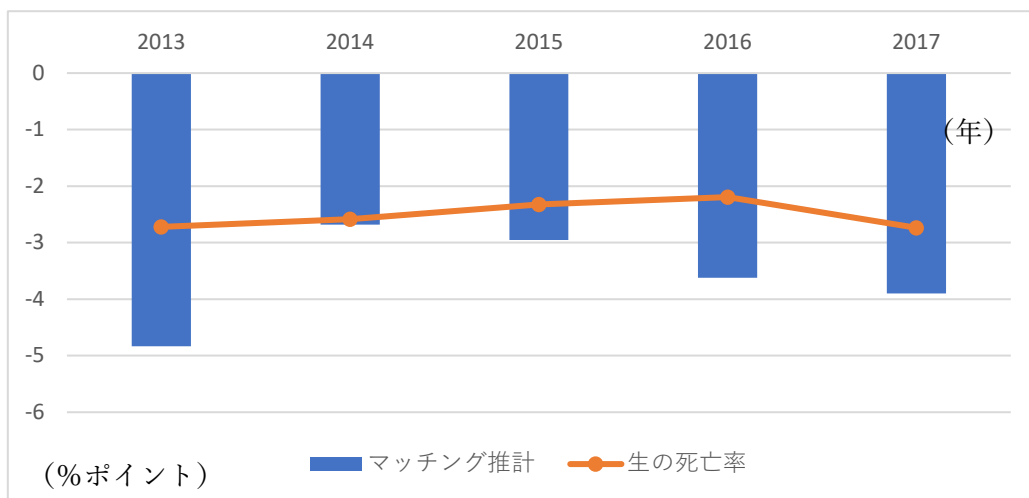
以上を各年度について見たものが第4図である。4疾患合計では(第4-1図)、先に見たように、調整後の死亡率の差は調整前よりも小さくなっている。年度別に見ても、セレクション・バイアスを調整した後も、DPCは医療の質を向上

させたことが確認された。ただし、死亡率の差は年とともに縮小している（2013年の1.29%から2017年の0.97%へ縮小）。これはやや不可解な動きである。マッチングによりDPC病院と非DPC病院は、DPC参加か否かという点を除いて同質となっているはずである。上記の結果は、DPC参加病院と非参加病院は同質にもかかわらず死亡率は異なる動きをしたということになる。新生物や消化器、循環器では、調整前の生の死亡率の差がほぼ横ばいとなる中で、調整後の死亡率の差は不規則な変動を見せている。特に呼吸器では推定期間の最初辺りに不規則な動きが見られる。これらの不規則な動きは、DPC選択関数が各年の選択行動を適切にとらえていないという限界に起因するものかもしれない。

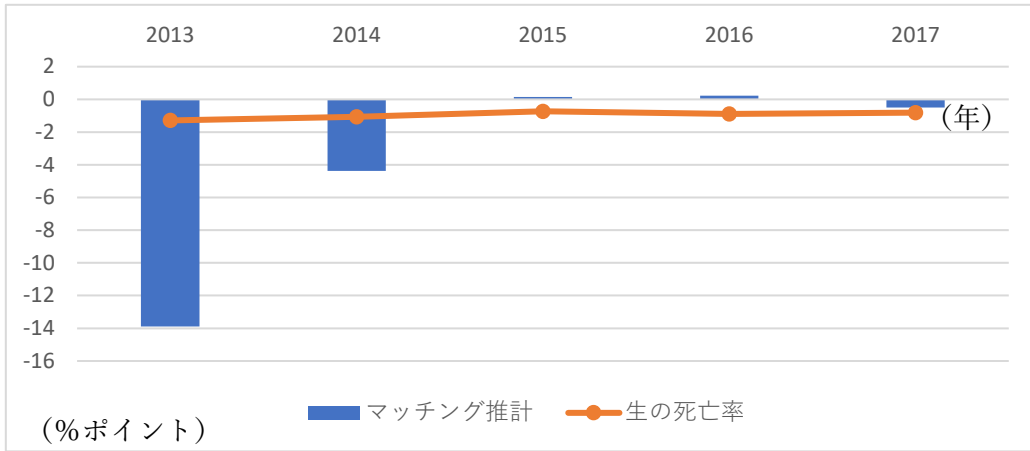
第4-1図 マッチング推定による死亡率の差：4疾患合計：各年度



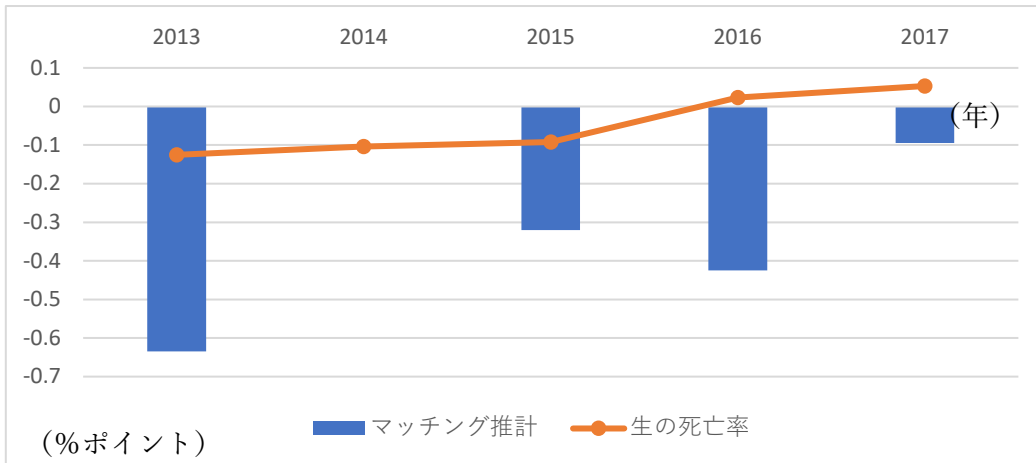
第4-2図 マッチング推定による死亡率の差：新生物：各年度



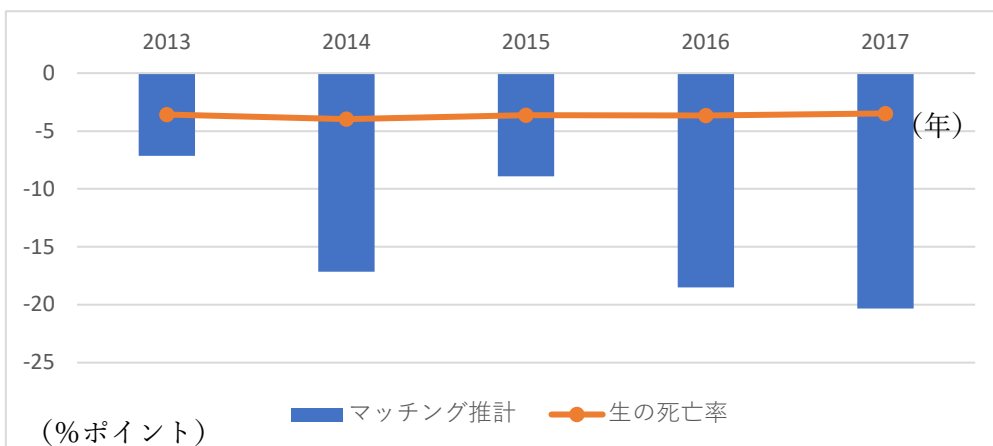
第 4-3 図 マッチング推定による死亡率の差：呼吸器：各年度



第 4-4 図 マッチング推定による死亡率の差：消化器：各年度



第 4-5 図 マッチング推定による死亡率の差：循環器：各年度



出所) 東京大学医療経済ビッグデータより作成

以上のように、DPCに参加するか否かの選択に起因するバイアスを調整するために、傾向スコアを利用したマッチングによって等質なサンプルを構成した。しかし、傾向スコアは観測される変数を使って計算するものであり、病院による選択が観測できる変数のみによって決定されるという仮定に依存している。本論文では選択関数の推定に当たって極めて豊富な情報を持つ観測データを利用して、こうした仮定が成立するように努めた。ただ、我々の利用したデータによってこの仮定が充たされているかどうかは先験的には決定できない。そこで、頑健性のチェックとして、観測できない要因がDPC選択に影響を与える可能性を考慮した上でDPCの効果を計測してみることにする。

実際の推定は、先に説明したように、選択関数の残差を死亡率関数の説明変数に加えるという形で実行する。死亡率関数の定式化は次のようになる。

$$\ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (18)$$

ここで、Yは死亡の時に1、生存の時に0をとる変数であり、 $p(Y=1)$ は死亡率を表わす。説明変数 x_i に含まれるのは、DPCレセプトか否かを示すダミー、レセプト種別で公費併用を示すダミー、同じく低所得を示すダミー、性別（女性か否か）、再入院有り、cci（チャールソン併存疾患指数）有り、年齢について3つのダミー（40歳から59歳、20歳から39歳、19歳以下）、FY2014年度ダミー*DPCレセプト、FY2015年度ダミー*DPCレセプト、FY2016年度ダミー*DPCレセプト、FY2017年度ダミー*DPCレセプト、傾向スコアの残差である。傾向スコアの残差を含む以外は(17)式と同じである。

第7表はこのように内生性を考慮した（残差を入れた）死亡率関数の4疾患合計の推定結果である。傾向スコアの係数は有意となり、説明変数では説明され尽くされない観測されない要因がDPC選択に影響を与えていることが分かる。問題は、これがDPCの効果の計測結果にどの程度の影響を及ぼすかである。第8表に、死亡率の水準の推定値を示した。4疾患合計のDPC参加病院についてみると、調整前の生の死亡率が0.54%であるのに対して、内生性を考慮した推定値は0.64%となっている。マッチングによる推定値は0.62%であったから（第2図）、残差を入れた場合とマッチングとでほとんど同じである。

第7表 内生性を考慮した死亡率関数の推定結果

	推定値	標準誤差	p 値
切片	-3.92	0.05	0.000
DPCレセプト	-1.03	0.06	0.000
レセプト種別_公費併用	0.11	0.04	0.013
レセプト種別_低所得	0.32	0.05	0.000
性別_女	-0.19	0.02	0.000
再入院有り	0.76	0.02	0.000
cci 有り	1.03	0.02	0.000
age_40 歳から 59 歳	-0.40	0.03	0.000
age_20 歳から 39 歳	-1.74	0.05	0.000
age_19 歳まで	-1.78	0.05	0.000
傾向スコアの残差	0.30	0.04	0.000
FY2014 年度	-0.10	0.05	0.062
FY2015 年度	-0.19	0.05	0.000
FY2016 年度	-0.27	0.05	0.000
FY2017 年度	-0.28	0.05	0.000
DPCレセプト:FY2014 年度	0.01	0.08	0.873
DPCレセプト:FY2015 年度	0.02	0.07	0.805
DPCレセプト:FY2016 年度	0.08	0.08	0.278
DPCレセプト:FY2017 年度	0.01	0.08	0.917

出所) 東京大学医療経済ビッグデータより作成

それでは、DPC 参加病院と非参加病院の死亡率の差についてはどうであろうか。第5図を見ると、残差を入れた場合とマッチングとで大きくは変わらない。すなわち、DPC の死亡率に対する影響の計測結果は、マッチング推定では内生性の影響は残るものの、それがマッチング推定の妥当性を失わせるようなものではないと言えよう。

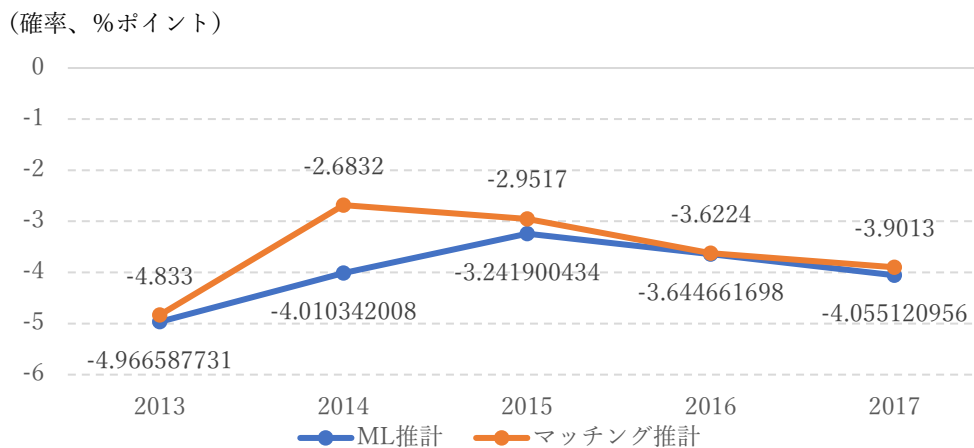
第8表 内生性を考慮した死亡率の推定結果（全年度）

	内生性を考慮した推定値	調整前の死亡率	参考：死亡率
全疾病	1.01	0.74	0.74
DPC参加病院	0.64	0.54	0.54
DPC非参加病院	1.64	1.09	1.09
新生物			
全病院	4.9	2.20	2.20
DPC参加病院	1.2	1.64	1.64
DPC非参加病院	5.1	4.13	4.13
呼吸器			
全病院	8.1	0.42	0.42
DPC参加病院	6.2	0.24	0.24
DPC非参加病院	16.4	1.17	1.17
消化器			
全病院	0.41	0.29	0.29
DPC参加病院	0.22	0.26	0.26
DPC非参加病院	0.62	0.31	0.31
循環器			
全病院	19.30	2.06	2.06
DPC参加病院	5.94	0.87	0.87
DPC非参加病院	47.07	4.53	4.53

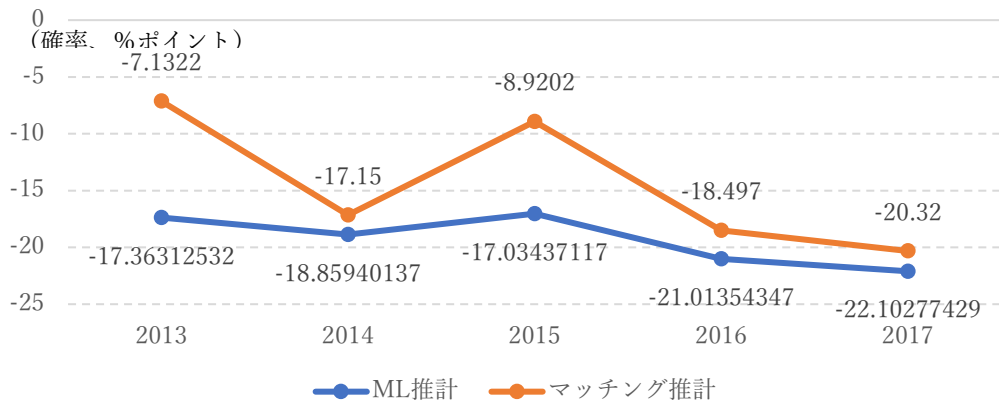
出所) 東京大学医療経済ビッグデータより作成

第5図 内生性を考慮した死亡率の差の推定結果（各年度）

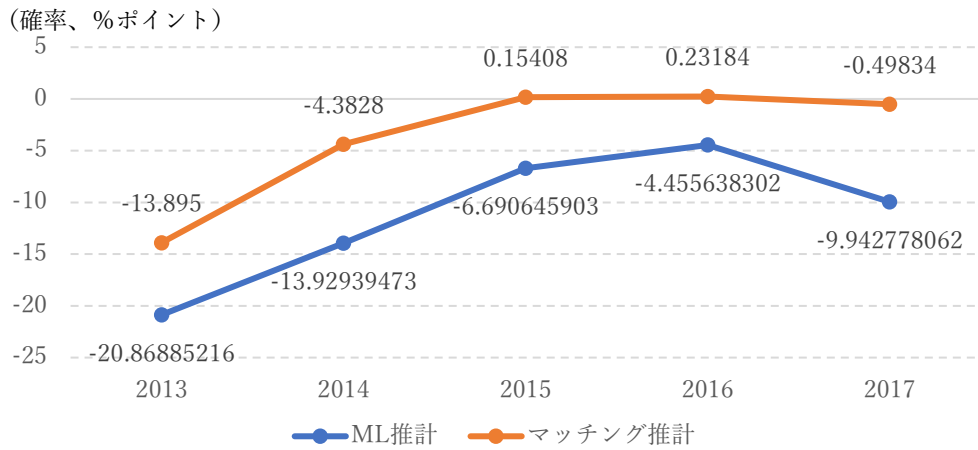
第5-1図：新生物



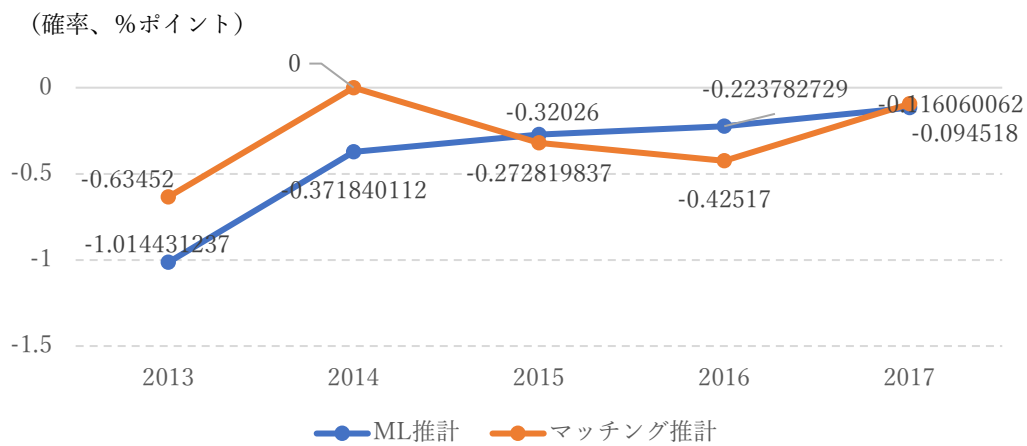
第 5-2 図：循環器



第 5-3 図：呼吸器



第 5-4 図：消化器



出所) 東京大学医療経済ビッグデータより作成

5. 終わりに

本論文では、包括支払方式（DPC）への参加が医療の質（死亡率）に与える影響を検証したが、DPCの参加の選択が内生的に行われる可能性を考慮して、傾向スコアによりセレクション・バイアスを調整した。その際、傾向スコアの推定が観測される要因のみに基づくという制約を緩和するために、病院についての豊富な情報を持つデータをDPC選択関数に説明変数として加えた。分析対象としては、新生物、呼吸器系、消化器系、循環器系、及びその合計とした。

4疾患合計についての結果は、DPCへの参加は死亡率を1%ポイント程度低下させることが確認された。これは、大きな改善効果と言える。また、バイアスを調整しない場合にはこの倍程度と過大評価になっており、マッチング推定の有効性が示された。個別の疾患における死亡率低下効果は、消化器系が0.3%ポイント、呼吸器系が2.7%ポイント、新生物が3.5%ポイント、循環器系が15.0%ポイントと、DPC参加の効果が大きい疾患と小さい疾患のばらつきがあったが、やはりすべての疾患でDPC参加は医療の質を向上させることが確認された。

さらに、DPC選択が観測されない要因によって影響を受ける可能性を考慮した頑健性のチェックを行ったが、マッチング推定には内生性の影響は残るものの、それがマッチング推定の妥当性を失わせるようなものではないことが確認された。

以上のように、DPC選択に影響を与える要因について十分なデータを用いて推定を行えば内生性をかなりの程度調整できることが示された。ただ、特に個別の疾患についての推定において、調整した方がDPCの効果が大きく計測されるという予想と異なる結果となったこと、また、マッチングによってDPC参加病院と非参加病院は等質になっているにもかかわらず、年とともに両グループの死亡率の差が変化（縮小）するという現象が見られたことなどから、セレクション・バイアスの調整方法についてさらに研究する余地は大きいと言えよう。例えば、本論文で採った内生性の調整方法は、選択関数の残差を死亡率関数の説明変数に加えるというややアドホックなやり方を採った。しかし、内生性の調整にはShared parameter modelを含む様々な方法が提案されており、そうした代替的な手法を検討していくことは有意義な方向性であろう。

参考文献

田中隆一 (2015) 『計量経済学の第一歩—実証分析のススメ』 有斐閣。

野口晴子・泉田信行・堀口裕正・康永秀生 (2010) 「診療の医学的・経済的側面に対する DPC 導入によるマネジメント改善効果の実証的検証」『医療と社会』 vol.20, no.1。

Angrist, Joshua, and Joern-Steffen Pischke. (2015) *Mastering Metrics*. Princeton University Press.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey and James Robins. (2018) Double/Debiased Machine Learning for Treatment and Structural Parameters. *Econometrics Journal*, volume 21, pp. C1–C68.

Cutler, David (1995) The Incidence of Adverse Medical Outcomes Under Prospective Payment. *Econometrica*, vol. 63, no. 1, pp. 29-50.

Daniels, Michael, and Joseph Hogan. (2008) *Missing Data in Longitudinal Studies, Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman and Hall/CRC.

Imbens, Guido, and Donald Rubin. (2015) *Causal Inference for Statistics, Social and Biomedical Sciences*. Cambridge University Press.

Kahn, Katherine, Emmett Keeler, Marjorie Sherwood, William Rogers, David Draper, Stanley Bentow, Ellen Reinisch, Lisa Rubenstein, Jacqueline Kosecoff and Robert Brook. (1990a) Comparing Outcomes of Care Before and After Implementation of the DRG-Based Prospective Payment System. *Journal of the American Medical Association*, vol.264, no.15, pp.1984-1988.

Kahn, Katherine, William Rogers, Lisa Rubenstein, Marjorie Sherwood, Ellen Harrison, Emmett Keeler, David Draper, Jacqueline Kosecoff and Robert H. Brook. (1990b) Measuring Quality of Care with Explicit Process Criteria Before and After Implementation of the DRG-based Prospective Payment System. *Journal of the American Medical Association*, vol.264, no.15, pp. 1969-1973.

Kosecoff, Jacqueline, Katherine Kahn, William Rogers, Ellen Reinisch, Marjorie Sherwood, Lisa Rubenstein, David Draper, Carol P. Roth, Carole Chew and Robert Brook. (1990) Prospective Payment System and Impairment at Discharge, The ‘Quicker-and-Sicker’ Story Revisited. *Journal of the American Medical Association*, vol.264, no.15, pp.1980-1983.

Little, Roderick, and Donald Rubin. (2019) *Statistical Analysis with Missing Data*. Third Edition. Wiley.

Molenberger, Geert, and Michael Kenward. (2007) *Missing Data in Clinical Trial*. Wiley.

Newhouse, Joseph. (2004) *Pricing the Priceless: A Health Care Conundrum* (Walras-Pareto Lectures). MIT Press.

Rogers, William, David Draper, Katherine Kahn, Emmett Keeler, Lisa Rubenstein, Jacqueline Kosecoff and Robert Brook. (1990) Quality of Care Before and After Implementation of the DRG-Based Prospective Payment System, A Summary of Effects. *Journal of the American Medical Association*, vol.264, no.15, pp.1989-1994.

Rubenstein, Lisa, Katherine Kahn, Ellen Harrison, Marjorie Sherwood, William Rogers, Caren Kamberg, David Draper and Robert H. Brook. (1990) Changes in Quality of Care for Five Diseases Measured By Implicit Review, 1981 to 1986. *Journal of the American Medical Association*, vol.264, no.15, pp. 1974-1979.

Wooldridge, Jeffrey. (2010) *Econometric Analysis of Cross Section and Panel Data*. Second Edition. MIT Press.

Wooldridge, Jeffrey. (2005) Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity. *Journal of Applied Econometrics*. vol. 20, Issue 1, pp. 39–54.