

虚血性心疾患患者のレセプトデータに基づいた 高度医療行為及び技術の計量分析

今日はファイザーヘルスリサーチ振興財団からご援助いただきました国際比較研究の一つである、急性心疾患を中心とした計量分析の報告をさせていただきたいと思います。この研究は実はスタンフォード大学で行われております国際比較研究に参加したもののです。昨年の2月に一度、スタンフォードで国際比較に関するミーティングがありまして、そこで研究の成果の一部を発表いたしました。

実はそのときのことにも若干触れなければならないのですが、その際には、社会医療診療行為実態調査の公表データを使って、日米で比較研究をするということでスタートしたわけなのですが、昨年の発表ではお恥ずかしい結果になってしまいました。と言いますのは、社会医療の公表データは、私共が考えていたほど精度が高くないという大きな問題点があります。例えば、1人の患者が70回も病院に来たことになっているというような、一種のデータ作成上の問題点がありまして、我々がやるような計量分析理論にはとても適さないということがはっきりしました。

そこでは私共は、直接患者のレセプトデータを集めるという作業を開始いたしました。今日お話ししますのは、某大手電機メーカーの健康保険組合にお願いして集めた、レセプトデータを解析したものです。

国際共同研究は実は昨年で終わったわけではなく、また来年度もおそらく2回目の会合があると思いますので、今日発表させていただくデータを使ったものが、来年度以降また利用されることになると思います。

しかし、今回お話ししますのは、一応研究としては完結しておりますので、これを発表させていただきたいと思います。

【スライド1】

分析のサンプルについて簡単に申し上げますと、今申しましたように、これは大手の電機メーカーの傘下にある病院データでありますので、ひとつバイアスがあります。と言いますのは、このメーカーは日本の代表的ないわゆる超一流企業です。従いまして、そこで働いている人も初めから色々な検査（資格）をクリアしてきた人たちなものですから、相対的にみると非常に健康な人が集まっているということとして、一般国民よりは健康状態が良い人たちが集まっていると考えて良いかと思います。

このデータを使って、大学院博士課程に在籍中の菅原啄磨君に色々と協力していただいて、研究が完結いたしました。彼の貢



学習院大学経済学部

経済学部長

南部 鶴彦

スライド1

分析対象（データ抽出条件）
④ 1993年6月～1998年3月までのレセプト
④ レセプトは某大手電気機械メーカーの健康保険組合の収蔵のもの
④ 虚血性心疾患を主病とする入院患者
④ 1サンプルは入院から退院までの1profile

献が極めて大きかったことを申し添えておきます。

【スライド2】

このデータは、分析対象としまして、1993年から1998年にわたる310件のレセプトデータで、全て入院から退院まで1プロファイルで捉えられるものを使ってあります。310件は狭心症が191件、急性心筋梗塞が74件、慢性虚血性心疾患が45件という内訳です。

【スライド3】

さらにそのサンプルの中身がどのような検査ないしは手術を受けたかという形で分類すると、図1のようになります。

後で申し上げますが、我々は計量分析を主としておりますので、その前の段階で、どのようなものを計量分析のときの変数としてとるかということを考えるわけです。図1にありますのは、そのような意味で言いますと、治療行為の中身について、まず区分をしたものです。CATHとありますのが心臓カテーテル法の検査を、したかしないかというような基準の治療行為です。PTCAは冠動脈の形成術の手術があったかどうか、CABGは冠動脈のバイパス手術があったかどうか、STENTと書いてありますのはステント留置がなされたかどうか、それから血栓溶解剤が使われたかどうかがTHROMBOLYTICSという表示になっているわけです。

全体として見ますと、この310例では、心臓カテーテルの検査があったのが50.3%というようになりますし、その他CABGは極めて少なくて0.3%程度という形で、それぞれの症例がわかるわけです。

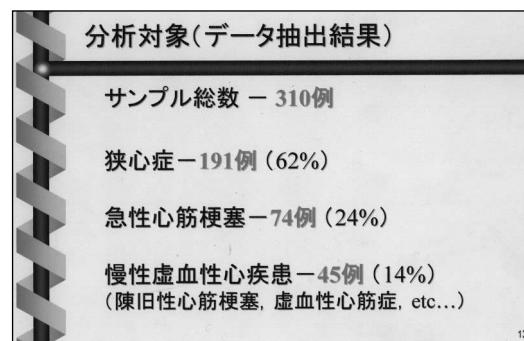
310例ですから、検査対象数が十分大きいとは言えません。しかし、実際に1プロファイルで入院～退院まで完結したデータを、かつ長期にわたって集めるということは極めて難しく、当面はこのデータに頼って分析をしたわけです。

分析に当たりまして、もう少しデータについて申し上げておきます。

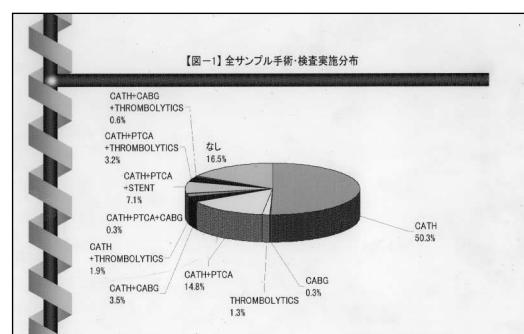
【スライド4】

これは今回のサンプルの年齢構成比です。急性心疾患ですから、基本的には高齢者が多いということになると思いますが、このように40歳以下はごくわずかで、51歳から

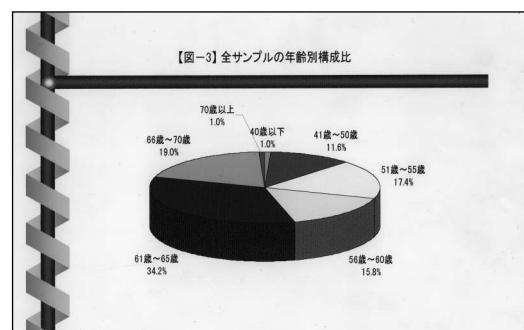
スライド2



スライド3



スライド4



それより上の人たちが圧倒的多数です。このような年齢構成が、計量分析の中にも一部取り込まれております。

【スライド5】

実は急性心疾患の病気だけではなくて、特に年齢層の高い患者さんに、合併症・併発症が必ずあるわけですので、こうした合併症・併発症がどのくらいあるかということを表にしたものです。

これも計量分析の中では、合併症・併発症があるかないかで当然診療点数が変わってくるので、その効果を見ようということでとりあげてあります。

【スライド6】

これは病院の種別構成比です。

何故これが入っているのかと申しますと、私共の分析は経済学の視点からやっていますので、いわば開設者あるいは病院のbehaviorというものが、結果として点数にどのくらい影響を与えるかということが重要な視点になるわけです。例えば国立病院と私立病院というような病院の組織の違いが、当然、医師の行動あるいはbehaviorに影響を与える、そしてそれが結果として点数にも影響を与えるのではないかという仮説が成り立つわけでして、そのために、ここにありますように、開設者別に病院の区分を見ました。

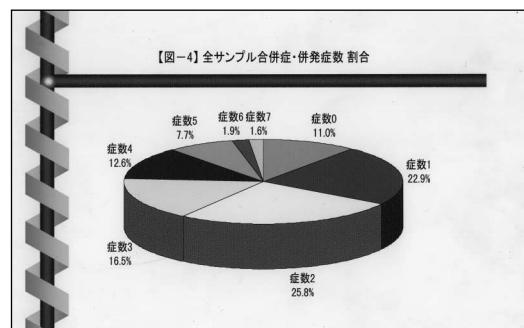
ご覧いただけますように、公的と呼んでいるものが23%、それから会社・個人というのが20%となっておりまして、その他日赤・済生会というようなものとか、公益法人、医療法人というものに、だいたい均等に分布しているというのが今回の姿です。

【スライド7】

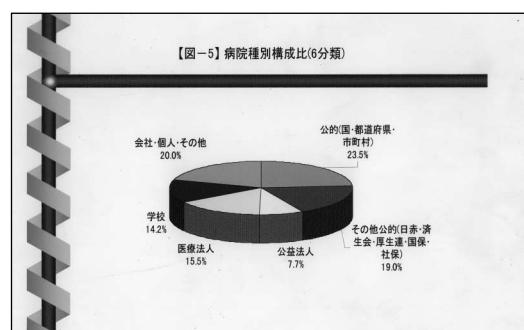
同じく分析の対象としましては、やはり病院の規模が違っているということが診療の行為あるいは内容に影響を与える可能性がありますので、病床数の区分をやっています。

そうしますと、ご覧のとおり、500床ないし700床以上という大きな病院が相対的に多い。これは病気の対象からして当然かと思いますが、このような病床数の区分がありまして、これも診療行為、結果としては点数に影響を与えるかどうかという観点での分析を行いました。

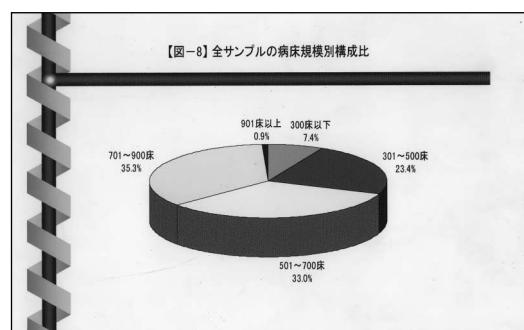
スライド5



スライド6



スライド7



【スライド8】

以上のような基本的な病院の属性、あるいは治療行為の中身、あるいは年齢その他のデモグラフィックなファクターというものを取り入れた分析をしようというのが、今回の目的です。

ご覧いただいております Y_i という式は、今回の推定にあたって使われた推定式です。時間がありませんので簡単にご説明しますと、左辺の Y_i は被説明変数と呼ばれるもので、右側のファクターがどれくらい左側の Y_i を説明しうるかということを示そうとしているわけです。そして、 α 、 β の次に $Tech_i$ というのがありますし、その次に γ があって $Patient_i$ 、それから μ があって $Hosp_i$ となっていますが、これは今回我々が分析の中で説明変数として取り上げようとした、それぞれの属性というものを表しているわけです。

まず最初の $Tech_i$ と書いてありますのは、治療行為ということに着目して、どのような治療行為が行われているのか、そのような治療行為が行われると点数にどれくらい影響を与えるのかというものを見るわけです。先ほど申し上げましたように、例えばカテーテル検査はあるかないか、PTCAが行われたかどうか、CABGが行われたかどうか、というような、それぞれの治療行為の有無という形で、変数が入ってまいります。これはダミー変数と経済用語で呼んでいるもので、つまりそのようなことが行われたら1、行われない場合は0というように、1と0という離散的な数字をとると考えて、そしてレセプトデータごとに順番にあるかないかを入れていくわけです。そうしますと、後で簡単に結果をご紹介しますが、例えば、CABGの手術がなされるということによって左側の Y_i がどれくらい動くのかということが読み取れるようになるわけです。

それから、 $Patient_i$ というのは患者の属性として、先ほど述べたとおり、年齢・性別のような本来患者が持っている属性が影響を与えるかどうかということです。結果を先に申し上げますと、今回の対象が急性心疾患でありますから、この患者属性は必ずしも有意に影響してしませんでした。

それから病院属性というのは、先ほど最後のところで申しましたように、例えば、開設者が一体誰であるかという病院の種類の違いがbehaviorに影響し、最終的には左側の Y_i に影響を与えるかどうかというものを見るための変数です。それから、病床の数。これは連続変数で、それぞれ200、300、400と変化するわけですが、それがやはり病院属性として入っています。

それぞれ、例えば $Tech_i$ と書いてあるところで言うと、あるTechnologyの*i*というものをやったときに、どれくらいそれがコストを上げるかというのが、*i*という数字で出てくるというように読み取ることになるわけです。

ここで、このような分析をやることの最も重要な点を申し上げておきますと、実はこういう形で分析をしますと、例えば $Tech_i$ について言えば、ダミー変数処理をしていることによって、非常にクリアに出てくることがあります。それは何かと言うと、CABGが一つの例だ

スライド8

推定モデルの設定	
$Y_i = \alpha + \sum \beta Tech_i + \sum \gamma Patient_i + \sum \varepsilon Hosp_i + \mu_i$	
【被説明変数】 Y_i :	
① 診療総点数、② 治療点数、③ 投薬・注射点数、④ 処置・手術点数、⑤ 検査・画像診断点数、⑥ 入院点数、⑦ 在院日数	
【説明変数】 $Tech$:「医療行為・技術」の適用 $Patient$:「患者属性」 $Hosp$:「病院属性」	
表1～7 にリスト 30	

と思いますが、CABG という診療行為をした結果として、左辺にありますY がどんなふうに動くかということがわかるわけですが、これは実は、例えばCABG の点数だけをとって何点であるというように見たのでは、到底知ることができないものを情報として与えるわけです。つまりCABG については、それが診療点数の上で何点と付いているわけですが、それに付随したさまざまな器具その他のコストがかかってくるわけです。この分析でいきますと、CABG については、そこだけ取り出したときに一体何点くらい、Y という形でそれが影響を受けるかというのが、計測されるのです。

ちょっと逆になりましたが、左辺のY について出てくるものを申し上げますと、1番目に診療総点数、それから治療点数という、今申し上げましたコスト的な要因、それから投薬・注射・処置・検査というように細かくブレイクダウンしたもの、あと在院日数もどのような影響を受けるかといいる形で説明変数に入っております。

以上のようなことで、主な結果だけ申し上げておきたいと思います。

【スライド9】

まず、カテーテルであります、心臓カテーテル法の検査をしたかどうかというのが、検査・画像の点数に当然影響を与えるわけで、このCATH があったかないかとの変数を見てみると、CATH だけをとったとき約2万点のコストをもたらすということが出てまいります。それからPTCA につきましては、治療総点数で9万点、診療総点数で10万点といいう形になっています。

この2つの区分は、診療総点数には食事代・入院費が入っていまして、治療総点数はそれらをとったものですから、若干の差がでてくるということです。

【スライド10】

それからCABG を見ていただきますと、これは今回の計測の中で、全ての点で統計的に有意な結果が得られました。診療総点数について言うと、CABG をやると約30万点、診療総点数については約36万点。あるいは在院日数についていうと、CABG の手術があると在院日数を35日延ばすというような結果が出ているわけです。ちなみにこの30万点、36万点というのは、CABG にまつわるほとんど基本的な全てのファクターを総合したものです。

我々は、この結果というのは、いわゆる包括化の議論と非常に結びついているのではないかと思っています。つまりCABG 関連の全てを足してみると、いま包括化でも36万点ほどつけたらどうかという案になっているわけですが、実はこの我々のレセプトデータの分析に

スライド9

分析結果－医療行為・技術①

- ⑦ 被説明変数に対し、8モデルを推定→【表-8】
決定係数、推定パラメーターのレンジから概ね結果は良好
- ⑧ CATH: 「検査・画像点数」(+約 2万点)
「入院」、「在院日数」がマイナス
検査入院が短期、CATHの非受検者の入院はそうではないケースに比して有意に長い
- ⑨ PTCA: 「治療総点数」(+約 9万点)
「診療総点数」(+約 10万点)

スライド10

分析結果－医療行為・技術②

- ⑩ CABG: 全モデルで有意
「治療総点数」 + 約 30万点
「診療総点数」 + 約 36万点
「在院日数」 + 約 35日
- ⑪ THROMBOLYTICS: 全モデルで有意
「治療総点数」 + 約 12万点
「診療総点数」 + 約 14万点
「在院日数」 + 約 17日

とっても、このようないわば丸めをするといったこととは関係なく、この結果が出ています。

あと血栓溶解療法につきましても、ここにありますように、それぞれ12万点、14万点という点数が得られるということです。

時間もございませんので、以上から結論を申し上げますと、この300例というそれほど数多くない事例ではありますが、このようなサンプルを使って、急性心疾患に関するコスト構造の分析がかなりの程度できるというように私共は確信いたしました。このデータ数を例えば3倍に増やし、かつアメリカと日本のデータをマージして、全体として一つのグループとして分析することができれば、より多くの情報が得られるのではないかと思っております。

これから一番大きな問題は、そのような良好なデータをどこまで集められるかということです。今回結果として、統計的に有意な結果が出てきた一番大きな理由は、私の直感では、やはり対象とした健康保険組合の病院データが非常に良かったということにあるのではないかと思います。そのようなデータ集めも、これからの大きな作業かなと思うわけです。

質疑応答

Q：（筑波大学社会工学部 山田先生）

大変興味あるご研究を有難うございます。

先ほどの鹿内先生のご研究で、近隣に病院があれば、産婦人科の患者行動が少し変わることがどうぞお聞きさせていただいたのですが、今の先生のご研究の病院の属性のところに、近隣にどのような病院（ベッド数）があるのかということで病院の行動がどのように変わるかということも、少しある関係があるような気がするのですが、そのあたりの属性はどのようにコントロールしておられるのでしょうか。

A：（南部先生）

今のご質問は、実は先ほど申し上げましたスタンフォードの国際比較研究では、アメリカ側はかなり重視したデータとして、説明変数に入れてあります。こちらでは、データ的にまだ病院数も多くなく、変数としては入っておりません。実は正確に申しますと、ある大手の電機メーカーの健保組合の、さらに東京を中心とした関東エリアと言いますが、特に東京近辺の病院だけがデータの中に入っているという特殊性がございます。ということはつまり、基本的には非常に競合があると言うのでしょうか、東京エリアですから、いくつかの大病院が一応選択可能性に入っていると考えて良いかと思うのですが、おっしゃるような形での説明変数はまだ入っておりませんので、今後入れていかなければならないかと考えております。

Q：（座長）

ちょっとそれに関連して、私から質問させていただきたいのですが、公的病院と私的病院

との間では点数に差はなかったのでしょうか。

A : (南部先生)

手元に資料がないので、あとで説明します。

Q : (大阪大学 中島先生)

計量分析の説明変数について教えていただきたいのですが。先生はTech・Patient・Hospitalと3つのグループに表していらっしゃいましたが、具体的にその変数はそれいくつあるのかということと、聞き落としたかもしれません、対象分析の人数を教えていただけませんでしょうか。

A : (南部先生)

まず対象(サンプル数)は310件(310人)のレセプトデータで、これが全て完結しています。それから説明変数ですが、数というのは例えば治療行為でいいと、カテーテルとPTCAというかたちで5種類とっているわけですが、そのような意味でしょうか。対象分析の人数については後ほどお渡しいたします。