

論文

医療の質を調整した医療アウトプットの計測

杉 原 茂¹⁾
川 渕 孝 一²⁾

概 要

本論文は、医療サービスのアウトプットの計測に医療の質を織り込む方法論を検討する。素朴な質の調整としては、まず、患者の入院時の併存症などのリスク要因を調整しない生（なま）の死亡率及び合併症率に基づく質の調整を施したアウトプット指数を作成する。しかし、患者のリスク要因は年々変化し死亡率／合併症率の変動の大きな要因となるため、次に、入院時のリスク要因を調整した死亡率／合併症率に基づく質の調整を施したアウトプット指数を作成する。分析結果は、リスク要因を調整しない場合でも医療の質の調整によってアウトプット指数の変化率は大きく変わるが、さらにリスク調整した場合にはリスク要因の変動によるノイズが除去されて医療のアウトプットをより正確に計測出来ている。以上より、医療のアウトプットの計測に当たってはリスク調整を実施した上で医療の質を調整することが必要である。なお、推定手法としては通常の最尤法のほかに、推定結果の極端な変動を回避するようにベイズ的階層モデルによる推定も試みたが、合併症のようにデータの不確実性が大きい場合には両者の結果に顕著な差が見られた。

I. はじめに

医療のような非競争市場型サービス部門におけるアウトプットとデフレーター計測は、価格と数量が競争的市場で決定されていないため容易ではない。更に問題を難しくしているのは、医療におけるサービスの質は非常に重要であるにもかかわらずその計測は非常に困難であることである。したがって、医療サービス部門におけるアウトプットを計測する際に医療の質を調整する手法を研究することが強く求められる。

近年、直接的計測法により非競争市場型サービス部門のアウトプットを計測することが活発に研究されている。Atkinson Review (2005) は、サービス生産のコストを集計するのではなく、サービスが提供される人数を計測することによりアウトプットを直接的に計測することを勧告しているが、それに加えて、サービスの質を計測することを推奨している。Eurostat (2001) も直接的計測と質の調整を勧告している。イギリス統計局は、公的部門について実際に直接的手法によるアウトプットを質の調整を施した上で計測し公表している (Office for National Statistics, 2007, 2008, 2015)。こうした動きは、イギ

1) 日本大学

2) 東京医科歯科大学

リスの研究者達の研究に依拠している (Dawson, et al., 2005). より最近では, ヨーロッパの研究者達は, 診療報酬の包括払制度である DRG システムを医療アウトプットの直接的計測にどのように利用することができるかを研究し始めている (Pierdzioch, 2008 及び Street, Scheller-Kreinsen and Busse, 2010). アメリカにおいては, デフレーター計測における質の調整に関して先駆的な研究が 1990 年代にとりわけ National Bureau of Economic Research を中心に行われた (Cutler and Berndt, eds. 2001). この研究は, 病院のプロファイリングの研究の伝統を踏まえて医療の質を計測しようとした. 国民所得統計における医療アウトプットの計測の実際的な方向性については Abraham and Mackie (2005) にとりまとめられている. こうした研究の流れに沿って, アメリカ商務省経済分析局は疾病の治療のデータに基づいて医療のサテライト勘定を構築しようとしている (Dun, et al., 2015).

上述の EU 諸国における医療アウトプットの直接的計測の方向性は, アクティビティに基づくアウトプット指数 (activity-based output index) と呼ばれるもので, 医療サービスを等質的なアクティビティに細分化し, 各アクティビティの対象となる人数を計測することで医療の質を調整したアウトプットを計測しようとするものである. 細分化されたアクティビティの内部で医療の質は一様であれば, こうした細分化により質をコントロールすることができると考えられている. しかし, こうしたアプローチは問題がある. 細分化されたアクティビティを集計するためのウェイトには医療サービス提供のコストが使われているが, これが社会的な医療サービスの価値を表わすと仮定されている. しかし, 価格が規制されたり非競争的な市場においては, 均衡における限界費用は消費者の限界便益とかい離する. 一つの改善策は, 医療の質を調整するために DRG システムを利用することであろう. DRG の各カテゴリーが, その内部で均質な医療サービスを表現していて, かつ, それらの診療報酬が社会的な限界費用に等しければ, DRG 分類を使って医療の質を調整することができる. DRG 分類はそうしたものとなるべく設定されたものであるが, 実際には問題がある. 特に留意したいのは, DRG 分類は手術や主要な処置を鍵にして設定されており, 医療提供者がそうした手術や処置を選択するわけだが, その選択の中で同じ分類に属する患者が必ずしも同じ属性や重症度を持つとは限らないということである. 逆に, 同じ重症度を持つ患者が異なる DRG 分類に振り分けられることもある. 更に言えば, 医療提供者がある患者を治療するに際して, DRG システムの中で高い価格付けを得ているがその患者にとって有効であることが証明されていないような治療パターンを選んだ場合, 医療サービスとしては社会的に価値が無くても計測されたアウトプットは増加することになる.

したがって, 明示的な質の指標を使ったより直接的な計測方法を研究することが望まれる. アクティビティに基づくアウトプット指数は先に述べたように細分化により質をコントロールしようとし, より直接的なアウトプット指数は明示的な質の指標を使って医療の質をコントロールしようとする. それぞれ利点と問題点があるが, とりわけ, 質の調整の精緻化と頑健性との間にはトレード = オフがある. アクティビティに基づくアウトプット指数は, 特定のモデルに基づくものではないので頑健であるが, 細分化により等質性が十分に確保されているか疑問が残る. 明示的な質の指標は, 患者のアウトカム (医療成果) に影響を及ぼすリスク要因についての詳細な情報を織り込むことが出来れば精緻な質の調整が可能と期待される. しかし, 十分な情報を利用できるかどうかということと共に, リスク要因の影響を計測する際に特定のモデルに依存することになり, 頑健性を確保することは簡単ではない.

本論文は, 直接的計測の中でも明示的な質の指標に基づいた指数を作成するための方法論を構築することを目指す. 明示的な質の指標を計測することは, アメリカにおいて盛んに研究された病院のプロファイリングの流れに沿ったものである (Bronskill, et al., 2002, Burgess, et al., 2003, Landrum,

Bronskill and Normand, 2000 及び Normand, Glickman, and Gatsonis, 1997). Berndt, et al. (2001) や Cutler, et al. (1998) は、主として急性心筋梗塞（AMI）について、医療価格指数に統計的に計測した医療の質を織り込もうという先駆的な研である。本論文でも、詳細な情報が利用できた急性心筋梗塞（AMI）患者の入院治療に指数作成を限定する。考慮される医療の質としては、死亡率と合併症の発症率を採る。こうした死亡や合併症を考慮することにより医療のアウトプットの計測がどの程度影響を受けるかを検証する。さらに、死亡や合併症を考慮する際に、患者の入院時の併存症などのリスク要因を調整しない生（なま）の死亡率や合併症率を質として使った場合と、患者のリスク要因を調整した上で計測した標準化死亡率等を質として使った場合とでの程度の違いが生じるかを検討する。これは、患者の予後についてのリスク要因が医療成果（死亡や合併症の発生）に影響するが、各年の患者のリスク要因の軽重（ケース・ミックス）は大きく変化するため、純粋な医療の質を計測するためにはリスク要因の変動の影響を除去する必要があるためである。

本論文の構成は以下のとおり。第Ⅱ節において、医療の質を調整したアウトプット指数の計測のための枠組みを説明する。第Ⅲ節では、リスク調整をしない生の医療成果（死亡／合併症）を使って医療の質を調整したアウトプット指数を作成する。第Ⅳ節においては、リスク調整をした医療成果（標準化死亡率／合併症率）を使って医療の質を調整したアウトプット指数を作成する。第Ⅴ節で結論を述べる。

Ⅱ. 医療の質を調整した医療アウトプット計測の枠組み

医療アウトプットは3つの要素から成る。計測単位、医療の質及び質の評価である。

$$\boxed{\text{アウトプット}} = \boxed{\text{計測単位}} \times \boxed{\text{医療の質}} \times \boxed{\text{質の評価}}$$

本論文では Dawson, et al. (2005) and Castelli, et al. (2007) に従って質を調整した産出インデックスを計測する。アウトプットには、複数の下位のカテゴリーがあり、下位カテゴリーのアウトプットを合計して全体のアウトプット指数を計算する。計測単位は患者であり、 x_{it} により t 期における i 番目のアウトプットの患者の数を表わす。医療の質は死亡率等が高い／低いといった医療成果で測ることとし、 q_{ijt} で t 期における i 番目のアウトプットの j 番目の医療成果の量（j 番目の医療成果が得られた患者の数）を表わし、 π_{ij} で i 番目のアウトプットの j 番目の医療成果の評価を表わす。質の評価は本来は金銭価値の評価であるが、本論文では QOL とする（すなわち、死亡／合併症発症の場合の QOL の値）。この QOL（効用値）に QOL 1 単位の金銭価値をかける必要があるが、QOL 1 単位当たりの金額が時間を通じて一定であればその変化はゼロであるので、計算から除外して考えることができる。アウトプット指数は次のように計算される。

$$I_{t+1} = \frac{\sum_i x_{i,t+1} \sum_j q_{ij,t+1} \pi_{ij}}{\sum_i x_{it} \sum_j q_{ij,t} \pi_{ij}} \dots\dots\dots \textcircled{1}$$

本論文では AMI 入院のアウトプット指数を計測するのでアウトプットの種類は $i=AMI$ だけであり、患者数 x_{it} で上式の 2 番目の総和の中に繰り込んで医療成果の量（ q_{ijt} : j 番目の医療成果が得られた患者の数）を割ると、上記の計算式は j 番目の医療成果が得られた患者の割合を各医療成果の QOL で加重和を採ったものとなる。

計測単位は入院エピソードの数とする。病院ケアと診療所のケアとの連携は医療の質の決定要因として重要であるが、こうしたケアの連携についてのデータは無いためケア全体の評価は諦めざるを得ない。医療成果としては、死亡、合併症を伴う生存及び合併症を伴わない生存の3つを考慮する。それぞれの医療成果の健康効用は次のように想定した：死亡はゼロ、合併症を伴う生存は0.7、合併症を伴わない生存は0.9。

医療の質となる死亡や合併症は、患者の入院時の併存症などのリスク要因によって大きな影響を受ける。各年の患者のリスク要因は大きく変動するため、リスク要因を調整した質の計測を行う。このリスク要因の調整は計測に不確実性が伴うので、統計的手法として最尤法（MLE）を使った推定とベイズ的な階層モデルを使った推定を試みる。すなわち、質の計測は、大きく分けて、リスク要因を調整しない場合（生[なま]の医療の質）とリスク要因の調整を加えた場合の2通りを計測し、統計手法としてはMLEと階層モデルの2種類、さらに医療属性として死亡率だけを考慮した場合と合併症も含めた場合の2つの場合を計測する。イメージとしては次のようになる。

$$\boxed{\text{リスクを調整した医療の質}} = \boxed{\text{生(なま)の医療成果}} \times \boxed{\text{リスク調整}}$$

死亡
MLE

合併症
階層モデル

リスク調整は、リスク要因が医療成果（死亡／合併症）に与える影響を計測するロジスティック回帰により行った。推定方法としては最尤法（MLE）が標準的な手法であるが、年ごとに変動する係数をサンプルを分割して推定すると、少ないサンプル・サイズのため極端な値を採りがちである。ベイズ的階層モデルにおいては、各年の係数を推定する際に、他の年の情報を「借りて」全体の平均に近づけるという「縮小（shrinkage）」効果を持ち、極端な推定値を避けて妥当な値となることが期待できる。このアプローチは、ランダム効果として表現される各年の係数が、同じ平均と分散を持つ共通の確率分布に従うという意味で交換可能であることを仮定している。MLEと階層モデルの相違は、次のような例で説明できる（Spiegelhalter, et al., 2004）。確率変数 c_i がある標準正規分布に従うとする： $c_i \sim N(0, \sigma^2)$ 。MLEは $\sigma^2 = \infty$ という場合に相当し、階層モデルは $0 < \sigma^2 < \infty$ という場合に相当する。分散が有限であるため、階層モデルにおける推定値は極端な値を採りにくい。

各病院で患者のリスク要因が異なるというケース・ミックスの調整をするためには、Timbie, et al. (2009)に従って間接的標準化（indirect standardization）を行った。これは、リスク要因と医療成果との関係を医療の質ととらえ、平均的な医療の質から想定される仮想的な医療成果（counterfactual outcomes）と現実の医療成果を比較して超過的な医療成果（超過死亡率／合併症率）を得て、それで平均的な医療成果を調整した標準化された医療成果（標準化死亡率／合併症率）を医療の質の指標とするものである。具体的な推定手順は後に述べる。

Ⅲ. 医療の質の調整Ⅰ：生の死亡率及び合併症率

本節では、リスク要因を調整しない生の死亡率及び合併症率を調整した2つのアウトプット指数を計測する。一つは、死亡率だけを調整したアウトプットであり、もう一つは死亡率に加えて合併症率も調

整したものである。最初に死亡率だけを調整した指数を作成するが、これは入院中に死亡した場合の効用をゼロとおき、生存して退院した人の効用は完全に健康な場合の効用1としたものである。次に、合併症も考慮した指数を作成する。生存して退院した患者は、その状態によって異なる生活の質（QOL）を持っている。これを医療の質に加味してアウトプットを計測する。退院時のQOLについてのデータは無いが、合併症についてのデータは存在する。合併症が発生するかどうかは入院中（や退院後）のQOLを規定する大きな要因の一つである。本論文では Timbie, et al. (2009) に従って、個別の合併症に効用値を割り当て、各患者がどの合併症を発症したかに応じて生存者の効用を推定する。

Ⅲ-1 データと基本統計量

データ収集の対象としたのは、調査に協力いただいた9病院において2004年4月1日から2007年3月31日までの間にAMIで入院した患者である。各病院は院内倫理委員会の承認を受けた。データ収集に当たっては構造化した調査票を作成した。質問票Ⅰでは患者の詳細な診断情報及び治療内容を、質問票Ⅱではレセプト情報及び担当医師のプロフィールを収集した。質問票Ⅲでは、当該病院におけるAMI治療に関する全体的な情報（年間症例数、PCIやCABGの施術数など）を収集した。質問票Ⅰは医師資格を持った東京医科歯科大学の非常勤講師が各病院の診療記録や退院録などから記入した。質問票Ⅱ及びⅢはレセプト情報へのアクセス権限のある各病院のスタッフが記入した。データの収集に当たってはグラント（JSPS KAKENHI 20K10372）を受けたことを記して謝したい。

本論文における分析対象は、サンプルの等質性を確保するためにST-Elevation AMIに限定した。実際、ACC/AHAによる診療ガイドラインも、ST-Elevation MIとNon-ST-Elevation MIとで別々のガイドラインとしている。

分析対象の病院は、すべての年で10症例以上のSTEMI患者を診療した病院に限ったため、9病院となった。サンプル数は全部で2631であり、各年の内訳は、2004年598、2005年は612、2006年は672、2007年は749である。

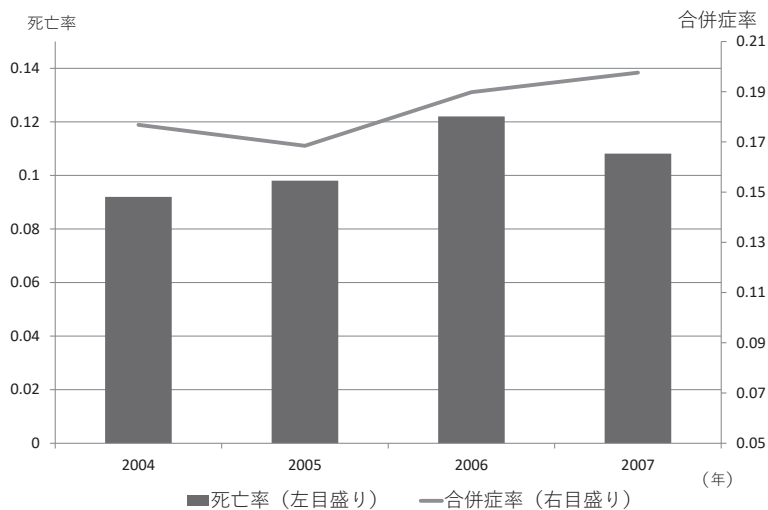
第1表は、全病院についての医療成果（死亡率及び合併症率）を示す。平均死亡率は10.6%、合併症率の平均は18.4%である。合併症の大部分は血管再建術の再施行である。生存患者の5%程度が心血管

第1表 医療成果：死亡率・合併症率

	全期間	2004	2005	2006	2007	(年)
死亡率	0.106	0.092	0.098	0.122	0.108	
合併症率	0.184	0.177	0.168	0.190	0.198	
合併症	0.046	0.042	0.024	0.053	0.061	
新たな梗塞	0.006	0.007	0.005	0.007	0.004	
心血管障害	0.033	0.026	0.013	0.044	0.045	
腎不全	0.010	0.011	0.007	0.005	0.015	
脳卒中	0.001	0.004	0.000	0.000	0.001	
心血管疎通術の再施行	0.183	0.176	0.169	0.192	0.192	
PCI再施行	0.164	0.153	0.142	0.181	0.177	
緊急CABG	0.022	0.029	0.032	0.010	0.016	
48時間以内の変化	0.002	0.002	0.005	0.000	0.000	
Cardiac arrest	0.000	0.000	0.000	0.000	0.000	
Shock	0.002	0.002	0.005	0.000	0.000	

(注) 合併症についての数値は生存退院のみに基づく。

第1図 死亡率と合併症率



障害や腎不全、新たな梗塞などの合併症を発症した。第1図はサンプル期間の死亡率及び合併症率を図で示したものである。院内死亡率は2004年から2006年まで上昇したが、2007年には若干低下した。合併症率は上昇トレンドを続けた。本論文において使用した合併症のリストは後に示す。

Ⅲ-2 生の死亡率及び合併症率に基づく医療の質の調整

リスク調整をしない医療の質を使った医療アウトプット指数を2種類作成する。一つは生の死亡率の変化のみを考慮したものであり、もう一つは生の死亡率に加えて生の合併症率を調整したものである。最初の生の死亡率だけを調整した指数においては、患者が入院中に死亡した場合の効用はゼロ、生きて退院した場合の効用は1とした。すなわち、生存退院において合併症が有るか無いかは区別していない。

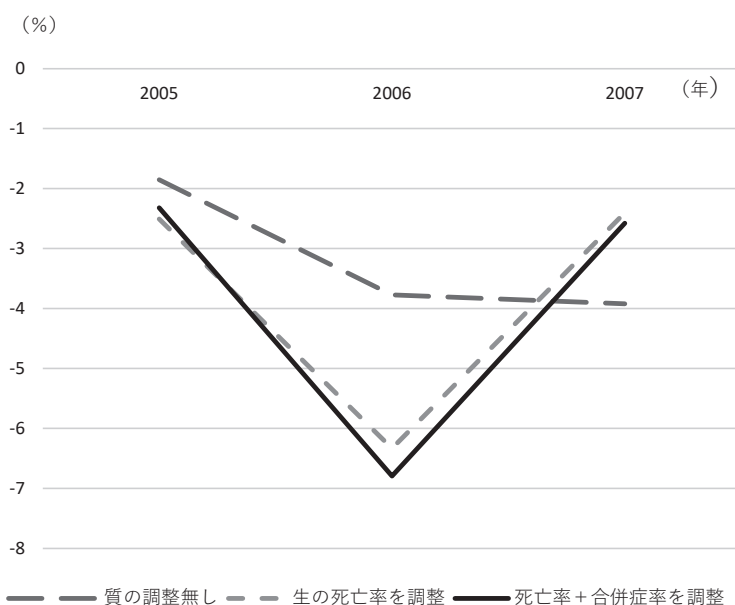
生の死亡率に加えて生の合併症率を調整した第2の指数においては、効用の値は次のように想定した(第2表)。院内死亡の場合はゼロ、生存退院の場合、合併症が無ければ0.9、合併症がある場合には0.7。合併症の無い生存退院の場合の効用値は Weintraub, et al. (2008) から採った。合併症が有る生存退院の場合は次のように推定した。計測に当たって考慮した合併症は、新たな心筋梗塞、脳卒中(stroke)、心血管障害(Cardiovascular disorder)、腎不全、PCI/CABGの再施行、及び、48時間以内の心停止もしくはショックである。脳卒中、腎不全、PCI/CABGの再施行の合併症があった場合の効用値は Timbie, et al. (2009) に与えられている。心血管障害及び48時間以内の心停止/ショックの合併症があった場合については良い推定値を見付けることができなかった。そこで、次のように推定した。まず、合併症として脳卒中、腎不全、PCI/CABGの再施行だけを採った場合の効用値を、各合併症を持つ患者数をウェイトとして計算した。その結果は0.77であった。次いで、心血管障害及び48時間以内の心停止/ショックがあった場合の効用値は、上記の限定された合併症を対象とした効用値より若干下方に修正して0.7とした。下方に修正したのは、心血管障害及び48時間以内の心停止/ショックは数は少ないが重篤な合併症と考えられるためである。

第2図は生の死亡率/合併症率を使って医療の質を調整したアウトプット指数の変化率を示したもの

第2表 合併症発症時の効用についての想定

	効用	参照文献
合併症無し	0.9	Weintraub, et al. (2008)
脳卒中	0.52	Timbie, et al. (2009)
腎不全	0.63	Timbie, et al. (2009)
PCI再施行/緊急CABG	0.78	Timbie, et al. (2009)

第2図 アウトプット指数の変化率：リスク調整をしない死亡率/合併症率による質の調整



である。質の調整をしていない指数の変化率も並べて示してある。質の調整は大きな影響を与えている。例えば2006年の変化率を見ると、死亡率で調整しない医療アウトプットは4%ほど減少したが、調整した場合には6%以上の減少となっている。合併症も併せて調整した場合を見ると、合併症を調整に含めない場合との差は大きくない。ただし、合併症有りの場合の効用値の推定が完全ではないことに鑑みると、合併症を調整することが重要ではないとは言い切れない。合併症の調整についてはさらに研究を進めるべきであろう。

少なくとも、死亡率について調整することは非常に重要であることは明白である。しかし、死亡率の変化は、医療の質の変化を反映するだけでなく、患者のリスク要因の変動にも大きな影響を受けている。そこで次のセクションにおいてリスク調整をしたアウトプット指数を作成する。その前に、第3表に入院時のリスク要因を示しておこう。全期間の平均では、平均年齢は68.9歳で、3分の1弱の患者が女性である。重症度分類であるKillip分類を見ると、半分程度がKillip分類1であり、4分の1がKillip 2、

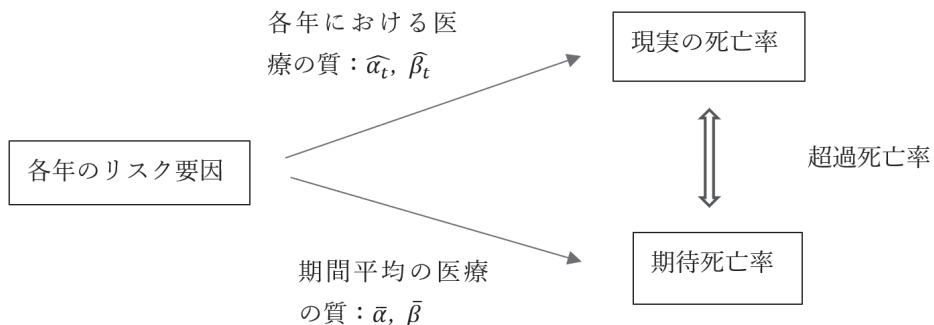
15%弱が Killip 3, 14%が Killip 4である。梗塞部位としては、左主幹部閉塞, 左脚ブロック (LBBB: Left Bundle Branch Block), 心室細動がそれぞれ4~6%程度を示める。半分以上の患者が高血圧であり, 40%弱が高脂血症, 3分の1弱が糖尿病の既往症があった。がん患者の割合は8%であった。以上のようなリスク要因が年々変動していることが読み取れよう。

IV. 医療の質の調整II: リスク調整

IV-1 推定モデル

入院時の併存症などのリスク要因を調整するため、医療成果を被説明変数とするロジスティック回帰式を推定した。推定方法としては、最尤法 (MLE) とベイズ的階層モデルの2つを試みた。先に触れたように、リスク調整の方法としては Timbie, et al. (2008) に従って間接的標準化 (indirect standardization) を利用した (第3図に間接的標準化のイメージを描いてある)。基本的な考え方は、現実の医療成果 (死亡率/合併症率) を期待される医療成果と比較することである。現実の医療成果は、当該年の患者のリスクの水準と、当該年の医療の質の水準から帰結することになる。期待される医療成果とは、当該年の患者のリスク水準に対して、ある仮想的な水準の医療の質を適用した時に得られると期待される医療成果である。ここでは、仮想的な医療の質としては、サンプル期間中の平均的な質を用いることにする。もし現実の医療成果が期待医療成果より良好であれば、医療の質は高いということになる。したがって、現実の医療成果と期待医療成果との差 (超過死亡率/合併症率) は医療の質の指標となる。間接的標準化の具体的な推定手順は次のとおりである。患者のリスクと対比した医療成果が医療の質であると考えられる。すなわち、リスクが高くても医療成果が高ければ、医療の質は高いと言える。患者のリスクと医療成果の関係を、医療成果を被説明変数としリスク要因を説明変数とするロジスティック回帰により推定する。各年ごとに回帰式を推定して得られた係数 (後掲②式の α_t 及び β_t) が各年のリスク要因と医療成果との関係であり、したがって、現実の医療の質を反映する。現実の医療の質は、この推定結果を用いて各患者についてそれぞれのリスク要因から医療成果の予測値を計算し、それを平均することにより得られる。もう一方の期待される医療成果は、医療成果の回帰式を各年ではなく全期間について推定してリスク要因と医療成果との平均的な関係を計測し (後掲③式の α 及び β)、それを各年の患者のリスク要因に適用して期待成果とする。現実の医療成果と期待される医療成果との差 (超過死亡率/合併症

第3図 間接的標準化のイメージ



率）が当該年の医療の質である。この超過的な医療成果（超過死亡率／合併症率）を全期間平均の医療成果に足すことにより標準化された医療成果（標準化死亡率／合併症率）を得ることになり、これを医療の質の指標として使うことが出来る。もし標準化死亡率／合併症率が低ければ、その年の医療の質は高いことになる。

(i) 最尤法（MLE）による推定モデル

間接的標準化を実行するためには2つの回帰式を推定する必要がある。一つは各年の係数が変化するモデルで、これが各年の医療の質を表わす。もう一つは全期間において係数が一定のモデルで、これが平均的な医療の質を表わす。医療成果 y_{it} は、もし患者 i が t 期においてイベント（死亡／合併症発症）があれば1、そうしたイベントが無ければゼロという値を採る。

各年の係数が変化するモデルは、次のようになる。

$$y_{it} = 1[\alpha_t + \beta_t \cdot x_{it} + u_{it} > 0] \quad \dots\dots\dots ②$$

ここで、 $1[\cdot]$ は指標関数で、その変数がプラスであれば1、マイナスであればゼロの値を採る。本論文では Timbie, et. al. (2009) に従って、リスク要因を表わす変数として重症度指数（severity index） x_{it} を利用する。これは、様々なリスク要因が医療成果に与える影響を集約的に表わす指数で、 $y_{it} = 1[\alpha_t + \sum_{p=1}^p \beta_p \cdot z_{ipt} + u_{it} > 0]$ というロジスティック回帰式を通じて計算する（ z_{ipt} は患者 i が t 期に p 番目のリスク要因を持っているかどうかを示すダミー変数）。重症度指数はこの推定結果から $x_{it} = \hat{\alpha} + \sum_{p=1}^p \hat{\beta}_p \cdot z_{ipt}$ として得ることが出来る。ここで用いたリスク要因は、年齢、性別、Killip 分類 2・3・4、左主幹部閉塞、左脚ブロック（LBBB：Left Bundle Branch Block）、左主幹部梗塞、心室細動、高血圧、高脂血症、糖尿病、心不全、心筋梗塞の既往、PCIの既往、CABGの既往、がん、出血傾向、腎不全、脳血管障害、動脈瘤、慢性閉塞性肺疾患（COPD：Chronic Obstructive Pulmonary Disease）である。これらのリスク要因は Krumholz, et al. (2006) で提案されたものに概ね沿ったものである。重症度指数の推定結果は A1 表に示した。

平均的な医療の質は、重症度指数 x_{it} を説明変数として、係数が一定の回帰式を通じて推定する。

$$y_{it} = 1[\alpha + \beta \cdot x_{it} + u_{it} > 0] \quad \dots\dots\dots ③$$

上記②及び③の回帰式を最尤法によって推定し、その推定結果から各患者の予測値を計算することにより、各年について、当該年の医療の質を反映した現実の医療成果と平均的な医療の質を反映した期待される医療成果を得る。両者の差分を各年について平均することにより各年の超過的な医療成果を計測し、それを平均的な医療成果に足すことにより各年の標準化された医療成果（標準化死亡率／合併症率）を得る。

(ii) ベイズ的階層モデル

上記②③と同様のロジスティック回帰式を、推定値の極端な変動を抑制するためにベイズ的階層モデルによって推定する。事前分布の想定は次のとおりである。2つの回帰係数 α_t 及び β_t は確率変数で、平均が μ 、precision matrix（分散・共分散行列の逆行列）が Σ^{-1} である2変数正規分布に従うと仮定する： $c_t \sim N(\mu, \Sigma^{-1})$ 、ただし、 $c_t \equiv \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix}$ 。各年のランダム効果 c_t は同一の確率分布に従うので、その値は全体の平均 μ に向けて縮小する（shrinkage）と期待される。

平均 μ は、平均0、分散100の正規分布に従うと仮定する。分散を100と仮定する理由は、Gelman and Hill (2007)の議論を踏まえて、いわゆる diffuse prior とすることを意図している。すなわち、分散が小さいと仮定すると、平均 μ の推定量は仮定された値(ゼロ)の近辺に分布することになる。これでは推定値をかなり決め打ちすることになってしまうので、事前分布としてはなるべく広い範囲で変動するように分散を大きく採ることが望ましい。Gelman and Hill (2007)の議論を要約すると次のようになる。ロジスティック回帰において、被説明変数の変動の程度は典型的には0.1とか1とかのオーダーであり、10とか100とかではない。したがって、推定される係数も絶対値で10を大きく上回ることは通常ない。彼らの分散のチョイスは 100^2 (標準偏差は100)である。これは、係数が概ね(-100, 100)の範囲に入ることを想定していることになる。本論文では、被説明変数の死亡率/合併症率自体0.1いくらかいかないかというほど小さいのでその変動のオーダーは通常0.1よりもかなり小さいことを考慮して、分散を 10^2 と仮定した。

さらに、precision matrix Ω は、scale matrix を持ち自由度が2のウィッシュャート分布に従うと仮定する： $\Sigma^{-1} \sim \text{Wishart}(\Omega, 2)$ 。自由度を2と仮定するのは、やはり diffuse prior という趣旨であり、また、 Ω は次数2の単位行列(I_2)と仮定した。

推定はマルコフ連鎖モンテカルロ法(MCMC: Markov chain Monte Carlo)によって行った。ソフトウェアとしてはWinBUGSを利用した。推定値が収束していることを確認するために、3つの平行する連鎖を走らせ、Gelman-Rubin 統計量を計算してチェックした。各連鎖とも最初に1万回繰り返しを行って収束させ、次いで事後分布から2万個のサンプルを採ってその平均値をパラメータの推定値とした。

IV-2 推定結果

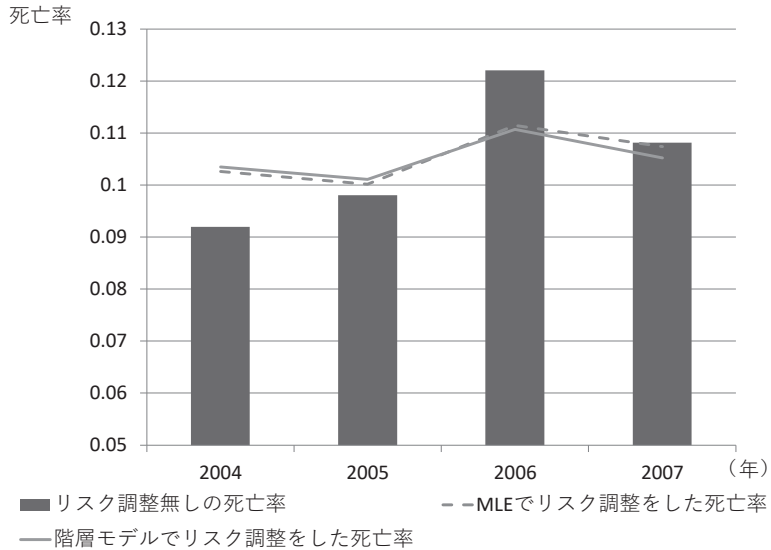
ロジスティック回帰式の推定結果を、MLEと階層モデルについて、死亡方程式は付表A2及びA3に、合併症方程式は付表A4及びA5に掲げてある。

第4図は、リスク調整をした標準化死亡率を生死亡率と並べて表示している。明らかに、リスク調整は死亡率の値に大きな影響を与えている。2004年及び2006年にはリスク調整前後の死亡率は1%ポイント以上異なる。リスク調整後の死亡率の変動は生の死亡率の変動よりも小さい。これはリスク調整がケース・ミックスのランダムな変動をうまく調整していることを意味する。MLEと階層モデルの推定方法の違いによるかい離は大きくはない。一つの可能性としては、各年のサンプル数が数百あるために、MLEによってもリスク要因の変動を十分に調整することが出来たということであろう。

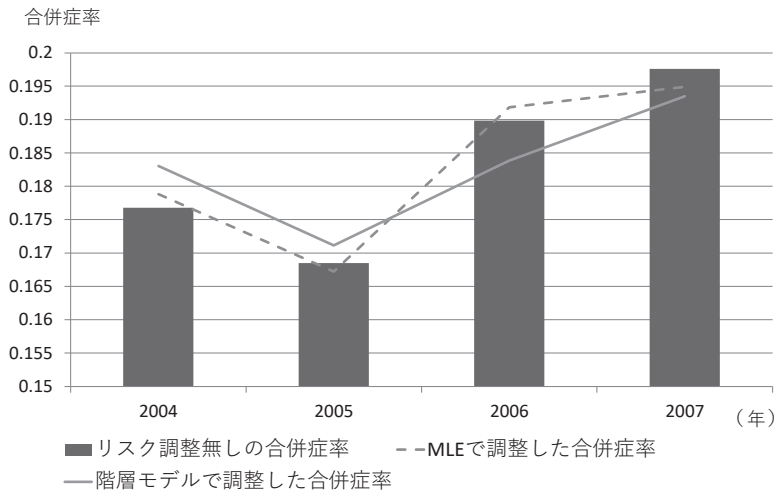
第5図は合併症率について、リスク調整をする前と後を比較している。MLEによるリスク調整は、合併症の場合には生の率とあまり変わらない。階層モデルによる調整では、リスク調整後の合併症率は生の率よりはるかにスムーズな動きをしており、リスク要因の変動の影響がきちんと除去されている。MLEと階層モデルのかい離を鑑みると、本論文における合併症の取り扱いはまだ初歩的なものである。ノイズの多いデータに対して階層モデルを使うことは意味があると考えられる。

第6図は、以上のリスク調整した標準化死亡率を使ってアウトプット指数を作成し、アウトプットの変化率を生死亡率を使った場合の変化率と並べて図示したものである。先の第4図からも予想されるように、リスク調整をした場合としない場合の変化率の差は大きい。例えば2006年には、リスク調整の有無でアウトプットの変化率は1%ポイント以上違ってくる。また、リスク調整をした変化率は生の変化率よりはるかにスムーズであり、リスク調整がきちんとなされていることを示している。リスク調

第4図 リスク調整前と後の死亡率

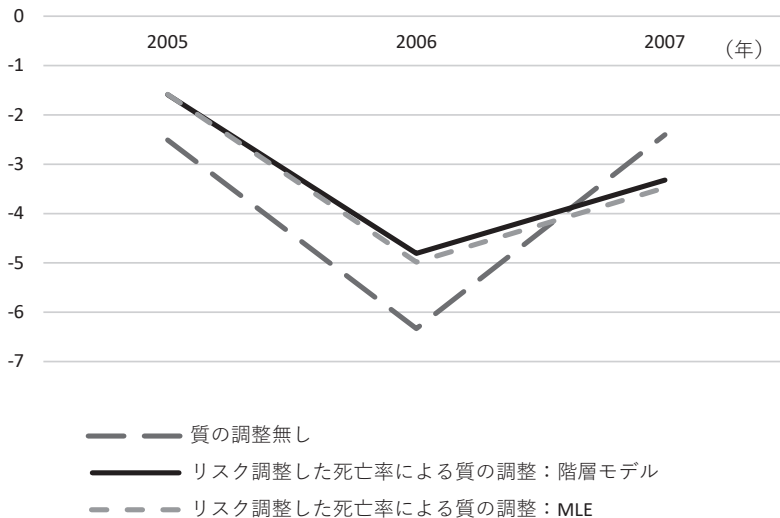


第5図 リスク調整前と後の合併症率



調整後の死亡率を使った場合のアウトプットの変化率の変動の程度は生の死亡率を使った時の変動の半分程度であり、生の死亡率を使った場合はアウトプットの変動は相当過大に推計されてしまうということになる。アウトプット計測においては、ケース・ミックスの変動の影響を除去するリスク調整が必須と言えよう。ただ、推定方法としてMLEを使った場合の変化率と階層モデルを使った場合の変化率では大きくは違っておらず、先に指摘したように、死亡率についてはこの程度のサンプル数があればMLEでも満足できる調整が可能かもしれない。

第6図 リスク調整した死亡率に基づく質の調整：階層モデルと MLE
(%)



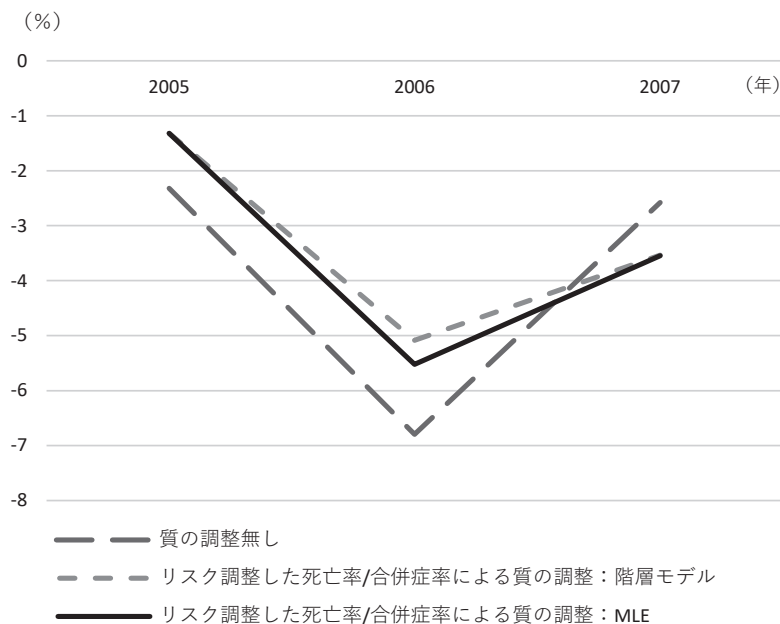
第7図は、死亡率だけでなく合併症率も合わせて調整した場合のアウトプット指数の変化率である。生の率との比較では死亡率の場合と同様にリスク調整が有効であることを示している。合併症を含めた場合、推定方法の違いによってアウトプット指数の変化率には違いが見られる。この点については第5図についてコメントしたとおりである。

第8図は、第6図と第7図を統合してどこで違いが生じているかを見やすくしたもので、リスク調整をした場合としない場合にアウトプットの変化率にどれだけ違いがあるか、そして、それぞれの場合に死亡率だけを使った場合と合併症も使った場合でどの程度違いがあるかを見たものである。図の下に大きく振れている2つの線は、2つともリスク調整をしていない場合の変化率である。上の方で振幅が小さい2つの線は、2つとも階層モデルによるリスク調整をした場合である。この2つのグループのかい離を見て分かるように、リスク調整の有無しが大きな影響を持っている。それに比べると、リスク調整の有無にかかわらず死亡率だけを考慮するか合併症も考慮するかはそこまで大きな影響はないと見られることもできるかもしれない。ただ、それでも変化率が0.5%ポイント程度は異なることもあるので、合併症をリスク調整の対象に含めることは重要な課題であると言えよう。

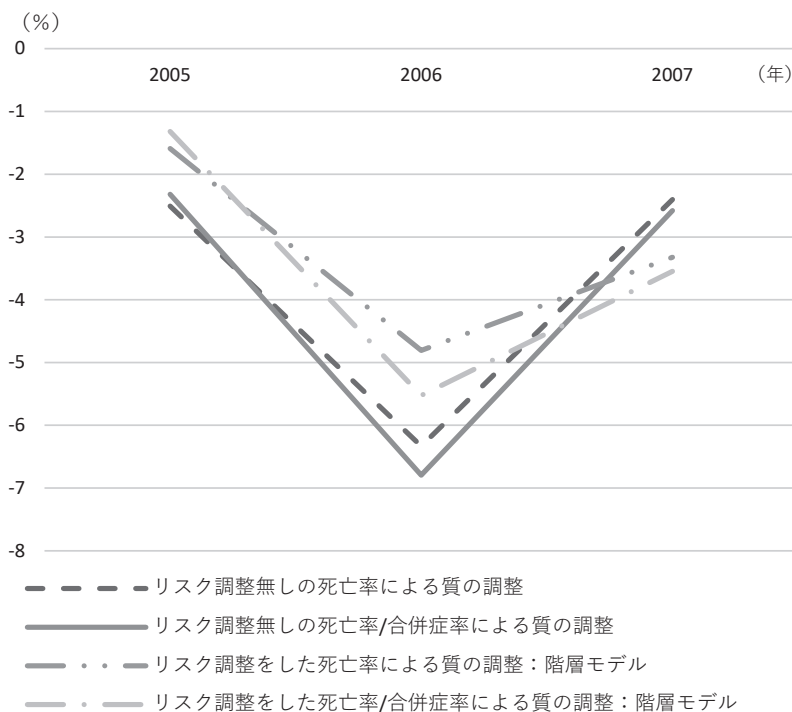
V. 結論

本論文では、医療サービスのアウトプットの計測に医療の質を織り込む方法論を検討した。素朴な質の調整としては、まず、患者のリスク要因を調整しない生(なま)の死亡率及び合併症率に基づく質の調整を施したアウトプット指数を作成した。生の医療成果に基づく医療の質の調整によってもアウトプット指数の変化率は顕著に変化し、質の調整が重要であることを示唆している。しかし、患者のリスク要因は年々変化し医療成果(死亡率/合併症率)の変動の大きな要因となるため、生の率の調整では

第7図 リスク調整した死亡率/合併症率に基づく質の調整：階層モデルと MLE



第8図 リスク調整の有無ごとの質の調整済みのアウトプット指数の変化率



不十分である。実際、生の医療成果に基づく質の調整では、アウトプット指数の変動が大きく、ケース・ミックスの変動によるノイズがかなり大きいことが窺える。

そこで、次に、リスク要因を調整した死亡率／合併症率に基づく質の調整を施したアウトプット指数を作成した。リスク調整はリスク要因と医療成果との関係を表現する統計的モデルを推定して実施したが、推定手法としては通常最尤法のほかに、推定結果の極端な変動を回避するようにベイズ的階層モデルによる推定も試みた。リスク調整した場合には生の医療成果に基づく場合に比べて遙かにスムーズな動きとなり、ケース・ミックスの変動によるノイズが除去されて医療のアウトプットをより正確に計測出来ていると言える。したがって、リスク要因を調整した医療成果に基づく質の調整を実施することが必須と言えよう。なお、医療成果として死亡率だけを調整した場合と合併症も併せて調整した場合とは違いは極端に大きなものではなかった（リスク調整をしない場合でも同様）。したがって、合併症についてのデータが得られない場合には死亡率だけでも調整することが必要であるが、それでも無視できない程度の違いは生じるので、可能な限り合併症も併せて調整することが望ましい。また、推定手法の違いについては、死亡率では最尤法と階層モデルでは大きな違いはなかったが、合併症のようにデータの不確実性が大きい場合には顕著な差が見られた。

参考文献

- Abraham, K. G., and C. Mackie, eds., *Beyond the Market: Designing Nonmarket Accounts for the United States*, The National Academies Press for the National Research Council, 2005.
- Atkinson, T., *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts*, 2005.
- Berndt, E. R., S. H. Busch and R. G. Frank, "Treatment Price Indexes for Acute Phase Major Depression," in D. Cutler and E. R. Berndt, eds., *Medical Care Output and Productivity*, University of Chicago Press, 463-505, 2001.
- Bronskill, S., S.-L. Normand, M. Landrum and R. Rosenheck, "Longitudinal Profiles of Health Care Providers," *Statistics in Medicine*, 21, 1067-1088, 2002.
- Burgess, J., C. Christiansen, S. Michalak and C. Morris, "Medical profiling: improving standards and risk adjustments using hierarchical models," *Journal of Health Economics*, 19, 291-309, 2003.
- Castelli, A., D. Dawson, H. Gravelle and A. Street, "Improving the Measurement of Health System Output Growth," *Health Economics*, 16, 1091-1107, 2007.
- Cutler, David and Ernst Berndt, eds. *Medical Care Output and Productivity*, University of Chicago Press, 2001.
- Cutler, D., M. McClellan, J. Newhouse and D. Remler, "Are Medical Prices Declining? Evidence from Heart Attack Treatment," *Quarterly Journal of Economics*, 113, 991-1024, 1998.
- Dawson D., H. Gravelle, M. O'Mahony, A. Street, M. Weale, A. Castelli, R. Jacobs, P. King, P. Loveridge, S. Martin, P. Stevens and L. Stokes, "Developing new approaches to measuring NHS outputs and productivity," *CHE Research Paper* 6, Centre for Health Economics, University of York, 2005.
- Dun, A., L. Rittmueller and B. Whitemire, "Introducing the New BEA Health Care Satellite Account," *Survey of Current Business*, January 1-21, 2015.
- Eurostat, *Handbook of Price and Volume Measures in National Accounts*, 2001.
- Gelman, A., and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, 2007.
- Krumholz H.M., Y. Wang, J. A. Mattera, Y. Wang, L. F. Han, M. J. Ingber, S. Roman, and S.-L. T. Normand, "An administrative claims model suitable for profiling hospital performance based upon 30-day mortality rates among patients with an acute myocardial infarction," *Circulation*, 113, 1683-1692, 2006.
- Landrum, M. B., S. E. Bronskill and S.-L. T. Normand S-LT, "Analytic methods for constructing cross-sectional profiles of health care providers," *Health Services and Outcomes Research Methodology*, 1, 23-47, 2000.
- Newhouse, J., *Pricing the Priceless: A Health Care Conundrum*, MIT Press, 2004.

- Normand, S.-L. T., M. E. Glickman and C. A. Gatsonis, "Statistical methods for profiling providers of medical care: issues and applications," *Journal of the American Statistical Association*, 92, 803-814, 1997.
- Pierdzioch, S., "Price and Volume Measurement of Hospital Services," Paper presented at the Working Party on National Account, OECD, 2008.
- Spiegelhalter, D. J., K. R. Abrams and J. P. Myles, *Bayesian Approaches to Clinical Trials and Health-Care*, Wiley, 2004.
- Street A., D. Scheller-Kreinsen, A. Geissler and R. Busse, "Determinants of hospital costs and performance variation: Methods, models and variables for the EuroDRG project," *Working Papers in Health Policy and Management*, 3, Universitätsverlag der TU Berlin, 2010.
- Timbie, J. W., J. Newhouse, M. Rosenthal, S.-L. Normand, "A Cost-Effectiveness Framework for Profiling the Value of Hospital Care," *Medical Decision Making*, 28, 419-434, 2008.
- Timbie, J. W., D. M. Shahian, P. Newhouse, M. B. Rosenthal and S.-L. T. Normand, "Composite measures for hospital quality using quality-adjusted life years," *Statistics in Medicine*, 28, 1238-1254, 2009.
- UK Department of Health, *Healthcare Output and Productivity: Accounting for Quality Change*, 2005.
- UK Office for National Statistics, *The ONS Productivity Handbook: A Statistical Overview and Guide*, 2007.
- UK Office for National Statistics, *Sources and Methods for Public Service Productivity: Health*, 2008.
- UK Office for National Statistics, *Public Service Productivity Estimates: Healthcare*, 2015.
- W. S. Weintraub, W. E. Boden, Z. Zhang, P. Kolm., Z. Zhang, J. A. Spertus, P. Hartigan, E. Veledar, C. Jurkowitz, J. Bowen, D. J. Maron, R. O'Rourke, M. Dada, K. K. Teo, R. Goeree, P. G. Barnett, "Cost-Effectiveness of Percutaneous Coronary Intervention in Optimally Treated Stable Coronary Patients," *Circulation Cardiovascular Quality and Outcomes*, 1, 12-20, 2008.

付 録

付表 A1 重症度指数の推定

	係数	標準誤差	t-値
年齢	0.051	0.008	6.5
性別	0.475	0.166	2.86
Killip2	1.663	0.369	4.5
Killip3	2.993	0.351	8.53
Killip4	3.668	0.355	10.33
左主幹部閉塞	0.230	0.278	0.83
左脚ブロック(LBBB)	0.300	0.225	1.34
心室細動	0.965	0.251	3.85
高血圧	-0.409	0.162	-2.53
高脂血症	-1.012	0.232	-4.36
糖尿病	0.149	0.168	0.89
心不全	0.040	0.230	0.17
心筋梗塞の既往歴	0.232	0.262	0.88
PCIの既往歴	-0.349	0.338	-1.03
CABGの既往歴	0.901	0.494	1.82
がん	0.728	0.226	3.23
出血	0.564	0.405	1.39
腎不全	0.478	0.195	2.45
脳血管障害	0.014	0.192	0.07
動脈瘤	0.574	0.372	1.54
慢性閉塞性肺疾患(COPD)	0.435	0.407	1.07
定数	-4.809	0.349	-13.8

サンプル数 = 2631
対数尤度 = -551.31298

付表 A2 死亡率関数の推定：MLE

(1) 各年についての推定

(年)	係数	標準誤差	p-値
2004			
定数	-4.169	0.413	0.000
重症度指数	1.179	0.143	0.000
サンプル数 =	598	対数尤度 = -104.47098	
2005			
定数	-4.141	0.385	0.000
重症度指数	1.153	0.132	0.000
サンプル数 =	612	対数尤度 = -114.54049	
2006			
定数	-3.614	0.309	0.000
重症度指数	1.014	0.103	0.000
サンプル数 =	672	対数尤度 = -146.82795	
2007			
定数	-3.169	0.234	0.000
重症度指数	0.804	0.084	0.000
サンプル数 =	749	対数尤度 = -181.08996	

(2) プールした推定

	係数	標準誤差	p-値
定数	-3.663	0.157	0.000
重症度指数	1.000	0.054	0.000
サンプル数 =	2631	対数尤度 = -551.31298	

付表 A3 死亡率関数の推定：ベイズ的階層モデル

(1) 各年についての推定

(年)				
2004	係数	標準偏差	2.5%	97.5%
定数	-4.078	0.353	-4.828	-3.439
重症度指数	1.149	0.125	0.919	1.411
サンプル数 = 598				
2005	係数	標準偏差	2.5%	97.5%
定数	-4.065	0.339	-4.780	-3.450
重症度指数	1.129	0.119	0.909	1.377
サンプル数 = 612				
2006	係数	標準偏差	2.5%	97.5%
定数	-3.673	0.287	-4.266	-3.139
重症度指数	1.032	0.097	0.850	1.230
サンプル数 = 672				
2007	係数	標準偏差	2.5%	97.5%
定数	-3.279	0.233	-3.751	-2.842
重症度指数	0.838	0.084	0.678	1.007
サンプル数 = 749				

(2) Pooled estimation

	係数	標準偏差	2.5%	97.5%
定数	-3.678	0.159	-3.997	-3.379
重症度指数	1.005	0.055	0.900	1.114
サンプル数 = 2631				

(注) 「2.5%」及び「97.5%」は、推定されたパラメーターの2.5パーセンタイル及び97.5パーセンタイルを示す。

付表 A4 合併症率関数の推定：MLE

(1) 各年についての推定

(年)				
2004	係数	標準誤差	p-値	
定数	-1.458	0.106	0.000	
重症度指数	0.150	0.049	0.002	
サンプル数 = 598 対数尤度 = -286.57969				
2005	係数	標準誤差	p-値	
定数	-1.538	0.108	0.000	
重症度指数	0.122	0.050	0.015	
サンプル数 = 612 対数尤度 = -286.8045				
2006	係数	標準誤差	p-値	
定数	-1.355	0.096	0.000	
重症度指数	0.032	0.042	0.437	
サンプル数 = 672 対数尤度 = -340.90053				
2007	係数	標準誤差	p-値	
定数	-1.336	0.090	0.000	
重症度指数	0.016	0.041	0.705	
サンプル数 = 749 対数尤度 = -383.16597				

(2) プールした推定

	係数	標準誤差	p-値
定数	-1.410	0.049	0.000
重症度指数	0.070	0.022	0.002
サンプル数 =	2631	対数尤度 = -1301.6275	

付表 A5 合併症率関数の推定：ベイズ的階層モデル

(1) 各年についての推定

(年)

年	係数	標準偏差	2.5%	97.5%
2004				
定数	-1.504	0.111	-1.724	-1.288
重症度指数	0.151	0.059	0.035	0.267
サンプル数 =	598			
2005				
定数	-1.585	0.112	-1.809	-1.369
重症度指数	0.088	0.060	-0.030	0.205
サンプル数 =	612			
2006				
定数	-1.486	0.108	-1.702	-1.279
重症度指数	-0.057	0.053	-0.162	0.045
サンプル数 =	672			
2007				
定数	-1.420	0.098	-1.614	-1.231
重症度指数	-0.027	0.049	-0.124	0.068
サンプル数 =	749			

(2) Pooled estimation

	係数	標準偏差	2.5%	97.5%
定数	-1.482	0.054	-1.589	-1.378
重症度指数	0.026	0.027	-0.027	0.080
サンプル数 =	2631			

(注) 「2.5%」及び「97.5%」は、推定されたパラメーターの2.5パーセンタイル及び97.5パーセンタイルを示す。