

ESRI Research Note No.12

非市場型サービス産業のアウトプット計測に関する研究のサーベイ
- 医療、教育、金融 -

乾 友彦・杉原 茂・川淵 孝一・空閑 信憲・池本 賢悟・石川 知宏

March 2010



内閣府経済社会総合研究所
Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

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。

なお、今後の修正が予定されるものであり、当研究所及び著者からの事前の許可なく論文を引用・転載することを禁止いたします。

(連絡先)総務部総務課 03-3581-0919 (直通)

非市場型サービス産業のアウトプット計測に関する研究のサーベイ：医療、教育、金融

内閣府 経済社会総合研究所 乾友彦、杉原茂、空閑信憲、池本賢悟、石川知宏
東京医科歯科大学 川淵孝一

1. はじめに

1990年代における日本経済の生産性上昇の停滞の主因は非製造業にあると、内閣府の『経済財政白書』(2001、2002)は指摘してきた。日本経済が直面している労働人口減少問題を乗り越えて、持続的な成長基盤を作り上げるためには、技術水準が世界のフロンティアにある製造業と比べて、未だキャッチ・アップの途上にあり、日本の国内総生産や総雇用の約8割を占める非製造業において、生産性を上昇させることが不可欠であることは言うまでもない。しかしその重要性に比べて、これまで非製造業(特に、サービス業)を対象にした緻密な実証研究はあまり行われてこなかった。世界的にみても非製造業に関する実証研究は発展途上にある(この問題は、Triplett and Bosworth(2004)、Berndt and Hulten(2007)における一連の論文で指摘されている)。サービス産業の統計整備や正確な生産性の計測はOECDやEU、米国においても最も重要な経済学、及び政策研究の課題の一つとして取り上げられている。しかし、サービス産業に関する統計的基盤が整備されてきていないため、現状の把握や政策立案などが適切に行うことができていない。

また、サービス産業の健全な発展と生産性の向上は、国民生活に直結する重要な政策課題である。例えば、最近の経済的・社会的問題をみても、金融バブルの崩壊、「医療崩壊」、ゆとり教育の見直しなど教育を巡る混乱など、サービス産業の動向が国民生活に極めて大きな影響を与えている。しかし、国民所得統計では、多くのサービス産業におけるアウトプットはインプットと等しいものとして計測されており、生産性上昇率は定義によりゼロとなるなど、アウトプットや生産性、価格が適切に計測されていない。その結果、医療、教育、金融などについては、そのサービスの質の評価がほとんどなされておらず、それらのサービス業がどれだけの価値を生み出しているのか、効率的な生産が行われているのか把握されていないまま、医療費の膨張など財政的な側面からのみ政策論議が行われる傾向が強い。

医療、教育、金融などの非市場型サービス産業は、アウトプットや生産性の計測が特に困難な分野である。非市場型サービス産業においては、そもそも市場価格が存在しないことや、価格が消費者の評価と乖離することがアウトプット等の計測を困難にしている。その背景にあるのは、情報の非対称性などの市場の不完全性、規制や公的保険等の存在、外部性などである。例えば、金融においては、情報の非対称性や預金保険等の公的保証の存在、医療においては、医師と患者の間の情報の非対称性、価格規制、公的保険の存在、教育においては、義務教育は無償、高等教育の授業料規制などがある。こうしたことから、

市場価格を使ったのではアウトプット等の計測が適切に行われなかったこととなる。なお、金融は通常は非市場型サービスには含まないが、金融機関等による資金仲介活動はそのリスクテイク行動の評価等が困難であり、生産活動を直接に観察することができないことから、本研究においては非市場サービスとして検討する。

近年、イギリスやアメリカにおいて、非市場型サービスのアウトプット等の計測を質の調整も織り込んで行う動きが活発化している。Atkinson Review(2005)においては、「インプット＝アウトプット」でなく、アウトプットを独立に計測し、質の調整もすべきだと提案された。その流れの中で、イギリスにおいては、先駆的に医療、教育、社会保護等幅広い分野についてアウトプットの直接的計測がSNA統計に取り入れられており、その後も質の調整方法について精力的に検討が進められている（UK Department of Health, 2005）。また、Eurostat(2001)でも非市場型サービスについても質の調整を行うこととされている。アメリカにおいては、医療サービスについて90年代後半以降NBERを拠点として、医療デフレーターや医療の生産性の分析に関して、死亡率や生活の質（QOL）を医療の質の指標とする意欲的なアプローチが開発され（Cutler and Berndt, eds., 2001）、現在、Bureau of Economic AnalysisでNIPAのサテライト勘定に取り込もうとする試みがなされている（Aizcorbe, Retus and Smith, 2007; Abraham, and Mackie, eds., 2005）。

ただ先述の通り、サービス産業は、先進国においてその重要性がましているのに拘らず、現在までの統計システムは製造業を主に想定してその活動を把握しており、サービス産業に関する本格的な統計整備は遅れてきた。そこで、日本のサービス産業の生産性が他の先進国に低い、またその成長率も低いと指摘されることが多いが、これは必ずしも十分に信頼できる統計（質の差異等を考慮した統計）に基づいて分析されているとは言い切れない。これまでサービス産業の生産性に関する研究があまり行われて来なかった理由は、Tily(2006)にも指摘されているように、データの不足、業種ごとの産出額の測定の難しさ、サービス産業の多様性などに原因があるものと考えられる。このような問題意識のもと、本稿ではサービス産業の多様性と質を考慮した生産性測定に関する研究がどこまで進展したかをサーベイする。特に米国、EU諸国で、サービス産業の生産性を実際にはどのように把握しているのかを纏める。ただし、個々のサービス産業をカバーする網羅的なサーベイではなく、非市場型であり今後の先進国経済において特に重要な位置を占めることが予想される産業である教育、医療、金融について生産額、生産性に関しての現在までの議論を整理し、今後の研究の方向性について、議論する。

本稿はこのサーベイを通じて、他の研究者の研究に資する情報を提供することに加えて、今後の研究の方向性について議論する。なお、本論文の構成は以下の通りである。2節、3節では、それぞれ医療、教育の生産、生産性の研究に関するサーベイ、第4節では現行の国民経済計算体系において先進各国が採用している金融の生産、付加価値を把握する「間接的に計測される金融仲介サービス」に関連する議論をサーベイする。第5節で、医療、教育、金融に関して今後の研究の方向性について議論する。

2. 医療のアウトプットの計測についてのサーベイ

2.1 はじめに

医療、教育、社会的保護（social protection）、治安・安全といった非市場サービスのアウトプットや生産性、価格を計測することは、国民の厚生水準を把握したり政府活動を評価するために非常に重要である。しかし、従来の SNA 等ではこれらサービスのアウトプット等を適切に計測できておらず、そのために現状の問題点が正しく理解されず、政策論議も混迷することがしばしば生じている。例えば医療についてみると、医療費の膨張ということのみが統計として提示され、それが生み出す効用や便益がどれほどの価値があるかは全く計測されていない。そのため、経済学的見地からすれば、限界的便益が限界コストを上回っているかどうか重要なはずであるが、ひたすら医療費を抑制すれば良いというような財政当局の見地ばかりが議論の焦点となり勝ちである。

しかし、例えば医療サービスについて、90 年代後半以降、アメリカの NBER を拠点として、医療価格や医療のコスト・ベネフィットの分析に関する意欲的なアプローチが開発され¹、現在、Bureau of Economic Analysis で NIPA のサテライト勘定に取り込もうとする試みがなされている。また、イギリスにおいても、Atkinson Review (2005) をスプリング・ボードとして、Department of Health が医療の質を取り込んだ医療アウトプットの計測に積極的に取り組んでいる。

非市場サービスのアウトプット/生産性/価格を適切に計測することは、理論的にも統計的にも非常にチャレンジングである。しかし、国民生活におけるこれらのサービスの重要性を考えると、この困難な課題を未解決のまま放置しておくことは許されない。この小論では、非市場サービスのアウトプット等の計測について、これまでの議論や試みのいくつかを医療を中心とし概観する。

2.2 非市場型サービスのアウトプットの計測の困難性²

一般に、アウトプットの計測には、(A)計測単位、(B)質の調整、(C)サービスの価値あるいは評価または支払い意思（value for money または willingness to pay）の 3 要素がある。すなわち、アウトプットは、図式的には、以下のように、3 要素の積として計測できる。

$$\boxed{\text{アウトプット}} = \boxed{\text{計測単位}} \times \boxed{\text{質の調整}} \times \boxed{\text{価値}}$$

¹ その集大成が Cutler and Berndt, eds. (2001) である。

² Berndt, Cutler, Frank, Griliches, Newhouse and Triplett (2001) 及び Triplett (2001) を下敷きにした。

特に、質の調整が不十分であると、次のような不都合が生じる。(i)質の上昇が価格の上昇として表れ、アウトプットや生産性の上昇にカウントされない。例えば、より高度の医療技術が利用可能になっても、単に高価格なだけで、実質的なアウトプットの増加としてはカウントされないという問題。(ii)質の低下がアウトプットの増加として表れる。例えば、医療における3分間診療は、患者数の増加をもたらすので、アウトプットの増加となってしまう。しかし、実は「粗診粗療」でサービスの質が低下し、アウトプットとして低下するように計測されるのが素直であろう。

非市場型サービスにおいては、特にこうした計測状の問題が顕著である。その理由としては、以下のようなことが挙げられる。

(1) 限界評価の問題

非市場型サービスのアウトプットの計測の困難さは、根本的には、消費者による評価(valuation)が得られないことにある。これには、次のようなケースがある。

(i)そもそも市場で取引されていないので市場価格がない。例えば、義務教育など無償で提供されるものがある。

(ii)価格(コスト)が消費者の限界的評価と乖離している

医療においては、情報の不完全性のため、患者でなく医師が治療方法を選択するが、医師による選択が患者の選好を完全に反映しているとは限らない(principal-agent problem)。また、健康保険により消費者が支払う価格が低下した場合、限界評価が非常に低くなる水準まで消費は増加する(モラル・ハザード)が、この限界評価は社会的なコストを下回っている。

(iii)外部経済/不経済の存在

治安・安全、伝染病の予防等公衆衛生などでは外部経済/不経済が存在するため、例えば、予防接種の料金は社会全体の便益を反映しない。

(2) 質の調整の問題

生産性や価格の計測に当たっては、質の調整をすることが重要である。例えば、新しい治療法のような技術進歩が生じると、同じ価格でもより効果的な治療(高い便益)が得られる。この場合、新しい治療法の価格が高くなっても、より高い効果が得られる分だけ価格は割り引いて計測する必要がある。

しかし、質の変化をどのように調整するかは難題だ。そもそも何をもって「質」とするか、その質をどのように計測するかは確立された定義や手法があるわけではない。この点については、後の2.3において、医療の質の計測を例にとって詳述する。

2.3 国際的ガイドライン、各種提案

非市場型サービスのアウトプットの計測については、国際的なSNAガイドラインや有識者等によりいくつかの提案がなされている。

(1) UN System of National Accounts (SNA 1993)

国連の SNA マニュアルにおいては、次のような指針が示されている。

- (A) Volume index は、財・サービスの量 (quantity) の加重平均とする。
- (B) ウェイトはそれぞれの価値 (value)。
- (C) Volume index は、実際に家計に提供される財・サービスの量であり、それから得られる便益や効用ではない。

(2) European System of Accounts (ESA) ; Eurostat Handbook Guidance

EU のマニュアルでは、Volume indicator として、次のようなものがあるとしている。

- (A) Input (例 : 雇用者数)
- (B) Activity (例 : 病院の手術数、警察の巡回数)
- (C) Output (例 : the amount of care received by a patient、the amount of teaching consumed by a pupil)
- (D) Outcome (例 : life expectancy、the level of education of the population)

EU マニュアルは、Volume indicator としては、上記のうち(C)Output が望ましいとしている。その際、次のような留意点を挙げている。

- (i)アウトプットの各タイプごとに基準年のコストをウェイトとして加重平均するべき。
- (ii)質を調整すべき。

(3) Atkinson Review (2005)

イギリス政府の諮問を受けて作成された Atkinson Review (2005) では、政府支出のアウトプットを質を調整して計測することが望ましいことを強調している。それは、次の原理として表明されている。

Principle B : 政府部門のアウトプットは質を調整して計測されるべきで、政府サービスのアウトカムに対する限界的貢献をカウントすべき。

Atkinson Review は Private services と Collective services の区別を取り上げているが、前者の Private services の質を計測する手法として、次のようなものを挙げている。

- (A)政府サービスを細分化する。これにより、質の変化は異なるカテゴリー間のシフトにより把握されることになる。
- (B)政府サービスの目的の達成度合いを計測する。
- (C)政府 activity の水準を、outcome への貢献度を調整して計測する。

後者の Collective services の方が計測は困難であるが、特に問題を複雑にするのは、民間サービスとの補完性を考慮すべきだとしていることであろう。これは、例えば、消防について考えると、人々が豊かになり家や家具等の価値が上昇すれば、消防サービスのアウトプットは上昇すると考えるのである。

Principle C : Complementarity between public and private output を考慮に入れるべき

なお、Atkinson Review は、以上のような一般原理に加え、個別分野として医療、教育、治安・安全、社会的保護について、Office for National Statistics に対して具体的な勧告を行っている。

2.4 医療の質の計測の困難性

医療の質の計測は非常に困難である。これには次のような理由がある。

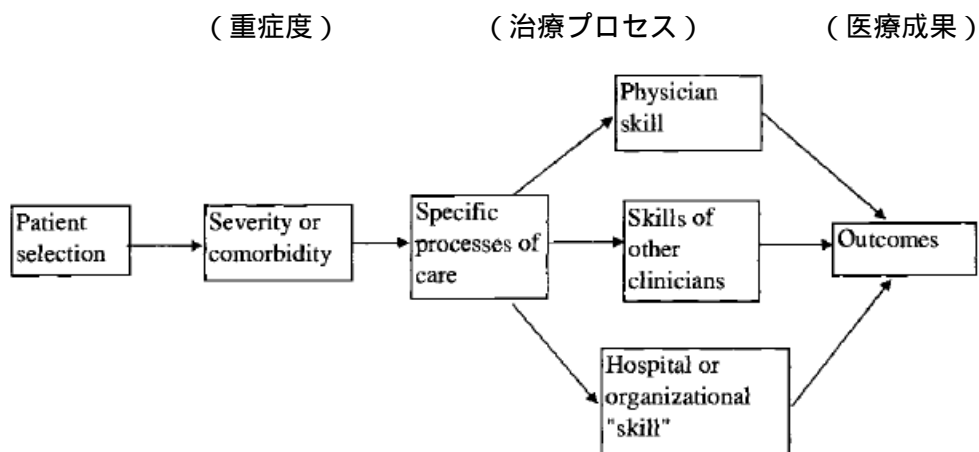
(i) 医療の質は、死亡率、QOL、満足度等々 multi-dimensional であり、何をもって医療の質とするか、また、これらをどのように評価し統合するかは合意がない。

(ii) 医療のアウトカムに影響する要因として、患者の属性、特に重症度がある。例えば、重症の患者であれば、いくら質の高い医療を提供してもかなりの患者が死亡してしまう可能性がある。一方、軽症の患者ばかりであれば、質の低い医療でもほとんどの患者が死亡しないであろう。このように、医療の質を計測するためには、重症度等を調整する必要がある。

(iii) さらに、医療成果（健康）に影響するものは、医療以外にたくさんある。すぐに思い付くものでも、生活環境、ライフ・スタイル、社会環境、経済状況・所得分配等がある。こうした複合的な影響から医療の貢献分だけを取り出すのは容易でない。

(iv) 臨床試験で死亡率等医療成果だけを計測することもあるが、臨床試験で計測されるのは、理想化された状況の下での医療成果であり、実際の医療の現場でそうした効果が実現している保障はない。一方、現実の医療現場でどのような効果が得られているかは、ほとんど計測されていない。

ここで、やや迂遠であるが、医療成果が産み出されるプロセスを考えてみよう。下図は、Halm, Lee and Chassin (2000)から採ったが、医療成果の決定要因を整理したものである。左端から見てみると、まず治療の対象となる患者が選ばれ、その患者はある重症度を持つ。当然ながら、重症であれば医療成果は悪くなる。次に、その患者に対して特定の治療法を適応するなどの治療プロセスが来る。ここでは、医師や医療組織の技術が治療効果を左右する。最終的に、それらの結果として医療成果が得られることになる。



以上を踏まえると、医療の質の調整においては、2つのことが重要である。一つは、どのように質を調整するかということであり、もう一つは、重症度を調整することである。

質の調整方法としては、直接に医療成果を計測するやり方と質の決定要因を利用して調整する方法がある。前者は、例えば、死亡率や QOL、患者満足度等を計測して医療の質とするやり方である。後者は、質を決定するような要因がどれだけ満たされているかを計測することにより、質の指標とするやり方である。例えば、AMI 患者に対して -ブロッカーを投与すれば予後が改善することが分かっているとすれば、 -ブロッカーを AMI 患者に投与した割合が高い病院ほど医療成果が良いと推測することができる。これは、アメリカなどで、プロセス指標として盛んに計測されている。プロセス指標については膨大な文献があるが、差し当たり、Spertus, et al. (2003) を参照。

また、質を計測する上で、重症度を調整することは重要である。重症であるほど予後は悪い。重症の患者を多く診察する病院と軽症の患者を多く診察する病院の死亡率等を単純に比較しても、正しく質の比較ができないことは明白であろう。

2.5 医療のアウトプットの計測例

以下では、医療のアウトプット計測の例を概観する。

(1) The OECD System of National Health Accounts

2000 年に取り組み開始した。現在、日本を含む 13 か国で作成している(第 1 表: Box 4)。項目としては、次のようなものを作成することとされている(第 2 表: OECD, 2000, Table 2)。

(A) Health care functions (ICHA-HC)

入院、デイケア、外来、在宅介護、処方薬、予防・公衆衛生、健康保険等。

(B) Health care service provider industries (ICHA-HP)

病院、介護施設、診療所(医科・歯科)、laboratory 等

(C) Sources of funding health care (ICHA-HF)

一般政府（社会保障とそれ以外）、民間部門（民間健康保険、自己負担、健康保険以外の対家計非営利法人等）

以上の OECD の体系は、概念的にはインプットである。また、医療の質は調整していない。例えば、3 分間診療で目一杯患者を診察すれば、医療アウトプットは増加する！

（２）イギリスの SNA における政府のアウトプットの推計

従前は「output=input」アプローチであったが、1998 年以降、アウトプットを直接計測するアプローチを相当取入れている（第 3 表：Atkinson Review, Table 2.1）。医療については、下記の(A)Cost-weighted output index を作成している（ただし、質の調整は行っていない）。

UK Department of Health. (2005) では、Atkinson Review を踏まえて、質の変化を調整して医療のアウトプットを計測している（この計測は Dawson, et al. (2005)に基づく。Castelli, Dawson, Gravelle and Street, 2007 も参照）、Castelli, Laudicella and Street. (2008)はさらに包括的な計測を試みている。

医療アウトプットの指標の種類としては、以下のようなものがある。

$$(A) \text{ Cost-weighted output index (CWOI): } I_{ct}^x = \frac{\sum_j x_{j,t+1} c_{jt}}{\sum_j x_{jt} c_{jt}}$$

これは、アウトプットの単位（ x ）をコスト（ c ）で加重平均したもの。使用するアウトプットの例としては、患者数（外来、入院）等がある。

$$(B) \text{ Quality-adjusted CWOI: } I_{qact}^x = \frac{\sum_j x_{j,t+1} \left(\frac{q_{j,t+1}}{q_{jt}} \right) c_{jt}}{\sum_j x_{jt} c_{jt}}$$

これは、質（ q ）を調整しコスト（ c ）で加重平均したもの。調整する質の例としては、死亡率、QOL、患者満足度、入院や手術までの待機日数等がある。

$$(C) \text{ Value-weighted output index (VWOI): } I_{yt}^x = \frac{\sum_j x_{j,t+1} \sum_k \pi_{kt} q_{kj,t+1}}{\sum_j x_{jt} \sum_k \pi_{kt} q_{kj,t}}$$

これは、アウトプットの単位（ x ）×質（ q ）×価値（ π ）として計測するもの。つまり、医療サービスの評価まで含めたものである。

主な計測結果は以下のとおり。

- (i) 医療の質として、死亡率や待機日数を使った Quality-adjusted CWOI を試算したもののうち、死亡率を使った結果は、第 4 表：UKDH, 2005, Table 1&Figure 2 にある。
- (ii) いくつかの手術について QOL を計測して、Value-weighted OI を試算したもの。
- (iii) 心血管障害について、Value-weighted OI を試算したものが第 5 表：UKDH, 2005, Table 2 及び Table 3 である。
- (iv) 以下のような patient experience を調整した Quality-adjusted CWOI を試算したものが第 6 表 UKDH, 2005, Table 5 である。具体的には、(i)Access and waiting; (ii)Safe, high-quality co-ordinated care; (iii)Better information, more choice; (iv)Building closer relationships; (v)Clean, friendly, comfortable place to be を調整している。なお、Public Services Agreement では、様々な指標もモニタリングしている（第 7 表：PSA, Annex A）。

以上のような計測の問題として、医療の質の改善は、医療支出だけによるのではないということがある（後述）。

（ 3 ） US Bureau of Economic Analysis

(A) Aizcorbe, et al. (2007)及び National Research Council(2008)

BEA は医療に関するサテライト勘定を作成しているが、Cutler 等の学術研究を踏まえ、より精緻な手法による計測を検討中である。特に、アウトプットを疾病レベルで把握することが眼目である。具体的には、デフレーターを計測に当たって、疾病の治療方法の代替を考慮することになる。ただし、治療による健康の改善は考慮しない。

(B) Cutler のプロジェクト：Rosen and Cutler (2007)

より意欲的に、Cutler を中心に、健康状態を計測して National Health Account を作成しようというプロジェクトも立ち上げられた（第 8 表：Rosen and Cutler, 2007, Table 1）。推計の概略は以下のとおり。

(i) Health impairments and symptoms から QALY を推計する。QALY 推計の手順は、Stewart, Susan, Rebecca Woodward, Rosen and Cutler(2007)に従う。Health impairments and symptoms についてのデータとしては、national survey が存在する。

(ii) 疾病ごとの支出を計測

(iii) Disease model を利用して、疾病レベルで医療の価値を計測し、それらの疾病を足し上げて、全体としての医療の生産性を推計する。Disease model とは、health inputs と outputs を関係付けるモデルと説明があるだけであるが、cost-effectiveness 分析のことと思われる。

(4) Cutler and Richardson (1999)

人的資本理論に基づき「医療の価値」を計測した。Health capital、医療支出、Net benefit の算式は下記のとおり。

$$Health\ Capital_t(s) = V \cdot \left[\sum_{k=0}^{\infty} \frac{H_t(s+k)}{(1+r)^k} \right]$$

V : Value of a year in perfect health

$H_t(s)$: Health of a person of age s in year t ,

where $H_t(s) \equiv \Pr[Alive\ at\ s] \cdot Q_t(s)$

with $Q_t(s)$: Average Quality of Life among those who are alive at age s in year t

$$Medical\ Spending_t(s) = \left[\sum_{k=0}^{\infty} \frac{\Pr_t(Alive\ at\ s+k | s) \cdot m_t(s+k)}{(1+r)^k} \right]$$

$m_t(s)$: Medical spending for people of age s in year t

Net Benefits of Medical Technology _{t_0, t_1} (s)

$$= \beta \cdot \Delta Health\ Capital_{t_0, t_1}(s) - \Delta Medical\ Spending_{t_0, t_1}(s)$$

ここで、 β : Share of changes in health capital resulting from changes in medical technology.

しかし、 β の値を知ることは難しい。そこで、次の指標を考える。

$$Effectiveness\ Ratio(s) = \left(\frac{\Delta Medical\ Spending_{t_0, t_1}(s)}{\Delta Health\ Capital_{t_0, t_1}(s)} \right)$$

この有効性指標が小さければ、医療技術はコストに見合うものである可能性が高い。

質の計測には、次の2つのアプローチを採っている。

(A) Years of Life approach : 生きていれば $H = 1$ 、死亡なら $H = 0$

(B) Quality-Adjusted Life Years approach : 疾病ごとの prevalence を QALY をウエイトとして加重平均 (第9表 : Cutler and Richardson, 1999, Table 5.2)。疾病ごとの QALY は、self-reported health (excellent/very good/good/fair/poor) を被説明変数とし、疾病を説明変数とする ordered probit に基づいて推計。

命の価値 V は 10 万ドルと仮定している。ただし、「命の価値」についてはコンセンサスがあるわけではないことに留意する必要がある。

計測結果は以下のとおり。

- (A) Effectiveness ratio は、YOL アプローチでも QALY アプローチでも 30%前後。
- (B) 疾患別にみると、心血管障害は、0 歳の人で 64%、65 歳の人で 8%、がんは、ともに health capital の変化がマイナスになる（死亡率が上昇する）ので、コスト効率的でない。

問題点としては、次のようなことが指摘できる。

- (A)医療成果の改善がすべて医療支出の増加によるものではない。例えば、食事、ライフ・スタイル、社会環境、経済状態や所得格差等も重要である。Cutler and Richardson (1999)の医療の価値は相当に過大評価されている可能性が高い。
- (B) また、医療成果は平均的には高くても、限界的には低い可能性が高い。
- (C) QALY の計測上の問題点

これらの問題点をより詳細に述べると以下のとおり。

(A)について

その後、医療の貢献部分だけを取り出して cost-effectiveness が計算されている。Cutler, McClellan and Newhouse (1999) 及び Cutler and McClellan (2001) において AMI 等の個別疾患について計測した後、全医療については Cutler, Rosen and Vijan (2006) が計測した。

ここでの鍵は、心疾患等の死亡率の低下に対する医療技術の寄与の計算である。例えば、Cutler, McClellan and Newhouse (1999) や Ford, et al. (2007)では、AMI 患者の死亡率の低下の半分程度が医療技術によるものとしている（第 10 表：Cutler, McClellan and Newhouse, 1999, Table 9）。この条件の下では、医療成果の向上によるベネフィットは医療コストを大きく上回る。また、Cutler, Rosen and Vijan(2006)は、医療による貢献を 50%として cost-effectiveness を計算し、医療支出は十分「割りに合う」(成果が見合う)ものとした³。

しかし、これらの研究においては、医療技術の寄与を計算する際、個々の医療技術の臨床試験における死亡率低下効果を足し上げて計算している。しかし、このような医療の貢献分の計測については以下のような問題がある。

- (i) 実際の医療の現場で臨床試験により検出された効果が現実に発揮されているとは限らない。これに関する研究は非常に多い。以下、若干のものを挙げる。
 - a. 病院や医師の医療成果を比較する Medical profiling⁴と呼ばれる研究分野における研究結

³ ただし、医療の貢献が 25%であれば、1990 年以降の医療支出は「割りに合わない」。また、Garber and Skinner (2008) は、貢献の割合が 50%であっても、将来の医療成果をディスカウントすれば（Cutler, Rosen and Vijan, 2006 はディスカウントしていない）、cost-effectiveness は大きく悪化すると指摘している。

⁴ Medical profiling について方法論的な議論をしているペーパーとして、Goldstein and Spiegelhalter (1996) や Burgess, Christiansen, Michalak and Morris (2003) などがある。実際の適用例はアメリカについては数多くあるので省略するが、日本についての例として川淵・杉原(2005b)がある。なお、McClellan and Staiger(1999)や Landrum, Normand and Rosenheck (2003) は、複数のアウトカム指標から潜在変数（因子分析）を用いて統合し、単一のパ

果によると、同じ治療方法でも、病院や医師ごとに医療成果にばらつきがあることが明らかになっている。例えば、ニューヨーク州の Cardiac Surgery/CABG の死亡率について、Marshall and Spiegelhalter (2001) 及び Bronskill, et al. (2002) 等、日本の AMI 患者に対するステント等高度医療技術について、川淵・杉原 (2005a) がある。

b. 病院や医師の手術等の量(経験)が医療成果に影響を及ぼす現象は量的効果と呼ばれる。この仮説が正しければ、量や経験が少ないと医療成果が悪化することになる。Hannan, et al. (1989) が草分けと言えるが、Birkmeyer, et al. (2002) 及び Birkmeyer, et al. (2003) が包括的に各種術式について分析した。日本については、Kawabuchi and Sugihara (2006) が AMI 患者に対する PTCA の成果を検証している。

c. さらに、そもそも適切でない患者に適用すると、どんな医療技術も有害無益である。これは、Appropriateness の研究と呼ばれ、Brook 等 RAND 研究所プロジェクトが代表的な研究である。その結果によると、Over-use(本来適用すべきでない患者に適用)が相当みられる。この場合、臨床試験で効果があると判定された医療技術も医療成果を改善しないことになる。

(ii) また、単純に個々の効果を足し上げているが、実は効果は additive ではないのではないか。例えば、治療方法 A 単独で死亡率を 5% 低下させ、B 単独で 5% 低下させるとする。A と B を併用した場合、死亡率は 10% 低下するかというと、そうでもない可能性が高い。

(iii) さらに、技術ごとにみると、死亡率低下に大きく寄与しているのはアスピリンや β -ブロッカーといった比較的安価な薬剤の使用率の上昇であり、PTCA や CABG といった高額の高高度治療方法が大きく寄与しているわけではない。したがって、より細かい個々の術式や薬剤レベルでは、コスト・ベネフィットが悪い可能性がある。

(B)医療技術の限界生産性の逡減について (第 1 図: Garber and Skinner, 2008, Figure 3)

新技術が開発された時、当初は選ばれた患者に適応されるので効果が高いが、その後普及してより広い範囲の患者に適用されるようになると、必ずしも理想的な適用対象ではないので、効果が小さくなる。こうした現象は、「flat-of-the-curve 効果」と呼ばれる。これについても膨大な研究があるが、以下、代表的なものを列記する。

(i) McClellan, McNeil and Newhouse (1994): 心血管のカテーテル検査について検証

(ii) Skinner, Staiger and Fisher (2006): AMI 患者の治療について検証

(iii) Skinner and Staiger (2009): 病院レベルで AMI 患者の治療について検証

(iv) Garber and Skinner (2008): 国際比較によって、必ずしもアメリカが効率性という点でアウトライヤーでないとした。しかし、これは他の国と同程度に限界効率性が低いということである可能性が高い。

特に、Skinner, Staiger and Fisher (2006)が、AMI 治療は 1990 年代前半まではコスト効率的

パフォーマンス指標を作成している。

であったが、90年代後半以降は効率的ではなくなったと指摘していることは重要である(第2図: Skinner, Staiger and Fisher, 2006, Exhibit 2)。Cutler, Rosen and Vijan (2006)の結果においても、70年代には cost-effectiveness 比率が低いが、80年代、90年代とかなり上昇している(第3図: Cutler, Rosen and Vijan, 2006, Figure 1)。これらの結果は、医療技術の適応範囲が拡大するとともに、限界的な便益が低下していることを示唆している。

いったん導入された治療方法は、その後は特段の検証を経ないで継続・拡大適用される傾向がある。医学における臨床試験や cost-effectiveness 分析は、通常、導入の時に計測されるだけである。医療の価値や生産性を継続的に計測する仕組みを構築する必要がある。

以上の(A)や(B)の要因を含めて、より多く医療費をかけても医療成果が改善しないということが繰り返し実証的に示されている。Regional variations の研究では、Fisher, et al. (2003) や Skinner, Fisher and Wennberg (2005) が代表的な成果である。DEA、確率的フロンティア・モデル等による効率性の計測については、Hollingworth (2003)が包括的にサーベイしている。

(C)QALY の計測については、次のような問題があると考えられる。

- (i) 個人ごとの scaling の違い: 同じ状態でも人により感じ方は異なる。
- (ii) 現状への心理的 adjustment: 病気の人、健康な人が推測するほど不幸ではない。Over time に QOL を計測する時には特に問題になり得ると思われる。
- (iii) Censored data となっている: 健康状態が excellent の人は、いくら医療支出を増やしても、それ以上改善しないことになってしまう。

以下、Cutler and Richardson (1999)に関連した文献をいくつか挙げる。

(A) Fukui and Iwamoto (2004)

Cutler and Richardson (1999)の手法を日本に適用した(Fukui and Iwamoto, 2004, Table 6)。結果は、日本全体として、医療成果はコストに見合うものとしている。ただし、この医療成果の中には、純粋に医療支出によるのでない部分が含まれている。

そこで、地域別データを使って推計している。これがこの論文の特長と言える。地域別のパネル・データを使うことにより、地域に固有の要因による医療成果の改善を除去できる。つまり、各県ごとの医療支出の変化と医療成果の変化との相関をみると、純粋に医療支出だけの効果を検証できると期待される。その結果は、医療支出と医療成果の間には相関がないということであった(第4図: Fukui and Iwamoto, 2004, Figure 6)。すなわち、医療の価値はゼロ!(ただし、まだ途中段階の結果だとしてあまり強調していない)。

この論文の問題点は、以下のとおり。

- (i) 主要部分については、Cutler and Richardson (1999)と同じく、医療以外の貢献分を含んでいることは、医療の価値を過大評価している。
- (ii) 最後の部分で、健康増進に対する医療の貢献分を統計的工夫で計測しようとしたところ

に特徴があるが、変化率の地域間のばらつきにより識別しようとしても医療以外の要素がかなり残ると考えられる。例えば、地域間で共通の健康改善要因やリスク・ファクター（生活水準の向上、ライフ・スタイルの変化、社会環境の変化等）

また、QALY ウェイトを計測する際に、Cutler and Richardson (1999)は障害レベルの推計を行っているのに対して、Fukui and Iwamoto (2004)では、おそらくデータの制約のためであろうが、疾病レベルで推計している。そのため、例えば、同じ糖尿病でも、失明や四肢切断等を併発した場合とそうでない場合では QALY に大きな差があると考えられるが、Fukui and Iwamoto (2004)ではそうした違いを考慮することができないのではないかと。

(B) Murphy and Topel (2003) : 健康と長寿の経済価値

健康 H_t が乗法的に効用 $u(c_t, l_t)$ を増加させる生涯期待効用関数 :

$\int_a^\infty H_t u(c_t, l_t) \tilde{S}(t, a) e^{-\rho(t-a)} dt$ から出発して ($\tilde{S}(t, a)$ は年齢 a の人が t 期まで生存する確率 :

生存関数) 健康と長寿化の価値を導出した。

各人は次の目的関数を最大化する。ただし、 μ は予算制約にかかるラグランジュ乗数、 $A(a)$ は a 歳の時の保有資産。

$$U(a) = \int_a^\infty \{H_t u(c_t, l_t) \tilde{S}(t, a) e^{-\rho(t-a)} + \mu [y_t - c_t] e^{-r(t-a)}\} \tilde{S}(t, a) dt + \mu \cdot A(a)$$

α を健康・長寿に影響を与えるファクター（「医療知識」）とする。医療の進歩の価値は、

$$U(a) \equiv \frac{U'_\alpha(a)}{\mu}$$

$$= \int_a^\infty \left[\frac{u(c_t, l_t)}{u'_c} - c_t + y_t \right] S(t, a) \Gamma_\alpha(t, a) dt + \int_a^\infty \frac{H'_\alpha(t) u(c_t, l_t)}{H_t u'_c} S(t, a) dt$$

ただし、 $\Gamma = \ln[\tilde{S}(t, a)]$ これを人口で合計すれば、社会的な支払意思となる。

Murphy and Topel (2003) はアメリカについて、河越 (2009) は日本について、長寿化の価値は医療費に比べて非常に大きいと結論している。例えば、河越 (2009) の場合、医療費は健康・長寿化の価値の 10 倍に相当するので、医療費の増加は費用便益分析上は十分合理的な支出増であったとしている。

ただし、両者とも、長寿化のうちどれだけが医療費の増加によるかを計測しているわけではない。すなわち、「健康の価値」ではあるが、「医療の価値」ではない。

また、計測したのは長寿化の価値のみで、QALY の上昇の効果は試算されていない。

(C) なお、Murphy and Topel (2003) と同タイプの試算として Nordhaus (2003) もある。

(5) WHO, Health Systems: Improving Performance (2000)

191 か国について、health system の目標達成度を示す指標を作成し、総合化した。

3 つの health system の目標と総合化に際してのそれぞれのウエイトは以下のとおり。

(i) Health (disability-adjusted life expectancy)	Total	50%
	Level	25%
	Distribution	25%
(ii) Responsiveness	Total	25%
	Level	12.5%
	Distribution	12.5%
(iii) Fairness	Total (=Distribution)	25%

このうち、Responsiveness のサブカテゴリーとそれぞれのウエイトは以下のとおり。

(i) Respect for Persons	50%
Respect for dignity	16.7%
Confidentiality	16.7%
Autonomy	16.7%
(ii) Clinical orientation	50%
Prompt attention	20%
Quality of amenities	15%
Access to social support networks	10%
Choice of provider	5%

上記のウエイトは、125 か国 1006 人 (半分は WHO 職員) に対するアンケート調査に基づく。

以上の達成度指標を被説明変数として、fixed effect “frontier” model (パネル・データの固定効果モデル) により各国 health expenditure の効率性を計測した (Evans, et al., 2000)。

$$y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 (X_{1it})^2 + \beta_4 (X_{2it})^2 + \beta_5 (X_{1it})(X_{2it}) + v_{it}$$

さらに、その推定結果に基づき、各国の医療システムの効率性をランキングした (第 12 表 : WHO, 2000, Annex Table 10)。

この分析に対する批判としては、fixed effect の中には、医療システムの効率性以外の様々な要因が含まれているということがある。そこで、Greene (2005) は、固定効果モデルに確率的フロンティア・モデルを組み合わせた true fixed effects stochastic frontier model を提唱した。

$$y_{it} = \alpha_i + \beta' x_{it} + v_{it} + u_{it}, \quad \text{where } u_{it} \sim |U_{it}| \text{ and } U_{it} \sim N(0, \sigma_v^2) \perp v_{it}$$

$$v_{it} \sim N(0, \sigma_v^2)$$

なお、WHO では、医療システムの効率性評価に関してさらなる精緻化を図っている (Murray and Evans, eds., 2003)

(6) 医療価格の試算

(A) Cutler, McClellan, Newhouse and Remler (1998)

彼らは、医療価格として、質を調整した Cost of Living Index を計測した。

効用関数を次ぎのように設定する。

$$U[H(m(t_1)), Y - p(t_1) \cdot m(t_1) - C] = U[H(m(t_0)), Y - p(t_0) \cdot m(t_0)]$$

ここで、 H : Health capital、 m : 医療支出 (実質)、 Y : 所得金額、 p : 医療価格、 C : 当期の価格で当期の医療サービスを受けることに対する (基準時の価格と医療サービスと比較しての) 支払意思。これは、補償変分である。

効用関数を t_0 の回りにテイラー展開して (x は消費を表す)

$$C = (U_H H_m / U_x) dm - d(p \cdot m)$$

Cost of Living Index は、基準時の所得で基準化して、次のように表される。

$$COL_{t_0, t_1} = 1 - C / Y_0$$

医療成果 (H) は死亡率とする。

試算すると、real Cost of Living Index (GDP デフレーターでデフレート) は年率 1% で低下しているという驚くべき結果が得られた (第 5 図 : Cutler, McClellan, Newhouse and Remler, 1998, Figure V)

(B) Berndt, Busch and Frank (2001) は、大うつ病の急性期治療の価格指数を算出した。

3. 教育のアウトプットの計測についてのサーベイ⁵

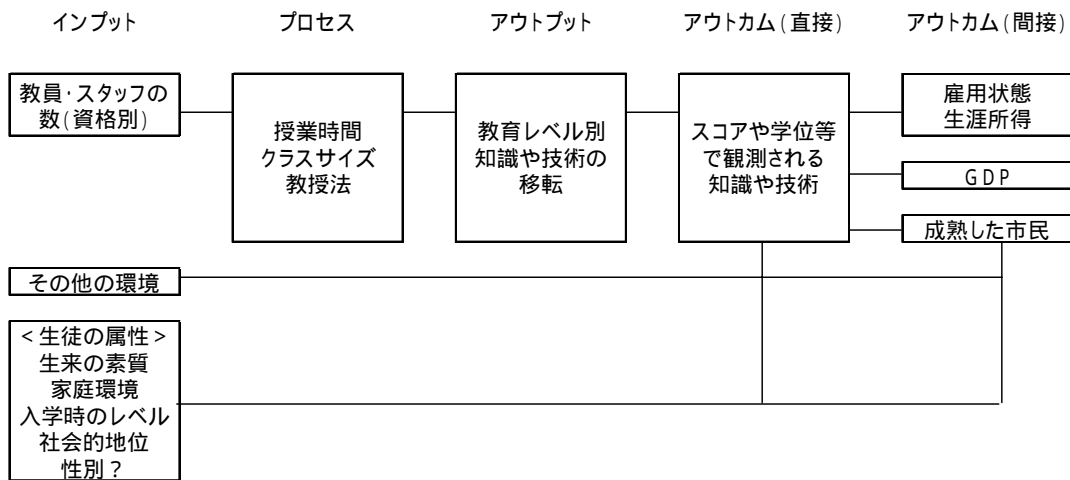
3.1 教育のアウトプットの内容

教育の生産性を計測するためには、アウトプットの内容を定義しなければならない。まず、国際機関でどのように定義されているかをみてみよう。

EUROSTAT (2001)では、テストスコアや進学率のように知識の移転が実現したかどうかを測った指標は、教育サービスに対するアウトカム（成果）であるとして、アウトプットとは区別している。そして、アウトプットを測る最もよい指標は、教育を受けるために児童生徒が過ごした時間（児童生徒数×授業時間）であるとする。他方、児童生徒の受けられる教育サービスの内容は、同じ時間数であっても、教員の経験年数や学校の設備等によって一様ではないと考えられる。そのため、アウトプットをより正確に算出するためには、教育内容の差異を反映させるような調整を行う必要がある。EUROSTAT(2001)では、こうした差異を反映させるに当たり、テストスコアや進学率のようなアウトカムに基づいた指標と、教育サービスの内容を直接評価した学校監査等の指標の両方の利用可能性を示している。

OECD(2008b)はこの質の調整の中身を、生徒のアウトカムに変化を与えた教育サービスの貢献部分とした。そしてEUROSTAT(2001)の考え方を発展させて、教育のインプット・アウトプット・アウトカムの流れを下図のように整理している。

<教育のインプット・アウトプット・アウトカム>



以上を踏まえると、教育のアウトプットの計測方法としては、次の3つがある。

- (A)学力など直接的な結果
- (B)所得など間接的な結果（人的資本アプローチ）
- (C)教育の質の決定要因を使った質の調整：クラス・サイズなど

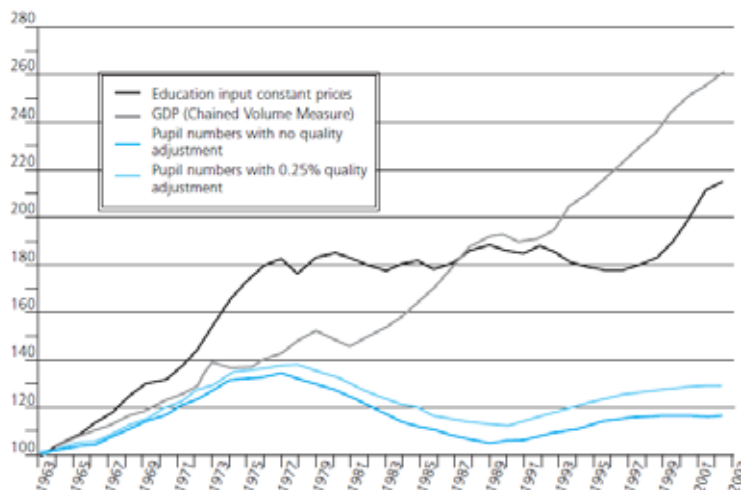
⁵本節は、主に澄田知子(2009)に基づく。

以下では、教育のアウトプット計測に関し先進的な取組みをしているイギリス、教育経済学の蓄積のあるアメリカを中心に、諸外国の状況を概観する⁶。

3.2 イギリスにおける公的教育のアウトプット計測

(1) 現状

イギリスでは1998年のSNA(国民経済計算体系)から教育のアウトプットの推計に産出指数を採用している。これは、公的教育部門に属する4つのタイプ(保育所、小学校、中学校、特殊学校)の生徒数に、学校のタイプ毎のコストでウエイト付けした指標である。さらに、小学校と中学校の生徒数に対して、年0.25%の質の調整を行っている。この調整レートは、イングランドで行われたGeneral Certificate of Secondary Education(GCSE)⁷の1994年から1997年の平均点の伸びを根拠にしている。



(出所) Atkinson, Tony (2005) *Atkinson Review*

2005年に公表されたAtkinson Reviewの中で、こうした教育のアウトプットの計測方法に関し、(A)産出指数の作成において生徒数をそのまま用いるのではなく出席率で調整すること、(B)質の調整は毎年の変化を反映できるよう改善すること、等が提示された。これを受け、(A)については2005年のSNAから出席率による調整が行われている。他方、(B)については、改善策が議論されているところである。

(2) 学力テストの利用方法の改善

以上のように、イギリスにおいては、主として学力という直接的な結果を利用してアウ

⁶ その他に、オランダ、ノルウェー、オーストラリアでも教育のアウトプットについての研究が行われている。これらについては、澄田(2009)を参照。

⁷ 義務教育の最終学年(11年生、16歳)で行われる到達度テスト。

トアウトを計測している。Atkinson Review での指摘を受け、Office for National Statistics(ONS)を中心に、質の調整方法を改善するための議論が行われている。以下に主なものを挙げる。

(A) 各年の変化の反映

(i) GCSE を用いた手法の改善

現行方式では、質の調整に固定レートが採用されているが、毎年の変化が反映されるよう、同じ GCSE のスコアを利用しつつも、(A) 毎年の好成績者が全体に占める割合の変化、(B) 毎年の平均点の変化、の 2 通りの方法が検討されている。

(ii) より細分化されたテストスコアの利用

11 年生での到達度しか反映されないという GCSE の問題を解消するため、イングランドで行われている義務教育の 4 つの段階⁸毎のテストスコアを利用し、それぞれの段階での到達度を反映させる方法が試みられている。さらにこの方法でも、それぞれの段階の最終年次での到達度しか測れないため、各段階のスコアの改善をそれに要した年数に振り分けて、毎年の質の調整レートを算出するという方法も試みられている。

(B) 教育水準局による学校監査結果の利用

イギリスでは、教育水準局(Office for Standards in Education, Ofsted)が 1992 年から学校監査を行っている。現在までに数次にわたって監査方式の変更があったため、時系列比較の可能な 2000 年以降のデータを用いて、学校の質の変化を計測することが試みられている。

図 1 は、以上のような様々な質の調整方法によるアウトプットの違いを示している。これを見ると、2001 年まではどの調整方法もほぼ同様の傾向(上昇)を示しているが、それ以降異なる動きとなっており、特に細分化されたテストスコアを用いた調整方法(KS テスト利用)はアウトプットの減少が見られる。

(3) 地域別データの導入

2007 年、ONS は地域別の教育水準を正確に反映させるため、スコットランドで行われているテストデータ等を利用して、教育のアウトプットの再推計を行った。その結果、公式統計である Blue Book 2006 で 9.6% 増とされていた 1996 年から 2005 年の教育のアウトプットは 6.6% 増と伸びが縮小した。これは、GDP を 0.2% 程度押し下げる規模に相当する。図 2 は、イギリスの地域毎のアウトプットを比較したものである。これにより、1996 年以降、地域によって教育のアウトプットに大きな違いが出ていることが分かった。

(4) 教育サービスの対象範囲の拡大

ONS は、2008 年、教育のアウトプットについて、従来の中高等教育以下に加えて高等教育にまで対象範囲を拡大し、再推計を行った。その結果が、図 3 である。1996 年から 2007 年の伸びは 5.3% と、高等教育を含まない場合(3.8%) に比べて高くなった。

⁸ KS1(5-7 歳)、KS2(7-11 歳)、KS3(11-14 歳)、KS4(14-16 歳)の 4 段階。

(5) 新たなアウトカムの定義の提案

Every Child Matters(ECM)に沿った、新たな質の評価についてのフレームワークの構築が提案されている。ECMは2003年9月、政府が議会に対して提出した緑書で、教育が達成すべき評価指標として、(A)健康であること、(B)安全であること、(C)楽しみかつ目標を達成すること、(D)前向きな活動に寄与すること、(E)経済的に幸せになれること、の5つが提示されている。イギリスでは、教育の成果としてテストスコアが過大評価されることに対する批判も高まっており、折しも全国テストの採点処理過程で問題が生じたことから、KS3(9年生)のテストが廃止されることとなった。教育のアウトプットの推計に関しても、授業時間以外に学校で過ごす時間の評価や、芸術・スポーツなどのカリキュラムの評価が課題となっている。

3.3 アメリカにおける公的教育のアウトプット計測

アメリカのSNAでは教育のアウトプットはインプット(教育予算)に基づき作成されているが、近年の教員の学歴や経験年数の変化、生徒数の伸びに対する予算の伸びの変化、クラスサイズの変化等に着目し、質を調整する試みが行われている。これらは、教育の決定要因を使って質を調整する試みと言える。

そこで、まず、教育の質に関連する要素についての研究結果を概観しよう。下の表は、Hanushek(2006)のサーベイの中で、教育の質に関連する資源投入の効果を調べた研究について、有意にプラスの効果を検出した研究とそうでない研究の分布をまとめたものである。これをみると、先生・生徒比率を始め、教育の質にプラスの影響を与えると確言できる政策は意外にないということが分かる。

Percentage distribution of estimated effect of key resources on student performance, based on 376 studies

Resources	Number of estimates	Statistically significant		Statistically insignificant
		Positive	Negative	
Real classroom resources				
Teacher-pupil ratio	276	14%	14%	72%
Teacher education	170	9	5	86
Teacher experience	206	29	5	66
Financial aggregates				
Teacher salary	118	20	7	73
Expenditure per pupil	163	27	7	66
Other				
Facilities	91	9	5	86
Administration	75	12	5	83

(出所) Hanushek(2006)

以下、いくつかの提案を概観しよう。

(1) 初等・中等教育のアウトプットの計測

Christian (2006)は、以下のような研究成果を基に、様々な調整方法でアウトプット計測した。

(A) 教員の質に関する研究

アメリカでは、経験が2年未満の教員の割合が、1980年の5.3%から2000年に8.8%へと上昇した。経験2年以上の教員が教えた生徒の方が、2年未満の教員の生徒よりテストスコアが0.12~0.19ポイント高くなるとの分析⁹や、5年以上の経験のある教員と比べて、全く経験のない教員の下では0.13ポイント、経験1~2年の教員の下では0.06ポイント、経験3~5年の教員の下では0.03ポイント、テストスコアが低くなる¹⁰等の研究がある。

(B) クラスサイズ、生徒/先生比率に関する研究

アメリカでは、1980年から2001年の間に、生徒/先生比率が18.7から15.9へ低下した。研究結果をみると、生徒/先生比率が18/1より高くなると生徒の成績が下がるとの分析¹¹や、低学年で経済的に恵まれない子どもでは特に、クラスサイズを17人以下にすると教育効果が高まるとする分析¹²がある。しかしその後は、クラスサイズは生徒の成績に有意な影響を及ぼさないとする結果も出ている¹³。その理由としては、クラスサイズはアットランダムに決まっているわけではなく、特別な配慮の必要な児童・生徒を小さなクラスに配置していること、クラスサイズを縮小するために、経験の浅い教員を雇わなければならないこと等が指摘されている。

(C) 両親の教育レベルに関する研究

生徒の達成度に関する学校以外の要因を排除する一つの方法として、National Assessment of Educational Progress(NAEP)のデータに基づき、両親の教育レベル(高校未満、高卒、専門学校卒、大卒、不明)によって調整した場合、アウトプットは0.07~0.22%低下した¹⁴。

以上のような研究成果を基に、様々な調整方法¹⁵でアウトプット計測した結果を比較したものが、次の表である。これを見ると、様々な調整をかけることによりアウトプットの伸びは高まるが、インプットの伸びに比べると約半分以下にとどまっていることが分かる。

⁹ Rivkin, Hanushek and Kain(2002)。

¹⁰ Rivkin et al(2005)。

¹¹ Ferguson(1991)。

¹² Finn(1998)。

¹³ Hanushek(1998, 2002)。

¹⁴ Christian(2006)。

¹⁵ 教員の質及びクラスサイズによる質の調整の下限値は「生徒/先生比率の10%減少または経験2年未満の教員の10%減少は、教育の質を1%向上させる」と仮定、上限値は「生徒/先生比率の10%減少または経験2年未満の教員の10%減少は、教育の質を3.3%向上させる」と仮定。

テストスコアによる質の調整の下限値は「17歳時点でのNAEPの数学のスコアが1ポイント改善した場合、教育の質が8.3%向上した」と仮定、上限値は「17歳時点でのNAEPの数学のスコアが1ポイント改善した場合、教育の質が27.5%向上した」と仮定。

<調整方法によるアウトプットの違い(1980-2001)>

(%)

インプット指標	
州及び地方政府の教育支出	2.41
アウトプット指標	
生徒数	0.73
生徒数(特殊教育を受ける生徒 = 一般生徒 × 2)	0.85
教員の質及びクラスサイズによる質の調整	
下限調整値	0.92
上限調整値	1.06
テストスコアによる質の調整	
下限調整値	0.97
上限調整値	1.22
下限調整値を両親の教育レベルにより調整	0.90
上限調整値を両親の教育レベルにより調整	1.00

(出所) Christian(2006)

(2) 高等教育に関するアウトプットの算出

高等教育について、予算に基づくインプット指数と、生徒数をはじめとするアウトプット指数とを比較したのが次の表である。初等・中等教育に比べてその差は小さいものの、やはりインプットの伸びがアウトプットの伸びを大幅に上回っている。

調整方法としては、まず生徒数につき、パートタイムをフルタイムの3分の1として計算しているが、結果は両者とも伸びが約1.2%で差異が見られない。生徒数に代えて学位取得数を使うと若干伸びが高まり、1.4%となる。学位に関し、その種類によってウエイトをかけたもの¹⁶も作成されているが、伸びは約1.4%と差異が見られない。生徒数と学位取得数を組み合わせた統合指数では、伸びは両者の中間となる。一般に、高等教育ではカリキュラムなど個々の生徒が受ける教育が一樣ではないので、質の調整が非常に困難であると言える。

<高等教育のアウトプット(1980-2001)>

(%)

インプット指標	
州及び地方政府教育予算	2.33
アウトプット指標	
生徒数	1.23
生徒数(パートタイム × 3 = フルタイム)	1.23
学位取得数	1.40
学位取得数(学位別調整)	1.39
生徒数及び学位取得数の統合指数	1.27

(出所) Christian(2006)

¹⁶ 学位取得数の学位別調整のウエイトは、associate's degrees=2、bachelor's degrees =4、master's degrees=4、first-professional degrees =6、doctoral degrees =8。

(3) 初等・中等・高等教育の統合指数

初等・中等・高等教育全体の動きを見たものが次の表である。インプットの伸びは2.5%であるのに対し、アウトプット指数で、単純な生徒数の伸びは1.1%、これに質の調整¹⁷をかけると、伸びは1.5%となる。

何らかの質の調整をかけると、単純な生徒数の伸びよりはアウトプットが上昇することが確認できるものの、インプット指数の伸びと比べるとかなり小さい。この理由としては、私立学校に進んだ生徒の数及び質の変化が反映されていないことなどが考えられる。

<初等・中等・高等教育の統合アウトプット>

(%)

	1980-1990	1990-2001	1980-2001
インプット指標	2.20	2.71	2.47
アウトプット指標			
生徒数	0.56	1.56	1.08
生徒数(質の調整)	1.01	1.86	1.45

(出所) Christian(2006)

(4) 人的資本理論に基づくアプローチ

アメリカでは、以上のような算出方法のほか、人的資本アプローチによるものもある¹⁸。ここでは、生徒数や授業時間は教育のアウトプットではなくインプットと考えられるべきであり、生徒数に対して質の調整を行ってもその結果は非常に小さく、教育の成果(人的資本の蓄積)は明確にならないとされる。そして教育のアウトプットをインプットと全く独立に計測するため、教育投資(教育のアウトプット)は生涯所得に変化を与えるものとして、受けた教育の違いによる生涯所得の差を推計している。

¹⁷ 初等・中等教育についてはNAEPの算数・数学のテストスコアで、高等教育については学位取得数で調整。

¹⁸ Fraumeni(2006)。

4. 金融¹⁹

4.1. 金融仲介業のアウトプット

資金の借り手と貸し手のニーズを仲介する金融仲介業を対象としたアウトプット計測手法である FISIM (Financial Intermediation Services Indirectly Measured : 間接的に計測される金融仲介サービス) という概念が 93SNA で提唱された。現在では、OECD 諸国においては概ね採用されているものの、我が国においては現時点では、まだ正式系列への採用は見送られており、国際比較という観点から GDP が過小評価されている状態である²⁰。しかし FISIM の現状の計測方法に関しては、リスクテイクやタームプレミアムなど金融サービスのアウトプットの範囲について論点も残されている²¹。

4.2. 金融のアウトプットおよびインプットに関するミクロ経済学的研究

ミクロ経済学の金融機関(銀行)の生産性、効率性に関する実証研究が積み重ねて来られており、そのなかで金融機関のアウトプット、インプットに関しては様々なアプローチが取られてきているが、代表的なアプローチは、Production Approach と Intermediation Approach の 2 つがあり、Intermediation Approach では更に 3 つのアプローチ ((A)Asset Approach、(B)User Cost Approach、(C)Value-Added Approach) に分類される。

Production Approach は、金融活動の金融仲介活動よりも、直接的なサービス活動に注目して、そのアウトプットを、預金の件数、融資件数等によって把握し、インプットに関しては通常の生産関数と同様に、労働、資本、中間投入として、預金はインプットとして考慮しない。

一方、近年の実証研究で多く採用されるアプローチが、Intermediation Approach で、金融活動の中心を金融仲介業と考えて、預金者と投資家の間を仲介する機能を重視する²²。Asset Approach においては、金融業のバランスシートの資産はアウトプット、負債はインプットとみなすアプローチである。そこで、融資や他の主要金融資産をアウトプットと考え、預金はアウトプットではなく、預金を含めた負債は全てインプットとみなす。これに労働や資本等を加えたものをインプットと考える。

User Cost Approach は、本節で議論する FISIM と似通った考え方であり、金融機関の収益に貢献するものがアウトプットとして定義する。金融資産の収益がその資産の機会費用 (reference rate) を上回るとき、あるいは金融負債のコストがその負債の機会費用を下回るときに、それぞれの資産、負債をアウトプットとみなす。また逆の場合はインプットと

¹⁹本節は、主に庄司啓史(2009)に基づく。

²⁰ 参考系列として採用されているが、統計委員会において、本系列への導入に関して議論が進行している。

²¹本稿における FISIM の考え方は、特に断りがない限り我が国の参考系列で準拠している、EU 統計局基準とする。

²² Intermediation Approach の詳細は、Berger and Humphrey(1992)を参照。

して考える。ただ、このケースはこの機会費用を定義し、実際に計測することが困難であり、またアウトプットとインプットの定義が安定しないという欠点がある。これに対して Value-Added Approach は、バランスシート上の資産、負債を金融業において産出される付加価値額（労働と資本のインプット量）に応じて配分する。特に負債に関しては付加価値が大きいことが推察される預金はアウトプットとして定義される。ただ、このアプローチの際は、バランスシートの各項目に対する金融業の付加価値の配分（労働や資本の配分）に関する情報が必要となる。

4.3. FISIM の概要と論点

4.3.1. FISIM の概念

多様化する金融仲介取引の実態を国民経済計算で把握・評価し、それらが果たす経済活動の役割を一つのサービス業として積極的に表現するために、金融仲介業の金利サービスによる付加価値を計測する手法として考案されたものが、FISIM である。

FISIM の簡単なイメージは、図 1 のとおりである。FISIM では、参照利子率という「“ノン・サービス”、“リスクフリー”利子率（リスク・プレミアムを最大限取り除き、さらにいかなる仲介サービス料をも含まない率）」と定義する利子率を導入する。そのうえで、金融仲介サービスを、(A)参照利子率よりも高い利率で資金を貸出で運用する活動（借り手 FISIM：資金の借り手である貸出先が受ける「情報生産（審査・モニタリング機能）に基づいた資金のオペラビリティ改善のための流動性の供給」等のサービス）、(B)参照利子率よりも低い利率で金融仲介機関が資金を調達する活動（貸し手 FISIM：資金の貸し手である預金者が受ける「保管（安全性）」及び「決済機能」等のサービス）、の 2 つに分け、アウトプットを発生させていると考える。

4.3.2. FISIM の主な論点

上記のように、金融仲介サービスのアウトプットや付加価値を積極的に評価し、国際比較上、GDP を精緻に計測するという観点からも、我が国での FISIM の正式導入が期待されるところである。但し、FISIM については、各国の事情に即した計測方法が採用されているとおり、国際的に議論が収束していない論点があることも事実である。93SNA 提案後、我が国でも FISIM 導入に向けて議論が重ねられてきたが、ここでは、(A)FISIM 対象商品、(B)FISIM の実質化、(C)参照利子率、について取り上げる。

4.3.3. FISIM 対象商品

FISIM 対象商品については、米国では債券を対象とし、欧州では対象にしないという相違がある。FISIM に債券を含むかどうかは、我が国のように多額の国債を発行し金融機関が

保有している場合に影響が大きい。債券が FISIM 対象外である場合に金融仲介機関の運用が貸出しから債券にシフトすると、図 1 右のように“運用残高<調達残高”という状態が生じる可能性がある。この場合、“金融仲介機関の利鞘< FISIM ”となり、過大推計された部分は主に調達サイド(貸し手 FISIM)のうち家計に配分され、最終消費、GDP が上昇する。他方、債券を計測対象に含めた場合は、大量の国債を金融機関が保有していると、運用サイドで借り手 FISIM が国債発行者である一般政府に配分される²³。

4.3.4. FISIM の実質化

EU 統計局基準では、運用資産残高と調達負債残高を金融仲介サービスの消費量ととらえ、運用利率と参照利率の率差及び参照利率と調達利率の率差を、サービスの価格ととらえ、それぞれを実質化することで、実質 FISIM の計算を行う。その際、残高は GDP デフレーター等で実質化し、率差は基準年の率差で固定する。

当該方式による実質化の定義には、(A)実質 FISIM の変動が実質残高の変動のみとなる、(B)率差固定の影響、という問題点がある。(A)については、単位残高当たりの FISIM (リスク・サービス)が時間の経過によらず不変となり、IT 化が進展する金融仲介業の提供サービスが量・質的に不変となる矛盾を伴う。(B)は固定基準年方式の実質化に共通の問題ではあるが、基準年の率差の大小が、推計期間を通じて実質 FISIM の大小に影響し続けるバイアスが発生する。また、2つの基準年の間に率差が大きく変動すると、基準改定によるブレが大きくなる。(A)及び(B)の問題点のほかにも、残高の実質化²⁴で使う GDP デフレーターが FISIM の上位項目デフレーターとなるため、イタレーションが発生するという問題もある²⁵。

4.3.5. 参照利率

(1) 現行の参照利率の算出手法

参照利率は、FISIM の根幹を成す概念であるが、論点も多く指摘されており、国際的にも議論はまだ収束していない。現在各国で採用されている参照利率には、大きく分けて、(A)金融仲介機関間取引レート(インターバンク・レート)方式、(B)国債利回り方式、(C)運用利率と調達利率の平均値方式、の3つのタイプがある。(A)はEU及び日本(参考系列)、(B)は米国、(C)はカナダやオーストラリアで採用されている。

²³なお、対象商品の問題は、FISIM の範囲を超えた金融サービス業のアウトプット計測とも関連し、MBS や CBO のような高度金融商品開発の GDP への反映、その際の直接的な手数料との関係など、今後の議論の深化が期待される。

²⁴米国の NIPA では、“the volume of banking transactions (小切手の振り出し、銀行間電子決済回数、預金及び引き出し数、貸出件数、ATM 利用回数など)に基づく算出の数量指数により、銀行サービスの実質算出を求めている (Moulton (2000)参照)。

²⁵カナダ、ドイツでは CPI 総合指数を採用している。

我が国への正式採用を考えた場合、(C)の平均方式は、運用利率と調達利率の間に必ず納まるため、マイナス FISIM が発生せず、国際比較が容易になるという利点はあるが、理論的根拠が非常に乏しい。(B)の国債利回り方式は、リスクフリーという定義をかんがみれば候補となり得るが、我が国においては、資産・負債の満期構成に合った国債利回りでの FISIM 計測が困難である。(A)のインターバンク・レート方式が最も現実的であろうと推察されるが、FISIM 対象機関のインターバンク取引における支払利子を調達残高で除した値として定義する場合は、必ずしも借り手と貸し手が対称ではないため、ノン・サービスとは言い切れず、ノンバンク等の存在により参照利率に上方バイアスもあるため、リスクフリーとも言い切れないという点は残る。また、1980 年代において、恒常的に運用利率と調達利率を上回っており、マイナスの借り手 FISIM が発生している。

(2) 高リスク先への貸出しによるマイナス FISIM の発生

マイナス FISIM は、マイナスのアウトプットを表す。先述の参照利率にインターバンク・レートを採用した際のマイナス FISIM の解釈としては、金融仲介機関の貸出運用が逆鞘であった可能性、データ及び計算上の問題が考えられる。それ以外に、高リスク先貸出しによるビジネスモデルにおいては、マイナス FISIM が発生し得る。金融機関において、貸出金利は、リスクフリー・レートにサービス料と貸出先の信用力に応じたリスク・プレミアムが上乗せされ、調達金利は、リスクフリー・レートに金融機関自身のリスク・プレミアムを上乗せした水準からサービス料を控除した水準で決まる(図2上段)。高リスクの与信先へ貸出する場合、貸出先及び金融機関自身のリスク・プレミアムは増大し、参照利率を調達利率が上回り、マイナスの貸し手 FISIM が発生し得る(図2下段)。

また、若干視点は異なるものの、ノンバンク等において、マイナスの貸し手 FISIM が発生することが知られている。ノンバンクは、主に銀行借入れによる資金調達なので、調達金利はリスクフリー・レートに調達機関自身の信用力に応じたリスク・プレミアムを使った銀行の運用金利で決まり、銀行を含めた1つの参照利率でアウトプットを計測すると、ノンバンク等側でマイナスの貸し手 FISIM が発生する。それをゼロとすると、その分銀行側との FISIM が二重計上の問題が発生する。

(3) 参照利率に関する今後の方向性

先述の高リスク先貸出しによるマイナス貸し手 FISIM 発生を無視すると、高リスク貸出しの増加が、デフォルトの存在を無視した、実態経済を反映しないアウトプット水準となる。この問題点を解決するためには、“リスク・プレミアム”をキーポイントとした、“参照利率の定義”が今後の研究課題となると思われる。

Wang & Basu (2008) 等は、図2の考え方を更に深化させ、図3右図にあるような借り手側、貸し手側に、別々の参照利率を設定する(複数参照利率方式)新しい銀行業アウトプット概念を示した。これは、図3左図のような従来方式では未分離であったリスク・

プレミアムと銀行業特有のサービス活動とを明確に区分することにより、リスクテイクがダイレクトに借り手 FISIM に計上されてしまうことを防ぎ、銀行業のサービスをアウトプットとして純粋に捉えようとするものである。Wang & Basu(2008)は、借り手 FISIM においては、金融仲介機関の機能の1つは、情報生産による貸出債権等の非市場性金融資産の評価であることから、貸出債権等の金融資産と同程度のリスクプロファイルを持つ市場性金融資産の利回りを基準とし、貸出金利との差をアウトプット計測として評価すべきとした²⁶。

一方の貸し手 FISIM においては、金融仲介機関の機能の1つはリスク負担（管理）であり、システミックリスクは、預金保険料により家計などの貸し手に転嫁されていることから、預金保険料が付いていない状態における、貸し手（預金者）の要求金利を基準とし、預金金利との差をアウトプット計測として評価すべきとした²⁷。

Wang & Basu (2008)の計測によれば、米商業銀行のアウトプット(1997 - 2007年の平均)が、従来の FISIM 手法と比較して 21%低減する結果となった（GDP もまた 0.3%が過大評価となる）。また、OECD（2008a）では、リスク・プレミアムをタームプレミアム及びデフォルトリスク・プレミアムに分解し、それぞれで参照利率を調整した場合における、ユーロエリアの FISIM（アウトプット）の計測を行い、2003Q1 - 2007Q4 の平均値で、タームプレミアムの存在により 24.3%、タームプレミアム及びデフォルトリスク両者の合計で 42.9%の FISIM の過大評価が発生しているという結果が示されている（図4）。

²⁶ 作間（2009）等は、一部金融仲介機関の参照利率へのアクセス不可能性(FISIM への批判)、中小企業のリスクプロファイル計測の不可能性(複数参照利率への批判)を主張し、費用測度(コストアプローチ)の採用を提唱する。

²⁷ (A)と(B)の差の部分は、銀行株主がシステミックリスクに対して要求する利回り（借手企業の付加価値の一部）で中間消費として扱う。

5. アウトプット計測の例:医療²⁸

5.1 はじめに

この節では、アウトプット計測のイメージを得るために、医療のアウトプットについて簡単に試算をしてみる。データは、川淵孝一東京医科歯科大学教授が収集した急性心筋梗塞(AMI)患者についてのデータである。このデータは協力していただける26病院から収集したもので、日本全体を代表するデータではないが、差し当たり手許にあるデータとして利用する。ここでの分析は非常に暫定的なものであり、あくまで、医療アウトプット計測のイメージということである。また、このAMIデータには医療費の情報も含まれているので、インプット(医療コスト)に対するアウトプットの関係、すなわち生産性の分析も可能であるが、それは今後の課題としたい。

このデータを分析する前に、まず、日本全体の心疾患の死亡や患者数について概観しておこう。

第1図は、日本全体について、人口10万人当たりの死亡率を死因別に描いたものである。悪性新生物による死亡が急激に増加しているが、心疾患も同じ程度のスピードで増加している。平成6年と7年に心疾患の死亡率が急減したのは、死亡診断書記入に終末期の状態としての心不全等を記入しないように徹底されたためであるので、トレンドとしては心疾患による死亡率は継続して上昇している。

第2図は、社会診療行為別調査により、虚血性心疾患の各年6月における病院への入院件数と診療実日数を図示したものである。驚くべきことに、件数は連続して減少し、診療日数も16年は増加したものの、17年は減少している。高齢化が進展し、患者数も死亡数も増加していると予想されるのに、意外である。

5.2 AMI調査データの特性

今回利用したAMI調査データは、平成15年から17年までに入院したAMI患者について、26病院から収集した3146症例である。個人が特定できないようコード化されている。

以下、このデータセットの特性を概観する。

(1) 年度ごとの推移

第3図は院内死亡率の年次推移を示したものである。平成15年度の約10%から、16年度9.1%、17年度7.6%とかなり急速に低下している。このような死亡率の変化は、重症度などの患者の属性の変化による面と病院における治療効果の向上による面がある。

第1表は、入院時のリスク・ファクターの割合を経時的に示している。年齢は年ごとにばらつくが、トレンド的に変化しているわけではない。Killip分類については、相対的に軽症のKillip1やKillip2の割合が上昇し、相対的に重症のKillip3やKillip4の割合が低下している。また、左室駆出率は上昇している。左房ブロックや心房細動を発生している患者の

²⁸ 本節の著者は、杉原茂・川淵孝一である。

割合も概ね低下している。心不全、陳旧性心筋梗塞、糖尿病を合併している患者の割合は低下しているが、高血圧、高脂血症、出血傾向、喫煙の割合は上昇している。総体的に言えば、入院時の重症度は低下しているとみられる。

第 2 表は、各種の治療法や検査を実施した割合を示している。心電図やアスピリン投与はほぼすべての患者に対して実施されており、経年変化はみられない。心血管造影や PCI はもともと高水準であったが、経年的にさらに上昇している。血栓溶解療法は低水準の中で徐々に上昇しているが、CABG はもともと割合が高くなかったのがさらに低下している。CABG から PCI への移行がみられる。

PCI の内訳をみると、単純バルーン (POBA) は低下傾向にある。ステントは 17 年度に低下しているが、これは薬剤溶出ステントを除いているためであり、薬剤溶出ステントは急速に増加しているので、これをあわせたステント全体も増加している。3 年間という短期間にも、単純バルーンからステントへ、さらに薬剤溶出ステントへという移行が観察される。

IVUS や心筋シンチグラムといった高度な検査の割合も上昇している。ただ、心 MRI は 17 年度の段階では非常に少ない。

第 3 表は、入院中に薬剤を投与された患者の割合を示したものである。アスピリンは当初から大半の患者に投与されている。β-ブロッカーは 15 年度から 16 年度にかけて上昇したが、17 年度は横ばいであり、継続的に上昇しているわけではない。カルシウム・チャネル・ブロッカーは上昇を続けている。ACEI は 15 年度から 16 年度にかけて上昇したが、17 年度は横ばいとなった。これは、継続的に上昇している ARB に代替されている面があるとみられる。スタチンは急速に上昇している。

同じ治療を実施するにしても、AMI の場合、迅速に実施することが必要である。実施までにタイム・ラグがあると、せっかくの治療も効果が減殺されてしまう。第 4 表は PCI を実施した患者について、各種の医療行為を実施するまでに要した時間を示したものである。極端に長い時間がかかった患者もいるので、平均値ではなく、中央値を表示してある。

心電図については、概ね最短の時間で実施されているものとみられる。アスピリン投与までの時間は 16 年度、17 年度と長くなっているが、トレンドとして長くなっているとも考えにくい。PCI までの時間は明確に短縮しているわけではないが、90 分程度というのは結構優秀な成績である。アメリカの AHA/ACC によるガイドラインでは、90 分以内に PCI を実施することが求められているが、実際にはアメリカでは 90 分以内で実施している割合は低い。アメリカと比較して日本の PCI の適時性は優れていると言えるが、年々改善しているというわけではない。

(2) 病院ごとの特性

以上は年度ごとの平均的な姿であるが、病院ごとにはかなりのばらつきがある。

第 5 表は、病院ごとに、院内死亡率と AMI 患者数を示したものである。病院ごとの患者数は大きく異なり、また、同じ病院でも年度ごとにもかなり変動する。特に患者数の少ない病院については、死亡率が偶然の変動によって大きく左右されるので比較には注意が必

要である。しかし、そうした偶然変動を割り引いて見ても、病院ごとの死亡率には相当のばらつきが存在する。症例数が 50 以上の病院に限って比較しても、院内死亡率は 2.5% (病院 25) から 16.0% (病院 23) までばらついている。

病院ごとの年次推移をみると、死亡率の動向は一様ではない。各年の患者数が非常に少ない病院もあるので比較は容易でない。差し当たり、各年の患者数が 30 以上の病院についてみると (第 4 図)、病院 20 及び 26 は経時的に死亡率が低下している。一方、病院 2 及び 17 は上昇している。病院 16 は横ばいと言えよう。

このような死亡率のばらつきの背景には何があるのであろうか。死亡率に影響を及ぼす病院側の要因としては、治療法の選択、治療の適切な実施などがある。

第 6 表は、治療関連の検査や術式の実施割合を病院ごとに比べたものである。心電図や心血管造影はほとんどの患者に対して実施されており、病院ごとのばらつきも小さい。PCI 実施割合もおしなべて高いが、最低の 59.6% (病院 23) から最高 100% までばらついている。薬剤溶出ステントについては、最低の 0% から最高の 35.4% (病院 15) までばらついている。血栓溶解療法は、多くの病院で全く実施していない一方、一部の病院 (病院 12、18、3) で高い割合となっている (それぞれ、60.6%、41.5%、33.3%)。CABG についても、全く実施しない病院と 10% を超える病院 (病院 3、6、13、21) に分極している。

治療の適時性についても大きなばらつきがある。第 7 表には、(1) 心電図取得までの時間、(2) アスピリン投与までの時間、(3) 心血管造影までの時間があるが、ここでは (4) PCI 実施までの時間をみてみよう。PCI 実施までの時間の中央値は、最短の 55 分 (病院 26) から最長の 192 分 (病院 19) までばらついている。

第 8 表は、病院ごとに薬物療法を実施した患者の割合を示したものである。アスピリンはほとんどの患者に対して投与されており、病院ごとのばらつきも小さい (ただし、病院 7 は例外)。β-ブロッカーについては、最低の 0% (病院 7 及び 11) から最高の 50.9% (病院 14) までばらついている。

第 9 表は、薬物療法の禁忌・適応別に薬物療法の実施率を示したものである。β-ブロッカーは、心機能が弱い患者に対しては禁忌とされている。ACEI は、心機能が弱い患者に適応があるとされている。左側の β-ブロッカーについてみると、多くの病院で、禁忌がない患者もある患者も同じような割合で投与されている。すなわち、薬剤の使用に当たって対象患者を適切に識別しているとは言えない。その中で、病院 6 や 12 は禁忌のない患者のみに投与し、禁忌のある患者には投与していない。右側の ACEI についても、概ね、適応の患者もそうでない患者も同じような割合で投与されている。細かくみると、病院 3 など適応患者への投与割合が高い病院がある一方、病院 12 や 14 のように適応のない患者にもっぱら投与している病院もある。

5.3 回帰分析

入院中に死亡する確率についてのロジスティック回帰を行う。時間変数 *time* を説明変数

に含めることにより、時間を通じての死亡率の変化を計測する。時間変数としては、タイム・トレンドと年度ダミーの 2 つの場合を試す。タイム・トレンドを使う場合は、時間を通じた死亡率の変化が毎年一定であると仮定することになる。年度ダミーを使う場合は、各年の変化の程度が異なることを許すことになる。

先にみたように、死亡率は病院によって相当異なる。そこで、定数項が各病院ごとに異なることを許すような定式化も試みる。各病院の定数項 a_i (i は病院の番号) は、共通の正規分布 (平均ゼロ) に従うと仮定する。こうしたモデルは、ランダム定数項モデルと呼ばれる。

さらに、各病院の死亡率の経時的変化も病院によって異なることを許すような定式化も試みる。この場合、時間変数 $time$ に係る係数が 2 種類出てくることになる。共通の係数 γ と各病院ごとの係数 c_i である。こうしたモデルは、ランダム係数モデルと呼ばれる。

以下に、各モデルの式を示しておく。 π_{ij} は、第 i 病院に入院した患者 j が死亡する確率であり、 x_{ij} はその患者の属性である。患者属性を含めない推定と含めた推定を行う。患者属性としては、差し当たり、年齢と Killip 分類を考える。

(1) 単純なロジット回帰

$$\text{logit}(\pi_{ij}) = \alpha + \beta \cdot x_{ij} + \gamma \cdot \text{time}_{ij} + \varepsilon_{ij}$$

(2) ランダム定数項モデル

$$\text{logit}(\pi_{ij}) = \alpha + a_i + \beta \cdot x_{ij} + \gamma \cdot \text{time}_{ij} + \varepsilon_{ij}$$

$$a_i \sim N(0, \sigma_a^2)$$

(3) ランダム係数モデル

$$\text{logit}(\pi_{ij}) = \alpha + a_i + \beta \cdot x_{ij} + (\gamma + c_i) \cdot \text{time}_{ij} + \varepsilon_{ij}$$

$$a_i \sim N(0, \sigma_a^2)$$

$$c_i \sim N(0, \sigma_c^2)$$

推定結果は第 10 表に示されている。これは、リスク・ファクターを説明変数に含めない定式化によるものである。単純なロジット回帰でタイム・トレンドを時間変数とした場合は、10%水準で有意となった。オッズ比は 0.86 となったが、これが意味することは、死亡率は毎年 14% ずつ低下していくということである。その右の、年度ダミーを時間変数とした場合には、16 年度ダミーは全く有意でないが、17 年度ダミーはある程度死亡率低下傾向

がみられることを示している。オッズ比が意味することは、死亡率が、15年度に比べて、16年度は10%低下し、17年度は25%低下するということである。

ランダム定数項モデルの推定結果がその下に示されている。このモデルでは、タイム・トレンドを時間変数とした場合も、年度ダミーとした場合も、有意性はかなり低くなってしまった。ランダム効果の分散は有意にゼロではない。すなわち、病院ごとのばらつきは存在すると言える。病院固有の効果を入れると時間変数が有意でなくなってしまうことからすると、全体としての死亡率低下が、全病院の平均的な低下というよりは、個別の病院の低下によるものであるという可能性がある。

ランダム係数モデルの推定結果がさらにその下に示されている。このモデルの結果はランダム定数項モデルと似たものとなった。ただし、ランダム効果の分散の推定値は、タイム・トレンドを時間変数とした場合、定数項ランダム効果は有意でなく、タイム・トレンドに係るランダム効果の分散が有意にゼロではない。

説明変数にリスク・ファクターを加えた推定結果が第11表に示されている。年齢、Killip分類ともに強く有意である。リスク・ファクターが死亡率に強く影響することが確認された。他方、時間変数はすべて有意でなくなってしまった。第3図でみたように経年的に死亡率が低下しているのは、リスク・ファクターの変化による見せかけの医療の質の向上であるのかもしれない。もしそうであるとすると、その間の医療費の増加は効率的なものとは言えないことになる。これについては、より長い期間のデータを使って詳細に検討する必要がある。

5.4 医療アウトプットの試算

ここでは、AMI入院のアウトプットを、生存して退院することと定義する。QOLの向上も重要な医療の質であるが、このデータベースではQOL指標を算出できないので、院内死亡率から計算できる生存退院率を質の指標とする。

第 t 期のAMI患者数を x_t 、生存退院率を s_t 、院内死亡率を d_t とする。第 t 期の質を調整したアウトプット（生存退院患者数）は $s_t \cdot x_t$ である。基準時点を第0期とすると、第 t 期

のアウトプット指数は、 $\frac{s_t \cdot x_t}{s_0 \cdot x_0}$ となる。

一方、生存退院率は、1から院内死亡率を差し引いたものである： $s_t = 1 - d_t$ 。したがって、上記のアウトプット指数は、次のように書き換えることができる： $\frac{s_t \cdot x_t}{s_0 \cdot x_0} = \frac{(1 - d_t) \cdot x_t}{(1 - d_0) \cdot x_0}$ 。

いま、第 t 期の死亡率は第0期に比べて r_t 倍になったとすると： $d_t = r_t \cdot d_0$ 。死亡率が経時的に低下していれば $r_t < 1$ となる。これを上記の式に代入すると、次のアウトプット指数が得られる。

$$\frac{(1 - r_t \cdot d_0) \cdot x_t}{(1 - d_0) \cdot x_0}$$

試算においては、基準時点は平成 15 年度とした。オッズ比 r_i は、ランダム係数モデルで年度ダミーを時間変数とし、リスク・ファクターを含まない定式化による結果を利用した。年度ダミーは必ずしも有意ではないが、ここでは、アウトプット計測のイメージを提供するという趣旨からあえて有意性にこだわらない。日本全体の患者数は社会診療行為別調査から採ったが、AMI 患者数は分からないので、虚血性心疾患の入院の件数とした。

結果は第 5 図に示されている。死亡率が低下している分、質の調整を行った場合の方が調整を行わない場合よりもアウトプットが多くなっている。ただ、その差はそれほど大きくないように見える。アウトプット指数の前年度からの変化率を第 6 図に示した。質の調整をした場合、調整をしない場合に比べて、平成 16 年度で約 1%ポイント、17 年度で 2%ポイント弱、減少率が縮小している。1~2%の変化は小さいとも考えられるが、医療費が巨額（全体では 30 兆円を超える）であることを考えると、それなりに大きなインパクトがあるとも言えよう。

5.5 病院ごとの分析

病院ごとに医療成果や治療内容が異なる。こうした病院ごとのばらつきを利用して、どのような要因が医療成果に影響を与えているかを検証することができる。医療の質の決定要因を分析することは非常に重要な課題である。ここでは、極く簡単な分析にとどめるが、今後さらに突っ込んだ分析を実施したい。

(1) 病院ごとのランダム効果の推定

死亡率に関する病院ごとの固有の効果は第 12 表に示されている。病院の固有の効果とは、病院固有の要因によってどれだけ死亡率が（平均よりも）高くなったり低くなったりするかを表すものである。プラスであれば平均よりも死亡率が高くなり、マイナスであれば低くなる。推定値を標準誤差で割った値をみると、病院番号の 22 や 23 が死亡率が有意に高く、2 や 20 が有意に低いと言える。ランダム効果をオッズ比に変換したのが第 7 図である。オッズ比は、各病院の死亡率が平均的死亡率より何倍高いかを示すもので、1 より高ければ死亡率が高く、1 より低ければ死亡率も低い。病院 22 や 23 は平均より 1.9 倍あるいは 1.6 倍も死亡率が高い。一方、病院 2 や 20 の死亡率は、平均より 30%程度低い。

(2) 治療プロセスと医療成果

このような病院ごとの死亡率の違いは、極めて多くの要因に起因する。特に重要と思われるのは、治療法の選択の違い（適切な治療法を選択しているかどうか）、治療の適時性などである。ここでは、差し当たり、PCI 施行までの時間との関係を第 8 図にプロットしてみた。PCI 施行までの時間は、質の指標としてアメリカなどで良く取り上げられる指標である。図をみると、かなりばらつきはあるが、概ね右上がりの関係が窺える。すなわち、PCI 施行までの時間が長い病院は死亡率が高く、PCI が適時に施行される病院は、死亡率が低いということになる。

病院ごとの医療成果を決定する要因としては、治療法の選択や治療の適時性など多くの

ものがある。AMI データベースにはこうした情報が豊富に含まれているので、今後さらに突っ込んだ検討を行うこととしたい。

6. おわりに

以上、医療、教育、金融についてそれぞれのアウトプット、生産性の計測についてサーベイしてきたが、それぞれの分野に関して今後の研究課題は以下のとおりである。

医療サービスのアウトプットを適切に計測することは、理論的にも統計的にも非常にチャレンジングである。しかし、国民生活における医療サービスの重要性を考えると、この困難な課題を未解決のまま放置しておくことは許されない。

今後の本分野に関する望ましい研究内容は、質を調整したアウトプット/生産性/デフレーターデフレーターの計測である。具体的には、UK Department of Health (2005) や Cutler and Richardson (1999) のラインに沿った計測が必要であると考え。前者による試算は現在の SNA の考え方の延長線上に構想されており、サテライト勘定などを通じた SNA への組み込み可能性が高い。また、後者による試算は、やや実験的色合いが強いが、医療の質や生産性についての情報を提供し、医療政策論議を evidence-based なものとするに大きく貢献する。

その際、これらの研究では実行されていないが、質の向上に対する医療支出の貢献分を純粋に取り出すことが重要な課題である。

教育のアウトプットや生産性の計測もまだ試験的な段階にあり、最も意欲的に取り組んでいるイギリスにおいてさえ、その手法には多くの改善余地が残されている。しかし日本でも、諸外国で行われているように、教育のアウトプットや生産性に関する研究を通じて様々なデータが蓄積され、より効果の高い教育の在り方について実証分析に基づいた議論が進展することが望まれる。

FISIM については、本稿で触れた論点の他に、QE (四半期データ)、国際収支統計との整合性及び暗黙の政府保証の問題等も課題として残されている。さらには、金融仲介機関の預金・貸出業務に限定した FISIM ではなく、広く金融業全体のアウトプットの計測を行うための研究を進めることも必要である。そのためには、理論的整理とその理論に沿うための精度の高いデータが欠かせず、公表されたデータだけではなく、金融当局や金融機関との協力により研究を進めていくべき分野であるものと考えられる。まずは、公表されたデータから日本の銀行業の信用コストスプレッドを推計することが必要であろう。