

An Alternative Analysis of Demand for Health Care

Nobuyuki Izumida, M.A.[†]

Abstract

This paper considers the demand for health care more comprehensively, by separating analysis of decision making on whether one should contact a physician or not, and analysis of choice of what kind of contact (outpatient, inpatient after outpatient service and inpatient) one should take.

We use pooled individual medical claim data with permission of three mutual health insurance associations. By aggregating the medical claim data by episode, we focus on empirical analysis of both how the "first contact" with a physician was affected by the rise of the co-insurance rate in September 1997, and what type of contact (outpatient, inpatient after outpatient service and inpatient) was chosen.

Our results show that the rise of the co-insurance rate in September 1997 (1) decreased probability of "first contact", but (2) had no effect on choice of contact type. These results indicate that the rise of the co-insurance rate decreases the demand for health care, but they do not indicate with statistical significance the worsening of health due to the rise of the co-insurance rate.

In our analysis, we compared the results by the simple use of dummy variables and the use of the difference in difference method. They give us relatively similar results when the data has a longer time horizon.

Keywords : Rise of co-insurance rate, Medical claim data, Episode probability of "first contact"

[†] National Institute of Population and Social Security Research

健康の不確実性が医療サービスの 受診行動に与える影響について*

山田 武†

本論文の目的は健康の不確実性が消費者の受診行動に与える影響について分析することにある。消費者は自分自身が病気なのかどうかを正確に判断することはできない。この意味で、消費者は健康の不確実性に直面することになる。Dardanoni and Wagstaff (1990) の仮定を若干修正した理論モデルの示すとおり、期待健康資本が上昇すると医療需要は減少し、健康の不確実性の度合いが増加すると医療需要は増加する。実証分析で使用したデータは1997年9月から1999年8月までの2年間分の約54万人の健康保険組合の加入者（被保険者男女と配偶者女性）の支払い情報データである。健康資本そのものを観察することはできないが、健康資本のフローである健康日数（1年で評価すると365日－外来日数－入院日数）は観察可能である。そこで、不確実性の指標として健康の期待値の代理変数として前年の月当たりの平均健康日数、健康の不確実性の度合いを示す代理変数として月当たりの健康日数の標準偏差を採用した。被説明変数はエピソード件数、説明変数は平均健康日数、健康日数の標準偏差などで、推定方法は Negative Binomial モデルを採用した。平均健康日数の推定値は負、健康日数の標準偏差の推定値は正でどちらも統計的に有意であった。このことは、消費者が治療を開始するかどうかの選択にあたっては、健康の不確実性が影響することを示しており、理論モデルと一致する。

キーワード 医療需要、不確実性、健康資本、健康日数

1. 目的

本論文の目的は、健康の不確実性が医療サービスの受診行動に与える影響について、実証分析を試みることにある。消費者の受診行動に健康が重

要な役割を果たすことはいうまでもない。病気の時に受診し、健康なときには受診しないと考えるのがごく自然である。しかし、消費者は自分自身がどの程度の健康なのかを確実に知っているわけではない。むしろ、健康の分布を知っているにすぎない。この意味で、消費者は健康の不確実性に直面することになる。医療経済学の分析では、健康そのものの不確実性よりも、医療支出の不確実性が強調されることが多い。消費者は保険に加入することによって、医療支出の不確実性を緩和することができるが、保険に加入することによって健康の不確実性そのものがなくなるわけではない。医療保険に加入したとしても、消費者は健康の不確実性のもとで、受診するかどうかを決定しなけ

*いくつかの研究会を通じて、故鶴田忠彦教授から研究内容にご指導を頂いただけでなく、研究活動と政策や社会との関わり、あるいは研究と教育の関係について深く考える機会をたびたび頂戴いたしました。特に、大規模な国民健康保険データを使った研究では、データベースの作成段階から研究に参加する貴重な機会を頂戴しました。本研究は厚生労働科学研究費補助金政策科学推進研究事業「生活習慣と健康、医療消費に関するマイクロ計量分析（H15－政策－012）」主任研究者小椋正立教授の研究成果の一部を加筆修正したものである。
†千葉商科大学商経学部

ればならない。

Grossman (1972) は健康資本という概念を導入したが、それにつづく多くの研究が確実な健康を前提として理論分析や実証分析を展開している。消費者は確実な健康を知っているわけではない点に着目して、健康の不確実性を明示的に取り上げた研究もある。たとえば、Cropper (1977), Dardanoni and Wagstaff (1990), Selden (1993), Liljas (1998), Picone, Uribe and Wilson (1998), Koç (2004) などをあげることができる。これらの研究はいずれも Grossman モデルを拡張して健康そのものの不確実性や健康の資本減耗の不確実性、治療成果の不確実性を考慮した医療需要の理論モデルを報告しているが、健康の不確実性を明示的に考慮した実証分析の先行例はみあたらない。

消費者が確実な健康状態を知っていると仮定している論文は、消費者が健康の期待値を知っていると解釈することもできる。しかし、消費者が健康の分布を知っているにもかかわらず、分散を無視して意思決定をするというのは不自然である。健康の期待値が同じであっても、分散が異なれば医療需要に影響するだろう。また、健康診断という私的な情報財を購入するのは、健康の不確実性があるからである (山田, 2003)。確実な健康を知っている消費者が費用を支払ってまで追加的な情報を入手する動機はない。しかし、多くの消費者が高額な人間ドックなどを受診するのは、不確実性を緩和するために情報を得ようとしているからである。したがって、消費者は医療サービスを必要する場合でも健康の不確実性を考慮すると考えられる。

本論文では Dardanoni and Wagstaff (1990) のリスクに関する定式化を若干変更した医療需要に関する理論モデルから次の結論を得た。健康の期待値が増加すると医療サービス需要が増加する。健康の不確実性が増加すると医療サービス需要は増加する。この結果は Dardanoni and Wagstaff

(1990) と一致する。

健康の不確実性を含む実証分析のためには、不確実性の具体的な定義が必要である。健康そのものを外部から観察することはできないが、健康資本の生産するサービスである健康日数は観察することが可能である。後半の実証分析では健康資本のフローである健康日数 (= 365 - 外来受診日数 - 入院日数) を使って健康を代理できるものとする。また、消費者は前年の経験から今年健康を推定すると仮定する。具体的には、健康の期待値を前年の健康日数の平均値 (= (365 - 外来受診日数 - 入院日数) / 12) から推定し、健康の標準偏差を前年の毎月の健康日数の標準偏差から推定すると仮定する。

これらの健康の不確実性の指標を使った実証分析の結果からもおおよそ理論モデルを支持する結論を得た。データは1997年9月から1999年8月までの111組合・約54万人の健康保険組合の加入者 (被保険者男女と女性配偶者) で、推定方法は Negative Binomial モデルである。被説明変数は1998年9月からの1年間のエピソード件数 (治療開始件数ただし過年度開始分を含む)、説明変数は1997年9月から1年間の平均健康日数、健康日数の標準偏差、受診日数がゼロであることを示すダミー、1998年9月から1年間の平均標準報酬、年齢階級ダミー変数である。推定結果はおおよそ理論モデルと一致し、期待健康日数の推定値は負、健康日数の標準偏差の推定値は正でどちらも統計的に有意であった。つまり、消費者は健康の期待値が低下する、あるいは、健康の分散が増加すると医療需要を増加させる。消費者が治療を開始するかどうかの選択にあたっては、健康の不確実性を考慮しているのである。

以下、第2節では Grossman の定式化に不確実性を取り入れた理論モデルを構築し健康の不確実性が受診行動に与える影響を理論的に説明する。第3節以下では、理論モデルをもとに実証分析を進める。第3節ではデータについて、特に、健康

の期待値と健康の不確実性を中心に説明し、第4節では回帰分析の結果を報告し、同時に考察を加える。第5節で本論文をしめくくる。

2. 健康の不確実性

Arrow (1963) も指摘しているように、消費者は医療支出の不確実性だけでなく、健康の不確実性にも直面する。消費者が自分自身の健康状態を確実に把握することができないため不確実性が生じる場合もあるし、外生的な要因による不確実性もある。また、治療そのものの成果に関する不確実性も存在する。医療保険を購入することによって医療支出の不確実性は緩和されるが、健康のそのものの不確実性がなくなるわけではない。消費者は健康の不確実性のもとで、受診するかどうかを意思決定しなければならない。

これまでの実証分析では、健康は確実なものとして扱われることが多かった。たとえば、さまざまな実証分析で使用されている国民生活基礎調査の健康票では健康状態を「よい」「まあよい」「ふつう」「あまりよくない」「よくない」のうちから選択するように求めている。これらの健康の評価は、回答者が確実な健康状態を知っていることを前提にしているとも考えられるし、健康の分布を考慮して健康の期待値を示していると思えることができるかもしれない。しかし、消費者が意思決定をする際に、期待値だけを利用するわけではないだろう。同じ健康の期待値であっても分散が異なれば消費者は異なるリスクに直面することになる。したがって、医療サービスの需要量は期待値だけでなく、分散にも依存すると考えられる。

一方、Willingness-to-Pay アプローチのように健康の不確実性を考慮して、健康か病気の2つの状態に限定して分析する方法もある。この場合には、健康は医療を必要としない状態であり、病気は医療を必要とする状態を示している。消費者はこのような2つの状態からなる不確実性のもと

で医療や消費を選択することになる。健康は不連続な2種類の状態であり、健康の平均値や分散はこれらのモデルでは重要ではない。

健康か病気といった分類では十分に分析出来ない場合には、健康を連続変数として取り扱う必要がある。これら分析には Cropper (1977), Dardanoni and Wagstaff (1990), Selden (1993), Liljas (1998), Picone, Uribe and Wilson (1998), Koç (2004) などが含まれる。これらはどれも Grossman モデルに健康の不確実性を付け加えた論文で、健康そのものや健康の資本減耗・治療成果に不確実性がある場合について検討されている。以下では、Dardanoni and Wagstaff (1990) の不確実性に関する仮定を若干変更して健康の不確実性が医療需要に与える影響をモデル化する。Dardanoni and Wagstaff (1990) では健康資本の分布関数の特性を使って、確率優位や平均保存的拡散などについて検討しているが、ここでは、健康資本の平均と分散を明示して分析する。

健康の不確実性を含まない医療サービスの消費に関する理論モデルは Grossman (1972) にならって、以下のように定式化することができる。

$$\max U = U(C, H)$$

$$\text{s.t. } Y = C + PM$$

$$H = H_0 + M$$

上の最大化問題で、 C ：消費財で価格は1、 H ：健康資本、 H_0 ：健康資本の初期値、 M ：医療サービス、 Y ：所得、 P ：医療サービスの価格である。消費者は消費と健康からなる効用を最大化する。ただし、予算制約と健康の生産関数の制約の下で効用を最大化することになる。上の最大化問題を医療サービスについて解くと、医療サービスの需要関数が得られる。一般的な仮定の下で、健康資本 H_0 の増加は医療サービスの需要を抑制することになる。

健康そのものに不確実性がある場合には上の最大化問題は次のように置き換えることができる。

$$\begin{aligned} \max U &= EU(C, H) \\ \text{s.t. } Y &= C + PM \\ H &= \bar{H}_0 + \alpha e + M \end{aligned}$$

または,

$$\max U = EU(Y - PM, \bar{H}_0 + \alpha e + M)$$

\bar{H}_0 は健康資本の初期値の期待値 (以下では健康の期待値と呼ぶ), e は健康の期待値からの乖離を示し, e の期待値は 0 と仮定する。 α は不確実性のスケールである。不確実性のない最大化問題との違いは, 健康が期待値に置き換えられていること, また αe によって健康の平均からの乖離が付け加えられている点にある。新しい最大化問題の 1 階の条件は(1)式で与えられる。括弧内は不確実性のない場合の 1 階の条件と全く同じであり, 均衡では限界代替率と価格比が一致することを意味している。

$$E[-PU_C + U_H] = 0 \quad (1)$$

また, 2 階の条件は 1 階の条件をもう一度医療サービス M で微分することによって得られる。(2)式の括弧内も不確実性のない場合の 1 階の条件と全く同じである。

$$E[P^2U_{CC} - 2PU_{CH} + U_{HH}] < 0 \quad (2)$$

以下では 2 階の条件が満たされると仮定する。具体的には

$$(\text{仮定 } A) \quad U_{CC} < 0, U_{HH} < 0, U_{CH} > 0$$

が満たされると仮定する。これらは消費や健康の限界効用は逓減し, また, 消費と健康は補完関係にあることを意味する。ここでの関心は健康の期待値 \bar{H}_0 や不確実のスケール α が医療サービス需要 M にどのような影響を与えるかである。1 階の条件(1)式を全微分して整理すると, 健康の期待値 \bar{H}_0 が医療サービス M に与える影響は(3)式のようなになる。

$$\frac{dM}{d\bar{H}_0} = \frac{E[PU_{CH} - U_{HH}]}{E[P^2U_{CC} - 2PU_{CH} + U_{HH}]} < 0 \quad (3)$$

符号を決定する際には仮定 A を使っている。健康の分布が右に平均移動した, つまり健康の期待値だけが增加した場合には, 医療サービス需要は

減少する。反対に健康の期待値だけが減少した場合には医療サービス需要は増加する。これは確実な健康が医療需要に与える影響の不確実性版である。次に, 不確実のスケール α が医療サービス需要に与える影響は次のようになる。

$$\frac{dM}{d\alpha} = \frac{E[PU_{CH}e - U_{HH}e]}{E[P^2U_{CC} - 2PU_{CH} + U_{HH}]} \quad (4)$$

仮定 A より分母の符号は負であるが, 分子の符号は定まらない。このままでは式全体の符号が確定しないので次のような仮定 B¹⁾を追加する。

$$(\text{仮定 } B) \quad E[U_{CH}e] \leq 0, E[U_{HH}e] > 0$$

この仮定 B を付け加えると, 不確実性のスケール α の増大によって医療サービス需要が増大することになる。

以上の不確実性を含む理論モデルからは, 健康の期待値が増大すると医療サービス需要が減少すること ((3)式), また, 健康の不確実性が増大すると医療サービスへの需要が増加すること ((4)式・仮定 B) が明らかになった。健康の期待値の変化の影響は不確実性のない場合と同じである。この意味で, Grossman モデルにあらわれる健康資本は健康の期待値と見なすことができるのである。一方, 不確実性が大きくなった場合には, 消費者は医療サービスの需要量を増加させる。これは不確実性が増大すると保険需要が増大するのと同じ現象と見なすことができる。

3. 推定方法とデータ

不確実性を含むモデルでも, 含まないモデルでも消費者が医療サービスの需要量を決定するという意味で, やや現実的ではないと思われる。多くの場合には医師が診断し, 治療内容を決定するから患者の選択の余地は限られている。患者が主体的に選択できるのは受診するかどうか程度である

¹⁾ この仮定は Sandmo (1970) によって貯蓄に関する 2 期間モデルの分析にあたって採用された仮定である。仮定 B の意味については酒井 (1982) を参照。

表1 変数の要約

変数名	内容
エピソード件数	1998年9月から1999年8月までのエピソード数。1998年8月以前に開始したエピソードを含む。
平均健康日数	健康の期待値の代理変数。1997年9月から1998年8月の外来受診日数と入院日数を365日から除き12で割ったもの。
健康標準偏差	健康の不確実性の度合いの代理変数。1997年9月から1998年8月までの各月の健康日数の標準偏差。
平均報酬	1998年9月から1999年8月までの標準偏差の平均値。単位千円。ボーナスは含まない。
受診日数0ダミー	1997年9月から1998年8月の外来受診と入院日数が0のときには1, それ以外は0
20-29歳ダミー	20-29歳: 1, それ以外: 0
30-39歳ダミー	30-39歳: 1, それ以外: 0
40-49歳ダミー	40-49歳: 1, それ以外: 0
50-59歳ダミー	50-59歳: 1, それ以外: 0

と考えられる。患者が受診するかどうかを決定し、治療内容を医師が決定するモデルは Two Part モデルと呼ばれている。以下の実証分析では Two Part モデルのうちのはじめのパート、すなわち患者が受診するかどうかを決定する段階を対象として分析する。

(1) 対象期間と対象

実証分析に使用するデータは1997年9月から1999年8月までの2年間に111健康保険組合に加入していた20歳以上59歳以下の被保険者とその配偶者（女性に限定）約54万人の支払情報データ²⁾に関する個票である³⁾。そのうち、被保険者男性が約28万人、被保険者女性が約10万人、配偶者女性が約15万人である。対象期間は1997年9月に被用者保険の被保険者の一部負担が変更になった直後で、この期間中には一部負担の変更はない。

²⁾ 支払い情報データとはレセプトに含まれる情報のうち、治療内容にあたる適用部分をのぞいたものである。なお、使用したデータのうち、組合や個人を特定する情報はすでにマスキングされているため、組合や個人を特定することはできない。

³⁾ データの使用にあたっては厚生労働省保険局のご協力をいただいたことに、感謝いたします。またデータの取扱については東京学芸大学鈴木亘助教授のサポートをいただきました。

また、60歳以上の加入者の中には退職後の任意継続が多数含まれる可能性があるため分析対象からは除外した。任意継続制度は退職後も在職時と同じ給付がうけられるが、保険料の対象となる標準報酬が実際の報酬とは異なること、また、保険料を全額支払う（事業者負担がない）ことなどの違いを考慮して分析の対象からは除外した。

(2) 変数

推定式の被説明変数は1998年9月から1999年8月までの1年間のエピソード件数、説明変数は、健康の期待値の代理変数として1997年9月から1998年8月までの平均健康日数（＝（365日－年間外来受診日数－年間入院日数）／12）、健康の分散の代理変数として1997年9月から1998年8月までの各月の健康日数の標準偏差を採用する。また、これ以外に1998年9月から1999年8月までの年間平均標準報酬、20-29歳ダミー、30-39歳ダミー、40-49歳ダミー、50-59歳ダミーなどである。（変数の要約は表1、基本統計量は表2を参照。）

(3) 被説明変数

被説明変数は1998年9月から1999年8月までの1年間のエピソード⁴⁾件数である。エピソード

表2 基本統計量

被保険者男性

変数名	サンプル数	平均	標準偏差	最小	最大
エピソード件数	284755	1.9671	2.0429	0	31
平均健康日数	284755	29.8084	1.2432	0	30.4167
健康日数の標準偏差	284755	0.7386	1.1558	0	57.69171
受診日数0ダミー	284755	0.2837	0.4508	0	1
平均標準報酬	284755	444.0678	156.6197	0	980
20歳代ダミー	284755	0.2098	0.4072	0	1
30歳代ダミー	284755	0.2893	0.4534	0	1
40歳代ダミー	284755	0.2720	0.4450	0	1
50歳代ダミー	284755	0.2289	0.4201	0	1

被保険者女性

変数名	サンプル数	平均	標準偏差	最小	最大
エピソード件数	104074	2.8792	2.4933	0	33
平均健康日数	104074	29.6305	1.2395	0	30.4167
健康日数の標準偏差	104074	0.9595	1.1802	0	18.39693
受診日数0ダミー	104074	0.1670	0.3730	0	1
平均標準報酬	104074	248.3675	102.0927	0	980
20歳代ダミー	104074	0.3550	0.4785	0	1
30歳代ダミー	104074	0.2114	0.4083	0	1
40歳代ダミー	104074	0.5196	0.4140	0	1
50歳代ダミー	104074	0.2139	0.4101	0	1

配偶者女性

変数名	サンプル数	平均	標準偏差	最小	最大
エピソード件数	146575	2.5823	2.4592	0	53
平均健康日数	146574	29.5676	1.5099	0	30.4167
健康日数の標準偏差	146573	1.0231	1.3522	0	37.66227
受診日数0ダミー	146575	0.2073	0.4054	0	1
平均標準報酬	146575	516.7036	154.2827	0	980
20歳代ダミー	146575	0.0856	0.2797	0	1
30歳代ダミー	146575	0.3306	0.4704	0	1
40歳代ダミー	146575	0.3653	0.4815	0	1
50歳代ダミー	146575	0.2185	0.4132	0	1

とは治療の開始から終了までを含む治療の単位である。日本では患者の一部負担をのぞく治療費は保険制度に請求される。このとき使われる請求書が通称レセプトで、受診した月ごとに作成される。一般に1エピソードは複数の月にわたるため、複数のレセプトを接合する作業が必要になる。ここでは、支払い情報データに含まれる診療開始日をつかって、複数のレセプトを1つのエピソードにまとめた。

理論モデルでは消費者は健康の不確実性のもと

で医療サービスの需要量を決定することになる。しかし、一般に実際の消費者は診断や治療内容を決定することはできない。消費者にできるのは受診(エピソード)を開始するかどうかだけである。具体的な治療内容は実質的に医師が決定すると考えられる。いわゆる Two Part モデルでは第1段階で消費者は受診するかどうかを決定する。消費者が選択できるのはエピソードを始めるかどうかで、エピソード件数はちょうど受診開始を決定した回数に一致する。エピソード件数は診療開始日をグルーピングしたものであるから、消費者が受診を開始した件数に他ならない。エピソードに

⁴⁾ エピソードの概念について詳しくは山田(2002)。

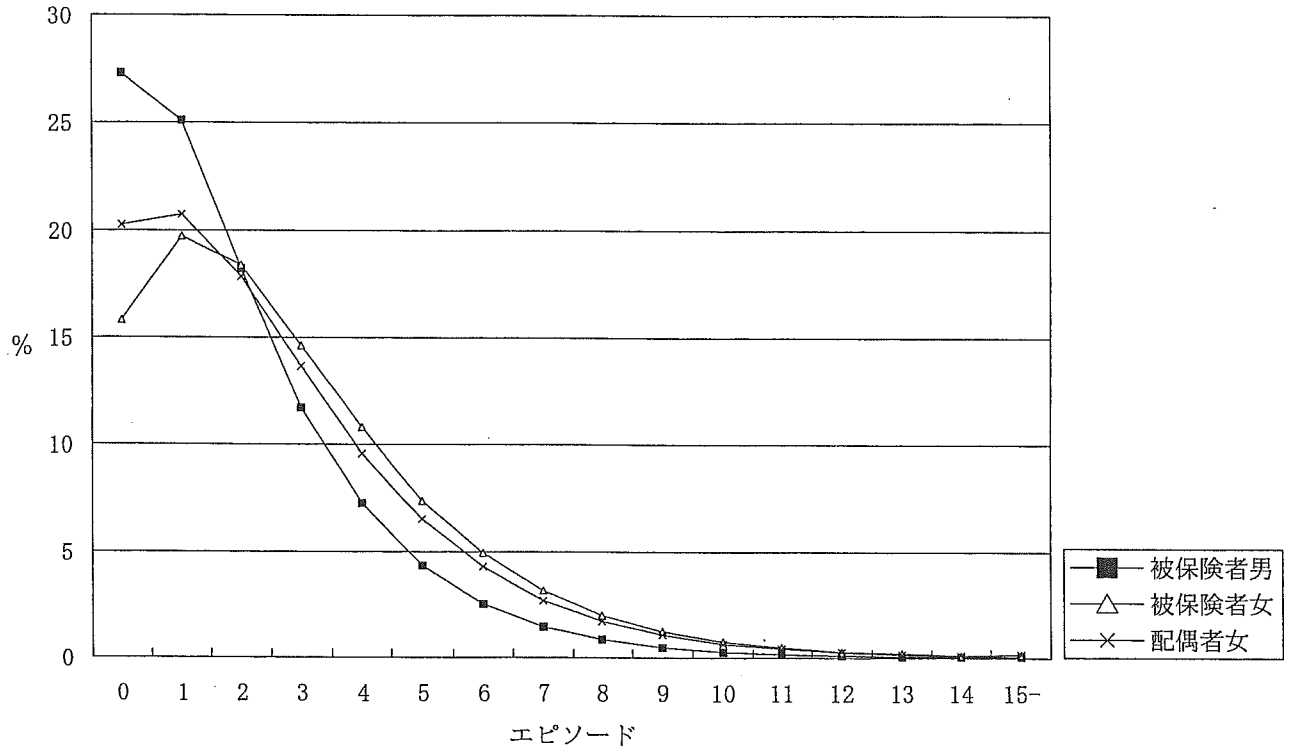


図1 エピソード件数の分布

は今期治療を開始したものと、前期以前に治療を開始したものの両方が含まれる。また、高血圧症などは治療が長期間にわたるものも、風邪のように短期間に治療が終了するものもある。それぞれのエピソードは質的に異なる側面を持つが、どちらも1エピソードとしてカウントした。

図1は1998年9月から1999年8月までの1年間のエピソード件数の分布を示している。図1は被保険者男性・被保険者女性・配偶者女性のそれぞれの分布が長い右裾を持っているという特徴を持っている。特に被保険者男性は0件が最も多く、女性の場合でも1件が最多である。以上の特徴から実証分析にあたっては、ポアソン回帰モデルやNegative Binomial (負の二項分布) モデルが考えられるが、平均値よりも分散が大きいであることを考慮してNegative Binomialモデルを採用した。

図2は年齢別の平均エピソードである。被保険者男性は35歳ごろまで緩やかに増加し、若干低下

するが50歳ごろから再び増加する。一方、同じ女性であっても被保険者女性と配偶者女性では異なる経路をたどっている。20歳代のうちは配偶者女性が被保険者女性を上回るが、35歳ごろからは被保険者女性が配偶者女性を上回る。また、男性と同じく35歳ごろからいったん低下するが、45歳ごろから再び増加する。なお、一貫して女性のエピソードは男性よりも高い。

(4) 説明変数

消費者は前年の健康から今期の健康を推測すると仮定し、健康の期待値の代理変数として平均健康日数を、健康の不確実性の指標として健康日数の標準偏差を使用する。なお、これらの指標の推定には1997年9月から1998年8月のデータを使用する。

a. 健康の期待値

ストックにあたる健康そのものを観察することはできないが、健康のフローを観察することは可

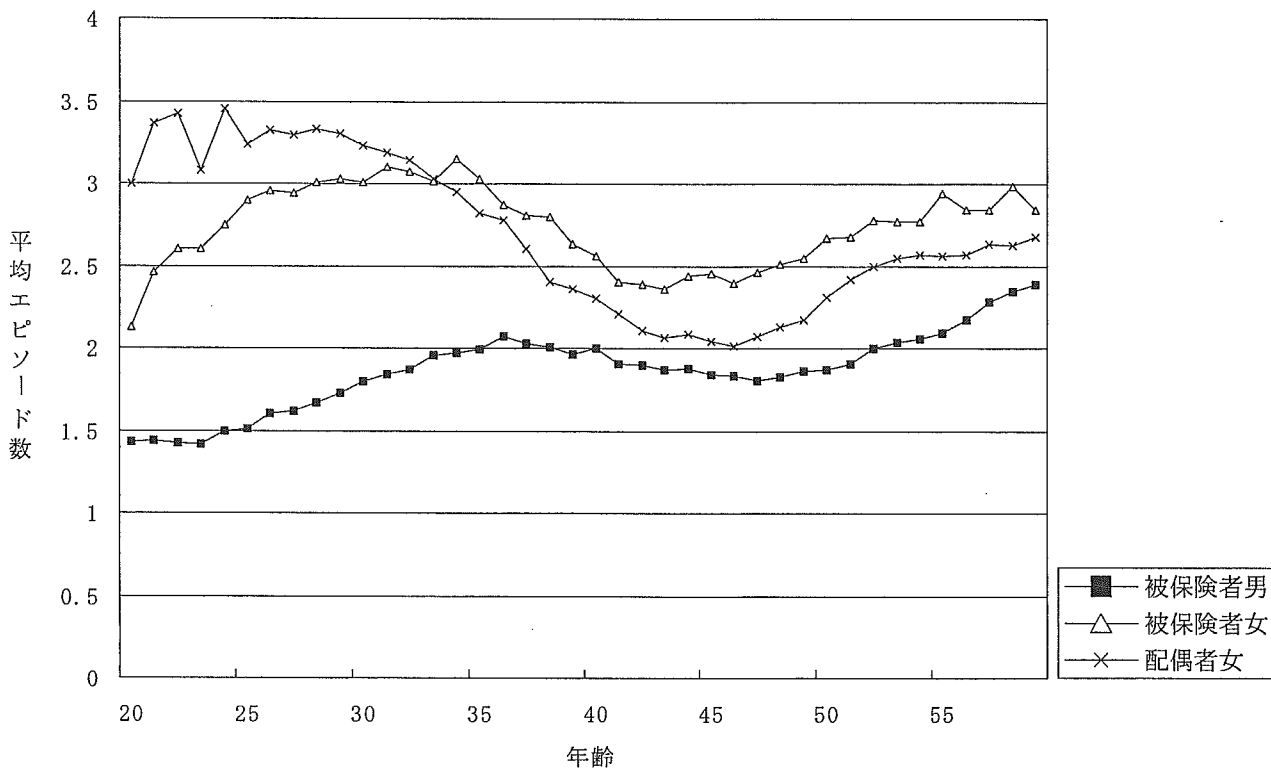


図2 エピソード件数 (年齢階級別)

能である。365日から外来受診日数と入院日数をのぞいた日数は健康のフローと見なすことができる。健康であったからこそ医療機関で受診のために時間を割く必要がなかったという意味で、ストックである健康からのフローのサービスということができる。本論文では年間健康日数を12で割ったものを健康の期待値の代理変数として利用する。

図3は1997年9月から1998年8月までの1年間の平均健康日数を示している。20歳以上では男女または被保険者と配偶者で健康日数には違いが見られる。まず、年齢を通じて被保険者男性がもっとも健康日数が多く、20歳以降徐々に低下する。35歳ぐらいまでは被保険者女性が配偶者女性よりも健康日数が多い。これには出産などが影響しているかもしれない。また、35歳以降はどちらも(被保険者も配偶者も)同じような経路をたどる。35歳以降若干持ち直すが、45歳ごろからは減少に転じる。

ここでは受診日数や入院日数だけを使って健康日数を定義しているが、この方法では過剰にある

いは過少に健康日数を推定する可能性がある。たとえば、インフルエンザで会社を休んだ3日のうち受診したのは1日だけであれば、健康のフローは1日ではなく3日減少したと考えるのが自然である。残念ながら支払い情報データには休業に関する情報はデータには含まれていない。その結果、平均健康日数は過剰に推定されている可能性がある。一方、平均健康日数は30.42日(=365/12)が上限で、それ以上の平均健康日数は存在しない。また、平均健康日数が30.42日の加入者も少ない。これらの加入者については平均健康日数が30.42日までに制限されているために、健康が過小評価されている可能性がある。その点を考慮して前年度の受診日数が0をしめす、受診日数0ダミーも付け加えた。受診日数を使う方法は支払い情報データから簡単に推定することができる簡便なアプローチであると考えられる。

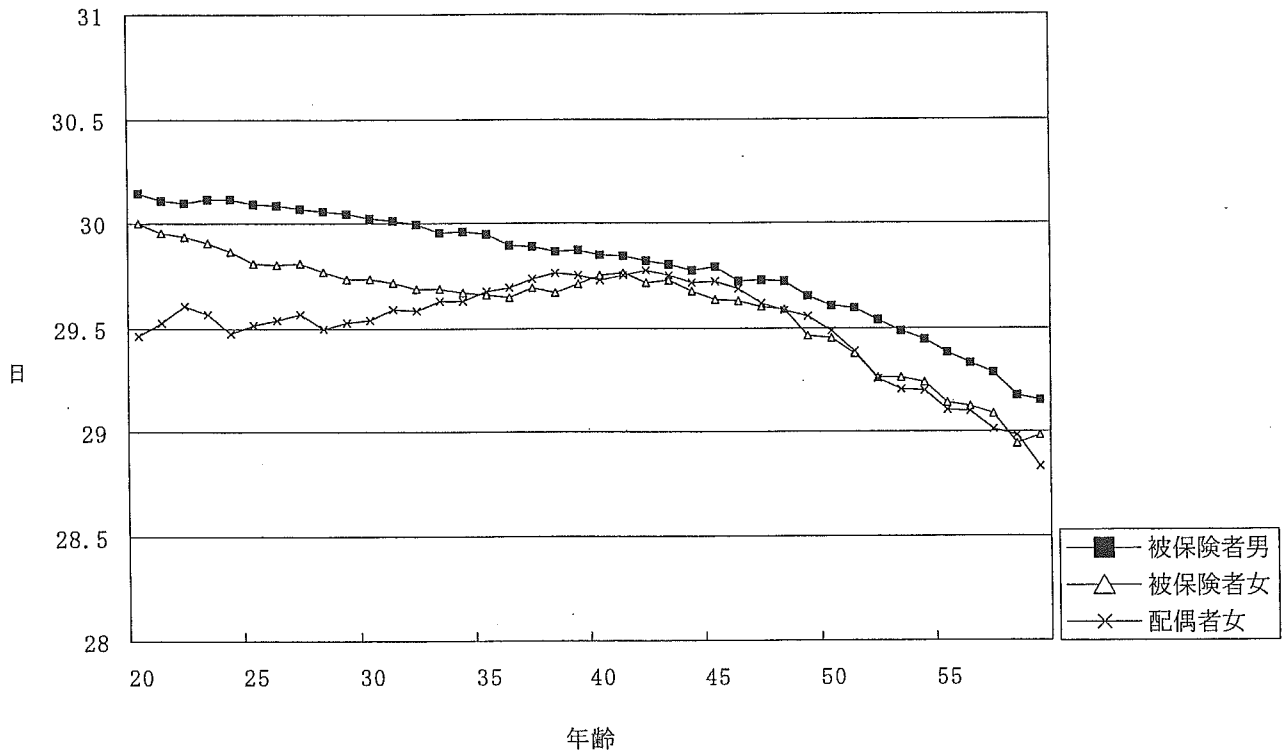


図3 平均健康日数

b. 健康の不確実性

一方、健康の分散を示す指標として、ここでは前年の月ごとの健康日数の標準偏差を採用した。たとえば、治療費は不確実性を質的な側面に踏み込んでいるという意味では望ましい性質を持っているが、消費者にとっては治療費を把握するのは意外とむずかしいと思われる。消費者は治療費の一部を自己負担する。ただし、消費者の自己負担には上限があり、また、本論文が対象とする健康保険組合の多くは付加給付制度を持っているため、実質的な自己負担率は、通常の一部負担率よりも低くなっていると予想される。その結果、治療費そのものの変動を理解しているとは考えられない。むしろ行動が制限される受診日数はわかりやすい指標と思われる。

健康日数の標準偏差は毎月の健康の不確実な変化、具体的には急性疾患などの突然の発生を反映することが期待される。反対に慢性的な治療が継続的に続けられる場合には標準偏差には影響しな

い場合もある。たとえば、糖尿病の治療で毎月4日定期的に通院しているとすれば、分散は0である。ところが、アトピー性皮膚炎で2月に1日受診する場合には、健康には変化がないにもかかわらず、分散が観察されることになる。

また、平均健康日数とのある程度の相関の発生もさげられない。図4はある健康保険組合の被保険者男性の健康日数と標準偏差の関係を示している。定義から平均健康日数が30.42日(=365日/12)の場合、標準偏差は0になる。散布図はこの点を頂点に点が広がっているように見える。実際、平均健康日数が減少しなければ大きな標準偏差はえられない。その結果、健康日数が30.42日を頂点として散布図は広がるのである。

もっとも、平均健康日数が減少したからと言って標準偏差が大きくなるとは限らない。たとえば、毎月3日ずつ通院した場合には平均健康日数は27.42日(=30.42-3)であるが、標準偏差は0である。しかし、健康日数が減少しなければ大き

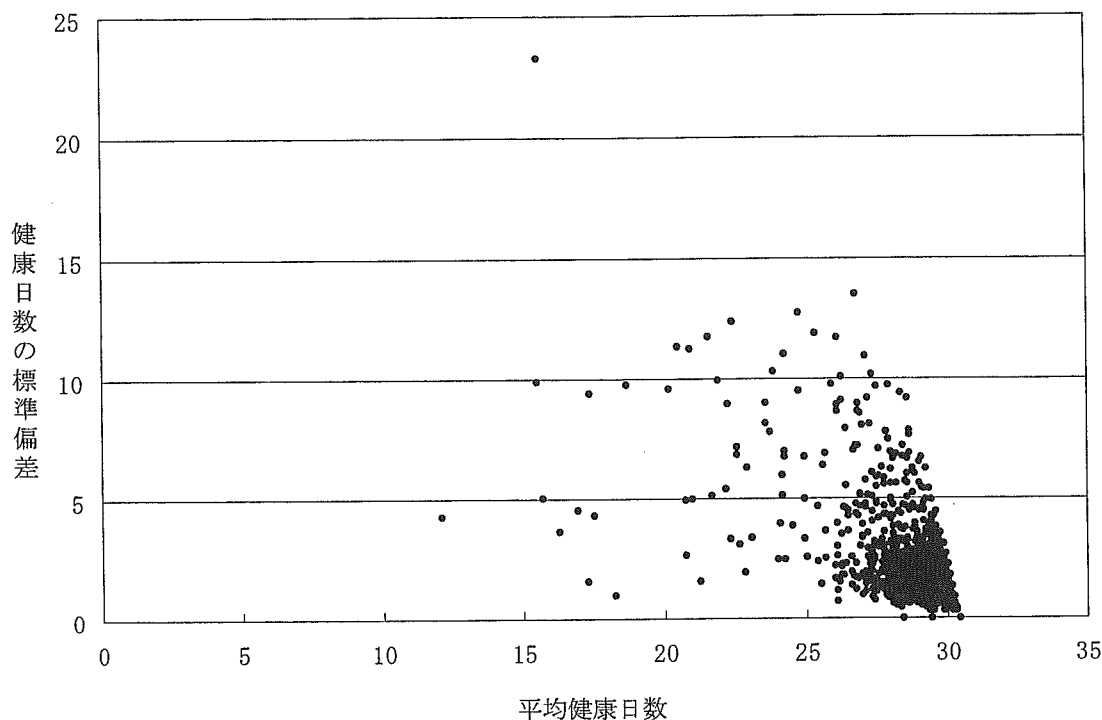


図4 平均健康日数と標準偏差

な標準偏差は発生しないのも事実である。全サンプルを使った場合でも健康日数と標準偏差の間には比較的高い相関が発生する。その結果、回帰分析の推定結果には多重共線性が影響していると考えられる。ただし、今回使用するサンプル数は非常に多く（被保険者男性が約28万人、被保険者女性が約10万人、配偶者女性が約15万人）、多重共線性の影響はある程度緩和されていると考えられる。

c. 受診日数0ダミー

前年度全く受診していない加入者の中もいる。健康日数は年間で365日に制限されてしまうから、それ以上の数字は存在しないから質的な違いを含んでいる。受診日数0ダミーは健康日数が365日以上である可能性を含むことを意味するダミー変数である。

d. 平均標準報酬

平均標準報酬は1998年9月から1999年8月までの12ヶ月間の標準報酬で、単位は千円である。ポー

ナスは含まれない。

e. 年齢ダミー

20-29歳ダミー、30-39歳ダミー、40-49歳ダミー、50-59歳ダミーは加入者の年齢がカテゴリーに一致するときには1、それ以外は0をとる2値変数である。医療サービスの需要に関する実証分析では、年齢または年齢階級ダミー変数は健康の代理変数として使われていることが多い。本論文では、健康の不確実性は平均健康日数や健康日数の標準偏差で表される。したがって、これらのダミー変数は健康以外の要因を反映することが予想される。

4. 推定結果

以下では、被保険者男性・被保険者女性・女性配偶者にわけて推定した。推定に当たってはまずポワソン回帰を実施したが、すべての推定式で推定結果のあてはまりは統計的に有意ではなかった。これは被説明変数の分散が平均を上回るためであ

表3 推定結果

	被保険者男性			被保険者女性		
	推定値	Z値	限界効果	推定値	Z値	限界効果
平均健康日数	-0.1412	-40.57***	-0.2480	-0.1732	-33.22***	-0.4605
健康日数の標準偏差	0.0080	2.75***	0.0141	-0.0023	-0.52	-0.0060
受診日数0ダミー	-0.9110	-163.83***	-1.3599	-0.8542	-86.66***	-1.7608
平均標準報酬	0.0005	33.75***	0.0008	0.0003	10.68***	0.0007
30歳代ダミー	0.0542	9.71***	0.0963	-0.0399	-5.99***	-0.1049
40歳代ダミー	-0.0769	-12.17***	-0.1327	-0.1807	-26.33***	-0.4572
50歳代ダミー	-0.0605	-8.81***	-0.1046	-0.1750	-25.11***	-0.4430
切片項	4.8332	45.6***		6.2766	39.6***	
n	284755			104074		
Log likelihood	-501561			-212129		

	配偶者女性		
	推定値	Z値	限界効果
平均健康日数	-0.1296	-28.2***	-0.3010
健康日数の標準偏差	0.0092	2.39**	0.0214
受診日数0ダミー	-0.9439	-110.53***	-1.7252
平均標準報酬	0.0002	12.47***	0.0005
30歳代ダミー	-0.1419	-17.94***	-0.3219
40歳代ダミー	-0.3529	-41.7***	-0.7856
50歳代ダミー	-0.2907	-32.25***	-0.6242
切片項	4.9980	35.67***	
n	146573		
Log likelihood	-288037		

- 1 ダミー変数の限界効果はダミー変数が0から1に変化したときの効果、それ以外はサンプルの平均値で評価した。
- 2 ***は1%で統計的に有意であることを示す。

る。そこで、分散が平均を上回る場合に対応する Negative Binomial モデルを使って推定した。推定結果は表3に示されている。

表3は被保険者男性、被保険者女性、配偶者女性について被説明変数をエピソード件数として推定した結果である。本論文の関心は、健康の不確実性が受診行動に与える影響について検討することである。理論モデルは健康の期待値（平均健康日数）の増加はエピソード件数に負の効果を持ち、健康の不確実性（健康日数の標準偏差）の増大はエピソード件数に正の効果を持っていることを示していた。実証分析の結果は、被保険者女性を除いて理論モデルの予想と一致する。以下では推定値そのものではなく、平均値で評価した限界効果を使って説明する。

平均健康日数について平均値で評価すると、平

均健康日数が1日増加すると、エピソードが0.2480件から0.4605件の間で減少する。平均健康日数は健康の期待値の代理変数である。この数字が高いほど、消費者は今年の健康の期待値は高くなると推測する。いいかえると前年医療機関に行かなかった日数が多いほど健康の期待値は高まるから、今年のエピソード（治療開始件数、ただし、過去からの継続分を含む）は減少することになる。この効果が最も大きいのは被保険者女性で、平均健康日数が1日増加すると、エピソード件数が0.4605件減少する。一方、効果が一番小さいのは男性で、平均健康日数が1日増加すると、エピソードが0.2480件減少する。また、同じ女性であっても、配偶者女性にくらべて被保険者女性の平均健康日数の効果が大きくなっている。

健康日数の標準偏差について平均値で評価する

と、標準偏差が1日増加すると、エピソードが0.0141件から0.0214件の間で増加する。健康日数の標準偏差は健康の分散の代理変数である。この数字が高いほど、消費者は健康が不確実性にさらされていると推測する。その結果、消費者はリスクをさけるために頻繁に受診するようになる。この効果が最も大きいのは配偶者女性で、標準偏差が1日増加するとエピソードが0.0214件増加する。一方、被保険者男性と被保険者女性ではほぼ同じように、0.0141件増加する。残念ながら、被保険者女性については統計的に有意な結果はえられなかった。

受診日数0ダミーは健康日数が365日で、前年度に一度も受診していないことを示す。健康日数は365日が上限であるという意味で、切断されている。健康日数が365日の加入者の中には365日以上に健康な消費者もいると考えられる。受診日数0ダミーはそのことを反映する。どの推定式でも統計的に有意で、受診日数が0日の加入者は他の加入者に比べて、エピソードが1.7608件から1.3599件減少する。

これらの結果は消費者が健康の不確実性を考慮して受診行動を選択していることをある程度裏付けている。もっとも、平均健康日数と健康日数の標準偏差の間には相関が認められるから、それぞれの推定値が多重共線性の影響を受けている可能性もあることに注意すべきである。ただし、サンプル数が非常に多いため、その影響は限定的であると考えられる。

平均標準報酬の推定値は正で統計的に有意である。つまり、報酬が増加するとエピソードが増加することになる。この効果は個人レベルでは非常に小さく、標準報酬が千円増加したときに0.0002件から0.0005件が増加するにすぎない。

年齢階級ダミーの推定値はどれも統計的に有意で、興味深い結果を示している。これらの推定のリファレンスグループは20歳代である。推定結果からは被保険者男性の30歳代ダミーをのぞいて、

すべての年齢階級ダミー変数は負になっている。

医療サービスの需要などの実証分析では年齢は健康の代理変数として使われることが多い。しかし、ここでは、健康の期待値や分散は別に取り上げられているから、年齢階級ダミーは平均健康日数や健康日数の標準偏差、受診日数0ダミーでは十分に反映することができなかった不確実性や、健康以外の要因を反映していることになる。たとえば、平均健康日数や健康日数の標準偏差が理論モデルの健康の期待値や不確実性を十分に表せず、高齢になるほど不確実性が高まるのであれば、これらのダミー変数は正になるはずである。ところが、実際には負の値をとっている。

これらの負の推定値が、健康への選好が変化したことを反映したとすれば、健康への関心が低下したと考えることもできる。しかし、一般的には40歳代・50歳代の健康への関心が低下するという説明には説得力が欠けると思われる。むしろ、被説明変数のエピソード件数に問題があるとも考えられる。年齢とともに生活習慣病などの治療が長期間にわたる疾病が多くなる。高血圧症で治療中に、風邪をひいたとしても、高血圧症を担当する医師の治療を受けたとすれば新しいエピソードは発生しない。慢性疾患を患っていなければ発生したはずの、新しいエピソードが慢性疾患のエピソードに埋没してしまう可能性がある。また、女性の場合には、20歳代から30歳代の出産期には比較的受診する機会が多くなる可能性もある。

5. まとめ

本論文の目的は健康の不確実性が消費者の受診行動に与える影響について分析することにある。実証分析の結果が示すとおり、健康の不確実性は消費者の受診行動に影響を与えている。健康の期待値が上昇するとはエピソード件数は減少し、健康の不確実性が増加するとエピソード数が増加する。したがって、消費者は健康の不確実性を考慮

して受診行動を決定していることが明らかになった。もっとも、健康の分散が与える影響は必ずしも大きくない。

今後いくつかの点については分析を深める必要がある。まず健康の不確実性を表す指標として健康日数が望ましい性質を持っているかどうかである。受診日数を利用した健康日数の推定は非常に簡便であるが、上限があることなど問題もある。また、受診日数のように目に見える指標だけでなく、同世代との比較や、過去の自分との比較を通じて健康を予想する可能性も考えられる。今後は健康の不確実性の指標そのものについて詳しく分析するべきだろう。

消費者の選好も重要な課題である。本論文ではすべての消費者は同じ効用関数を持っていることを仮定している。同じ不確実性のもとでは、同じ受診行動をとることが前提となっている。しかし、実際には、健康に対してより慎重な消費者もいれば、そうでない消費者もいるだろう。今回の分析は大規模なデータではあるが、横断面データであるために、個人効果には対処することができない。パネルデータによる分析では選好の違いを個人効果に還元できる可能性があるため、より詳細な分析が可能になる。

もう一つの課題は不確実性と医療支出の関係である。本論文で Two Part モデルのうちの第 1 段階のみを対象とした。消費者は健康の不確実性のもとで受診するかどうかを決定しなければならない。その結果、頻繁に受診する消費者が医療資源を消費するのかどうかは、マクロな医療費を考える上で重要な課題である。

参考文献

- Arrow KJ (1963) "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*. 53 : 941-973.
- Cropper ML (1977) "Health, Investment in Health, and Occupational Choice," *Journal of Political Economy*. 85 : 1273-1294.
- Dardanoni V and Wagstaff A (1990) "Uncertainty and the Demand for Health Care," *Journal of Health Economics*. 9 : 23-38.
- Grossman M (1972) "On the Concept of Health Capital and the Demand for Health Care," *Journal of Political Economy*. 80 : 223-255.
- Liljas B (1998) "The Demand for Health with Uncertainty and Insurance," *Journal of Health Economics*. 17 : 153-170.
- Koç Ç (2004) "The Effect of Uncertainty on the Demand for Health Insurance," *Journal of Risk and Insurance*. 71 : 41-61.
- Picone G, Uribe M and Wilson RM (1998) "The Effect of Uncertainty on the Demand for Medical Care, Health Capital, and Wealth," *Journal of Health Economics*. 17 : 171-186.
- Sandmo A (1970) "The Effect of Uncertainty on Saving Decisions," *Review of Economic Studies*. 37 : 353-360.
- Selden T (1993) "Uncertainty and Health Care Spending by the Poor : The Health Capital Model Revisited," *Journal of Health Economics*. 12 : 109-115.
- 酒井泰弘 (1982) 『不確実性の経済学』有斐閣
- 山田武 (2002) 「国民健康保険支払い業務データを利用した医師誘発需要仮説の検討」『季刊社会保障研究』38(1) : 39-51
- 山田武 (2003) 「健康診断の受診と情報としての健康診断の価値」『医療と社会』13(1) : 39-52
- 連絡先：山田 武
t2yamada@cuc.ac.jp

Uncertainty in Health Capital and Demand for Health Care

Takeshi Yamada[†]

Abstract

Negative binomial model was applied to the reimbursement records of large companies in Japan. The data include about 540 thousand male/female employee and spouse reimbursement records for 2 years (1997-1998). The means and variances in health capital could not be directly observed from reimbursement records. However, the number of healthy days, defined as the number of the days individuals spend without health care, could be obtained by subtracting the number of inpatient and outpatient days in each month. In the demand equation for health care, the dependent variable is the number of treatment episodes, and the independent variables include mean healthy days per month in the preceding year, while standard deviation of healthy days by month in the preceding year, in addition to age, wages, etc. Mean and standard deviation of healthy days in the preceding year were used as proxies of the expected health and the degree of uncertainty in health. Estimated results were consistent with the results of theoretical models which modify the hypothesis on the distribution of health capital in Dardanoni and Wagstaff (1990). Mean healthy days has a negative effect on the demand for health care. A one day increase in mean healthy days reduces treatment episodes by 0.09. Standard deviation in healthy days has a positive effect. A one day increase in standard deviation in healthy days increases treatment episodes by 0.02. These results show that patient behavior is related not only to the expected health but also to the variance in health.

Keywords : Uncertainty, Health capital, Demand for health, Healthy days

[†] Faculty of Commerce and Economics, Chiba University of Commerce

健康診断の検査は医療費の予測に有効か*

小 椋 正 立†

将来の医療費は、全般的な健康状態と現在の医療費によって予測できると仮定したモデルを用いて、健康診断の各検査の数値が予測にどれくらい有効かを検定した。健康状態の指標として、健診の各検査値のほか、今期の傷病名から外生性が認められたものを用いた。分析に使用したのは、1996年度から始まる数年間にわたる2つの健康保険組合の約2万人のデータセットである。1年先の医療費については、BMI、総コレステロール、最高血圧は、平均からの乖離は、プラス、マイナスどちらの方向でも、医療費の増加要因となる。これに対して、平均以上の血糖、クレアチニン、尿酸、ALP、赤血球は医療費を引き上げるが、ZTT、 γ GTP、GOTが平均以下であれば医療費は下がる。1年先の医療費については、クレアチニンの上方への乖離が最大の増加要因であるが、この結果は5年先までの医療費について持続する。このほかにも血糖と尿酸は同じような性向を示す。これに対して、血圧については、最高血圧は、上方、下方への乖離どちらも医療費の増加要因になる関係が、最低血圧については、下方への乖離が医療費の減少要因となる関係が、それぞれ安定的に観察される。

キーワード 企業の健康診断, 検査結果, 医療費の予測, 傷病ダミー

1. はじめに

わが国の企業は労働安全衛生法によって、毎年1回、従業員に対して決められた項目の健康診断を提供することを義務付けられており、多くの企業では毎年春または秋に、従業員が集団で企業内の医療施設や健診カーにおいて健康診断を受診する風景が繰り返されている。第二次世界大戦後に、わが国でこのような健康診断が要請されたのは、感染力の強い結核が職場を通じて蔓延するのを食い止めるためであった。しかし、1960年代に結核がほぼ完全に克服されてからは、職場健診の目的

は胃ガンをはじめとするガンの二次予防に移り、さらに最近ではそれが生活習慣病の二次予防に移行しつつあるように見える。もちろん職場での健康診断には、職場での「安全」の確保という目的もある。従業員の健康状態を考慮して作業の内容を決めることは、危険を伴う職場ではとくにその必要性が高いことは明らかである。一般の職場においても、従業員の体調を考慮して職務の負担が過重にならないように調整することは、過労死などの発生を未然に防ぐとともに、高い生産性を維持するために必要なことである。

しかし、健康診断にはかなりのコストがかかる。最近では、すべての就労者に対して、半強制的に実施する医学検査の効果について、さまざまな懐疑的な意見が表明され始めている。平均よりも健康に恵まれている、企業の従業員を対象として健康診断を実施することは、医療資源の配分の効率

* この研究は「医療費データと接合された健診データ等による健診の効果分析 (H15-政策-018)」平成15年度の成果に基づくものである。2004年度日本経済学会 (岡山大学) における柿原浩明教授の有益な示唆に感謝する。

† 法政大学大学院エイジング総合研究所

性と公平性の両面から問題視する見方も少なくない。あるいは、職場での半強制的な健康診断で早期に問題を発見しても、それが必ずしも医療機関で自発的な受診につながっていないとの指摘もある。

この反面、これまで継続的に実施されてきた健康診断はかなりの規模に達しており、多くの専門家や技術者を雇用した、競争的な産業となっている。検査項目の設定、質問票、判定結果の通知方法やそのフォローアップについても、従業員の健康診断の受診率を高めるために、さまざまなノウハウが活かされている。しかもわが国では従業員の健康管理を重要な経営戦略の一つとして位置付けてきた企業も少なくない。とくに最近、現在の医療保険制度の中で、生活習慣病については、個人にも、企業にも、医療保険制度にも、効果的に予防し、コントロールするためのインセンティブが不在ではないか、という指摘がなされている。したがって、ライフサイクルを通じた健康管理システムの中で、職場の健康診断がどのような役割を果たしているのかは、慎重に判断しなければならない。

職場健診も含め、わが国の健康診断の健康診断で実施されている上記の各検査について、二次予防に関する費用効果分析の観点から、現時点での医学と公衆衛生学の知見をサーベイしたすぐれた研究として矢野 他 (1999) がある。矢野ほかの研究は産業医の観点から、各検査の検査値の異常がどのような疾病の発見に結びつき、それについて果たして効果的な介入が存在するかどうかを念頭に置きながら、それぞれの検査について費用効果分析を試みたものである。これに対してこの論文の目的は、職場の健康診断において一般的に実施されている検査の結果に、将来の医療費に関する有効な情報が含まれているかどうかを分析することである。わが国では、毎年、数千万人の労働者がこうした検査を受けているが、その情報は、もっぱら個々の検査結果が正常範囲にあるかどうか

かを判断するために利用されている。個々の検査についてはどのように正常値の範囲を設定するかについては、多くの医学的な分析の蓄積があり、検査値が正常値の範囲を越えていれば、より精密な再検査や医療機関を受診することを前提として、そうでないものについてその情報が利用されることはほとんどない。しかし、もし正常の範囲にあるデータも含めて、すべてのデータから将来の医療費の増加を予測できれば、現在の検査情報の利用方法は効率的ではないと言えるのではないだろうか。

この論文の結論は次のとおりである。すなわち健康診断の検査結果には、数年先までの医療費の予測に有効な情報が多く含まれており、その内容は、糖尿病、高血圧、動脈硬化、腎臓病等の生活習慣病に関わると考えられるものである。

2. 医療費モデルの定式化

誤差を無視すれば、今期の医療費を $Y(t)$ とすると、来期の医療費 $Y(t+1)$ は今期の全般的な健康状態を表す指標関数 $\phi(t)$ と $Y(t)$ の積として表すことができるものと仮定する。すなわち、

$$Y(t+1) = \phi(t) \cdot Y(t)$$

が成り立つが、これを対数表示すると

$$\log Y(t+1) = \log Y(t) + \phi(t) + \varepsilon_t$$

と書くことができる。ここで $\phi(t) = \log \phi(t)$ であるが、指標関数としてはもっぱら $\phi(t)$ を用いることにする。

さらにこの式を少し一般化して、

$$\log Y(t+1) = \theta \log Y(t) + \phi(t) + \varepsilon(t)$$

と表すことにする。この θ は今期の疾病が期中に完治せず、来期まで繰り越すことや、加齢により自然に疾病が重症化することなどによって決まる、個人のいくつかの属性の関数である。また指標関数の ϕ は個人の全般的な健康状態を表しており、具体的には健診検査の検査結果と個人の属性の関数である。

表1 各年度在籍者の1996年から2001年間の通算検査回数の分布と有効標本数

	通算検査回数	1996	1997	1998	1999	2000	2001	Total
	A 組合	0	2,586	2,714	1,998	1,623	1,707	2,389
1		2,308	2,252	1,823	1,656	3,478	4,045	15,562
2		2,074	2,214	2,317	2,188	1,773	3,521	14,087
3		2,633	3,024	3,226	3,008	2,512	2,091	16,494
4		4,799	5,112	5,188	5,193	4,805	4,422	29,519
5		7,957	7,994	7,995	7,994	7,996	7,758	47,694
6		3,489	3,489	3,489	3,489	3,489	3,489	20,934
在籍者合計 有効標本数		25,846 10,900	26,799 11,019	26,036 11,180	25,151 10,523	25,760 9,968	27,715 5,996	157,307 59,586
B 組合	通算検査回数	1996	1997	1998	1999	2000	2001	Total
	0	11,459	10,874	10,462	9,907	9,873	10,288	62,863
	1	2,153	1,678	1,587	1,534	1,550	1,762	10,264
	2	1,784	1,901	1,575	1,510	1,561	1,399	9,730
	3	2,089	2,269	2,476	2,299	2,014	1,792	12,939
	4	1,811	2,132	2,447	2,460	2,084	1,844	12,778
	5	2,660	2,996	2,987	3,008	3,003	2,588	17,242
	6	5,721	5,722	5,718	5,718	5,718	5,718	34,315
	在籍者合計 有効標本数	27,677 8,841	27,572 9,100	27,252 9,294	26,436 0	25,803 0	25,391 0	160,131 86,821

したがって来期の医療費は、 θ と $\phi(t)$ に関連した個人属性を $X(t)$ 、健診の検査結果を $I(t)$ とすると、

$$\log Y(t+1) = \alpha \log Y(t) + \beta' X(t) + \gamma' I(t) + \varepsilon(t)$$

として表すことができる。私たちの関心は $I(t)$ の内のどの検査結果が次の年の医療費の予測に有効であるかを見出すことにある。

また、この定式化を

$$\log Y(t+k) = \alpha \log Y(t) + \beta' X(t) + \gamma' I(t) + \varepsilon(t)$$

$$k = 1, 2, 3, 4$$

として、同じようにどの検査結果が将来の医療費の予測にとくに有効であるかを検討する。

3. データおよびデータベース

ここで用いたデータセットは2つの健康保険組合(A, B)の1996年度から始まる数年分の各従業員の健診結果に関する数値データを各年度の医

療費レセプトデータから計算した総医療費と連結したものである。この間に健診を受けた従業員の実数はA組合では3万人弱、B組合は2万人であり、この人たちが分析の対象となる。それぞれの組合について、受診した健康診断の回数の分布を表1に示している。この中から、分析に必要なデータが完備している従業員と年度だけを選び出したが、その数は1996年度から2000年度までの期間は約1万人、2001年度のみ5千人、B組合では1996年度から1998年度までの期間で約9千人である。ここでの分析に必要なデータが完備しているのは、健診を受診した年度に分析の対象となる検査を受診しており、その年の医療費レセプトデータと標準報酬データが揃っていること、医療費レセプトデータに第一傷病名の記載がある従業員である。

(1) 職場健診の検査項目

職場の健康診断の検査項目は、労働安全衛生法第66条、労働安全衛生規則第44・45条に決められている。それによれば企業は常時使用する労働者

に対し、1年ごとに1回（ただし深夜業労働者等は6ヶ月ごとに1回）、定期的①既往歴および業務歴の調査、②自覚症状および他覚症状の有無の検査、③身長、体重、視力及び聴力の検査、④胸部エックス線検査および喀痰検査、⑤血圧の測定、⑥尿中の糖及び蛋白の有無の検査、⑦貧血検査、⑧肝機能検査、⑨血中脂質検査、⑩血糖検査、⑪心電図検査を実施しなければならないとされている。これは職場の健康診断のいわばミニマムであるが、それでも医師が必要ないと認めるときは、身長（20歳以上）、聴力（45歳未満）、喀痰検査（胸部X線で異常がない場合）、尿検査（血糖検査で異常がない場合）のほか、貧血検査、肝機能検査、血中脂質検査、心電図検査、血糖検査は40歳未満について、それぞれ省略することが認められている。しかも、多くの職場ではこれを超える範囲の検査が行われており、それが年度ごとに変化し、個人の病歴や自覚症状等によっても変えられているのが実情である。

ここではそれぞれの組合において、一次健診の身体測定、血液検査、尿検査の中から、できるだけ多数の従業員が受けている検査項目で、検査結果が数値化されているものを選択した。表2 Aに、それぞれの組合において比較的、実施頻度が高い検査項目とそれぞれの検査結果に関する記述等計量と、それぞれの検査の正常値を示している。ここで分析に用いたのは、Aについては8万人回以上、Bについては6万人回以上の実績があつて、かつ結果が数値化されている検査項目（*を付したもの）に限った。さらに表2 Bには有効標本におけるこれらの検査項目の記述等計量を示している。

(2) 検査結果の変換

上に述べたように、原則として検査結果が数値化されているものを選択した。しかし、検査によっては、健康状態が検査数値の単調な関数ではないものがある。たとえば血中の総コレステロールの数

値が高ければ血管の梗塞のリスクが高くなることは良く知られているが、コレステロールは血管の保護作用も持つため、反対に低すぎる水準も医療費を高める可能性があるといわれている。このようなケースを考慮して、私たちはすべての検査結果の数値をプールした標本の総平均からの乖離（偏差）の絶対値で表すことにしたが、総平均よりも大きいものは変数名の頭に+を付けて表し、総平均よりも小さいものは変数名の頭に-を付けて表した。たとえばBMIであれば、総平均よりもBMIが大きい人は+_bmiがプラスの数値を取り、-_bmiはゼロである。これに対して総平均よりもBMIが小さい人は-_bmiがプラスの値を取り、+_bmiはゼロとなる。これを式で書くと、

$$+_bmi = \max(0, bmi - \text{mean}(bmi))$$

$$-_bmi = \max(0, \text{mean}(bmi) - bmi)$$

である。

このため医療費の推計式には、1つの検査に対応する変数が2つ存在しており、もし両方の符号がプラスであれば、医療費から見て、標本の平均値が最も理想的な数値であることを示している。他方、+_変数の係数がプラス、-_変数の係数がマイナスであれば、検査数値が高いほど医療費が上昇することを示し、その反対に+_変数の係数がマイナス、-_変数の係数がプラスであれば、検査数値が高いほど医療費が低下することを示している。このように、この定式化は、検査数値が健康状態と単調に対応していない場合にも対応できる利点があるが、真の構造が単調な場合には、*t* 値の低下を招くことは避けられないのが難点である。今回はかなりの標本数が得られたため、原則としてこのような変数の分割を行った。しかし、2値変数の場合や推計結果が変則となる場合には、例外的に元の変数のままとしたものもある。

(3) 診断病名の外生性の検定

公的医療保険の医療費は診断病名と処置内容と

表2 A 検査項目の実施頻度と記述統計量（正常値のMは男子）

検査項目	正常値	A組合					B組合				
		Obs	Mean	Std.Dev.	Min	Max	Obs	Mean	Std.Dev.	Min	Max
身長		109026	167.747	7.135	106	200	76580	165.704	7.976	132.3	198.9
体重		109006	64.927	10.732	27.5	168.5	76573	63.057	10.486	29.3	154.6
視力裸眼右		74214	0.879	0.779	0	9.9	48273	0.836	0.43	0	7.22
視力裸眼左		74212	0.89	0.77	0	9.9	48305	0.855	0.436	0	2
聴力1000Hz右	1	84214	1.024	0.152	1	2	69965	1.044	0.204	1	2
聴力1000Hz左	1	84207	1.024	0.153	1	2	69961	1.045	0.206	1	2
聴力4000Hz右	1	83980	1.135	0.342	1	2	69918	1.09	0.286	1	2
聴力4000Hz左	1	83965	1.14	0.347	1	2	69914	1.093	0.291	1	2
血圧（1回目）高	<140	108915	124.538	16.567	0	250	76554	123.114	16.643	74	230
血圧（1回目）低	<90	108917	76.385	11.794	0	180	76547	75.619	11.592	0	148
総コレステロール	120~219	94959	200.043	35.178	55	611	76426	201.448	35.657	57	2217
中性脂肪	30~140	94949	123.638	101.577	6	3856	76422	125.156	104.433	10	5108
HDL コレステロール	M40~70	93551	53.908	16.088	1	464	71806	56.83	15.721	1	771
尿素窒素	8~23	90793	14.933	3.971	0	177	33762	15.204	4.29	1.2	237
クレアチニン	0.5~1.2	90612	0.948	0.392	0	50	72246	0.883	0.318	0.1	17.2
尿酸M	3.8~7.5	93545	5.732	1.343	0	58	70448	5.701	1.432	0.1	62
尿糖（定性）	1	108078	1.077	0.477	1	6	75490	1.068	0.439	1	6
尿蛋白（定性）	1	108109	1.104	0.474	1	6	75496	1.124	0.492	1	5
尿潜血	1	101876	1.168	0.64	1	6	70036	1.275	0.795	1	5
GTT 血糖（負荷前）	70~110	89503	98.255	22.16	0	557	61406	98.546	19.928	40	540
HbA1C	4.3~5.7	57825	5.028	0.772	0	51	3484	5.581	1.172	3.7	14
ZTT	4~12	57131	2.032	0.856	1	6	64020	5.943	2.944	0	51.8
GOT	5~40	95876	24.075	12.686	4	797	66241	23.694	12.733	4	596
GPT	5~35	95875	27.237	21.631	0	712	66251	25.839	19.695	1	790
γGTP	<60	95880	40.407	50.987	2	1724	66118	36.302	50.98	2	2330
ALP (IU/l)	80~240	86956	173.245	61.506	0.1	1440	62488	164.343	65.639	1	1969
総ビリルビン	0.2~1.2	89905	0.742	0.332	0	7.6	4975	0.8	0.51	0.1	18.6
総蛋白	6.7~8.3	88790	7.349	0.411	0	14.7	24199	7.317	0.417	1.5	19
血清アミラーゼ	55~175	80023	92.018	40.358	7	742	2630	116.461	72.742	6	1450
赤血球	M430~570	95675	485.333	39.545	13	696	62220	467.706	41.544	70	667
ヘモグロビン	M13.5~17.5	95670	15.022	1.277	2.9	47.7	62184	14.602	1.444	3.1	70
ヘマトクリット	M42~53	95010	45.327	3.648	13	67.9	62020	44.203	4.042	2.4	64.8
白血球	3300~9000	94911	5637.659	2405.379	48	30000	31504	6175.762	1786.936	34	70000
血小板	12~35	88074	24.133	5.481	2.3	154.6	26344	23.622	5.427	2.5	155.1