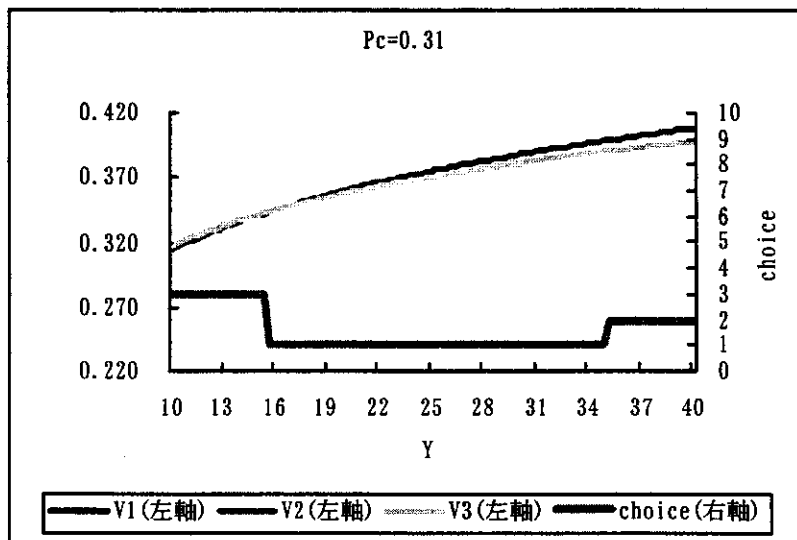
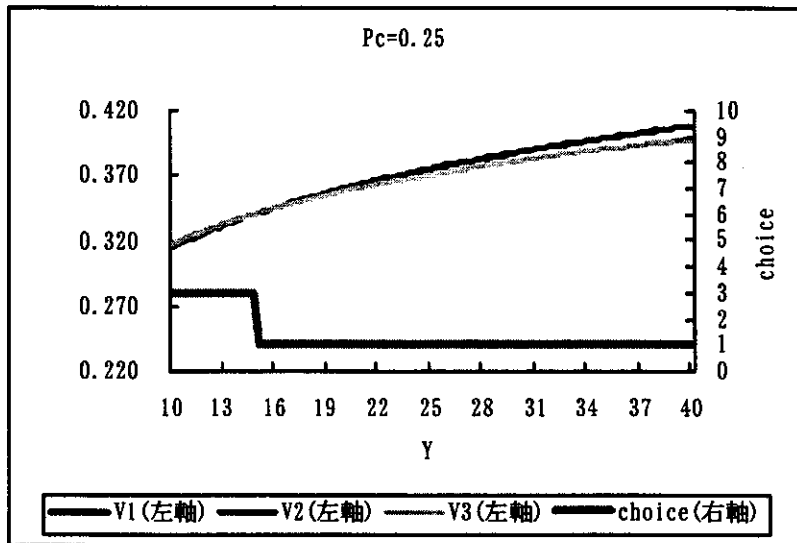
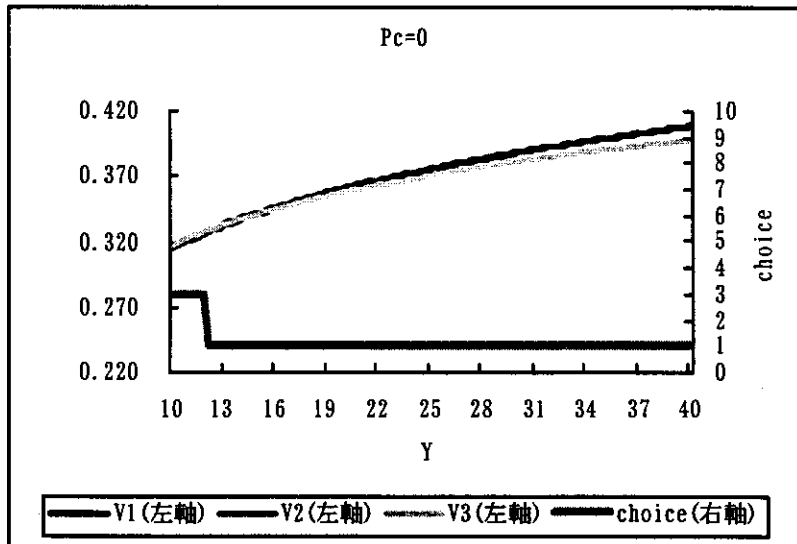
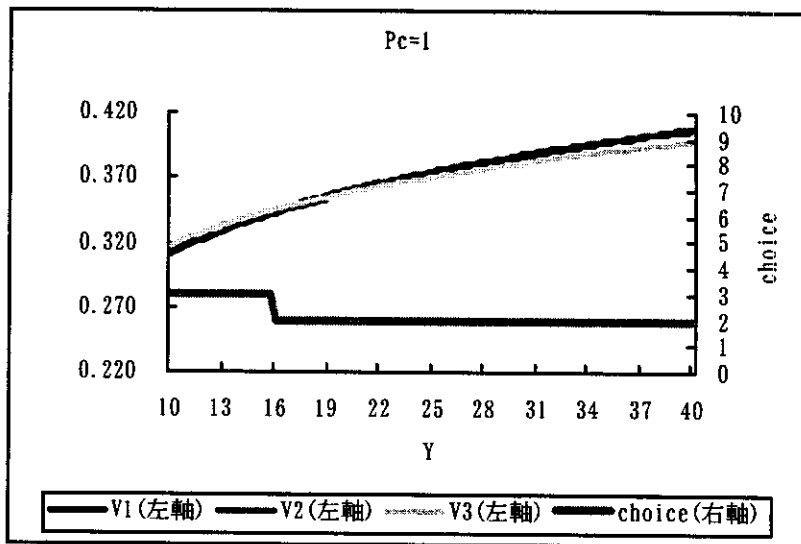
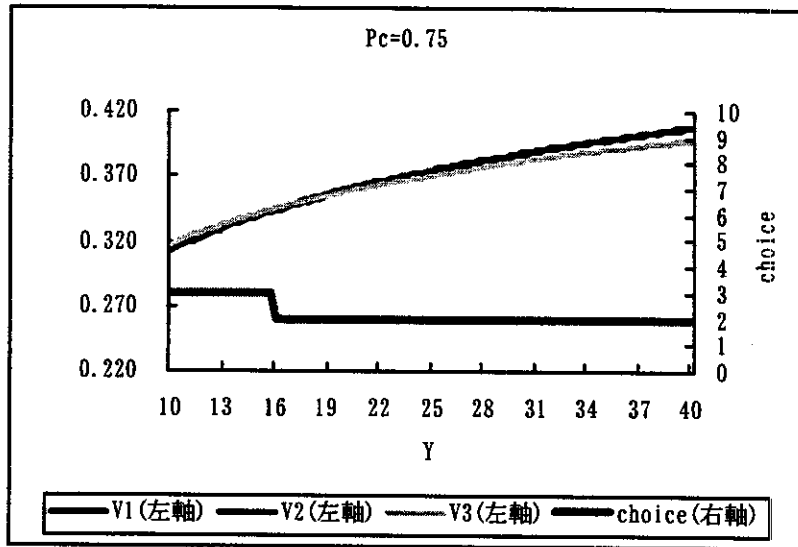
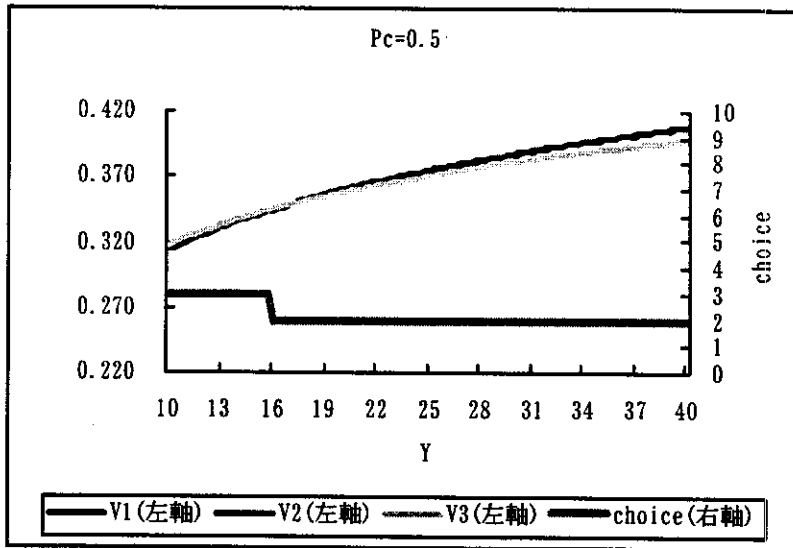


図表 12 所得(Y)の影響 1



図表 12 所得(Y)の影響 2



厚生科学研究費補助金（政策科学推進研究事業）

（分担）研究報告書

健康診断の分析の基礎となる患者受診動態の分析

分担研究者 泉田信行 国立社会保障・人口問題研究所

本研究では健康保険組合のデータを用いて患者受診行動をエピソードベースで把握したパネルデータを作成し、それによって患者受診行動の要因について実証的に検討する。これによって、個別の患者の属性の効果を考慮した患者受診行動の分析を行い、健康診断受診と医療機関受診行動の関係の分析の一助とする。

健康保険組合から提供された患者受診行動に関するデータを利用し、患者受診行動をエピソードベースで測定したパネルデータを作成した。パネルデータを用いた Poisson Regression を行い、次のような結果を得た。(1)男性に比して女性の方がエピソードの発生頻度が高く、(2)40歳を超えるとエピソード数が減少すること、(3)家族の方が本人よりもエピソード数が多いこと、(4)同一世帯の家族構成員数が多いほど個人のエピソード数が増加すること、(5)世帯所得が高いほど個人のエピソード数が多いこと、を示唆する結果を得た。これらの結果はこれまでの先行研究において得られてきた結果と整合的な結果である。

これまで日本で行われてきた医療機関受診に関する分析に対して、より精密なデータを用いて分析を行った結果、これまで得られてきた結果が確認されただけでなく、新しい知見も得られたため、この新しいデータセットに健康診断データをマッチングすることにより、健康診断受診と医療機関受診の間関係を分析することはこれまでの医療機関受診の分析を踏まえた上でより新しい知見が得られると考えられ、有用であると言えよう。

A.研究目的

1997 年以来医療制度の抜本改革に対して財政収支を改善することを目的とする患者自己負担率の引上げが行われてきた。その政策の理論的・実証的背景となるべき研究、自己負担（率）の患者受診行動に与える分析、は日本でも行われてきたが、未だ不十分であると考えられる。そこで、本研究では健康保険組合のデータを用いて患者受診行動をエピソードベースで把握したパネルデータを作成し、それによって患者受診行動の要因について実証的に検討する。これによって、個別の患者の属性の効果を考慮した患者受診行動の

分析を行い、健康診断受診と医療機関受診行動の関係の分析の一助とする。

B.研究方法

3 つの健康保険組合から提供された患者受診行動に関するデータを利用し、パネルデータを作成する。この際に、患者受診行動はエピソードベースで測定する。提供されたデータは 5 年間の受診行動を追跡できるものであるが、途中から加入資格を得た者や加入資格を失った者はサンプルから除外した。この結果、サンプル数は 3 組合で 201,350、41,690、90,100 となった。分析対象となる実人員はこ

れをそれぞれ5で除した数となる。

被説明変数を各個人の各年度のエピソード数とした。エピソードとは治療の開始から終了までを指す。エピソードの数は、治療日数や医療費と異なり、個人がコントロールできるものではないと考えられる。それゆえ、パネルデータを用いたPoisson Regressionを用いた。説明変数は利用可能な情報から女性ダミー、20歳階級を基準とする年齢階級ダミー、家族ダミー（加入者本人を基準とする）、家族規模（人数）変数、1人当たり所得変数、とした。なお、医療価格変数については、ランダムに発生するエピソードにおいて受診を決定する際に、価格情報を事前に用いることは不可能であると考えられるため、説明変数に含めなかった。

C. 研究結果

サンプルの属性については後掲の表1から表7にまとめられている。パネルデータを用いたPoisson Regressionの推定結果については表8にまとめられている。それによれば、(1)男性に比して女性の方がエピソードの発生頻度が高く、(2)40歳を超えるとエピソード数が減少すること、(3)家族の方が本人よりもエピソード数が多いこと、(4)同一世帯の家族構成員数が多いほど個人のエピソード数が増加すること、(5)世帯所得が高いほど個人のエピソード数が多いこと、を示唆する結果を得た。これらの結果はこれまでの先行研究において得られてきた結果と整合的な結果である。

D. 考察

本研究で行った分析はこれまで日本で行われてきた医療機関受診行動の分析をさらに進めるものである。医療機関を受診するか否か

という2値的な選択ではなく、確率的に発生すると考えられるエピソードがどのような個人属性によって影響を受けるのかを明らかにした。さらに、個人属性のみならず世帯所得や家族の規模などのひとまとまりの世帯としての要因も考慮に入れた。このような特性を持つパネルデータを用いた分析を行っている。この点で方法論的にも使用したデータとしても新しい研究であると考えられる。

E. 結論

これまで日本で行われてきた医療機関受診に関する分析に対して、より精密なデータを用いて分析を行った結果、これまで得られてきた結果が確認されただけでなく、新しい知見も得られた。この新しいデータセットに健康診断データをマッチングすることにより、健康診断受診と医療機関受診の間の関係を分析することはこれまでの医療機関受診の分析を踏まえた上でより新しい知見が得られると考えられ、有用であると言えよう。

F. 研究発表

なし

G. 知的所有権の取得状況

なし

資料 1 (泉田信行 国立社会保障・人口問題研究所)

エピソードベースで測定したパネルデータの属性と
Poisson Regression の推定結果

資料 2 (佐藤雅代 国立社会保障・人口問題研究所)

「分析に使用する個票データの概要について」

資料 1 データの属性と Poisson Regression の結果

表1 性別サンプル数

性別	62	68	991
男性	114850	22705	42215
女性	86500	18985	47885
合計	201350	41690	90100

表2 本人家族別サンプル数

本人家族別	62	68	991
本人	86030	17205	48810
家族	115320	24485	41290
合計	201350	41690	90100

表3 年齢階級別サンプル数

年齢階級	62	68	991
0	14405	2165	6930
10	33385	7520	8810
20	29080	6595	9225
30	34485	5595	26680
40	32155	7925	13825
50	43000	8125	12990
60	8060	2695	9030
70	3055	505	1515
80	3160	470	790
90	565	95	305
合計	201350	41690	90100

表4 世帯規模

世帯規模	組合名		
	62	68	991
1	26145	5480	23460
2	23565	5580	14825
3	42155	8020	18075
4	71115	15140	23860
5	29925	6235	8175
6	6770	970	1465
7	1450	240	210
8	225	25	30
合計	201350	41690	90100

表5 エピソード数

エピソード数	組合名		
	62	68	991
0	43776	9554	18101
1	58645	12441	25260
2	54039	11096	25010
3	32893	6377	15818
4	10489	1924	5163
5	1426	282	707
6	82	16	41
合計	201350	41690	90100

表7 一人当たり所得記述統計

組合名	標本数	平均	標準偏差	最小	最大
62	201350	157.456	86.25069	0	980
68	41690	194.0386	100.6444	0	980
991	90100	210.0546	119.5875	0	980

表6 年齢階級別エピソード数

組合62 エピソード数	年齢階級											合計
	0	10	20	30	40	50	60	70	80	90	合計	
0	1310	3970	7324	6716	7289	11173	2482	1558	1660	294	43776	
1	2699	8143	8849	10457	10109	13751	2529	931	986	191	58645	
2	4473	10483	7263	9867	8739	10688	1806	340	328	52	54039	
3	4088	7844	4163	5570	4536	5474	920	161	118	19	32893	
4	1600	2611	1279	1667	1290	1647	280	53	53	9	10489	
5	223	322	189	198	182	248	40	11	13	0	1426	
6	12	12	13	10	10	19	3	1	2	0	82	
合計	14410	33385	29080	34485	32155	43000	8060	3055	3160	565	201350	

組合68 エピソード数	年齢階級											合計
	1	10	20	30	40	50	60	70	80	90	合計	
0	198	894	1702	1224	1908	2170	865	278	267	48	9554	
1	444	1813	2042	1751	2559	2663	848	152	139	30	12441	
2	675	2410	1664	1557	2169	1951	568	49	42	11	11096	
3	589	1718	891	817	1013	1007	303	20	14	5	6377	
4	229	602	252	221	243	277	87	5	8	0	1924	
5	29	79	44	24	30	52	22	1	0	1	282	
6	1	4	0	1	3	5	2	0	0	0	16	
合計	2165	7520	6595	5595	7925	8125	2695	505	470	95	41690	

組合991 エピソード数	年齢階級											合計
	1	10	20	30	40	50	60	70	80	90	合計	
0	674	1087	2051	4549	2716	2882	2913	681	385	163	18101	
1	1365	2047	2740	7436	4104	4050	2705	448	277	88	25260	
2	2142	2726	2486	7912	3947	3466	2006	211	78	36	25010	
3	1902	2113	1409	5003	2270	1914	1039	124	30	14	15818	
4	745	752	477	1546	673	591	315	44	16	4	5163	
5	99	78	60	228	109	78	45	6	4	0	707	
6	3	7	2	6	6	9	7	1	0	0	41	
合計	6930	8810	9225	26680	13825	12990	9030	1515	790	305	90100	

表8 推定結果

説明変数	係数	標準誤差	z	P> z	係数	標準誤差	z	P> z	係数	標準誤差	z	P> z
女性ダミー	0.1347949	0.0069469	19.4		0.1261204	0.0152095	8.29		0.1841934	0.0091929	20.04	
0歳階級	0.4349656	0.0114778	37.9		0.404806	0.0274649	14.74		0.4097316	0.0180785	22.66	
10歳階級	0.3049873	0.0094269	32.35		0.2809142	0.0199547	14.08		0.3002773	0.0173562	17.3	
30歳階級	0.0508001	0.0097182	5.23		0.0662195	0.0218653	3.03	0.002	0.0677487	0.0137768	4.92	0
40歳階級	-0.039676	0.0099884	-3.97		-0.06053	0.0209406	-2.89	0.004	-0.027674	0.0158065	-1.75	0.08
50歳階級	-0.092007	0.0095169	-9.67		-0.097273	0.0208746	-4.66		-0.101862	0.0159867	-6.37	
60歳階級	-0.162088	0.0154587	-10.49		-0.132423	0.0282038	-4.7		-0.197086	0.0172684	-11.41	
70歳階級	-0.660713	0.0263576	-25.07		-0.841018	0.0672713	-12.5		-0.455593	0.0351705	-12.95	
80歳階級	-0.720598	0.026474	-27.22		-0.888033	0.0706459	-12.57		-0.679733	0.050912	-13.35	
90歳階級	-0.761938	0.0616991	-12.35		-0.719439	0.1446655	-4.97		-0.714501	0.082323	-8.68	
家族ダミー	-0.14892	0.0091101	-16.35		-0.028999	0.0195679	-1.48	0.138	-0.205848	0.0121736	-16.91	
家族規模	0.030287	0.0030935	9.79		0.0614979	0.0069107	8.9		0.046213	0.0043517	10.62	
1人当たり所得	0.0001055	0.0000445	2.37	0.018	0.0007079	0.0000811	8.72		0.0003708	0.0000397	9.35	
定数項	0.300058	0.0167206	17.95		-0.013897	0.0356197	-0.39	0.696	0.2355378	0.0205479	11.46	
/lnalpha	-2.019828	0.0152991	-2.049813	-1.989842	-2.052209	0.034802	-2.120419	-1.983998	-2.094508	0.0234152	-2.140401	-2.048615
alpha	0.1326783	0.0020299	0.1287589	0.136717	0.1284509	0.0044703	0.1199813	0.1375183	0.1231308	0.0028831	0.1176077	0.1289133
obs	201350				41690				90100			
Wald	chi2(13) =	5986.49			chi2(13) =	1379.68			chi2(13) =	2419.27		
Log	likelihood =	-299854.4			likelihood =	-61397.45			likelihood =	-136188.2		

「分析に使用する個票データの概要について」

佐藤雅代

国立社会保障・人口問題研究所研究員

1. はじめに

本プロジェクトでは、3つの健康保険組合の、被保険者本人及び被扶養者の①適用データと②医療機関受診のレセプトデータ、被保険者本人の③健康診断データを使用している。データの対象とする期間は、1996年4月から2001年3月までの5年間である。これらのデータについてその概要を述べると同時に、データクリーニングの必要性について検討することを、本稿の目的とする。

3つの健康保険組合は、062組合、068組合、991組合と番号で識別され、順に、機械産業、石油産業、証券産業の健保組合である。しかし、これらのデータは、あくまでも匿名企業の匿名社員のデータであり、マッチングのために同一個人を識別する方法は、第三者機関が任意につけたシーケンス番号が同一かどうかだけであり、匿名性は完璧に確保されている。

①適用データとは、各健保組合の被保険者に関するデータである。個人識別のための情報と、被保険者本人に関しては毎月の標準報酬の額が入力されている。②レセプトデータとは、各月の被保険者の受診行動がわかるもので、適応データと同様の個人識別のための情報に加えて、医療機関ごとの受診に関する情報が入力されている。ただし、すべての項目が入力されているわけではなく、5年間にまったくレセプトがない被保険者もいるし、逆に一ヶ月に複数枚のレセプトがある被保険者もいる。③健康診断データは、基本的には被保険者本人を対象とする健康診断の際のデータである。大きくは、血圧等の各検査の数値を示す結果データ、各検査の数値をもとに医師が下した判定を示す判定データ、自覚症状や日常生活に関する自己申告や診察所見等を示す健康調査データ、既往歴に関する既往歴データの4つに分かれている。以上の3つのデータについて、以下で、それぞれ概要を述べる。

2. 適用データについて

①適用データとは、各健保組合の被保険者に関するデータである。個人識別のためには、世帯番号、本人家族の別、続柄、性別、生年月日、資格取得日、資格喪失日が入力されており、被保険者本人に関しては、毎月の標準報酬の額が入力されている。概要を表1に示す。

提供をうけたファイルのレコード数はそれぞれ表1に示した通りだが、062組合についてのみ、世帯番号から資格喪失日までの個人識別キーがまったく同じレコードが3つあったため、これをエラーデータとして除外している。

本来、適用データは、被保険者1人につき1レコードであるが、提供を受けたデータについては、1. 本人の転勤の度に世帯番号が変更されている、2. いったん資格を喪失してからまた資格取得する個人がいる、ことから、必ずしも1人①レコードではなかった。そこで、それらを調整して、あらためて被保険者数を計算したところ、当初のレコード数に対して65.7%、64.5%、62.0%まで個人数が絞り込まれた。なお、被保険者数に対する本人数の割合は40.0%、38.7%、55.0%である。実数からすると、062組合、991組合、068組合の順に規模が大きな組合である。

被保険者数がわかったところで、この中での死亡および出生の数を計算する。これは、別途提供を受けた、組合健保からの現金給付に関するデータより導出した。死亡については埋葬費が給付された件数を、出生については出産手当金が給付された件数を計算している。被保険者本人のうち、0.44%、0.46%、0.37%が死亡していることがわかった。

以上、個人についてみてきたが、次はそれら個人の集まりである世帯について見てみる。まず、世帯番号の変更回数は0回～5回であるが、これは被保険者本人の転勤の回数とほぼ同値である。062組合では最高5回の世帯が1世帯あり、068組合及び991組合での最高は4回である。また、世帯規模に関しては各組合にそれほど差がなく、最高規模が9人と8人であるが、4人までの世帯が90.9%、90.7%、96.4%と9割以上を占めている。ただし、単身世帯については、991組合の62.0%が、他の36.9%や31.6%に比較して非常に高い割合となっている。

3. レセプトデータについて

②レセプトデータとは、各月の被保険者の受診行動がわかるもので、適応データと同様の個人識別のためのデータに加えて、診療年月、一般・老人の別、診療開始日、診療区分、都道府県、医療機関、疾病大分類、疾病119分類、日数、点数、食事療養費、一部負担額、薬剤一部負担額、食事療養費標準負担額、自己負担額、高額療養費、附加金、公費負担額が入力されている。ただし、すべての項目に数値が入力されているわけではない。また、レセプトデータは、医療機関を受診した者についてのみ作成されているものであるため、5年間にまったくレセプトが無い被保険者もいるし、逆に一ヶ月に複数枚のレセプトがある被保険者もいる。概要を表2に示す。

適用データの際と同様に、まず提供を受けたレコード数を示し、続いてエラーレコード数を示している。ここでは、前述適用データに該当する個人のいないレコードをエラーレコードとしている。各組合とも0.16%、0.04%、0.05%とエラーの発生は非常に低率であったため、以後該当レコードは削除するものとする。レコード数からこのエラーレコードを削除した値が、レセプト総枚数である。

通常、レセプトを用いた分析は、医療機関を受診した者についてのみしか対象とすることができないが、このデータは適用データとレセプトデータをマッチングすることにより、まったく受診していない被保険者を特定することができる。その数詞を5年間無受診の被保険者数、及び

うち本人数として示している。本人については被保険者の3.5%、3.4%、3.9%が無受診であった。つまり、残りの被保険者本人は5年間に最低1回以上医療機関において保険診療を受けていることになる。

レセプト1枚当たりについて平均値、最小値、最大値を示している。レセプトでは1点当たり10円であるので、医療機関に発生したコストは総点数×10円と考えることができる。これらの費用のうち、一定割合が保険負担、残りが自己負担、そしてまた場合によっては高額療養費や公費負担等で賄われることになる。

次に、受診した個人単位でみると、一人あたりのレセプト枚数は平均して33.7枚～37.0枚となっている。これは、概算すると1年あたり6.7～7.4枚のレセプトとなり、2ヶ月に1回程度の受診が平均であることを意味する。ただし、これはあくまでも平均値であり、最大受診回数は477回、470回、763回となっており、これは1ヵ月に8件以上の医療機関に受診したことを意味する。ただし、この点については、頻度や点数の水準をみた上で検討するべきであるので、これは今後の課題とする。

個人単位での総点数や自己負担額については表に示す通りである。最小値と最大値に非常に大きな乖離があることから、平均値で議論することが妥当であるとは言いきれないが、平均値のレベルでは各組合にそれほどの差はないようである。個人単位の最大総点数を金額ベースでみると、1.6億円、0.8億円、1.2億円と非常に高い水準となっているが、自己負担額の最大レベルをみると、323万円、302万円、251万円であり、高額療養費等の制度により、自己負担が総点数に対して1.97%、3.60%、2.07%と低水準に抑えられていることがわかる。

4. 健康診断データについて

③健康診断データは、基本的には被保険者本人を対象とする健康診断の際のデータである。大きくは、血圧等の各検査の数値を示す結果データ、各検査の数値をもとに医師が下した判定を示す判定データ、自覚症状や日常生活に関する自己申告や診察所見等を示す健康調査データ、既往歴に関する既往歴データの4つにわかれている。健保組合ごとに実施する健康診断の回数や項目が異なるが、重複するものも多い。概要を表2に示す。

まず、結果データに関して、提供をうけたレコード数をカウントした上で、エラーレコードを示す。ここでは、適用データに該当する個人のいないレコードをエラーレコード①とし、健診区分が不明なレコードをエラーレコード②としている。991組合のエラー発生率が①について2.2%、①+②で4.0%と若干高めであるが、残りの組合については0.56%、0.63%と非常に低率であったため、以後該当レコードは削除するものとする。レコード数からこのエラーレコードを削除したレコードを、以後の分析に用いる。

レセプトデータの時と同様、5年間、まったく健康診断を受診しなかった被保険者をカウントした。なお、当該健康診断は、本人を対象とするものであり、レコードはすべて本人のものであ

る。068組合において被保険者本人のうち1.7%が健康診断無受診であったのが、医療機関無受診の割合より低かったのに対して、062組合と991組合においては8.2%、8.5%と医療機関無受診の倍以上の人が健診無受診であることがわかった。

受診回数についてみると、組合ごとに差があることがわかる。これは、提供される健診の数に違いがあることと、健診受診に対する組合の熱心さの違い、また在職期間の長短に起因するものと考えられる。受診頻度に関しては、在職期間との関係で議論すべきであるが、これは今後の課題とする。なお、健診区分については、受診者数とあわせて表4に示している。

続いて、判定データについてだが、レコード数、受診者数に関しては、上述の結果ファイルと同じであった。念のため、レコード数とエラーレコード数と判定レコード数についてのみ表3に示している。

健康調査ファイルは、自覚症状、日常生活に関しては健診受診者の自己申告、診察所見と検査指示項目については医師の判断が、それぞれデータとして入力されている。自覚症状については、しこりがあるや、めまいがするといった健康に関する異常に関する自己申告を、日常生活に関しては食欲や睡眠といった日常生活に関する質問事項に加え、タバコや酒といった嗜好品に関する質問事項に関する回答が入力されており、本プロジェクトで注目頻度の高い部分である。今までと同様に、レコード数を示した。すると、結果データや判定データと比較して、レコード数が少ないことがわかったが、これは自己申告に基づくためである。エラーレコードは、適用データに該当する個人のいないレコードである。エラーデータを除いた健康調査レコード数は、結果レコードや判定レコードの16.9%、62.2%、49.5%であった。

最後の既往歴ファイルは、該当する疾患のコードと転帰とその疾患にかかった年齢を複数個示したファイルである。健康調査ファイル同様自己申告であることに加え、とくに既往歴が無い場合にはレコードが無いため、4.16%、7.45%、6.35%と結果レコードや判定レコードに対する割合が低くなっている。

5. おわりに

以上、提供をうけたデータを、主にレコード数の観点から概観した。非常に整備されたデータではあるが、エラーレコードが散見されたことから、分析にあたっては精緻なデータクリーニングが必要である。

特に、被保険者である期間に関しては、より細かくチェックする必要がある。なぜなら、適用データの中で、1996年4月から2001年3月までの対象期間中に、資格取得と喪失を繰り返す被保険者が存在するためである。資格取得及び喪失に関しては、1日単位でわかるため、どのレベルでコントロールするかをコンセンサスを研究班内にとる必要があるであろう。つまり、資格喪失と取得の間がどの程度なら、継続した資格保有者とするのか、といった点である。

医療費データと健診データすなわち組合の健康管理データの接合という経済学的には初めて

の試みを成功させるためにも、データの精査が今後とも必要であることが予想される。エラーレコードの抽出については、現段階ではまだ初歩的な段階に過ぎないが、エラーレコードは、今後分析を進めるにあたって次々と抽出されていくものとする。研究班の研究が進むたびに、エラーをチェックし、それらを除いたもので再度計算しなおし、最終的に妥当なデータを用いたしっかりした研究結果を作り上げることを、今後の課題としたい。

厚生科学研究費補助金（政策科学推進研究事業）

（分担）研究報告書

喫煙、飲酒など生活習慣間の相関とこの生活習慣の自覚症状に及ぼす効果の検出

分担研究者 河村 真 法政大学経済学部

1996年から2000年にかけて行われた3社の健康診断個票データの中で、喫煙、飲酒、自覚症状など生活習慣や検診受診時の健康状態を問診した健康調査票を用いて、喫煙と飲酒など生活習慣間の相関や喫煙、飲酒など生活習慣と検診時の健康状態（自覚症状）の相関を見た。

まず、生活習慣間の相関に関する結果を要約すれば、特に、飲酒と喫煙の相関が高いと見られ、さらに、喫煙者に限れば、喫煙の頻度が高ければ、飲酒の頻度が高いことが顕著に見られる

次に、自覚症状と生活習慣の相関の結果を要約する。検診時の健康状態を問うため設けられた自覚症状の項目は、36項目に上るが、当該自覚症状有りと回答する可能性が高い自覚症状を4つ選び、喫煙、飲酒など生活習慣と自覚症状の相関を見た。喫煙が、これら4自覚症状に関して、ありと回答する可能性を高めることが特徴的である。しかし、飲酒やカフェインの摂取は、それぞれ、自覚症状ありと回答する可能性を高める効果を一貫して有すとはいえない。他に、これらいづれの自覚症状に関しても、自覚症状ありと回答する可能性を高めるものに、受診者の高齢と良好でない睡眠状況が挙げられる。

A. 研究目的

われわれが利用可能な3社の健康診断個票データは、喫煙、飲酒など日常の生活習慣、検診者が認識する健康状態（自覚症状）、検診結果さらに医療機関への受診情報を突合させて見ることができる。その中で、生活習慣間の相関や生活習慣と検診者が認識する健康状態（自覚症状）や検診結果との相関を見ることにより、喫煙、飲酒など習慣性の高い行為が健康状態に影響を与えるとする仮説をわれわれのデータセットを用いて確認できる。さらに、これら習慣性の高い日常行為が、検診者本人の健康状態に関する認識に影響を及ぼしているかを確認できることもわれわれのデ

ータセットを用いる際には興味深い。われわれのデータセットでは、会社により程度は異なるが、健康診断受診者の1割から5割程度のみが生活習慣や自覚症状を聞く健康調査票に回答していない。そのため、本稿では、健康調査票を回答した標本に限って、生活習慣間の相関そして生活習慣と受診者が認識する健康状態の相関を見ることに絞って、喫煙、飲酒など習慣性の高い行為が検診者本人の健康状態に関する認識に影響を及ぼしているかを確認する。

B. 研究方法

習慣性の高い生活習慣として、喫煙状況およ

び飲酒状況別に、その他の生活習慣（カフェインの摂取状況、睡眠、食欲、便通の状況、）や受診者の年齢階層の相対頻度分布をとり、それらを比較することにより、飲酒状況や喫煙状況と他の生活習慣との相関を推測する。さらに、受診者が認識する健康状態（自覚症状）を「せき、たんが多い」、「胃が不快、痛む、もたれる、はる」、「背中や腰が痛い」、「だるい、疲れやすい、酒に弱くなった」の4項目に限り、飲酒状況、喫煙状況など生活習慣とのクロス-tabュレーションをとり、これら自覚症状の有無別にこれら各生活習慣状況の相対頻度を取り、その違いを見ることで、自覚症状と生活習慣の相関を推測する。さらに、受診者の年齢、性別など個人属性や他の生活習慣状況が各自覚症状の有無に与える効果を除いた上で、当該生活習慣の状況と各自覚症状の有無の相関を見るため、プロビット分析を行う。

C. 研究結果

生活習慣間の相関をクロス-tabュレーションによりみれば、喫煙の有無と飲酒の有無に強い相関があると見られ、喫煙の頻度が高ければ、飲酒の頻度も高くなると見られる。同様に、非喫煙者に比べ、喫煙者は、カフェインの摂取量が高く、さらに、喫煙者の中でも、喫煙の頻度が高い者は、カフェイン摂取の頻度が高まることも認められる。喫煙と飲酒ほど明確な相関ではないが、飲酒の頻度が特に高い者は、カフェイン摂取の頻度が飲酒をほとんどしない者に比べ高いことも認められる。

次に、生活習慣と自覚症状の有無との相関に関して、プロビット分析の結果によれば、4自覚症状を訴える可能性に喫煙の状況が有意に正の効果をもたらすことが示された。すなわち、非喫煙に比べ、喫煙、さらに、喫煙の

頻度が高いことは、自覚症状を訴える可能性を高めることが示された。さらに、カフェイン摂取の頻度が高ければ（コーヒやお茶の高頻度の飲用も）これら4つの自覚症状を訴える可能性を高める。他に、年齢が高いことや、睡眠が良く取れないことも4つの自覚症状を訴える可能性を高めることが示された。

D. 考察

これらの観測結果をつなぎ合わせれば、高頻度の喫煙、飲酒、カフェイン摂取がここで選んだ4つの自覚症状を訴える確率を高めると言える。しかし、本稿の分析では、検診結果をリンクさせて分析していないため、高頻度の喫煙、飲酒、カフェイン摂取が直接健康状態を悪化させるまたは疾病発生確率を高めることまでは言えない。プロビット分析の結果では、飲酒状況が自覚症状の有無に与える効果はさほど有意に示されない。喫煙状況と飲酒状況に相関が高いためとみられ、これらの相関を考慮して、飲酒状況の効果を識別できるようにプロビット分析の改善が必要と思われる。

E. 結論

高頻度に喫煙する者は、飲酒やカフェイン摂取も高頻度に行っている可能性が高い。これらのグループは、ここで分析した4つの自覚症状を訴える可能性が高い、つまり、自らの健康状態を良くないと判断する可能性が高い。

F. 研究発表

無し

G. 知的所有権の取得状況

無し

資料（河村真 法政大学経済学部）

「喫煙、飲酒など生活習慣間の相関とこの生活習慣の
自覚症状に及ぼす効果の検出」

「喫煙、飲酒など生活習慣間の相関とこの生活習慣の自覚症状に及ぼす効果の検出」

河村 真

法政大学経済学部教授

1 はじめに

われわれが用いる職場健康診断の個票データには、健康診断および人間ドックの結果に加え、喫煙、飲酒などの生活習慣に関する問診結果も利用可能である。これは、健康調査票と呼ばれ、職場健診を受ける前に、受診者に、生活習慣や受診時の身体状況（自覚症状など）を自己申告により記入させるものである。

われわれが用いた職場健診のデータは、平成8年度から平成11年度にわたる3社の職場健診の適用データ、健診結果データ、健康調査票データから構成される。これら3社のうち、会社コード062の会社は、製造業の会社であり、コード068の会社は、第二次産業の会社、コード991の会社は、第三次産業に分類される。会社コード062の健診結果票の標本数は、109940、068の会社では、54547、991の会社では、133396である。これに比して、健康調査票の標本数は、062の会社で、18447、068の会社で、33843、991の会社で、65721である。062の会社では、健診受診者の17%しか回答していないが、068の会社では、受診者の62%が回答している。受診者の健康調査票回答率には、会社間でかなりばらつきが認められる。

ここでは、特に、健康調査票から得られる生活習慣のうち、喫煙、飲酒、お茶・コーヒーの飲用状況にしばり、これらが受診者の健康状況の認識（自覚症状）に影響を与えているかを記述統計およびプロビット分析により明らかにする。ほかに、これら喫煙、飲酒、お茶・コーヒーの飲用状況の間での相関も記述統計により明らかにする。

2 生活習慣の項目に関する記述統計

これら3社の健康調査票で得られる質問項目には、食欲、睡眠、便通、生理、喫煙、飲酒、お茶・コーヒーの飲用の状況がある。われわれのデータで特定できる受診者の属性は、性別と年齢である。他に、健診の受診年度や調査票が春・秋の定期健診または人間ドックよりいずれにより得られたかをしめず健診区分を知ることが出来る。

食欲、睡眠の状況を、受診者に良、やや不良、不良の3選択択より選ばせている。表2-1に、性別、年齢階層別に良、やや不良、不良の標本数分布および相対頻度分布を掲げている。食欲に関しては、やや不良、不良と不調を訴えた標本は、4-7%であり、大部分の標本が食欲は良と回答しており、性別、年齢階層別でも、3社の標本で共通して顕著な傾向は見出せない。

次に、喫煙の状況を性別、年齢階層別に見てみる。表 2-2 に示す。会社 991 および会社 068 の健康調査票では、喫煙の状況を 4 選択肢（1 日 20 本以内で喫煙、1 日 20 本以上で喫煙、非喫煙、前喫煙）で回答させている。会社 062 では、喫煙および非喫煙の 2 選択肢で回答させている。喫煙と回答した標本数の相対頻度は、会社 991 で 36%、会社 068 で 48%、会社 062 では、77%となる。ただし、標本には、同一個人が、異なる年度に回答した結果を異なる標本として含めているため、喫煙率と対応しない。しかし、会社 062 の喫煙の相対頻度が著しく高い。これは、会社 062 の健康調査票の標本では、男性の標本数が 16524、女性が 1923 である。この標本の 9 割近くは男性の調査結果で占められている。さらに、会社 062 が製造業の会社であり、工場勤務の個人が含まれているなどのためと見られる。

性別で見れば、会社 991、会社 068 で、男性と女性で、喫煙、非喫煙の相対頻度に大きな違いが見られ、男性が喫煙している可能性が高いことを示す。女性で非喫煙と回答する可能性は 8 割強にも及ぶ。さらに、20 本以上の喫煙と回答した標本の相対頻度に最も顕著に違いが見られる。

年齢階層で見れば、1 日 20 本以内または 20 本以上喫煙と回答した標本の頻度は、35-44 歳の年齢階層で最も高いが、50 歳台、60 歳台になるとその頻度が若干下がる。

次に、飲酒の状況を性別、年齢階層別に見てみる。飲酒の状況は、3 社とも、ほとんど飲まない、時々飲む、ほとんど毎日飲むの 3 選択肢から受診者に回答させている。ほとんど飲まないと回答する標本の相対頻度は、会社 991 で 29%、会社 068 で 19%、会社 062 で 33%であり、毎日飲むと回答する標本のそれは、会社 991 で 23%、会社 068 で 37%、会社 062 で 38%である。会社 991 では、時々飲むと回答した標本が多く、次いで、ほとんど飲まない、毎日飲むという順である。会社 068 では、時々飲むと回答した標本が多く、ついで、毎日飲む、ほとんど飲まないの順である。会社 062 では、毎日飲む、ほとんど飲まない、時々飲むの順であるが、それぞれに回答した標本数にあまり大きな違いはない。

性別に見れば、喫煙同様性別の違いは、顕著である。会社 991、会社 068 では、ほとんど飲まないと回答した女性の標本は、女性の標本の 4 割程度であるが、同一の回答をした男性の標本の比率は 17%程度であり大きく異なる。一方、毎日飲むと回答した標本は、男性では 4 割程度であるが、女性では 1 割未満である。すなわち、ほとんど飲まないと回答する標本は、女性で相対的に多く、毎日飲むと回答した標本では、相対的に男性が多いと見られる

年齢階層別に見れば、特に、会社 068 で、毎日飲むと回答した標本の相対頻度が、年齢階層が高まるにつれ、高くなることが示される。しかし、会社 991、会社 062 では、年齢階層別に明確な傾向はさほど見られない。

他に、お茶やコーヒーの飲料状況を性別、年齢階層別に見てみる。コーヒーおよび濃いお茶を飲用する頻度を聞いているのは、カフェインの摂取量の程度を聞いていると見られる。回答の選択肢は、ほとんど飲まない、時々飲む、1 日 5 杯以上飲むの三つである。コーヒーおよび濃いお茶を時々飲むと回答する標本は、会社 991 および会社 068 で 3/4、会社 062 で 7 割弱である。会社 991 および会社 068 では、コーヒーまたは濃いお茶を 1 日 5 杯以上飲用すると回答

した標本は、全体の13-14%であるが、会社062では、22%と若干高い。

性別に見れば、喫煙の状況または飲酒の状況に比して、その違いは顕著でない。ほとんど飲まないと回答した標本が占める比率は、男性の標本に比べ、女性の標本で高い。1日5杯以上飲むと回答した標本の比率は、女性の標本に比べ、男性の標本でより高い。すなわち、女性に比べ、男性が、かなり高い頻度でカフェインを摂取する可能性が高いことが示唆されると見られる。

年齢階層別にみれば、1日5杯以上コーヒーまたは濃いお茶を飲むと回答する相対頻度が、会社991および068では45-54歳の年齢階層で高く、会社062では35-44歳で高い。そして、これら年齢階層を境に、年齢階層が高くなればこの数値は下がり、年齢階層が低くなるにつれ、この数値は、やはり下落することが示される。

3 生活習慣間での相関

次に、喫煙の状況または飲酒の状況別にそれ以外の生活習慣の状況の相対頻度分布が異なるかを見ることにより、喫煙状況または飲酒の状況とそれ以外の生活習慣の回答に相関があるかを見ることにする。

表3-1には、喫煙状況別に食欲、睡眠、便秘、コーヒー・お茶の飲用、飲酒の状況に関する標本数および相対頻度をそれぞれ掲げた。表3-2には、飲酒状況別に、食欲、睡眠、便秘、コーヒー・お茶の飲用の状況の標本数および相対頻度をそれぞれ掲げた。

食欲の状況に関しては、各会社の標本ともに、「やや不良」、「不良」といった異常を訴える相対頻度は、いずれの喫煙の状況においても、10%未満であり、顕著な違いが見受けられにくい。睡眠の状況に関しては、「やや不良」、「不良」といった異常を訴える相対頻度は、20-25%である。食欲の状況に比べれば、睡眠の状況の相対頻度は、喫煙の状況別に異なるが、3社の標本共通の傾向を見出せない。

そこで、飲酒の状況およびコーヒー・お茶の飲用の状況の相対頻度分布を喫煙の状況別に絞って見る。さらにその結果を図3-1、図3-2、図3-3に示す。図3-1、図3-2によれば、会社991、会社068の標本ともに、1日20本以上喫煙する標本では、ほとんど飲酒しない相対頻度が低く、ほとんど毎日飲むと回答した標本の頻度が高い。一方、非喫煙の標本に限れば、ほとんど飲酒しない標本の相対頻度が高く、毎日飲酒の標本の相対頻度が小さくなる。他に、現在喫煙しないが以前喫煙していた標本に関しては、毎日飲酒する標本の相対頻度が意外に高く、ほとんど飲酒しないと回答する標本の頻度が低い。これは、禁煙しても、飲酒の習慣まで変えることは難しいことを示唆して居るかもしれない。図3-1から図3-3の結果を要約すれば、非喫煙者に比べ、喫煙者は飲酒の頻度が高まる、または、アルコールの摂取量が高いと見られる。さらに、喫煙本数が多くなれば、さらに飲酒の頻度が高まる。他に、以前喫煙し禁煙しても、飲酒の頻度は下がらないことも示された。

喫煙の状況別にコーヒー・お茶の飲用状況に関して相対頻度分布をとり、図 3-4、図 3-5、図 3-6 に示す。飲酒の状況ほど顕著に違いが見られないが、喫煙者のほうが、コーヒーや濃いお茶の飲料の頻度が高くなり、喫煙本数の多いものほど、コーヒー・濃いお茶を飲用する頻度が高くなる。一方、非喫煙者では、ほとんどコーヒーや濃いお茶を飲用しない者も少なくないことが示された。

次に、飲酒の状況別に、他の生活習慣の状況に関して相対頻度をとり違いを比べてみる。その結果を表 3-2 に示す。食欲の状況、睡眠の状況、便通の状況に関しては、喫煙の状況別に見て時と同様に、飲酒の状況別に 3 社の標本で共通な傾向が明確には見受けられない。そこで、飲酒の状況別に、コーヒー・お茶の飲用状況に関して相対頻度分布をとり、その違いを図 3-7、図 3-8、図 3-9 を用いて比較する。3 社の標本共通して見られる傾向は、お酒をほとんど飲まないと回答した標本では、他に比べ、コーヒーや濃いお茶をほとんど飲まないと回答した標本の相対頻度が高く、時々、または毎日飲酒をする標本では、コーヒーや濃いお茶を飲まない相対頻度が低い。すなわち、ほとんどお酒を飲まない者は、少なくない確率でコーヒーや濃いお茶もほとんど飲まない。しかし、飲酒の頻度が高くなるにつれ、コーヒーや濃いお茶を頻繁に飲用するとは言えない。飲酒の頻度とコーヒー・お茶の飲用頻度の相関は、喫煙と飲酒ほど明らかではない。

4 自覚症状と生活習慣の相関

3 社の健康調査票には、最近の健康状況という質問項目があり、36 の自覚症状の有無についてたずねている。36 の自覚症状の一覧を表 4-0 に示す。自覚症状の項目の中には、自覚症状有りとは回答する標本数が著しく低い特殊なものも含まれている。そこで、自覚症状有りとは回答した標本の相対頻度が高い順に、4 つの自覚症状に絞って見ることにする。選んだ自覚症状は、「せき、たんが多い」、「胃が不快、痛む、もたれる、はる」、「背中や腰が痛い」、「だるい、疲れやすい、酒に弱くなった」の 4 症状である。

表 4-1 には、自覚症状「せき、たんが多い」の有りと回答した標本の相対頻度を年齢、性別、喫煙状況、飲酒状況、コーヒー・お茶の飲用状況別で比較している。性別で見れば、男性の方が、「有り」の相対頻度が 3 社の標本ともに高い。年齢階層で見れば、年齢階層が高まれば、この自覚症状有りとする相対頻度が高まる。喫煙の状況では、1 日 20 本以上の喫煙する標本で、この自覚症状の頻度が高い。飲酒およびお茶・コーヒー飲用の頻度が高まれば、この症状有りとする頻度が高まることが示される。

表 4-2 には、自覚症状「胃が不快、痛む、もたれる、はる」有りの相対頻度を年齢、性別、喫煙状況、飲酒状況、コーヒー・お茶の飲用状況別で比較している。性別では、女性で自覚症状有りの相対頻度が若干高いが、その他では、表 4-1 と同様な傾向が見られる。

表 4-3 では、自覚症状「背中や腰が痛い」有りの相対頻度を年齢、性別、喫煙状況、飲酒状