

# ナショナルデータベースにおける 特定健診・レセプトデータの突合精度

国立保健医療科学院 上席主任研究官  
岡本 悦司

## 【ポスター -1, 2】

報道もされているように、メタボの人はそうでない人に比べて1人当たり医療費が9万円も違う。この数値は単なる報道ではなく、大臣告示にも入っており、今年度から始った2期計画の医療費見通しにも使われています。これが本当だとすれば、保険料もそれだけ差を設けなくてはならないことになるでしょう。

このようなメタボ対策の評価のために国はレセプトや健診のデータを匿名化してデータベースを作っているわけですが、評価のためにはレセプトと特定健診データを「これとこれが同じ人間のものだ」とくっつけることが大前提になります。しかし、このくっつけることが100%完璧ではない。レセプトがない人はいいのですが、レセプトがあっても暗号（ハッシュ関数という暗号を使っています）が違ったらくっつかないので、突合率は低いということが示されています。

では、どの程度突合されているのか。さらにそれが性・年齢階級別に違いがあるのかないのか、を分析してみました。

使ったデータは4つの公表されているデータを組み合わせました。

細かい数値を言うより、かいつまんで要点だけ説明します。

## 【ポスター -3】

2009年度に特定健診は2,159万人が受けています。荒っぽく言うと、全国民の6人に1人が健診を受けたわけです。

## ポスター 1

**背景**

2009年度の特定健診結果と、10年度の医療機関の治療実績を示したレセプト(診療報酬明細書)を突き合わせることで約269万人について、性別、年齢階級別に分析した。(2013年4月7日 03時06分 読売新聞)

**メタボの人 医療費9万円高**

## ポスター 2

**レセプト突合ができなかった者**

第7回保険者による健診・保健指導等に関する検討会(2012年2月24日)資料1

- ①平成22年4月～23年3月にレセプト情報がない者(医療機関を受診していない者)
- ②レセプト情報はありますが特定健康診査情報ファイルとの突合ができなかった者(加入保険が変更になったこと等)
  - ハッシュ値による突き合わせの精度には限界があり、21年度に特定健診データが存在し、かつ22年度のレセプト情報と突き合わせができる者は21年度受診者の男9.8%、女15.7%であり、年齢階級別にその割合は区々である。
  - 同一人であっても氏名の表記が異なったり、変更されたりした場合においては突き合わせができなくなる。
  - 21年度の特定健診データを22年度のレセプトデータと突き合わせる理由は、21年度においては特に診療所レセプト電子化率が30→60%まで1年間で急激に伸びている時期であることを考慮したもの

レセプトと突合してみますと、医療費は7,554億円でした。国民医療費が40兆円くらいあって、6人に1人が受けて7,554億円。「少ないな」という感じをもたれると思います。「少ないな」というのでは研究になりませんから、もしこれが100%突合できていれば、医療費はいくらだったはずだ、ということを出して比較すれば、医療費面で見るとレセプトの何%が把握されているかを出せるということです。

### ポスター 3

#### 方法

- 約2159万人の2009年度特定健診受診者の医療費(医科+調剤)の推計額(A)
  - 性・年齢階級別の一人当たり医療費(医療給付実態調査)
  - 医療費は健診非受診者>受診者(たとえば入院患者は健診を受けられない)を加味する(協会けんぽデータ)
- 特定健診受診者の実際に突合された医療費総額(B)
  - 電子化率を加味する
- B/Aが突合率(理想的には100%かそれに近いはずだが・・・)

結論から言いますと、2,159万人の医科と調剤レセプトが突合できれば5兆円あったはずである。それで7,554億円ということは、医療費でいうと14.9%しか把握できなかった、という結論です。

5兆円はどうやって算出したかということ、6人に1人といっても単純に40兆円を6分の1にするわけにはいきません。医療費は年齢・性別によって違うことがわかりただけだと思いますが、より重要な点は、健診を受けた人と受けなかった人とでは医療費は違います。たとえば入院中や治療中の人は健診を受けないでしょうから、単純に健診受診者と非受診者を比較すると受診者の方が一人当たり医療費は低いのです。ではどれだけ低いのか？ このデータはこれまでなかなか得られなかったのですが、幸い全国健康保険協会が独自に調べて公表していました。それによると、男の場合は健診受けている人は受けていない人に比べて3分の2くらいということでした。

また医療費には当然男女差がありますが、社会医療調査も医療給付実態調査も以前は性別データを出していませんでした。そこヘラッキーなことに去年から出すようになりましたので、男女別比較ができるようになりました。

全体の突合率が14.9%だったのですが、さらに今回の分析で明らかになったことは、性・年齢階級別に見てみると、(突合率でいうと本来なら性・年齢とは関係がないはずだから全部14.9%のはずなのですが)明かな差が見られました。第一に、男の突合率は女より悪い。さらに、年齢で見ると高齢者の突合率は良いのだけれど若年者は悪い。特にメタボ対策はどちらかというとな女より男、特に現役の男を対象にしています。一番評価で重要になってくる40歳～64歳の現役の男に関していうと、10%にも行っていない。一桁台です。

では、なぜこういう傾向があるのかということなのですが、私の研究では突合率を見ているだけなので、なぜこんなに差があるのかまではわかりません。色々考えると、若い男が多い健保組合の匿名化が上手くいっていないとか、データにバラツキあるとかが考えられるのですが、それは憶測の域を出ません。

結論から言って、9万円差があったということは、現役の年齢で言うとレセプトの1割以

下をくっつけて出した結論であるということです。

これは私の希望なのですが、まだ4、5年以内の新しいデータだから保険者には残っているはずなので、ちゃんと突合をやり直す。日本医療データセンターなどの民間のデータベースを見ると90%以上の突合率を持っているので、不可能ではないだろう。

同時に、なぜかという原因をはっきりさせる。一つは名前の入力ミスもあるでしょう。安倍首相が福島に行った時に東電が作った名前入りの防護服で、その名前の「安倍」が間違っていました、そういったことも考えられます。

とにかく早く原因をはっきりさせて、可能ならできるだけデータを再収集して、きちんとした評価をすべきではないかということで、結論とします。

ポスター 4

### 1) 受診率を算出(人口を分母)

**[表1] 2009年度特定健康診査受診率**

年齢	男			女				
	2009年10月		受診率 R (受診者数 / 人口)	2009年10月		受診率 R (受診者数 / 人口)		
	現在人口 (千人)	対象者数 (千人)		現在人口 (千人)	対象者数 (千人)			
40~44	4,323	4056351	2208376	51.1%	4,258	3851465	1380709	32.4%
45~49	3,932	3685567	2054324	52.2%	3,894	3552116	1315061	33.8%
50~54	3,863	3542461	1903919	49.3%	3,877	3473490	1293219	33.4%
55~59	4,517	4011840	1975968	43.7%	4,616	3989531	1419481	30.8%
60~64	4,603	4078432	1565725	34.0%	4,810	4373575	1488006	30.9%
65~69	4,005	3484940	1214405	30.3%	4,380	3922897	1481643	33.8%
70~74	3,199	2840267	1019997	31.9%	3,712	3346803	1270050	34.2%
total	28,442	25699858	11942714	42.0%	29,547	26511877	9646169	32.6%

出典: 2009年度特定健康診査・特定保健指導の実施状況(確報値)  
source: [http://www.mhlw.go.jp/bunya/shakaihoshu/iryusei/01/d/inf03\\_h21\\_03.pdf](http://www.mhlw.go.jp/bunya/shakaihoshu/iryusei/01/d/inf03_h21_03.pdf)

ポスター 5

### 2) 健診受診者と非受診者の医療費を比較

**[表2] 特定健康診査受診者・非受診者間の一人当たり年間医療調剤医療費(円)**

年齢	男			女		
	受診者	非受診者	比 (非受診者/ 受診者)	受診者	非受診者	比 (非受診者/ 受診者)
40~44	68460	83017	1.21	80391	93937	1.17
45~49	89120	112220	1.26	92159	107617	1.17
50~54	117000	148787	1.27	109917	128035	1.16
55~59	149394	197421	1.32	128347	154232	1.20
60~64	191084	257593	1.35	158692	195420	1.23
65~69	235556	339828	1.44	199147	258205	1.30
70~74	332376	512231	1.54	290377	403938	1.39
total	129273	188335	1.46	114226	147770	1.29

出典: 全国健康保険協会平成20年度の健診データと医療費データの分析  
source: <http://www.kyokaikenpo.or.jp/media/Files/honbu/cat740/2506/250611003.xls>

ポスター 6

### 3) 全被保険者の一人当たり医療費を推計

**[表3] 全医療保険加入者の性・年齢階級別一人当たり医療調剤年間医療費(2011年度)**

AGE	男			女		
	被保険者数(N)	医療調剤医療費 総額(円)(C)	一人当たり 医療費 P(C/N)	被保険者数(N)	医療調剤医療費 総額(円)(C)	一人当たり 医療費 P(C/N)
40~44	4006878	453169049750	113098	3926809	473955768170	120697
45~49	3321980	503331494490	151516	3287134	491671766320	149575
50~54	3138226	624634900530	199041	3139869	584106152970	186029
55~59	3465152	908033409780	262047	3518876	811274890500	230549
60~64	4780875	1650648766370	345261	4960455	1411115393690	284473
65~69	3459967	1643229952330	474927	3777606	1421169561630	376209
70~74	3019459	1953067825520	646827	3483921	1825207400040	523895
total	25192537	7736115198770	307080	26094670	7018500933320	268963

出典: 2011年度医療給付実態調査  
<http://www.e-stat.go.jp/SGL/eStat/GL02020101.do?metho=download&fileId=000006435815&releaseCount=1>

ポスター 7

### 4) 突合された医療費を算出

**[表4] 特定健康診査受診者と突合された医療調剤レセト医療費**

年齢	男			女		
	レセトと突合された特定健康診査受診者数(A)	一人当たり医療調剤医療費(円)(B)	突合された医療調剤医療費(円)(A*B)	レセトと突合された特定健康診査受診者数(A)	一人当たり医療調剤医療費(円)(B)	突合された医療調剤医療費(円)(A*B)
40~44	96704	14419	13943766739	89837	13499	12126810616
45~49	97584	17113	16699638200	94102	15335	14430862278
50~54	119628	20626	24674486348	120840	16688	20165611842
55~59	155984	24559	38308534776	163497	18647	30487271483
60~64	107668	25308	27248692485	241881	21153	51165276230
65~69	299373	34212	102421567712	424601	28157	119539750511
70~74	295569	40104	118534614534	378241	33409	126367197103
total	1172510	29154	341831300794	1512999	24739	374296780063

出典: 第7回保険者による健診・保健指導等に関する検討会(平成24年2月24日)  
source: <http://www.mhlw.go.jp/stf/shingi/2/98520000023mfr-att/2/98520000023mhb.pdf>

### 5) 突合できた医療費/推計額を算出

表51 特定健診受診者の医科調剤医療費の推計額とレセプトと突合できた医療費額の比較

年齢	特定健診 受診者数 N	受診率 (受診者 数/人口) R	比 (非受診者 /受診者) r	全医療保 険加入者 一人当り 医療費 P (R+r-Rr)		特定健診受診者 一人当り医療 費推計額(円) P (R+r-Rr)		特定健診受診者の レセプトと突合でき た医療費 N*P/(R+r-Rr) 94.8%で補正)		突合で きた医 療費/ 推計額
				男	女	男	女	男	女	
40~44	2208376	51.1%	1.21	113098	102443	226,231,656,926	14,708,614,704	6.5%		
45~49	2054324	52.2%	1.26	151516	134827	276,978,115,892	17,615,652,110	6.4%		
50~54	1903919	49.3%	1.27	189041	174938	333,067,001,956	26,027,939,186	7.8%		
55~59	1975968	43.7%	1.32	282047	221915	438,496,975,016	40,409,846,810	9.2%		
60~64	1565725	34.0%	1.35	345261	280776	439,617,687,810	28,743,346,503	6.5%		
65~69	1214405	30.3%	1.44	474927	362972	440,795,159,536	108,039,628,388	24.5%		
70~74	1019997	31.9%	1.54	646827	472625	482,076,124,377	125,036,513,222	25.9%		
計	11942714	42.0%	1.46	307080	242744	2,899,018,186,819	360,581,540,922	12.4%		
40~44	1390709	32.4%	1.17	120697	106359	149,612,721,709	12,791,994,321	8.6%		
45~49	1315061	33.8%	1.17	149575	134820	177,033,700,183	15,222,428,563	8.6%		
50~54	1293219	33.4%	1.16	186029	167616	216,764,167,320	21,271,742,449	9.8%		
55~59	1419481	30.8%	1.20	230549	202297	287,156,175,368	32,159,569,075	11.2%		
60~64	1486006	30.9%	1.23	284473	245248	364,439,955,855	52,971,810,269	14.8%		
65~69	1481643	33.8%	1.30	376209	314494	465,967,106,619	126,111,551,172	27.1%		
70~74	1270050	34.2%	1.39	523895	416691	529,218,233,237	133,298,731,121	25.2%		
計	9646169	32.6%	1.29	268963	224549	2,166,033,679,871	394,827,827,071	18.2%		
男女 計	21588883	37.2%	1.38	287686	232281	5,065,051,866,690	755,409,367,993	14.9%		

- ### まとめ・結論
- NDBはレセプト14.9%しか突合できていない
  - 突合率は男<女, 若年<高齢(特定保健指導の主対象である65歳未満では**10%未満**と何らかのシステムティックなバイアスを示唆する(ランダムではない)。
  - 考えられる原因・・・入力ミス(例:菊池と菊地)全角と半角, 生年月日はレセプトでは和暦, 特定健診では西暦, 氏名はレセプトでは姓名間にスペースをいれる。カタカナも。(国保データベースでは修正済)
  - すみやかにハッシュ関数暗号化プログラムを修正し, 過去データを再収集できないか。

## 質疑応答

**会場:** 素朴な質問で申し訳ないのですが、これは暗号化の設計のミスではないか。一言で言うと、取り出したい情報を戻せるようにするための暗号化ですから、それを戻せなかったという設計エラーではないかという疑問があるのですが、いかがでしょうか。

**岡本:** その通りです。私が設計したわけではないのですが、亡くなられた開原先生(座長)以下、私も委員として加わった研究利用のための検討会で、ハッシュ関数の議論もありました。私も素人なので「ハッシュ関数はこんなものか」、「こうすればくつつくのか」という程度で聞き、検討会ではむしろ「プライバシーがどうこう」という議論が主体で、「これで100%突合できるのかチェックをしろ」という議論にはなりませんでした。

現に企業でハッシュ関数できちんと100%近い突合率でやっている(そのかわり手間をかけています)ところもあるわけです。もちろんそれは簡単なことではなく、最初の段階から100%は無理とわかっていまして、80~90%いけば何とか思っていたのですが、今回の結果を見て愕然としました。原因は何か。はっきり言って、ソフトにバグがあったのです。

まさに鋭いご指摘です。申し訳ございません。私が作ったのではないですが…(笑)。

ここで誰の責任かというのではなく、早くバグを修正しなくてはならない。データがなくならないうちに、早く。20年、30年前ではなく4、5年以内のデータなので、「悪いけれど、もう一度出し直しをして下さい」と言えばすむのではないか、私は考えております。

**会場：** 医療費と健診データの突合が悪いということで、それに関するものは良くないと思うのですが、ではレセプト単体の分析ということになるとNDB（ナショナルデータベース）は大丈夫なのかということが1つ。2点目は、20件近く厚生労働省はこのデータベースからデータを出していて、査読付きの論文誌に発表されているものもある。これに対する影響というのは先生はどうお考えになるのか、教えていただきたいと思います。

**岡本：** まず、今回評価したのは健診とレセプトの突合です。正直な話、レセプトの生年月日は昭和や平成という和暦で、健診は西暦なのです。だからくっつかない。少なくともソフトを作る人はそれくらいの判断はつくだろうと思っていましたが、どうも考慮せずに作成したようです。

新しい国保のデータベース(KDB)はちゃんと処理しています。(先ほどの「安倍」のように字が異なるとどうしようもないという問題は残りますが)

ですから、レセプトと健診の突合率は悪いが、少なくともレセプトとレセプト、健診と健診はもう少しいいのではないかと考えられるのですが、それは評価していません。

国際誌査読についてですが、突合率は「無条件で100%」とは絶対に考えられないので、学術誌に投稿するときはどれだけの突合率かということはレビューワーとしても明確にすべきと思います。

#### 座長

今回のこのセッションは、今日的なテーマや、直接住民の方にどう還元していけるかというヘルスリサーチが狙っている研究の成果を、的確な内容で皆さんに短時間で発表していただいて、有効な時間を過ごせたと思っています。

これを基に皆さん、どんどん研究費助成にも応募していただきたいし、研究を発展させていただけたらと思います。有り難うございました。