

3から9までの質問は、質問2で「1委託している」と回答いただいた方のみご回答ください。

4) レセプト情報はどのような媒体で委託先に提供しましたか。

1. レセプトの原本(紙)
2. レセプトのコピー(紙)
3. レセプトの原本から必要な項目のみ転記
4. 電子媒体
5. その他 ()

5) 提供情報に次の項目を含めましたか。

- i. 氏名..... (1. 含めた 2. 含めなかった 3. 分からない)
- ii. 記番号..... (1. 含めた 2. 含めなかった 3. 分からない)
- iii. 生年月・年齢.... (1. 含めた 2. 含めなかった 3. 分からない)
- iv. 性..... (1. 含めた 2. 含めなかった 3. 分からない)
- v. 保険医療機関番号 (1. 含めた 2. 含めなかった 3. 分からない)

6) レセプト情報の分析を外部に委託したことをどのように被保険者に伝えましたか。該当するものすべてに○をつけてください。

1. 調査分析前に広報等に掲載した。
2. 調査分析後に広報等に掲載した。
3. 被保険者には伝えていない。
4. その他 ()

7) レセプト情報の分析結果をどのように被保険者に伝えましたか。該当するものすべてに○をつけてください。

1. 広報等に掲載した。
2. 報告書を希望者に配布した。
3. 特に伝えていない。
4. その他 ()

8) 実際に、保険者と委託先との間で委託契約を文書で取り交わしましたか。

1. 行った
2. 行わない

9) 前問で「1. 行った」と回答された方にお伺いします。委託契約の中に、守秘義務についての項目が含まれていましたか。

1. 含まれていた
2. 含まれていない
3. その他 ()

これより先の質問は全保険者にご回答をお願いします。

II 今後のレセプト情報を利用した研究について

10) 今後レセプトを用いてどのような調査研究が必要と考えますか。それぞれの項目について

該当する番号に○をつけてください。

- | | | | |
|----------------------------|-------|--------|----------|
| i. 特定の疾患に関する医療費..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| ii. 特定の疾患の発生頻度..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| iii. 特定の診療行為の実施頻度..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| iv. 薬剤の処方頻度..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| v. 医療費と健康診査や健康教育の関連..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| vi. 医療費と生活習慣の関連..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| vii. 特定の治療法に関する費用効果分析..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| viii. 受療行動の実態把握..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| ix. 医療費の地域格差の要因..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| x. 医療機関別の分析..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| xi. レセプトコンピューターの開発..... | 1. 必要 | 2. 不必要 | 3. 分からない |
| xii | そ | の | 他 |
| (| | |) |

III 貴保険者の概要について記入してください。

11) 加入者数(家族含む)について該当する番号に○をつけてください。

1. 1000人未満

2. 1万人未満
3. 10万人未満
4. 30万人未満
5. 30万人以上

12) 老人保健制度の加入率について該当する番号に○をつけてください。

1. 30%以上
2. 20%以上 30%未満
3. 10%以上 20%未満
4. 5%以上 10%未満
5. 5%未満

13) 保険の種類について該当する番号に○をつけてください。

1. 市町村国保
2. 国保組合
3. 単一健保組合
4. 総合健保組合

14) 所在地について該当する番号に○をつけてください。

1. 北海道 ・ 東北
2. 関東
3. 東海 ・ 甲信越 ・ 北陸
4. 近畿
5. 中国 ・ 四国
6. 九州 ・ 沖縄

15) レセプト情報の利用のあり方や本調査に関してご意見があればご記入ください。

本調査の結果を必要とする方は、下記にお名前と資料郵送先をご記入ください。本調査の結果については <http://resept.com/>にて 2003 年 5 月頃に掲載する予定です。

郵便番号：

ご住所：

お名前：

E-mail：

最後までご回答ありがとうございました。ご協力に心よりお礼申し上げます。

分担研究報告書

レセプト傷病分析の原理と手法--PDM法について

分担研究者 岡本悦司(国立保健医療科学院技術評価部)
研究協力者 畑 栄一 (同研修企画部)

研究要旨

レセプト電算化によりレセプト情報の利活用は飛躍的に進展すると期待できるが、最大の障害は傷病分析である。従来の人々の作業に依存した主傷病分類法では、客観的な分類が不可能であるだけでなく、個人情報保護の面でも好ましくない。そこでレセプトに記載された複数傷病名を客観的かつ自動的に分析する原理(PDM法)を考案し、パソコン上で使用できるプログラムを開発した。本稿では、PDM法の原理と概念、ならびにその精度を向上させるための補正法やシミュレーションデータを使った検証法を解説し、その実際の応用例として宮城県名取市におけるインフルエンザワクチン効果評価事業について紹介する。

個人情報保護しつつ、レセプト情報の有効な利活用をはかるためには、含まれる情報が人手を介することなく、全て機械的に処理されることが望ましい。紙レセプト主体の現在では、レセプトのパンチ入力に外部委託され、とりわけ傷病分類は人が手と目で判断することがふつうである。レセプトの電子化がすすめば、パンチ入力の委託は必要なくなるが、傷病分類が人手に依存するものであることに変わりはない。

また従来の人々の判断で複数傷病の中から任意に主傷病を選択するいわゆる主傷病法では、レセプト情報を定量的かつ客観的に処理することは困難である。

こうした問題を解決し、個人情報保護をはかりつつ電子化レセプトを有効活用する目的で、複数傷病の記載されたレセプトの日数、点数といった情報を客観的かつ自動的に分析する原理を考案し、それをWindowsパソコン上で使用できるプログラムを開発した。

【推計法の必要性】

複数傷病の記載されたレセプトにおいて、傷病ごとの日数、点数が求められるとすれば理想である。しかしそれは不可能であり、唯一可能なのは真に近い値を「推計値」するしかない。したがって真の値はわからないが限りなく近い「推計法」を開発することしか方法はない。

その根拠は

方程式の解が求まるのは変数 \leq 方程式でなければならず、変数 $>$ 方程式では求まらない。という数学の公理にある。レセプトには常に1以上の傷病名が記載されている。つまり傷病名数 $>$ レセプト件数となる。傷病名数=レセプト件数(つまり全レセプトに傷病名が一つしかない)ならそもそもPDM法など不必要だから、複数傷病名レセプトを分析するには何らかの方法で推計するしかない。

[説明]

3件のレセプトを1,2,3で、傷病名A,B,Cの一日当点数をa,b,cで、日数をN,点数をPであらわし、レセプト1にはA,B、レセプト2にはA,B,C、レセプト3にはA,Cの傷病名が記載されている。す

ると3つの等式ができる。

$$P1=N1*(a1+b1)$$

$$P2=N2*(a2+b2+c2)$$

$$P3=N3*(a3 +c3)$$

これら3つの式で a1,a2,a3,b1,b2,c2,c3 の7つの数値が求められるか?

7つの変数を求めるには式は7つ以上いる。しかるにレセプト件数(式の数)は常に変数の数より少ない。このため7つの数値を求めることは不可である。

もっとも各レセプトの各傷病ごとの点数や日数が同一である場合は事情が異なる。たとえば3つの傷病が同じ点数、日数で a,b,c と3つの変数なら

$$P1=N1*(a+b)$$

$$P2=N2*(a+b+c)$$

$$P3=N3*(a +c)$$

となり、必ず解が求まる(傷病名数が119もあっても一般化逆行列を用いれば瞬時に求まった)。

【重回帰分析】

多変量解析の観点からレセプトをみると、点数、日数は目的変数、多数の傷病名は説明変数となる。重回帰分析を用いない理由について説明する。

日数や点数を目的変数に、119の傷病分類ごとの傷病名の出現数(無ければゼロ、1,2,3---)を説明変数に重回帰分析を行うことは可能である。そして各傷病分類の重回帰係数は日数や点数に対する各傷病の寄与度を意味しており「重み」として使うことは考えられる。しかしこの試みは頓挫し断念した。その理由は以下のとおり。

●重回帰係数は必ず負値がでる

重回帰分析は可能だが、119傷病の各係数の相当数が負値になる。点数を比例配分するPDM法では負値はつかえず、このことを数学上では「モデルが破綻している」と呼ぶという。おびただしい負値の理由としては多重共線性の問題や線型性への疑問が考えられる。ただ、医療機関の種類や患者属性等を厳格に補正すれば全て正の係数になる可能性は残されている。今後厳格な条件をつけて重回帰分析を試みる余地は残されているかもしれない。

●重回帰係数は件数が多くなければ信頼できない

重回帰分析は、説明変数よりもケース数が相当大きくなければならない。100件のレセプトを119の説明変数で分析することは困難である。むしろ統計ソフトにデータをつっこむと答えはでるがp値が0.8にもなったりして「こんな数値が出ましたが当たるも当たらぬも八卦」となる。事実、あるデータセットでは極端に大きな係数がでた傷病が別のデータセットではマイナスになったりと、とても安定した分析に耐えなかった。

●原点を通る回帰では正しい推計ができない

通常重回帰分析では

$$Y=B0+B1X1+B2X2+---+BnXn (B0は定数項)$$

という回帰式を考え、119傷病について得られた係数 B1~B119 に各傷病の出現数をかけ、定数 B0 を足すと合計点数に一致する。つまり傷病名が全く無いレセプトでも B0 点はあると考える。

しかしレセプトでは傷病名ゼロなら点数もゼロと考えるので(実際には傷病名の無いレセプトは

たまにあり、そんなレセプトでも点数はちゃんとある・・・もつともこれらは記入もれとおもわれる)

$$Y=B_1X_1+B_2X_2+\dots+B_nX_n$$

と定数項(intercept)の無い回帰を使わなければならない(SPSS でも Excel でもデフォルトは上記のような「定数項有りの回帰」だが「原点を通る回帰」をクリックすれば原点を通る回帰を行える)。

定数項の無い「原点を通る回帰」では、負値の係数は少なくなる(しかし決してなくなる)。ところが定数項有りの通常回帰式のように、各傷病の係数に出現数をかけて合計しても合計点数 Y に一致せず、通常 20~30% 小さい数字にしかならない。

この理由がわからず SPSS に問合わせしたところ、SPSS 国際サポート本部に照会していただき、以下のような回答をいただいたので参考まで転載する。

SPSS 国際サポート本部からの回答

n 個の点 $\{(x_i, y_i)\}$, $i=1$ to n , に直線をあてはめた場合、その直線と点との残差の合計は 0 になります。これは Draper と Smith の「Applied Regression Analysis 第 3 版」の 26 ページで証明されています。予測値と残差を保存した場合、従属変数、予測値、残差 は定義上次の式に従います。

$$\text{従属変数} = \text{予測値} + \text{残差}$$

これら 3 つの変数をそれぞれ合計した場合、残差の合計が 0 になるという事実から、従属変数の値の合計が予測値の合計と一致するという結論に至ります。

しかしながら、定数項を回帰から除外した場合、これはあてはまりません。定数項を除外することは、あてはめの線が原点を通るようにすることです。仮に次の 3 つのデータポイントがあるとします。

x	y
80	100
90	110
340	130

これに最小二乗法による線をあてはめると、傾きは 0.1 切片が 96 あたりになります。この線を原点を通るようにするには、x の平均あたりで回転させ傾斜を急にするので、最初の 2 点では予測値を下げ、3 番目では予測値を上げるという事になります。モデルは最初の二つの x に対する y を過小評価し、最後の x に対する y を過大評価しています。これにより、残差の合計は正の値となり、予測値の合計は 観測された従属変数の値の合計よりも少なくなります。

It is true that if you fit a straight line to n points $\{(x_i, y_i)\}$, $i = 1$ to n , then the sum of the residuals about the fitted line is 0. This is proven in Draper and Smith, Applied Regression Analysis, 3rd edition, page 26. If we save the predicted and residual values, the dependent, predicted, and residual variables by definition obey the formula

$$\text{dependent} = \text{predicted} + \text{residual}$$

If we sum the three variables, the fact that the residuals sum to 0 leads to the conclusion that the sum of the dependent values equals the sum of the predicted values.

THIS IS NOT TRUE, HOWEVER, IF ONE LEAVES OUT THE CONSTANT TERM FROM THE REGRESSION.

Leaving out the constant term forces the fitted line to pass through the origin. Consider three data points

x	y
80	100
90	110
340	130

If you fit the least squares line to these points, the slope is about 0.1 and the intercept

around 96. Forcing the line to pass through the origin steepens it by rotating it about the mean of the x's, having the effect of decreasing the predicted values at the first two points and increasing it at the third. The model underestimates y for the first two x's and overestimates it for the last, with the net effect that the sum of the residuals is positive and the sum of the predicted values is less than the sum of the observed values.

というわけで、目的変数(従属変数)の値の合計は定数項ありの場合 予測値の合計と一致しますが、なしの場合は原点を通るように強制するのでずれが生じ、残差の合計が 0 にならないために一致はしない、ということになります。

●線型計画法

変数にたとえば5以上という条件をつけて行う線型計画法という手法も検討したので付言する。専門家の意見では、偏回帰係数の解釈まで要求されるような回帰方程式で、100変数で実行して、解釈可能な結果は出ない、ほぼ必ず多重共線性や抑制効果が生じる、一説によると説明変数の数は2つか3つまでしか無理だという。単に予測値と実績値が近ければよいのであれば、説明変数は100でも200でも構わない。でも線型計画なら符号条件が付けられるから納得できる結果が得られるのではないかと期待しても結局のところ、分析者が課した範囲条件の境界に解が定まる。つまり、非負と制約すれば0となり、10以上という制約をおけば10になる・・・結局のところ分析者の指定をそのまま出力しているのと同じ結果になる。ということでレセプト傷病分析への線型計画法の適用は見送った。

●重回帰分析とは目的が異なる

重回帰分析も林の数量化I類も全て個々のケースの目的変数を推定するための手法です。たとえば傷病A,B,Cがある15日の入院レセプトの予想点数は？といった個々のレセプトの点数の予測のための手段が重回帰分析であり数量化理論といえる。しかしここでの目的は、個々のレセプトの点数を予測するための手段ではなく、レセプトの集合(データセット)における傷病別割合を推計することであり、傷病名A,B,Cのレセプトの点数は何点かを予測するための手法ではない。したがって重回帰分析とは目的が異なる。

【PDM法の原理】

PDM(Proportional Disease Magnitude)法はレセプトの点数や日数を、それに記載された全ての傷病名に一定の「重み(magnitude)」を与えて比例配分してゆく分析法である。PDM法によれば、客観的かつ自動的な傷病分析が可能となり、もし共通の「重み」を用いれば、たとえばA市とB町国保ではどちらが糖尿病の医療費がかかっているか、またC村で糖尿病対策をする前と後とで糖尿病の医療費が増えたか減ったか、等を客観的に比較することも可能となる。裏返せば、では傷病ごとの「重み」をどう決めるか？が課題となる。

重みは基本的にどのように決めてもよく、全ての重みを1(つまり全ての傷病を等しいとする)としてもかまわない。そうなると単純に傷病名の出現数に比例した配分となる。重みの決め方としては、たとえばDRGも一種の重みといえるし、外国では専門医にアンケートして重みを測定しようという試みもある。PDM法では、これまで患者調査で得られた「主傷病となる確率」や社会医療診療行為別調査で得られた傷病ごとの「一件当たり点数・日数」を用いたこともある。

【PDM法開発の経緯】

「全ての傷病名を分析し、点数や日数といった医療の資源消費を定量的かつ自動的に傷病分類できないか」そう考えた筆者は94年頃「診療報酬明細書による傷病構造の解析」研究に着手(94年度文部省科学研究費奨励研究)。95年6月米国、シカゴで開催された第12回 Association of Health Service Researchにおいてその基本概念を発表し、同年10月山形で

開催された日本公衆衛生学会で開催された「第 1 回レセプト情報の活用を考える自由集会」において BASIC で組んだ初歩的なプログラムを公表した。PDM(Proportional Disease Magnitude)法と名付けられた本手法の概念は、厚生統計分野の代表的な学術誌である「厚生指標」96年6月号に掲載された。

PDM 法の最大のネックは、記載された多数の傷病名をコード化し入力しなければならない「手作業」にあった。レセプト電算化も期待されたほどには進展せず、実用化へは遠い道のりの状態が続いた。1999 年には健康保険組合連合会による「レセプト分析による医療機関評価」研究が行なわれ、そこで PDM 法により傷病構造を補正して医療費の額や在院日数の長短を異なった医療機関間で比較することが試みられた。この時作成された ExcelVBA プログラムはインターネット(resept.com)上で公表され、関係者の関心を集めた。しかし、このプログラムも誰でも気軽に使用できるほどユーザーフレンドリーではなかった。

2001 年、岡本は旧国立公衆衛生院に移り、愛媛県、宮城県の国民健康保険団体連合会が実施している全疾病入力レセプトの解析に着手。また本研究に分担研究者として参画、その助成によりそれまでコンピューターに強い人でなければ使用できなかったプログラムが Windows 上で誰でも簡単に使用できるプログラムとして完成をみた。

2002 年 10 月 24 日埼玉県で開催された「第 5 回レセプト情報の活用を考える自由集会」で本プログラム(Ver.1)は公表され、その後の研究の進展をふまえて、本年度中には Ver.2 が完成する見通しである。

【PDM 法の個人情報保護上のメリット】

PDM 法は傷病分類を自動化することにより、手作業の労力と時間を節約するのみならず、個人情報保護の面でも効果がある。データ入力作業はレセプトが完全電算化されれば不必要になるが、傷病分類だけは自動化は不可能で、これまで人間がレセプトを見て判断するしかなかった。熟練したレセプト点検職員は1日に1000件近いレセプトを分類できるが、それでももし年間12億件のレセプト全件を分類するとなると、のべ120万日・人もの人にレセプトを「見せる」必要がある。

どんなに契約書で守秘義務を課してもプライバシー漏洩の危険は、関わる人が多くなると必然的に大きくなる。プライバシー保護を100%確保するには、医療機関から提出された電子化レセプトが、誰からも中身をみられることなく、コンピューター内部で全て処理される、のが理想である。PDM 法により、単純集計はもちろん、傷病分類まで、人手を介することなくブラックボックスの中で処理され、結果の表だけが出力されるようになり、プライバシー保護は万全になる。

【日数と点数の扱い】

PDM 法ではレセプト総点数を日数と一日当点数に分離して分析している。その理由は日数は1~30(大の月は31)のどの数値でもとりうる自然現象(確率事象)であるが、点数は自然現象ではなく、人為的に決められた数値であることにある。

診察料を例にとると病院の初診料は250点、診療所は270点と決められており、病院のレセプトであれば初診料は必ず250点であって270点である可能性はない。また初診料は一回しか請求できず同一レセプトに2回の請求はない。他の診療行為はもっと複雑だが、とことん細分化してゆくと必ず点数表で決められた点数の和になる。このように日数と点数とは異なった性質の情報であることから分離して分析することがベターといえる。

研究の形態としても、日数は受療状況であって疫学研究に該当し、点数は金額であって経済

研究に当たる。たとえばインフルエンザワクチンの有効性評価は疫学研究であって経済研究ではない。したがってインフルエンザワクチンの有効性評価は日数で行うべきであって点数で行うべきではない。この区分は2002年7月より疫学研究倫理指針が施行され、疫学研究については経済研究とは異なった扱いが求められるようになっただけに重要である。

レセプトが日数という疫学情報と点数という経済情報を合わせもっている点はきわめて豊富な情報源であるといえる。

【平均値を重みに使うことについて】

PDM法プログラムでは「重み」として日数も点数も傷病ごとの単純平均値が用いられている。すなわちある傷病Aの「重み」はその傷病を含む全てのレセプトの日数及び点数の単純平均を算出し、それを重みとして各レセプトごとに比例配分するわけである。いささか乱暴な感じもするが、レセプト全体における特定の傷病の日数点数の合計値だけを知りたいのであれば単純平均でよい。

その理由は、単純平均に個数をかければ、分布がどのようになっていようと同じ合計値になるからである。たとえば人口1万人のA市とB市が一人当たり所得がいずれも100万円であったとしても、所得分布もA市B市同じとは限らない。平均は同じでもA市の方がバラツキ(分布)が大きい、すなわち貧富格差は大きいかもしれない。「所得を比較する時に単純平均で比較するな」は経済学の大原則であり、これが単純平均に対する不信となっているようだ。しかし、A市とB市の所得分布がどうであれ、A市B市の所得総額が100億円であることは間違いない。したがって両市の所得総額を知りたいのであれば単純平均でよく、むしろ中央値や最頻値は不適当なのである。

傷病名のような名義変数を数値化(数量化)する分析法に林の数量化I類がある。これは、たとえば数学が好き・嫌い、といった生徒の属性で数学の得点を予想する多変量解析法である。予想される数学の得点は外的基準、好き・嫌いという属性はアイテムと呼ばれる。PDM法の重みも、たとえば糖尿病の重みは、あるレセプトで糖尿病という傷病名が有る・無い、で点数がどうか予想するものである。119分類の傷病名について重みを算定することは、ある傷病名が有る・無い、をアイテムとして林の数量化I類を119回くりかえすこととに相当する。

林の数量化I類は多変量解析の手法であって、たとえば数学が好き・嫌い、国語が好き・嫌いというふうには複数のアイテムで予測を試みるが、アイテムが一つの場合、たとえば糖尿病が有る・無い、でレセプトの点数を予測する場合は、単純に糖尿病の有るレセプト、無いレセプトの平均値をそのまま使うことになる。「常識で考えてもこれは妥当であろう」とされる(藤沢偉作「楽しく学べる多変量解析法」現代数学社63頁)。

ちなみにある市の外来レセプトについて、119の傷病分類ごとに、その傷病の有るレセプト、無いレセプトの点数をプロットしたのが【図1】である。明らかに、ある傷病の有るレセプトの点数はそれぞれ大きく異なっているが、ある傷病の無いレセプトはおしなべて全体平均に一致している。すなわち、ある傷病を持つレセプトが全体に比べて少数である場合は、その平均値は全体平均からの相対的な乖離を示していると考えられることができる。

もし全傷病が同じと仮定するとその重みは当然ながら全体平均となるが、その逆、つまり「全傷病は同じではない」とすればこれら傷病別平均値を重みとすることは妥当であろう。

PDM法はあくまでレセプトの集合の中のある疾病の総額を求めるものであって、個々のレセプトにおける特定の疾病の額を推計することは、計算の過程でむろん可能だが、その値はあまり信用できないだろう。それは平均所得100万円のA市内で誰か市民をつかまえて所得を調べても

100万円でない可能性の方が大きい。所得は決して正規分布しないからである。

傷病ごとの点数も正規分布するかどうかはまだ十分には解明していない。治療内容が定型的な傷病なら正規分布しPDM法による個々のレセプトにおける推計値もある程度信頼できるであろうが、バラツキの大きい傷病では重みによって比例配分された点数がどれだけ信頼できるかは疑問である。さらに複雑なことはバラツキの小さい傷病も大きな傷病も同じレセプトに「同居」しているということである。極端な場合、レセプトに10の傷病が記載されていても医療費の大半はうち一つの傷病に費やされている、といったことも多々あるだろう。

したがって現時点ではPDM法はレセプト全体における傷病の合計値の推計法であって、個々のレセプトの傷病の点数や日数を推計する上では慎重であるべきと考えられる。

【平均値の算出法】

ある傷病名の記載されたレセプトを全て集めて平均値を出す。傷病名とはよく似た傷病名は同じ傷病分類に一括する。たとえば糖尿病、糖尿病性腎症、糖尿病性網膜症はそれぞれ異なった傷病名だが、119分類ではいずれも糖尿病[402]として一括される。したがって、これら3つの傷病が記載されたレセプトは、402,402,402と同じ分類番号が3つ並ぶ。この場合、PDM法では402というひとつの傷病名にまとめることはせず、402が3回出現するものとして重複カウントする。その根拠を【例題】を用いて説明する。

【例題】診療科別の医師所得を推計したい。一人の医師が単一の診療科を標榜している場合はその医師の所得をそのままその診療科の所得として問題がない。しかしわが国では自由標榜制をとっているので一人の医師が複数の診療科を標榜しているのがふつうである。

そこで、診療科ごとの医師の平均所得を求め、複数標榜の医師の所得はそれに比例して按分する手法をとる。

そのためにはまず診療科ごとの平均所得を算出しなければならない。

以下の例を考える。

表 A	標榜科	所得
Dr.田中	循環器内科	1000 万円
Dr.山本	消化器内科	2000 万円
Dr.佐藤	循環器内科	3000 万円
	消化器内科	

循環器内科と消化器内科の平均所得はそれぞれ $(1000+3000)/2=2000$ 万円、 $(2000+3000)/2=2500$ 万円である。

Dr.佐藤の所得 3000 万円は両科の平均値で比例配分され、以下のようなになる。

表 B	標榜科	所得
Dr.田中	循環器内科	1000 万円
Dr.山本	消化器内科	2000 万円
Dr.佐藤	循環器内科	1333 万円
	消化器内科	1667 万円

よって3人の医師の合計所得 6000 万円の診療科別内訳は、循環器内科 2333 万円、消化器

内科 3667 万円と推計される。

以上が PDM 法プログラムで用いている平均値計算法である。

しかし以下のような計算法も考えられる。

表 C	標榜科	所得
Dr.田中	循環器内科	1000 万円
Dr.山本	消化器内科	2000 万円
Dr.佐藤	循環器内科	1000 万円
	消化器内科	2000 万円

よって3人の医師の合計所得 6000 万円の診療科別内訳は循環器内科 2000 万円, 消化器内科 4000 万円と推計される。

つまりそれぞれの診療科の平均値は, それぞれの診療科単独標榜の医師だけでだし, 複数標榜科医師の所得は単独の平均値で配分するというものである。この計算法は理想的だが, PDM 法では採用していない。その理由を説明する。

単独のみの診療科(あるいは傷病名)の「純粋」な平均値を計算するのは, 医師(あるいはレセプト)の大半は単独標榜で一部のみ複数科標榜という状況では有効であり用いるべきであろう。しかしレセプトでは複数傷病名が多く(だからこそ PDM 法が必要になる)むしろ傷病名一つのレセプトの方が例外的である。

この計算法をとるためには 119 傷病分類の全てについて単独傷病名が相当数(少なくとも 10 件以上)ある必要になる。高血圧や糖尿病のような頻度の高い傷病ならいざしらず, 119 全てについて 10 件以上の単独傷病レセプトを確保するには相当規模のレセプト件数が必要になる。小規模な市町村や組合, また単月レセプトやもっと狭い地区のレセプト分析は不可能になる。100 件程度の件数でも比較的安定した分析をしなければならない PDM 法の要請を満たせない。

医師が 10 人おり, 循環器内科, 消化器内科単独標榜の医師は各一人で残り8人は複数標榜, という状況を考えれば, たった2人の所得の割合で残り8人の所得を配分することの無謀さが理解できる。レセプトの状況はまさにこれに近い。むしろ表3の計算法を否定しているのではなく, レセプト電算化がすすみ都道府県単位の多数のレセプトを分析できるようになればこの計算法の可能性もある。

次に重複カウントの問題に移る。

分類上, 循環器内科も消化器内科もともに「内科」というカテゴリーにくられる。すなわち以下のようなになる。

表 D	標榜科	所得
Dr.田中	内科	1000 万円
Dr.山本	内科	2000 万円
Dr.佐藤	内科	3000 万円
	内科	

PDM法プログラムでは傷病名を重複カウントしており, これより内科の平均所得は $(1000+2000+3000+3000)/4=2250$ 万円となる。これは先に出した循環器内科, 消化器内科別の

平均2000万円と2500万円のさらに平均となる。

一見するとDr.佐藤について内科3000万円を2回カウントするのはおかしいように見える。「分類がくられて同一分類になったのだから重複カウントはおかしい」とDr.佐藤の内科は1500万円として計算すべきのようにも思われる。

これに対する説明は、分類上3人の医師は同一診療科になったが、実際には3人の医師は異なった診療科であり、Dr.田中とDr.山本は違う診療科だし、Dr.佐藤は他の2人の診療科両方の患者をみていることである。もしDr.佐藤は2つの診療科を標榜しているから、とって診療科当たりの所得1500万円として計算すると内科の平均所得は $(1000+2000+1500+1500)/4=1500$ 万円とおかしなことになる。

【平均値の補正】

上の説明は単純平均を重みとしてPDM法を適用した時のバイアスについても示唆を与えている。3人の医師の所得6000万円の内訳は循環器内科2000万円、消化器内科4000万円がもし真の値とすると、PDM法による推計結果はそれぞれ2333万円、3667万円となった。つまり少ない方の循環器内科は過大評価し、大きい方の消化器内科は過小評価された、すなわち上下の格差が縮小し、均等に近づいた。

これは平均値を用いてPDM法を適用する場合の系統的なバイアスといえ、事実、平均値を重みとしてPDM法で傷病別点数を推計した場合、腎不全のような高額傷病の医療費は過少に、逆に皮膚炎のような低額傷病は過大に評価される傾向にあることが当初から気付かれていた。

そこで平均値を重みとして用いる場合、一定の補正が必要になる。

ここでレセプトの傷病数をNとし、点数をPとすると傷病当たり平均点数は P/N 。そのうちの一つの傷病の点数が ΔP だけ大きく(小さく)なってもレセプトの点数Pは $\Delta P/N$ しか増加(減少)しない。

したがって全体平均と同じ傷病については、その傷病の記載されたレセプトの点数の平均をそのまま使っても変化しないが、平均値より大きい傷病は過少に、逆に小さい傷病は過大に評価され、その影響はレセプトの平均傷病数Nによって決まる。

一般にレセプト平均傷病数Nの場合、ある傷病の「真」の平均点数をX、ある傷病の記載された全レセプトの平均点数をYとすると両者は次のような関係になると考えられる。

$$Y = X/N + b \quad (b \text{ は定数})$$

定数bは、全体平均値 μ で $Y=X$ と交わるように定まる。かりに傷病別点数の全体平均 μ を500とすると、上式は以下のようなになる。

$$N=2 \text{ なら } Y = 0.5 * X + 250$$

$$N=4 \text{ なら } Y = 0.25 * X + 375$$

$$N=8 \text{ なら } Y = 0.125 * X + 437.5$$

このようにレセプトの傷病名数が多くなるほど直線は平べったくなってゆく。ちなみに外来レセプトの平均傷病名数は約4、入院レセプトは約8である。

知りたいのは各傷病の「真」の平均点数Xであるが、実際のレセプト分析ではXは不明で、わかるのはYと μ だけである。そこでYと μ だけからXをできるだけ正確に予測するため、補正してできるだけXに近いY'を求める以下の補正式を考案した。

$$Y' = Y * (Y/\mu)^k \quad (k > 0)$$

これはYが全体平均 μ と同じであればそのまま、もしYが全体平均 μ の2倍であれば4倍に、逆に2分の1であれば4分の1にする補正をk回くりかえす式である。k=0ならY'はYそのまま、つまり無

補正ということになり、 $k=1$ を1次補正、2なら2次補正と呼ぶことにする。その概念をグラフで示した【図2】。そして、この補正を何次(k 回)まで行えばよいかを、レセプト傷病数 N の数に応じてシミュレートした【図2-1~3】。

その結果、 $N=2$ すなわち平均傷病数が2つの時は1次補正($k=1$)で十分であるが、 $N=4$ では2次補正まで必要であり、 $N=8$ では5次補正まで必要であることがわかった。

以上をふまえ、PDM法Ver.2には補正式のオプションを加え、外来レセプトでは2次、入院レセプトでは5次補正が望ましいとした。

この補正は、点数についてのみ行い、日数については行なわない(理由は後述)。

【繰り返し法】

PDM法は、まず傷病ごとに平均値を算出し、それを重みとして各レセプトの点数を記載された傷病ごとに配分し、さらにレセプトごとに傷病ごとに配分された点数を傷病ごとに合計する、という演算処理である。したがって演算処理の過程で、レセプトごとに傷病別の平均値で比例配分された新しいレセプトデータがまるまる生成される。

このように処理過程で生成された新しいレセプトデータを再度PDM法にかけ、その過程で生成されたレセプトデータをもういちどPDM法にかける…という処理を繰り返すと一定条件の下では最適な値に収束することが明らかになった。最尤推定法(Maximum Likelihood Estimator)という、推定式を繰り返せば最も尤もらしい値に収束する、という手法の考えに近いものである。PDM法は、各傷病の点数の平均値を計算し、それを補正式で補正したが、繰り返し法では補正式を使うのではなく、傷病別の平均値を出しては同じ処理を繰り返す。

傷病別の平均値は繰り返すごとに微妙に変化してゆく。しかし繰り返しを重ねるごとに変化は小さくなってゆき、やがてはほとんど変化しなくなる。コンピュータープログラムにおいては、たとえば回を重ねるごとの変化がたとえば0.01より小さくなったら「繰り返し打ち切り」とプログラミングしておく。

この方法は、とくにレセプト件数や傷病分類数が多いとコンピューターに大きな負担を強いるのでPDM法にはまだ組み込まれていないが、将来コンピューターの性能が向上すれば可能性の高い手法である。ただし、この手法ではどのような条件でも収束するとはかぎらず、あくまで加法原理(後述)が成立していることが条件となる。

【シミュレーションデータによる検証】

PDM法開発の上でこれまで最大の障害は、その妥当性の検証ができない、ということであった。すなわち実際のレセプトデータの傷病別の「真」の日数や点数は不明であり、適用して分析結果を出しても、それがどれだけ真実に近いのか検証できなかった。

それが、シミュレーションデータを用いることによって部分的には解決した。シミュレーションデータとはすなわちあらかじめ傷病別の点数等の「正解」を設定し、それに合わせて生成したデータのことである。この「正解」のわかっているデータをPDM法にかけて、推計結果と「正解」とがどれだけ一致しているかで妥当性を検証できる。

社会医療調査や国民健康保険医療給付実態調査等を元にわが国の外来レセプトにできる限り近づけて作製したレセプト1000件分のシミュレーションデータで検証した結果が【図3-1,2】である。ヨコ軸にはシミュレーションデータの傷病別点数(正解、 x)、タテ軸にそれぞれの方法による傷病別点数の推計値(y)を示す。もし完全に一致する $y=x$ になり決定係数(R^2)は1となる。

図からも明らかなように平均値を補正する方法でも繰り返し方法でもきわめて良好な近似が

得られた。しかしながら、日数については平均値をそのまま使う方法では良好な近似であったが、繰り返し法では収束はしたものの、その近似はきわめて不良であった【図4-1~3】。繰り返し法でなぜ日数でいい結果がえられなかったか、その理由は現時点では断定できないが、ひとつの仮説としては以下に述べる「加法原理が日数ではあてはまらない」ことが考えられる。

これにより日数で評価するインフルエンザワクチンの効果測定は繰り返し法ではなく平均値そのまま重みとして用いる方法で実施した。

(繰り返し法のコンピュータープログラミングでは横山徹爾主任研究官の協力に感謝する)

【加法原理と最大原理】

点数の推計では傷病数で割り、平均値を補正式で補正したが、日数の推計ではそのような処置は行っていない。また繰り返し法では日数については適切な値に収束しなかった。その理由を考察する。

点数については、たとえば費用1万円を要する傷病Aと2万円の傷病Bが合併するレセプトの医療費は3万円とみなしている(加法原理)。それに対して、月4回通院を要する傷病Aと2回ですむ傷病Bを合併するレセプトの日数は6日ではなく4日とみなしている。明らかに、異なる傷病を合併しても同時に治療が可能であり、たしあわされるのではなくその最大日数が受療日数になる、と考えられる(最大原理)。

したがって傷病別の点数を推計するためには合計点数を傷病数でわる必要があるが、日数についてはそのような処理は行わない。そして傷病数でわる場合は、それによって得られる平均値は傷病数による補正が必要となるが、日数についてはそのような補正は必要とされない。

もっともこのような最大原理と加法原理が厳密になりたつわけではなく、月4回の通院を要する傷病Aに別の傷病が合併すると通院日数は平均して4回よりも多くなる可能性はあるし、また1万円の傷病Aと2万円傷病Bとが合併しても3万円になるとはかぎらず、たとえば診察料のように傷病数とは無関係に一定額しか請求できない診療報酬もある。投薬や検査の中には複数の傷病を適応症にしているものもある。これまで漠然と点数とは総点数を指してきたが、点数をさらに診察料、投薬料、検査料と細分化してPDM法で分析する場合、最大原理と加法原理とは適宜使いわける必要がある。

ちなみにPDM法Ver.2では、傷病数でわらず、一件当点数をそのまま使うオプションも加えた。これは手術料等単一の傷病に月一回しか行なわれない診療行為の報酬分析を意図している。

現時点では、上に述べたような「わりきり」を行ったが、日数と点数が傷病との関連で最大原理、加法原理がどれだけ通用するかは、今後の研究課題である。

【PDM法プログラムの概要】

本研究の最大の成果物は、以上のPDM法の原理をウィンドウズパソコン上で行えるコンピュータープログラムである。そのVer.1が2002年10月埼玉で開催された日本公衆衛生学会総会における第5回レセプト情報の活用を考える自由集会で発表された【マニュアルを参考資料に添付】。その後、シミュレーションによる補正式の開発等ととりくみ、近くVer.2が完成する見通しである。

Ver.2の概要を【表1】に示すとともに、それを実際にある市の外来レセプト15771件に適用し、点数および日数の傷病別割合を出した結果を示す【図5】。

【PDM法の応用例】

PDM法を具体的に応用した例として、インフルエンザワクチン接種の有効性評価をあげる。

本研究は宮城県名取市の委託によるもので報告書が2003年2月名取市に提出された【資料2】。

ワクチンの有効性評価で最も適切なものは無作為割付比較試験であるが、高齢者へのインフルエンザワクチンの有効性は無作為割付比較試験で既に確立されているため、評価目的に一般人口に行うことは倫理的に許容されない。そこでインフルエンザワクチン接種の有効性評価をレセプト情報を用いて行ったところ、無作為割付比較試験と同様の結果が得られた。

名取市の65歳以上高齢者のレセプトをワクチン接種者台帳と個人情報でリンケージし接種、非接種者にわけた。しかしそれだけでは、分母がわからないため発生率を比較することはできない。そこで接種、非接種者両群の総受療日数に占めるインフルエンザ及び関連呼吸器疾患の「割合」を測定することによって比較する。【日数と点数の扱い】でも述べたようにワクチンの有効性は疫学情報である日数で評価すべきであって、経済情報である点数で評価すべきではない。そして日数については補正式を用いることなくPDM法で傷病別日数を推計すべきである。

接種、非接種者の疾病リスクは受療日数に反映されるとみなし、もしワクチンの効果がなければ、総受療日数に占めるインフルエンザの割合は両群で同じであると仮定する。もしワクチンの効果があれば、それはインフルエンザによる受療日数の減少として現れ、結果として総受療日数に占めるインフルエンザの割合は減少する。こうすることによって、発生率が不明でも、接種、非接種者両群のリスクを補正しつつ、ワクチンによるインフルエンザ受療日数の減少を客観的に測定できる。そして得られた結果は既に確立された無作為割付比較試験の結果ときわめて近似していた。

インフルエンザによる受療日数を客観的にきりわけるPDM法により、たとえ発生率がわからなくてもレセプトだけで簡便にワクチン効果を測定できる手法が確立された。これにより、今後全国の自治体はインフルエンザワクチン接種事業の有効性をレセプトだけから行えることになり、行政評価の観点からも、疫学研究の観点からもきわめて有意義な成果であった。

【添付資料】

【資料1】PDM法使用マニュアル2002年10月24日公表

【資料2】名取市「高齢者インフルエンザ予防接種の効果分析」報告書2003年2月

【図1】傷病名の有無による日・傷病当点数の平均

某市外来レセプト

腎不全

- 当該傷病名のあるレセプト
- 当該傷病名の無いレセプト





