

(a)合計と食事・生活療養の3ヶ月間の平均 (b)1レセプトあたりの合計と食事・生活療養の平均
図3 合計, 食事・生活療養の比較

③レセプト種別に基づいた種別による受診行為の検討

図4にレセプト種別による診療日数の解析を行った。図は左側が3ヶ月間の累計結果、右が1レセプトあたりの平均値を示した。この結果より、レセプト種別による、すなわち、高齢者や世帯主、未就学、さらに1種、2種での医療費の割合を推定できると考えた。

図4-1、4-2には入院外の結果を示した。結果より、どの群も1種に分類される対象者の累積診療回数が多いことがわかる。一方で1レセプ

トあたりの平均からどの群も月に1~3回の受診回数であることがわかった。

図4-3、4-4に入院の診療日数の結果を示した。結果より、外来と同じようにどの群も1種に分類される対象者の累積診療回数が多いことがわかった。外来も入院も後期高齢者群の一般・低所得者に分類される対象者の診療回数が多いことがわかった。一方で、1レセプトあたりの診療回数はどの群も概ね20~25日程度であることがわかった。

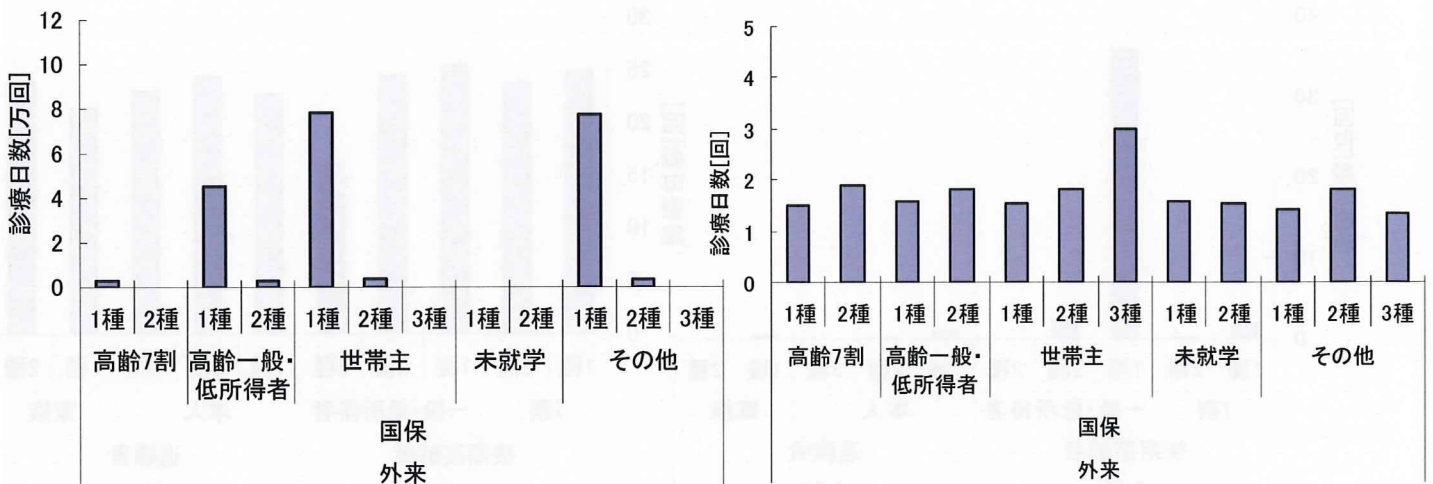


図4-1 国保(外来)の診療日数の累計と1レセプトあたりの平均

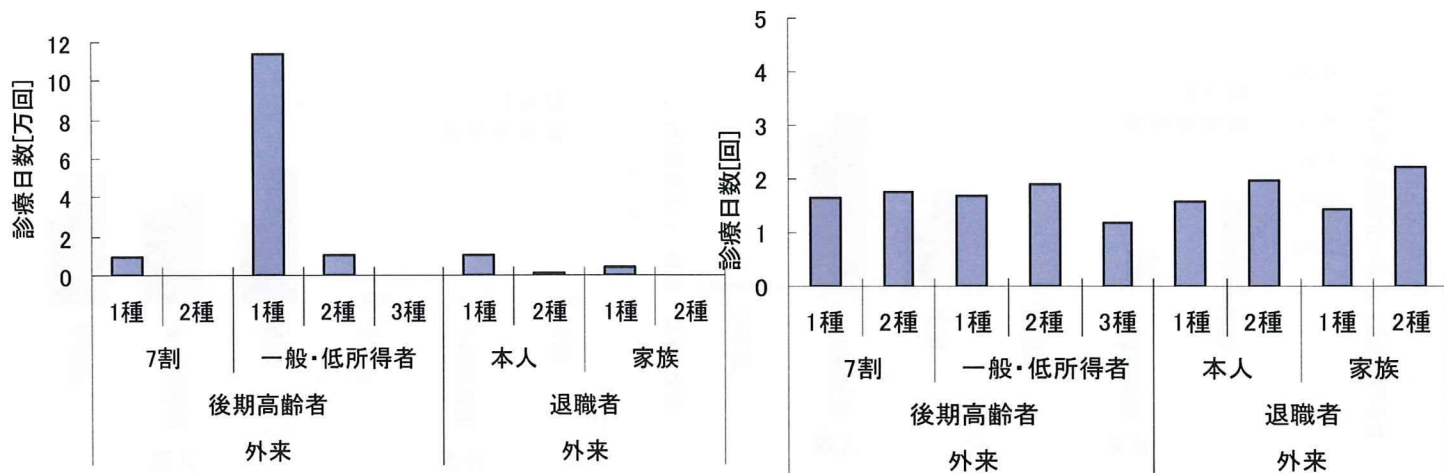


図 4-2 後期高齢者・退職者（外来）の診療日数の累計と1レセプトあたりの平均

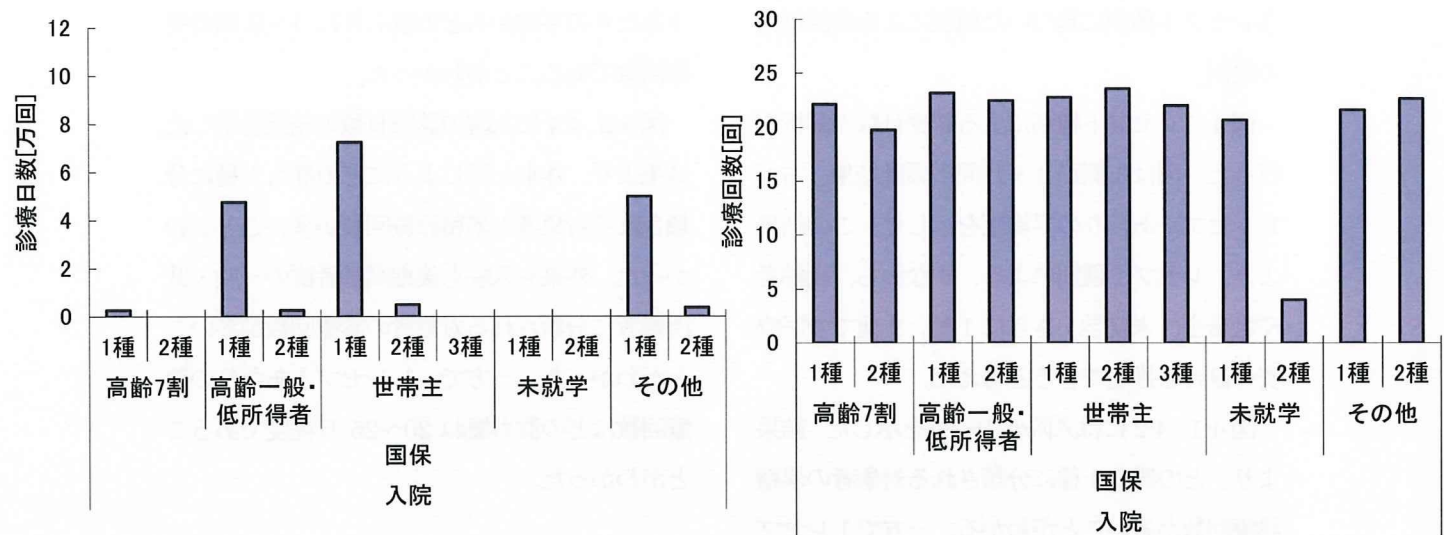


図 4-3 国保（入院）の診療日数の累計と1レセプトあたりの平均

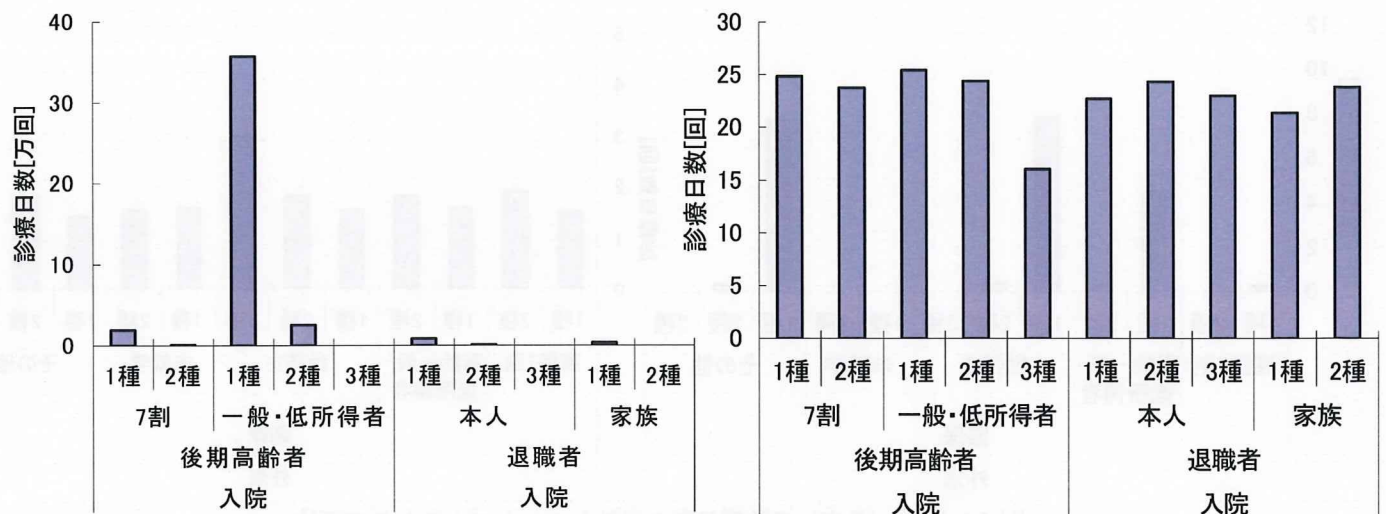


図 4-4 後期高齢者・退職者（入院）の診療日数の累計と1レセプトあたりの平均

D. 考察

①難病者の国保に対する医療費と受診行為に占める割合の考察

本データは国保中央会における医療費全体の割合を比較した。難病者とそれ以外の人（健常者を含む様々な対象者）の医療機関等を利用する年間の傾向が同じだと考えると、国保の医療費において、難病者は全体の約1%を占め、退職者、後期高齢者は1~1.2%前後を占める。国保全体に対する影響は前期高齢者（65~74歳）が全体に対して49%を占め（H21年10月）、後期高齢者が国保全体と同額（国保の1.1倍）を占めることを考えると、難病者が占める割合は高くない。

図1より、医療費の合計をレセプト件数で除して1レセプトあたりの医療費単価を3カ月の平均として求めると、国保該当者群では58,888円、退職者群では59,288円、後期高齢者群では112,955円であった。比較した国保全体では、国保該当者群で19,784円、退職者群では20,903円、後期高齢者群では30,852円であった。つまり、国保該当群と退職者群では難病者の方が約3倍、後期高齢者群では、3.7倍の医療費がかかっていることになる。

もちろん総医療費は“受診患者数”×“1人あたりの平均医療費”であるので、絶対数の違いが医療費全体に与える影響に表れているが、難病者の病態を勘案しても難病者の医療費が難病ではない身体機能を持つ人に比べて3倍程度の範囲であることがわかった。

以上のことより、ここからは医療費の高低（高い、安い）のみを議論するのではなく、“生活”、“支援”という観点から議論する必要があると考えられる。

難病者の国保全体の入院・食事療養・生活療養に占める割合は3~4.5%であり、入院外の0.6

~1.2%に比べ高いことがわかった。この傾向は退職者や後期高齢者で高い割合となることから、一般健常者と同様に、加齢により入院・食事・生活療養の必要性が高いことが考えられる。レセプト件数は、難病者は0.3~0.4%であり、入院、食事療養・生活療養では1.2~1.7%、入院外は0.4~0.7%であった。診療日数は0.6~0.9%、入院、食事療養・生活療養で1.6~2.5%、入院外は0.3~0.6%であった。以上より、入院外に比べて入院の費用が国保該当者で1.3倍、退職者で1.7倍、後期高齢者で5.2倍大きいことがわかった。

この結果は各病名に分けて解析したものではなく、法別番号51に該当する平均値である。そのため、疾病によって医療の必要度は異なると考えられる。ここでは大枠のみ議論したが、疾病による必要な医療行為や生活支援のあり方を今後は詳細に議論したい。

②国保、退職者、後期高齢者群での全体の医療費と1レセプトあたりの医療費と受診行動の解析の考察

図2(a)の入院・入院外と国保・後期高齢者・退職者群の3ヶ月間の平均の結果より、外来の診療日数では国保該当者群、後期高齢者群、退職者群の順であった。入院に関しては、後期高齢者群、国保該当者群、退職者群の順であった。図3(a)より、これにかかわる外来、入院の医療費、食事・生活療養費も同じ傾向であった。一方で、図2(b)より、1レセプトあたりの1ヶ月の平均の診療日数は、外来で約1.7日、入院で21.5~24.2日であり、3群間で大きな差は見られなかった。

医療費全額補助の対象となっている難病者の自己負担はここからは議論できないが、自己負担が発生した場合にも同様の医療行為を受ける場合についても考慮が必要となる。疾病の進行

状況や投薬の必要性によって医療費に大きなバラつきが出ることは間違いないが、難病にかかわる生涯医療費を罹患年数で除したものが1年にかかる平均医療費となる。ここでは、疾病ごとの検討を加えていないが、軽度から重度まで幅広く含めた難病者群の平均医療費（1レセプトあたり）として、外来では退職者の3583点/月、国保該当者の2609点/月、後期高齢者の2009点/月が明らかになった。入院では国保該当者で62000点/月、退職者の60000点/月、後期高齢者の50000点/月がわかった。食事療養・生活療養では、どのグループも約40000円/月であった。

今後、各疾病で生涯の平均医療費を算出し、罹患期間を推定することで難病者や家族が負担する医療費の推定を行う。また、治療行為のみではなく、介護や入院、在宅のそれぞれで必要となる負担を算出し、公平な支援のあり方について議論したい。

また分子標的薬などの使用により、患者の病気の進行を抑え、在宅や仕事などの社会的活動を維持できることもある。したがって、医療費の適正使用、適切な支援という観点について、単なる医療費やその使用者をデータのみから議論することはできない。個々の患者の状態や支援のあり方を見据えた検討が必要である。合わせて、調剤とその効果まで含めて検討しなければならないと考える。

③レセプト種別に基づいた受診行為の検討の考察

高齢者や低所得者、世帯主などのレセプト種別に基づく診療日数の解析を行った。その結果、全体の合計では、どの群も第1種に該当する対象者の割合が高くなった。しかし、1レセプトあたりの診療日数は各群で大きな差は見られなかった。国保中央会における難病者のレセプト

種別中の人数が明らかでないため、受診頻度等は推定できないが、1種に該当する人数が他群よりも多いか、受診頻度が高いことがこの結果に表れていると推定される。

E. 結論

本解析より次の点が明らかになった。

- ・ 国保の総医療費において、難病者は全体の約1%を占め、退職者、後期高齢者は1～1.2%前後を占める。前期高齢者（総医療費の約50%）や後期高齢者（国保の総医療費と同額）に比べて、難病者が占める割合は小さい。
- ・ 3ヶ月間のデータの平均値、標準偏差に大きなばらつきは見られなかった。
- ・ 1レセプトあたりの医療コストの平均は、難病者は非難病者に比べて約3倍であった。医療ニーズを考えた場合、これはそれほど大きい値ではないため、生活や支援という立場からどのような医療構造を計画することが望ましいかを議論する必要がある。
- ・ 軽度から重度まで幅広く含めた難病者群の1レセプトあたりの平均医療費として、外来では約2000～3600点/月、入院では50000～62000点/月かかっていることがわかった。食事療養・生活療養では約40000円/月であった。補助対象にならない難病者の場合、大きな自己負担が長期的に加わる可能性が示唆された。そのため、生涯医療費という観点からどのような支援が公平であるかを生活や在宅というキーワードを含めて今後議論する必要があると考えられた。
- ・ レセプト種別における解析から、医療費の合計では、第1種に該当する対象者の割合が高かった。しかし、1レセプトあたりの診療日数は各群で大きな差は見られなかつ

た。

2. 学会発表 なし

F. 健康危険情報：なし

H. 知的財産権の出願・登録状況

G. 研究発表

1. 特許取得：なし

2. 実用新案登録：なし

1. 論文発表 なし

3. その他：なし

厚生労働科学研究費補助金（難治性疾患克服研究事業）
分担研究報告書

原発性胆汁性肝硬変および自己免疫性肝炎の医療費に関わる調査

研究分担者 渋谷 明隆 北里大学医学部医療安全・管理学 准教授

研究要旨

難治性肝疾患である原発性胆汁性肝硬変(PBC)および自己免疫性肝炎(AIH)はともに自己免疫現象を背景にした特定疾患であるが、PBCには医療費助成があり、AIHにはない。DPCデータの分析から、両疾患の入院に関わる医療費について比較した。その結果、平均在院日数はPBCよりのAIHの方が長い。1入院にかかる医療費はPBCでは181,667円、AIHは120,662円であった。PBCでは生体肝移植例を除くと医療資源費用は1入院あたり130,362円となり、AIHと差はなかった。PBCの生体肝移植8例の総医療資源費用は110,178,320円とでありPBC総額の17.0%を占めていた。以上より、PBCとAIHの1入院あたりの医療費に差はないが、PBCの生体肝移植例では著しく高額な医療費がかかっていた。医療費助成のあり方を再検討する必要があると思われた。

A. 研究目的

原発性胆汁性肝硬変(PBC)と自己免疫性肝炎(AIH)はともに自己免疫性肝疾患であり、生涯にわたる診療が必要であるが、医療費助成の対象となるのはPBCのみである。DPCデータの分析から、両疾患の入院に関わる医療費について調査し比較した。

B. 研究方法

(株)グローバルヘルスコンサルティング・ジャパンの保有するDPCデータを分析し、PBCとAIHに関わる医療費構造を比較した。調査病院数はDPC導入311病院で、2007年7月から2009年10月までに退院した症例のうち「もっとも医療資源を投入した疾患」としてPBCまたはAIHを選択した症例を対象とした。

C. 研究結果

PBCは1232例(男性:女性=18.1%:81.9%)、AIHは1376例(男性:女性=17.2%:82.8%)が対象となった。患者の平均年齢はPBC 62.1歳、AIH 61.2歳である。平均在院日数はPBC 15.3日、AIH 19.3日であった。入院中の死亡率はPBCでは5.8%、AIHでは1.7%であった。1入院に関わる医療費はPBCで181,667円、AIHで120,662円であった。ただし、PBCでは8症例の生体肝移植があり、これらを除くと医療資源費用は1入院あたり130,362円となり、AIHと差はなかった。生体肝移植例を除外して両疾患の医療費の内訳をみると、PBCでは投薬、注射、処置、検査、画像がそれぞれ17,977円、36,251円、8,544円、52,598円、14,992円であり、AIHでは15,125円、21,565円、6,733円、61,766円、15,434円と、両疾患で費用分布、金額に大きな違いはなかった。投薬、注射

のうち、両疾患とも血液製剤が高額医療費の上位を占めていた。PBC の総医療資源費用は 648,655,776 円で、AIH の総額は 700,777,508 円であった。PBC の生体肝移植 8 例(0.65%)の総医療資源費用は 110,178,320 円とであり PBC 総額の 17.0%を占めていた。PBC のうち、肝硬変の重症度を示す Child 分類が明らかな症例は 288 例であり、Child A, Child B, Child C の医療資源費用の総額はそれぞれ 32,855,326 円(一人当たり 217,585 円)、40,095,974 円(一人当たり 668,266 円)、93,227,776 円(一人当たり 1,210,750 円)であった。

D. 考察

PBC と AIH はともに自己免疫性肝疾患であり、生涯にわたる診療が必要であるが、医療費助成の対象となるのは症候性 PBC のみである。しかし、実際の診療の現場では症候性 PBC と無症候性 PBC を厳密に区分することは困難で、多くの無症候性 PBC が医療費助成を受けている。我々のこれまでの検討で無症候性 PBC の生命予後は良好であり、むしろ長期のステロイド療法を必要とする AIH にこそ入院・外来を含めた医療費助成の必要性があると考えている。今回は入院例に限った検討であるが、PBC と AIH では生体肝移植例を除いた 1 入院に関わる医療費がほぼ同額であることが明らかになった。在院日数が AIH でより長いのはステロイド療法の影響があると考えられる。

一方、PBC の生体肝移植例では高額の入院医療費がかかるうえに、生涯にわたる免疫抑制剤の服用や再発の可能性などから十分な助成がなされるべきと考えられる。また、肝硬変化した PBC では Child A, Child B, Child C と重症化するに従い医療費は高額になり、患者負担が増加していた。今後、両疾患の生涯医療費について

て外来費用を含めた検討が必要である。

E. 結論

PBC と AIH では生体肝移植例を除いた 1 入院に関わる医療費がほぼ同額であることが明らかになった。PBC 1232 例中 8 例(0.65%)の生体肝移植例で PBC 医療費全体の 17%を占めていた。

〈謝辞〉

本研究・調査にご協力いただいた(株)グローバルヘルスコンサルティング・ジャパンに感謝します。

〈参考文献〉

Akitaka Shibuya, et al. Hepatocellular carcinoma and survival in patients with primary biliary cirrhosis. *Hepatology* 35:1172-1178, 2002

F. 健康危険情報：なし

G. 研究発表

1. 論文発表

1. Akitaka Shibuya, Keiko Hanai, Yumi Arai, Kazui Soma : The cost of adverse events in hospitalized patients. *Kitasato Medical Journal*, 39:165-170, 2009.
2. 渋谷明隆、小林弘祐 : DPC データ分析からみた北里大学 4 病院の位置づけ. *北里医学*, 39 : 117-128, 2009.

2. 学会発表

渋谷明隆、中澤貴秀、荒井康夫 : 一般病院と特定機能病院での肝癌治療の収支実態の比較. 第 45 回日本肝臓学会総会, ワークショップ「DPC 時代における肝細胞癌治療戦略」, 2009, 京都. (肝臓 50 Suppl.2:A483, 2009.)

H.知的財産権の出願・登録状況

1. 特許取得：なし
2. 実用新案登録：なし
3. その他：なし

厚生労働科学研究費補助金（難治性疾患克服研究事業）
分担研究報告書

Rを用いたシミュレーションによるサンプルサイズの算出

研究分担者 森實 敏夫 国際医療福祉大学塩谷病院内科 教授

研究要旨

ある疾患の患者一人当たりの医療費は同じ医療機関の中でも、さまざまであり、一定の分布に従うことが想定される。正規分布に従うとした場合、母集団の平均値と標準偏差を少数サンプルの値から想定し、それをもとに、任意に設定した最大過誤 E の範囲にサンプルの平均値が収まるようなサンプルサイズの算出が必要である。フリーの統計ソフトである R を用いたシミュレーションによるサンプルサイズの算出のための関数を作成した。1 医療機関の場合および3 医療機関の場合の関数を作成し、その計算結果が適切であることを確認した

A. 研究目的

ある疾患の患者一人当たりの医療費は同じ医療機関の中でも、さまざまであり、一定の分布に従うことが想定される。正規分布に従うとした場合、平均値 μ と標準偏差 σ の2つのパラメータでその分布を表わすことができる。

そこから、サンプルサイズ n のサンプルを抽出した場合、すなわち n 名の患者の医療費の額のデータが得られた場合、そのサンプルの平均値は平均値 μ 、標準偏差 σ/\sqrt{n} の正規分布に従うことは中心極限定理を適用して知ることができる。言い換えると、サンプルサイズ n のサンプルを抽出することを何回も繰り返すと、毎回異なる値が得られ、その値は一定の正規分布に従うということである。したがって、1回のサンプル抽出で得られた平均値はその中の1つであり、 μ そのものではない。

そこで、サンプルの平均値をある任意の精度で求めたい場合、容認しうる最大過誤を設定して、

必要なサンプルサイズを算出することが可能になり、この手法は正確度分析（Precision analysis）と呼ばれている¹⁾。

1つの医療機関を対象にする場合には、理論的な計算式で計算することによって、必要なサンプルサイズを容易に求めることができるが、複数の医療機関を対象にする場合には、それぞれの医療機関の特性により、患者1人当たりの医療費の分布は異なるので、医療機関の間のばらつきを考慮する必要があり、計算法は複雑になる。

PCの機能向上に伴い、モンテカルロ・シミュレーションにより、一定の分布の母集団からのサンプルを多数得て、その特性を確認することが容易にできるようになった。Rという統計解析ソフト²⁾はフリーで公開されており、ランダムサンプルのための関数が用意されているだけでなく、使用者が目的に応じた関数を作成することが可能である。今回、Rを用いて、サン

ルサイズを算出する方法を開発した。RはThe Comprehensive R Archive Network (CRAN)のホームページ(<http://cran.r-project.org>)からダウンロードできる。

B. 研究方法

理論的に算出した場合とシミュレーションによる算出の比較

最初に、1つの医療機関のデータについて、すなわち、単一群で平均値 μ と標準偏差 σ が分かっている場合のサンプルサイズの算出を理論的な計算式により算出した場合と、Rを用いたシミュレーションにより算出した場合を比較して、その妥当性を検討した。

理論的な計算式は次のようなものである。最大過誤Eとした場合、サンプルの平均値が $(1-\alpha)100\%$ の範囲に入るようにするためのサンプルサイズは次の式で算出される：

$$n = \frac{Z_{\alpha/2}^2 \sigma^2}{E^2}$$

この式は次の式から展開したものである：

$$(1-\alpha)100\% = \bar{X} \pm Z_{\alpha/2} \sigma / \sqrt{n}$$

ここで、たとえば、95%信頼区間であれば、 α は0.05となり、 $Z_{\alpha/2}$ は1.96となる。すなわち、標準正規分布に変換して、 $0 \pm 2SD$ の範囲にほぼ相当する。 \bar{X} はサンプルの平均値のことである。したがって、最大過誤Eは $\mu - \bar{X}$ である。

次にRを用いたシミュレーションであるが、

$rnorm(n, m, s)$ の関数で、平均値 m 、標準偏差 s の分布の母集団から、 n 個のランダムサンプルを得ることができる。さらに、Rには繰り返しの関数として、`replicate(回数, expression, simplify=TRUE)`という関数がある。`expression`を回数分実行し、ベクトルに収めるということを実行してくれる。

3つの医療機関のデータを想定したシミュレーション

次に、3つの医療機関のデータについて、それぞれの医療機関のその疾患の患者総数、医療費の平均値と標準偏差が分かっている場合、ある任意のサンプル数を調査した場合の、平均値の精度、すなわち95%信頼区間を、Rを用いたシミュレーションにより算出する。

なお、上記いずれの場合も、母集団の平均値と標準偏差を正確に知ることはできないので、少数例のサンプルから求めた平均値と標準偏差の値を用いることとする。なぜならば、ランダムサンプルであれば、その平均値も標準偏差も母集団の平均値と標準偏差に近似するからである。また、標準偏差については、その平均値から一番値の小さい例の値の差を求め、それを1.96で割り算した値を標準偏差として用いることも可能である。後者は、95%のデータが含まれる範囲の下限値が一番値の小さい例の値であるとみなして、算出する方法である。

3つの医療機関のデータに対するサンプルサイズ算出の関数

さらに、3つの医療機関の患者数の整数比、それぞれの医療費の平均値と標準偏差の値から、

シミュレーションに基づき、必要なサンプルサイズを算出する関数を作成した。

C. 研究結果

理論的に算出した場合とモンテカルロ・シミュレーションによる算出の比較

たとえば、平均値 12000 (円)、標準偏差 1000 の正規分布に従う母集団を想定して、式 (1) を用いて、最大過誤 $E=200$ の場合で、 $\alpha=0.05$ 、すなわち 95% の確率でサンプルの平均値が 12000 ± 200 の範囲に入るサンプルサイズを計算してみた。

$$1.96^2 \times 1000^2 / 200^2 = 96$$

つまり、96 症例ということになる。

同じ母集団から 96 個のランダムサンプルを得て、その平均値の分布を、R を用いたシミュレーションで見てみた。平均値を算出し、それを 10 万回繰り返し、変数 `m_dist` に格納することを行うには次のように記述する。ここでは、平均値、標準偏差、サンプルサイズ、回数をそれぞれ `m`, `s`, `n`, `k` という変数にいったん代入してから計算を実行している：

```
> m = 12000
> s = 1000
> n = 96
> k = 100000
> m_dist = replicate(k, mean(rnorm(n, m, s)))
```

これで、ランダムサンプル 96 個の平均値 10 万個が変数 `m_dist` にベクトルとして格納される。その 95% 信頼区間は `quantile()` 関数を用いて

求めることができる。

```
> quantile(m_dist, c(0.025, 0.975))
      2.5%      97.5%
11800.30 12200.04
```

上記の式 (1) による計算と同じ値である。すなわち、平均値 12000、標準偏差 1000 の母集団から、96 個のサンプルを得て平均値を求めた場合、95% の場合は、11800 から 12200 の間に入る事が分かる。

次に、R を用いたシミュレーションで、任意の最大過誤 E を設定し、サンプルサイズを算出するための関数を作成した。平均値 m と標準偏差 s は固定して、 n をある値から 1 つずつ増加させ、`quantile(変数, 0.025)` の値が $m - E$ を超えたときの n の値にすればよいと考えられる。

R では自分で関数をテキストファイルとして書いて、それを呼び出して実行させることが可能である。まず、適当なフォルダ名で新しいフォルダを作成する。R の File メニューから、Change dir... を選び、そのフォルダを作業フォルダに指定する。以下に示す内容をメモ帳などのテキストエディタで書きこんで、テキストファイルとして、そのフォルダに保存する。ファイル名は `samp_size_m_sd_s.R` のようにする。すなわち、拡張子として R を付ける。なお、ファイル名は自分で分かりやすい他のものでも構わない。

```
sample.size.m.sd=function(m,s,E)
{
  lowlimit = m - E
  maxn = 1000
```

```
k = 100000
for (n in 1:maxn)
{
  m.dist = replicate(k, mean(rnorm(n, m, s)))
  if (quantile(m.dist,0.025)>= lowlimit)
  break
}
return(n)
}
```

このプログラムをRから呼び出して、実行させることができる。このプログラムは、 m, s, E の3つの引数を受け取り、95%信頼区間の下限値として、`lowlimit` という変数に $m - E$ の値を代入する。 m は平均値、 s は標準偏差である。サンプルサイズの上限值として、1000 を `maxn` という変数に代入する。`rnorm()`関数を用いて n 個のランダムサンプルを得る作業を100000回実行させるが、この回数の値は k という変数に代入する。 n の値を1から開始し、1ずつ増加させながら、平均値 m 、標準偏差 s の正規分布に従う母集団からのランダムサンプルを n 個抽出して、その平均値を求める作業を100000回行い、95%信頼区間の下限値が `lowlimit` の値を超えた時点で、この繰り返し作業を中断し、その時の n の値を返り値として戻す。以上である。

Rでは次のように記述する：

```
> source("samp_size_m_sd_s.R")
```

これにより、上記のプログラムを使用できる状態にする。これにより、関数 `sample.size.m.sd`(平均値、標準偏差、最大過誤値)を用いることができるようになる。この関数には、カッコ内の3つの値をコンマで区切っ

て入力する。

```
> sample.size.m.sd(12000,1000,200)
[1] 96
```

この例では、平均値が100、標準偏差が10の母集団からのランダムサンプルの平均値の分布で、95%信頼区間の下限値が12000-200以上となった際の値が得られる。たしかに、96という値が得られた。式(1)で算出した値と同じであり、最初のシミュレーションの結果とも一致している。

なお、 $k=100000$ と10万個のサンプルにすると誤差が小さくなるが、計算に時間がかかるので、 $k=10000$ 程度でも実用にはそれほど支障がない。実際に $k=10000$ と1万個のサンプルに基づいたシミュレーションではサンプルサイズは94となった。

以上、理論的な計算値とシミュレーションによる値が一致することが確認された。

3つの医療機関のデータを想定したモンテカルロ・シミュレーション

各医療機関を受診しているその疾患の患者数は異なるので、それぞれの医療機関の平均値を単純に平均することはできない。各医療機関のその疾患の患者数の比のデータが必要になる。できるだけ小さな値で、整数で比を表す必要がある。

それでは、表1に示す、3つの医療機関を調査するとして、すなわち、それぞれ異なる分布の3つの層からのランダムサンプルの平均値の

分布をみる作業である。

表 1. 3 医療機関のデータ。

	患者数(N _i)	整数比(n _i)	医療費平均値(m _i) (円)	医療費標準偏差(s _i)
医療機関 1	40	2	7000	650
医療機関 2	60	3	8000	1200
医療機関 3	100	5	9500	1000

それぞれの医療機関からの症例数として、2 : 3 : 5 になるように、10、15、25 例を調査する
としよう。サンプルの抽出は、1 万回行う設定
とした。なお、k = 100000 とすれば、10 万回
になり、より精度を上げることができる。
h_average という変数に、それぞれの病院の 10、

15、25 例の平均値の値が、1 万個、計 3 万個の
値がベクトルとして代入される。その 3 万個の
値の 95%信頼区間を quantile()関数で、平均値、
中央値などの値を summary()関数で表示した。
さらに、hist()関数でヒストグラムを表示した。
R で以下のように記述する :

```
> k = 10000
> n1 = 10
> m1 = 7000
> s1 = 650
> n2 = 15
> m2 = 8000
> s2 = 1200
> n3 = 25
> m3 = 9500
> s3 = 1000
> h_average = replicate(k, mean(c(rnorm(n1, m1, s1), rnorm(n2, m2, s2), rnorm(n3, m3, s3))))
> quantile(h_average, c(0.025, 0.975))
2.5%    97.5%
8269.641 8822.601
> summary(h_average)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 8052   8455   8551   8551   8649   9094
> hist(h_average)
```

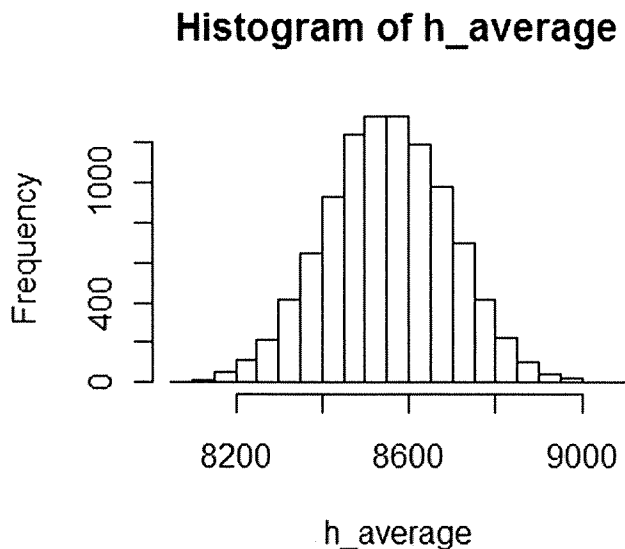


図 1.それぞれ異なる分布の 3 つの層からのランダムサンプルの平均値の分布。

平均値が 8551 円、95%信頼区間が 8270～8823 円という結果が得られた。すなわち、それぞれの医療機関から、10、15、25 例ずつのランダムサンプルを抽出して、平均値を求めることを行った場合、95%の確率で、この範囲の値が得られることが分かる。すなわち、これらのサンプルサイズの場合、実際のデータとして、3 医療機関全体の平均値は 8551 となる確率が最も高いが、もし偶然ずれたとしても、小さい場合に 8270、大きい場合に 8823 となり、これらの範囲から外れることはほとんどないと考えてよいということになる。もし、この程度の精度で十分であるといえるなら、それぞれの医療機関から、10、15、25 例ずつの調査をすれば十分であると言える。

3 つの医療機関のデータに対するサンプルサイズ算出の関数

3 つの医療機関の患者数の整数比、それぞれの

医療費の平均値と標準偏差の値から、シミュレーションに基づき、必要なサンプルサイズを算出する関数を作成した。以下の内容をテキストファイルとして、ファイル名は `samp_size_m_sd_m.R` として保存する。

```
sample.size.m.sd.m=function(m1,s1,n1,m2,s2
,n2,m3,s3,n3,E)
{
gm = (m1*n1 + m2*n2 + m3*n3)/(n1 + n2 +
n3)
lowlimit = gm - E
maxn = 1000
k = 3000
for (n in 1:maxn)
{
num1 = n1*n
num2 = n2*n
num3 = n3*n
```

```

h_average = replicate(k, mean(c(rnorm(num1,
m1, s1), rnorm(num2, m2, s2), rnorm(num3,
m3, s3))))

if (quantile(h_average,0.025)>= lowlimit)
break
}
return(n)
}

```

この関数は、医療機関1の平均値 m_1 、標準偏差 s_1 、整数比の値 n_1 、医療機関2の平均値 m_2 、標準偏差 s_2 、整数比の値 n_2 、医療機関3の平均値 m_3 、標準偏差 s_3 、整数比の値 n_3 とし、全体の平均値の最大過誤を E として代入する。返り値 n をそれぞれ n_1, n_2, n_3 に掛け算した値が、それぞれの医療機関でのサンプルサイズとなる。ランダムサンプルの平均値の計算は k で設定するが、3000回としている。計算時間を長くして精度を高めたければ、 $k = 100000$ と大きな値としてもよい。

上記の例のデータを用いた場合、Rでは次のように記述する。 E は $8551 - 8270 = 281$ として試算してみる：

```

> source("samp_size_m_sd_m.R")
> m1 = 7000
> s1 = 650
> n1 = 2
> m2 = 8000
> s2 = 1200
> n2 = 3
> m3 = 9500
> s3 = 1000
> n3 = 5

```

```

> E = 281
> sample.size.m.sd.m(m1,s1,n1,m2,s2,n2,m3,
s3,n3,E)
[1] 5

```

実際に $n = 5$ という値が得られたので、それぞれの医療機関でのサンプルサイズは、2、3、5に5を掛け算した値、すなわち、10、15、25例ずつということになる。このサンプルサイズで調査すれば、95%の確率で、全体の平均値は 8551 ± 281 に含まれるということになる。上記の例と同じ値である。

D. 考察

Rの `rnorm()` 関数を用いると、平均値と標準偏差、サンプル数を指定することによって、ランダムサンプルの抽出をシミュレートすることができる。これを `replicate()` 関数で任意の回数繰り返すことで、モンテカルロ・シミュレーションにより、サンプルサイズを算出できることを示した。

今回、3医療機関の場合の、サンプルサイズ算出関数を作成したが、関数を書き換え、これをさらに任意の個数に拡張することも可能である。

サンプルサイズの算出は調査開始前に行うものであり、それぞれの医療機関のある疾患の患者の医療費の月額あるいは年額の分布や平均値、標準偏差の情報は、少数例のデータに基づいた想定値を使用せざるを得ない。しかしながら、実際に調査が終了した時点では、調査で得られたデータに基づいて、95%信頼区間を計算することができるので、データの精度はそれで判断することができる。

また、医療費の分布は疾患によっては、正規分布に従わない場合もあることがすでに示されている。低い方の値の度数が多く、高くなるに従って、度数が小さな分布の者が示されている。このような場合は、医療費の値をログ変換などによって、変換することで、正規分布に従う可能性が高い。今後、適切な変換法についても、解析を進めたい。

E. 結論

〈参考文献〉

- 1) Chow SC, Wang H, Shao J: Sample Size Calculations in Clinical Research, Second Edition. 2007, Chapman & Hall, Boca Raton, FL, USA.
- 2) R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical

Computing, Vienna, Austria. ISBN 3-900051-07-0,
URL: <http://www.R-project.org>.

F. 健康危険情報：なし

G. 研究発表

1. 論文発表

森實敏夫：Rを用いたシミュレーションによるサンプルサイズの算出。あいまつく 2010;31(1):10-14.

2. 学会発表

なし

H. 知的財産権の出願・登録状況

1. 特許取得：なし

2. 実用新案登録：なし

3. その他：なし

Ⅲ. 株式会社 PRRIZM 報告書

難治性疾患の医療費構造に関する研究

(平成20年度データによる粗集計—基金と国保のデータを中心として)

平成21年7月19日

(株)健康保険医療情報総合研究所

I. 収集されたデータについて

II. 粗集計項目

基金

医療費の概要(医科と調剤)

外来における疾患別件数、および請求額・自己負担額・公費負担額の合計と平均
外来における疾患別各金額のパーセントイル値
外来における上位4疾患における階層区分別箱ひげ図
外来 疾患件数上位4件 都道府県別件数比較
外来における上位4疾患の平均請求金額に対する自己負担額等の割合

入院における疾患別件数、および請求額・自己負担額・公費負担額の合計と平均
入院における疾患別各金額のパーセントイル値
入院における上位4疾患における階層区分別箱ひげ図
入院 疾患件数上位4件 都道府県別件数比較
入院における上位4疾患の平均請求金額に対する自己負担額等の割合

入院/外来の疾患、階層区分別 ①件数 ②合計額 ③平均額

国保

疾患別、①件数 ②合計請求額 ③平均請求額 ④平均診療実日数

入外別件数と平均請求点数
入院/外来における上位4疾患の合計請求額の箱ひげ図
入院/外来 疾患件数上位4件 都道府県別件数比較
入院/外来における被保険者数10000人あたりのレセプト数比較
入院/外来における疾患別の診療区分別の点数の平均

施設D (省略)

階層区分別 ①件数 ②合計請求額 ③平均請求額 ④診区別平均額
階層区分別、診療実日数
階層区分別、受療率

施設E分析 (省略)

患者数(6ヶ月データで発生した患者数)
疾患別で発生した患者数(外来/入院)
受診件数(疾患別、外来/入院レセプト患者数)
受療率
外来実日数平均値
入院LOS
疾患別医療費(外来/入院)

大学間比較

クローン病
潰瘍性大腸炎

階層化の類推の妥当性検討(施設Aデータと、基金データの比較)

各データの特徴

	1.国保	2.基金	3.施設D	4.施設E	5.施設B	6.施設A	7.施設E
期間	1ヶ月 (H21年3月審査分)	3ヶ月 (H20年11月12月、 H21年1月診療分)	6ヶ月 (H20年7月～12月)	6ヶ月 (H20年7月～12月)	1ヶ月 (H20年7月)	4ヶ月 (H20年7月～10月)	4ヶ月 (H20年7月～10月)
データ一覧	・レセプトデータ	・特定疾患治療研究事業(法別51)に係る連名簿データ	・EFファイル ・公費患者一覧 ・疾患患者一覧	・特定疾患リスト			
外来／入院	外来／入院	外来／入院	外来／入院	外来／入院	外来／入院	外来／入院	外来／入院
分析対象レセプト数(外来)	192,897	209,935(医科) 352,404(調剤)	16,785	653	120	1,266	1,290
分析対象レセプト数(入院)	21,533	29,160	610	55	3	17	21
疾患分類の可否	○	○	×	○	○(クローン、潰瘍性大腸炎のみ)	○	○
患者単位の集計(患者IDの有無)	×	×	○	○	×	○	○
階層化の可否	×	○(負担額から類推)	○	×	○(類推した結果、AIに集中した為)	○	○(本院のみ)
診療区分	○	×	○	×	×	×	×
性別	×	×	×	○	×	×	×
年齢	×	×	×	○	○	○	○
都道府県	○	○	(特定疾患番号3、4 桁目)	×	×	×	×
その他		階層化について、精度を上げる為に、限度額に一致するデータのみの使用		入院日と診療日数(在院日数)が不一致、金額の合計値が不一致等のデータあり、分析対象から外した			