

平成 30 年 6 月 28 日現在

機関番号：82602

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00050

研究課題名(和文) 全国介護保険レセプトデータ・国民生活基礎調査の整備と代表性に関する研究

研究課題名(英文) Study for representativeness and maintenance of Comprehensive Survey of Living Conditions and Survey of Long-term Care Benefit Expenditures

研究代表者

高橋 秀人 (TAKAHASHI, Hideto)

国立保健医療科学院・その他部局等・統括研究官

研究者番号：80261808

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：国民生活基礎調査(大調査3年毎、中の2年は簡易調査)について、単位区の設定が国勢調査(5年毎)になっている。調査区(単位区)の大きさは50世帯程度であるが、拡大係数により結果が増幅される。そのため人口移動に伴う世帯数の変化の結果への影響が無視できないと考える。介護給付費等実態調査は、基本情報(H1)、集計情報(T1)、詳細情報(D1)、居宅サービス計画費情報(D5)、受給者台帳(M)からなる。個人×利用事業所数×月(H1)、サービス種類×利用事業所数×月(T1)、サービス番号×利用事業所数×月(D1)、個人×月(M)となる構造を基に、月単位で必要データを抽出する。

研究成果の概要(英文)：For Comprehensive Survey of Living Conditions (CSLC), large-scale survey every three years and Small-scale survey in each interim year of Large-scale survey year, survey districts were reviewed to National Census every 5 years. The size of survey districts was around 50 households and inflation factor amplified estimation error. So the influence on the results of movement of young people and single-household cannot be ignored. For Survey of Long-term Care Benefit Expenditures (SLCBE), data are provided the set which consist of basic information (H1), collective information (T1), detailed information (D1), cost plan for household services (D5), Recipient ledger (M). We extracted the data on a monthly basis by managing them with the structure, individual × the number of applied facilities × months (H1), the kind of service × the number of applied facilities × months (T1), (service id) × the number of applied facilities × months (D1), individual × months (M).

研究分野：統計学, 福祉疫学

キーワード：政府統計 代表性 国民生活基礎調査 国の検討会 集落抽出法 拡大係数 後置番号

1. 研究開始当初の背景. 統計法の改正により, 公的統計の体系的整備, 統計データの利用促進と秘密の保護, 等が定められ, 公的統計は「行政のための統計」から「社会基盤としての統計」と, その性質が大きく変化した. しかしながら, 研究者にとって「公的統計」は必ずしも使いやすい状態にはなっていない. データ申請の手続きが煩雑であるとか, データ申請から許可までの時間が長いということに加えて, データの精度, 代表性などの問題もある.

2. 研究の目的. 国民生活基礎調査(以降, 基礎調)は国民の保健, 医療, 福祉, 年金, 福祉等に関わる「国の基幹統計」である. 様々な問題を含んでいるということは知られているが, どこにどのような問題があるのかについては, あまり知られていない. 本研究では, いままでの国や研究者の著作から, 統計学的に見て, 国民生活基礎調査のどこに問題があって, 現在どのように改善が考えられているのかを整理することを目的とする. 加えて, 介護保険給付費等実態調査のデータの抽出方法についてまとめた.

3. 研究の方法. 基礎調を統計学的観点からどこにどのような問題があるのかを精査する. 検討資料として, 厚労省 電子公開資料, 岩崎報告書, 厚労省 統計委員会(人口・社会統計部会, 匿名データ部会), 橋本英樹, 厚労省報告書「平成 20-21 年度厚生統計研究事業「国民生活基礎調査の新体系の構築に関する調査研究」, 橋本英樹ら 今後の国民生活基礎調査のあり方についての一考察, 矢田晴那 政策分析ツールとしてのマイクロ・シミュレーションの研究とした.

介護給付費等実態調査については基本情報, 集計情報, 詳細情報, 居宅サービス計画費情報, 受給者台帳からなるデータを, 受給者居住地域保険者番号と被保険者番号を元に突合する過程をまとめた.

#### 4. 研究成果

[1] 国民生活基礎調査について(1) review

1 「国民生活基礎調査」についてより抜粋  
<http://www.mhlw.go.jp/toukei/list/20-21tyousa.html#anchor02>.

調査の対象は(1)大規模調査(3年に1回: 直近 H28 年), (2)簡易調査(大規模調査年以外)とも全国の世帯及び世帯員を対象. 世帯票及び健康票については平成 22 年国勢調査区(約 101 万地区)のうち後置番号 1 及び 8 から層化無作為抽出した(1)5,410(2)1106 地区内のすべての世帯(約(1)29(2)5.9 万世帯)及び世帯員(約 71(29)万人), (1のみ)介護票については前記の 5,410 地区内から層化無作為抽出した 2,446 地区内の介護保険法の要介護者及び要支援者(約 8 千人), 所得票・貯蓄票については前記の(1)5,410(2)1106 地区に設定された単位区のうち後置番号 1 から層化無作為抽出した(1)1,963(2)500 単位区内のすべての世帯(約(1)3(2)0.9 万世帯)及び世帯員(約(1)8(2.3)万人)を調査客体とした

((1)平成 28 年大規模調査)(2)簡易調査). (注: 「後置番号」とは, 国勢調査区の種類を表す番号であり, 「1」は一般調査区, 「8」はおおむね 50 人以上の単身者が居住している寄宿舍・寮等のある地域.  
<http://www.pref.tottori.lg.jp/secure/109753/yougo.pdf>.

国民生活基礎調査の統計学的検討課題について, H-1 岩崎報告書(2011) 国民生活基礎調査の標本設計・推定手法等に関する研究会  
<http://www.mhlw.go.jp/toukei/list/dl/20-21ronbun00.pdf> より抜粋. 標本設計について(1)国民生活基礎調査で標本設計上目的とする表章単位について, 大規模年については, 世帯票, 健康票は, 都道府県・指定都市別に表章し, 所得票, 貯蓄票, 介護票は全国で表章, 中間年については, 世帯票, 所得票を全国表章である. (2)集落抽出について 調査方法について, 所得票・貯蓄票と介護票の調査対象地区は重ならない. 世帯のランダムサンプリングについて, ランダムサンプリングのためには, 国勢調査地区内の新たに世帯名簿を作成する必要がある. 国勢調査地区内の世帯の名簿情報が正しく得られなければ, その名簿に基づくランダムサンプリングを行うことができず, 新たな非標本誤差が発生する. 集落抽出のメリットとデメリットについて, ランダムサンプリングに比べて標本数が同じならば母集団の推計精度が劣る. 最大のメリットは調査が効率的に行える. 集落抽出法は集落内の抽出についてランダムサンプリングを行っていないので調査方法として劣っているというような見方をされることが多いが, 全く見当違いである. 国民生活基礎調査では, 集落抽出のメリットやデメリット等を総合的に判断して, 調査方法として集落抽出が採用されている. (3)国勢調査地区の問題点について, 例えば, 調査地区内に大規模なマンションが建設された場合, 調査世帯数が大きくなるが, 調査区当たりの世帯数が安定しないのは実査上も精度の確保の観点からも好ましいことではない. 一つの国勢調査地区内の調査対象世帯数は 50 前後で安定していることが望ましいことから, 世帯数が大幅に増加した場合は, 単位区の方式を参考に 50 世帯前後に絞り込むことも検討すべき. (4)母集団の推定方法について, 世帯票, 健康票は世帯員数を補助変数とする比推定, 所得票, 貯蓄票は抽出率による比推定により母集団推定を行っている. これは調査開始から変更していないが, 近年, 回収率の低下の問題があり, 母集団推定においてもこの問題は無視出来なくなってきた. 無回答がランダムに発生していれば, 標本数の減少による精度の低下はあるものの, 母集団の推定値は偏らない. しかしながら, 都市部と地方では都市部の方が回収率は低く, また, 世帯構造別では単身の若年層の回収率が低いことなどが観測さ

れている。都市部と地方の回収率の違いについてみると、世帯票と健康票は都道府県別に比推定を行っていることから、都道府県ごとの回収率の違いについてはある程度の補整が働いている。一方、所得票と貯蓄票は回収率の影響をそのまま受け、推定値は回収率のより高い地方の影響を受けた値となっている。世帯構造別の違いについてみると、回収率の低い若年の単独世帯が少なく推定され、国勢調査との世帯数の違いの問題が表面化している。また、世帯員数を補助変量とする比推定を行う際に、国民生活基礎調査の調査対象からは施設入所者等が除かれているのにもかかわらず、人口全体に合わせにしている。従って世帯数総数は大きめの推定となっているはずであるが、若年の単身世帯の回収率が低いことによる世帯構造別の回収率の違いにより世帯数総数が少なく評価される影響の方が大きく、世帯数総数への影響はあまり問題とはなっていない。なお、高齢の女性の施設入所者が多いことから、国民生活基礎調査の性別、年齢構成と全人口の性別、年齢構成を比べると、国民生活基礎調査では高齢の女性が少なくなっている。調査結果の母集団推定における質の向上について、(6)非標本誤差の縮小として、国民生活基礎調査では、回答率の偏りによる非標本誤差が無視出来ないため、その縮小策の検討が必要である。第1の対応策は、回収率の改善である。調査環境の悪化は著しいものがあり、回収率の改善はかなり困難であるが、改善に向けて取り組んでいく必要がある。第2の対応策は、母集団推定方法の改善である。これについては、後で具体的に検討するが、無回答者の情報が何も無ければ対応は困難である。(7)施設入所者の扱いについて、国民生活基礎調査は世帯を対象とした調査であり、母集団推定の際に施設入所者の扱いを改善することが考えられる。現在は、施設入所者等の割合は少ないため、影響は一部の統計に留まっているが、今後の施設入所者等の増大を考えると対策を検討する必要がある。なお、施設入所者等の情報は、国勢調査でしか分からないため、施設入所者等の推計方法についても検討する必要があるが、実際にはかなり難しい問題である。

## H-2 橋本英樹研究班

平成 21 年度 厚生統計研究委託金 研究事業「国民生活基礎調査の新体系の構築に関する研究事業」より抜粋

H-2-1 山岡研究、国民生活基礎調査におけるサンプリングの問題点を検討する。国民生活基礎調査大調査年のサンプリングの特徴は、都道府県表章の推計誤差を一定にすることを目的として、都道府県ごとにほぼ均等の地点数を無作為抽出したうえで、地点内の世帯について悉皆調査を行うという点にある。国民生活基礎調査室に赴き、同調査を担当する職員からサンプリングの手

法・拡大乗数の計算方法などについて聞き取った。1)調査方法に関する問題点として地点の抽出、比推定について、抽出された単位区内に含まれる全世帯を対象に実施された調査の結果に基づき、各都道府県の母数推計は比推定により求められている。都道府県の下位になる地区を単純無作為抽出し、その人口(国勢調査や人口動態などに基づく)とその特性(世帯票の世帯類型など)の出現頻度を調べ、調査から得られた特性ごとの比率を人口あたりで求めて、これを都道府県人口に掛け合わせている。この場合3つの前提 ある都道府県について人口が事前に分かっている その都道府県を世帯数がほぼ等しいと思われる地区に分画できている 調査対象となった特定の地区について全数が調査できている、が必要。調査単位区内の全世帯・全数を調査できているわけではない。おそらく若年者の回答率が悪く、在宅の可能性が高い高齢者がオーバーサンプルされていると思われる。こうした未回答による非系統誤差が生じている場合、観察標本だけから得られた比率をそのまま都道府県人口に拡大してしまうので、バイアスを拡大することになる。

H-2-2 池田研究国民生活基礎調査(大規模年)の抽出方法に関する一考察では、国民生活基礎調査(大規模年)では、集落抽出法と代替法としての層化二段抽出法による結果比較した。その結果両者による違いはほとんどなかった。

H-2-3 橋本先生 厚生指標では、健康票にこころの健康度調査項目 k-6 に関する検討がなされた。

矢田晴那 政策分析ツールとしてのマイクロ・シミュレーションの研究(財務省財務総合政策研究所「フィナンシャル・レビュー」(2011.2)より抜粋)では下記の点が指摘された。データセットの補正として、「国民生活基礎調査」のデータでは、全国の世帯数を(理論的には)復元できる乗数として 拡大乗数(10 桁)が提供されている。公表されている「国民生活基礎調査」の集計値は、この乗数を使用して集計した値が主である。本稿で使用しているデータは集計客体数からさらに 3,514 世帯を除外しているが、統計の抽出方法から、拡大乗数適用後の数値の方が日本の縮図に理論上近い形であるので、拡大乗数を適用した後のデータセットを基本形とすることとした(以降、拡大乗数を使用した場合を「基本データ」、使用しない場合を「元データ」とする)。調査の回収率が 100%であった場合、結果に対して拡大乗数をかけることで、理論的には全国の世帯を集計したことになる。しかし、拡大乗数は調査客体に対する乗数であるので、回収客体数、集計客体数、モデル上の使用世帯数で計算すると結果が異なることになる。近年個人情報保護の考え方の高まりなどの影?により、回収率は年々低下しており、特に若年層と単身世帯

の(01 収率が極端に低い傾向にあるという指摘もある。拡大乗数を使用した場合、使用しない元のデータと比較してどのような影響が出るのかを見るため、可処分所得の10分位集計を行った。拡大乗数を使用した基本データでは、元データと比べて平均世帯収入と平均可処分所得が全階層で増加している。世帯収入が高い世帯では低い世帯に比べて、拡大乗数が高くなっている。所得票では、拡大係数の差が広がる。

以上から、国民生活基礎調査の(1)サンプリング「集落抽出法」については、集落抽出法自体はランダムサンプリングに比べてやや効率が下がるものの、悪いデザインではない(コストの面からは優れたデザインである)。集落抽出法は、サンプルになった地域をもって、全体を代表させる方法なので、その地域の特性に依存する結果になる。そのため、経時的な結果を解釈するとき、経時的な差なのか、サンプルになった地域差なのかについて、注意する必要がある。(2)代表性について、大規模調査 世帯票・健康票 後置番号1,8のみ(2-7が除かれる) 後置番号4?社会施設・病院(介護施設,社会福祉施設など)、後置番号4が全体の中でどの程度の割合を占めるかにも依存するが、「健康票の要介護者や患者の情報」を扱うときに代表性が崩れる可能性がある(大規模調査所得票後置番号1のみ)、どの都道府県、市町村でも集落の単位区の大きさはそろっている(50世帯程度)しかし、都道府県の代表値を作成する過程で、県の人口サイズ(世帯員数)と県内調査区総世帯員数の比(拡大係数)でもって、結果を増幅する。そのため、総世帯員数の小さい結果がより大きく反映される。そのような地域の欠損や外れ値には注意する。つまり、国民生活基礎調査(大調査3年毎、中の2年は簡易調査)について、単位区の設定が国勢調査(5年毎)になっている。調査区(単位区)の大きさは50世帯程度であるが、拡大係数により結果が増幅される。そのため人口移動に伴う世帯数の変化の結果への影響が無視できない、となる。

[2] 国民生活基礎調査について、国の検討状況、調査方法と解析方法の改善に向けての動き。国民生活基礎調査の調査方法もしくは解析方法(拡大乗数の使用法も含む)の改善に向けて、公的な委員会での議論、研究班としての調査研究、また、個々の研究者による解析上の工夫など、様々なレベルで努力がなされていることに関するレビューを行う。期間としては、第1回統計委員会(平成19年10月)以降の議事録や資料からわかる範囲とする(代表性の保たれた推定値を算出するための方法についての議論に焦点を絞り、調査項目等についての議論は取り扱わない)

1. 公的委員会、研究会等における動き。(ア)平成20年より以前の動き。統計委員会の前身

である統計審議会での議論は直接的にはホームページ上で確認できないが、平成19年以降の統計委員会、人口・社会統計部会の議論の中で、以前の情報がわかる部分もある。例えば、第34回人口・社会統計部会の議事録からわかることとして、平成12年の統計審議会において層化多段抽出の導入を検討したが、名簿作りに労力・コストがかかるため導入が困難だという理由により、翌年の答申において集落抽出が妥当であるという結論になったとのことである。また、第18回・第19回の人口・社会統計部会の議事録より、平成15年度に「国民生活基礎調査における無回答データ等の影響を考慮した調査設計に関する研究会」という研究会が開かれ、無回答データを考慮した母集団推定方法について議論されていることがわかる。この議論をもとに、世帯員名簿の精緻化を図ることを目的として、平成17-19年調査において、調査員による住民基本台帳の情報収集と近隣世帯からの無回答世帯についての聞き取り調査を行った。しかし、結果としては、調査員への負担が大きすぎる、情報にうまくアクセスできないなどの理由で正確な世帯数の推定はできなかったとのことである。平成19年国民生活基礎調査に係る統計審議会答申については第18回人口・社会統計部会の資料として確認することができる。この答申では、「本調査において都道府県別の推計人口による比推定により母集団を推定する方法が採られていることについては、より適当な推定を行う観点から、例えば、他の標本調査で採られている推定方法との比較を行うなどの研究を行うことが必要である。」との記載がある。(イ)平成22年大規模調査に向けた議論。平成22年大規模調査に向け、諮問第21号「国民生活基礎調査の変更について」によって統計委員会(第27回、平成21年10月)に諮問され、その後人口・社会統計部会に付議された(第18回、平成21年11-12月)。母集団の推定方法に関しては、上記の平成17-19年の調査などをもとに議論を行ったが、非標本誤差の縮小は当面困難であるという結論となり、回答率を上げることによる誤差の縮小を図ることとなった。第30回統計委員会(平成22年1月)にて答申案が議論され、その中で非標本誤差の縮小に向けて平成22年度内に研究会を開催することが表明された。そのほか、回収率改善にむけていくつかの方策が取られた。(ウ)平成25年大規模調査に向けた議論。平成25年大規模調査に向け、諮問第45号「国民生活基礎調査の変更について」から始まる統計委員会(第58回：平成24年10月、第61回：平成25年1月)及び人口・社会統計部会(第34回・第35回：平成24年11月、第37回：平成24年12月)で調査内容の変更についての議論が行われた。平成22年には国勢調査と国民生活基礎調査が同じ年に行われたため、国勢調査で捉え

られた世帯数に対する国民生活基礎調査の結果として公表されている推計世帯数の割合(捕捉率)を属性ごとに計算したところ、男性・若年者・低学歴の捕捉率が低いことが判明した。一方で、前回の答申を受けて、平成 22 年度に「国民生活基礎調査の標本設計・推計手法等に関する研究会」が開催され、この中で、傾向スコアを用いた解析を十数パターン行い、一部不詳データに向けた補正を行ったが、傾向スコアの使い方によって結果に差があるものがあり、どの手法がもっとも適切かを判断することはできなかった。結果、傾向スコアのどの方法にも一長一短があり、どれがベストか定まらないということで、傾向スコアでの調整は行わない方針となった。また、キャリブレーション法についても議論の俎上に載ったが、こちらはまだ研究が進んでいる段階ということで公的統計の報告に加えるのは時期尚早との判断になった。中・長期的には、非標本誤差を解析し、集計値を補正する理論の利用可能性について、引き続き取り組む必要があるとの文言が、答申の中に含まれた。また、前回の答申以降に回収率改善のために取られた対策をレビューした結果、より一層回収率を向上させるのは現行の調査内容や方法のままでは難しいとの結論となった。さらなる改善のためには、調査事項の大幅削減、郵送調査の導入、コールセンターの設置など、調査全体に関わる見直しが必要と考えられた。この見直しにむけて、平成 23 年に予定されていた(予算の関係で遂行できなかった)国民生活基礎調査の試験調査と同様の調査を、平成 26 年に行う予定となった。(工)平成 28 年大規模調査に向けた議論。平成 28 年の大規模調査に向けた調査の変更のために、諮問 82 号「国民生活基礎調査の変更について」をうけて、統計委員会(第 92 回:平成 27 年 10 月,第 94 回:平成 28 年 1 月)と人口・社会統計部会(第 62 回・第 64 回:平成 27 年 11 月,第 65 回・第 68 回:平成 27 年 12 月,第 70 回:平成 28 年 1 月)にて議論が行われた。一連の人口・社会統計部会において、非標本誤差の検討に関して非常に充実した議論が行われた。結果として、平成 29 年度をめどとして、非標本誤差に関する新たな検討会を立ち上げるなど、今後の検討・検証の工程表が提示された。美添先生が議論をリードし、今後の課題をまとめたメモなどを議事資料として提示されている。また、調査設計等に関する情報提供をより一層行なっていくべきという意見が答申に盛り込まれた。(オ)国民生活基礎調査の非標本誤差の縮小に向けた研究会における議論。前回の統計委員会答申を受け、2017 年 7 月より国民生活基礎調査の非標本誤差の縮小に向けた研究会が発足し、2018 年 1 月時点で 3 回の会合が行われている。(カ)横断的課題検討部会における議論。平成 28 年 5 月より不定期に開かれている統計委員会の

下部組織である横断的課題検討部会は、公的統計に関する横断的な課題を扱うこととなっている。その中で、統計制度に関する計画的な検査等を行うべき課題については「統計制度工場及び推計方法改善ワーキンググループ」が設置され、こちらで議論されることとなっている。「母集団情報の検証・整備」、「回収率向上方策の推進」、「欠測値、外れ値への対応」、「母集団推定における補正」、「精度検証の定期的実施と結果の公開、調査方法等改善」などの課題が挙げられており、まさに国民生活基礎調査が直面している課題と一致している。議事録で確認できる範囲では、未だ国民生活基礎調査は対象となっていない。人口・社会統計部会でも、横断的課題検討部会についての話題が挙げたが、あくまで厚生労働省として改善に取り組んだ上で、その過程で生じたデータを用いて部会・ワーキンググループにも評価を行ってもらおう、というスタンスが確認されている。

## 2. 拡大乗数の補正に関する研究者の対応

国民生活基礎調査に付されている拡大乗数については、調査票の回収率が 100%ではないため、拡大乗数を使用したとしても偏りが補正されないままになってしまう。厚生労働省が行なっている研究デザイン上での方策を講じるのとは別に、データの二次利用を進める計量経済学者等が解析を行う上で、拡大乗数を補正する試みが各研究のなかで行われている。

### [3] 韓国の 2 次データ利用の紹介

#### 1. National Health Insurance Claims Data (「健康保険標本コーホート」)

資料提供主体:韓国国民健康保険公団のビッグデータ運営室

[https://nhiss.nhis.or.kr/bd/ab/bdaba000\\_eng.do](https://nhiss.nhis.or.kr/bd/ab/bdaba000_eng.do). サンプルング基準:2002 年健康保険資格対象者(約 100 万人). 年度:2002~2013 年, 内容:「資格 DB (社会・経済的な資格の要因(障害および死亡含む))」、「診療 DB (医療利用(診療の現況))」、「健康検診 DB」、「療養機関 DB (病院の現況)」。

#### 2. 国民健康栄養調査 (Korea National Health and Nutrition Examination Survey, KNHANES, since 1998) 資料提供主体:韓国保健福祉部, 疾病管理本部

<https://knhanes.cdc.go.kr/knhanes/eng/index.do> 対象者の選定, サンプルング:第 1 期(1998)から第 3 期(2005)まで国民健康栄養調査は 3 年周期 2~3 カ月短期調査体系に実施した。第 4 期(2007~2009)から年中調査体系に改編されることによって 3 つの年度がそれぞれ独立的な 3 つの循環の標本に全国を代表する確率の見本となるように循環標本調査(Rolling Sampling Survey)方式を導入。調査期間, 国民健康栄養調査'は専門調査随行者チームが毎週(年 48 週間)4 地域(年 192 つの地域)を調査している。一地域(分担区)ごと

に3日間調査が行われているが、移動検診車両が該当地域を訪問して検診や健康アンケート調査を実施する。成人を基準にしたとき、1当たり計1時間30分~2時間がかかる。

調査内容-検診の調査(身体計り、肥満、高血圧、糖尿病、肝炎、肝機能、慢性腎臓病、貧血、重金属、閉鎖性肺疾患、虫歯、歯周疾患、視力、屈折異常、騒音露出アンケート、結核、握力、ビタミン等)、-健康アンケート調査(世帯調査、喫煙、飲酒、肥満および体重調節、罹患、医療利用、予防接種や健康検診、活動制限や生活の質、損傷(事故及び中毒)、安全意識、精神健康、女性の健康、教育や経済活動、口腔健康等)、-栄養調査(食品や栄養素摂取状況、食生活形態、食餌のサプリメント、栄養知識、食品安定性、授乳の現況、理由補充食、食品摂取頻度等)

#### [4] 介護給付費実態調査(全国介護レセプト)の整備

【データの概要】介護給付費実態調査は、介護事業所等から提出された介護給付費明細書等(介護レセプト)のコピーを、国民健康保険中央会の取りまとめのもと、厚生労働省に提出されたものである。ただし、レセプトの二次利用を拒否した保険者(通常は市町村)のデータは利用できない。なお、利用可能な市町村数は、全1,742区市町村中1,630区市町村(93.6%)である(2014年3月31日現在)。

【データの構造】データ構造は、基本情報(H1)、集計情報(T1)、詳細情報(D1)、居宅サービス計画費情報(D5)、受給者台帳(M)からなり、これらは別ファイルとして固定長のテキスト形式(ASCII)。月単位の個人集計。H1に生年月日、被保険者番号、要介護状態区分、認定有効期間、公費受給者番号、請求事業所情報等が、D1に受給したサービス番号、単位数、回数または日数、公費分等が、T1にD1からサービス種類ごとに集計された単位数、回数または日数、公費分等が記載されている。レコード単位は、(H1)原則的に個人×利用事業所数×月、(D1)サービス番号×利用事業所数×月、(T1)サービス種類×利用事業所数×月、(M)受給者台帳なので、個人×月。これらをレコード単位として、各情報に対し、月単位でファイルが提供される。記載項目の多くは金額にまつわるもの。個人属性は、(M)性、生年月、要介護度、要介護認定期間等。分析には(M)と(T1,D1,H1)とを突合が必要。介護保険受給者居住地域保険者番号と被保険者番号を連結キーとする。被保険者番号は同一保険者内でのみユニークであり、異なる保険者での同一被保険者番号は通常は別人である。つまり他の保険者地域に転出した場合追跡できない。

【突合方法】「M、D1」「M、T1」について。レコード数はD1が最多、T1は項目数が多い、ので同一情報種別かつ同一年月であっても、複

数のファイルに分割して提供される。D1、T1には過誤レセプトが含まれている可能性には要注意。変数定義書には過誤レセプトを表すフラグの存在が確認できるものの、提供を受けたデータではそのカラムはブランクとなっており、どのレセプトが過誤であるかを別のロジックで特定する必要がある。

【H1の問題】同一被保険者、同一月、同一事業所では1レコードであるが複数レコード存在したもあり利用を控えた。

【利用上の課題】1.保険者間の移動に対応困難であること。移動により、被保険者番号が変わってしまうため、データ上は別人としてカウントされてしまう。2.現状では、仮に将来NDBのような医療レセプトが利用可能になっても、介護レセプトとの突合は事実上困難である。3.所得に関連するような情報が(ほぼ)皆無であること。たとえば、被保険者が市町村で支払う保険料段階等が利用可能であれば、より多彩な分析が可能になる。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 0 件)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

#### 6. 研究組織

(1)研究代表者 高橋秀人

研究者番号：80261808

国立保健医療科学院 統括研究官

048(458)6148

〒351-0917 埼玉県和光市南 2-3-6

(2)研究分担者 田宮菜奈子

研究者番号：20236748

筑波大学医療医学系 教授

029(853)8324

〒305-8575 茨城県つくば市天王台 1-1-1

(3)研究協力者 杉山雄大

研究者番号：20725668

国立研究開発法人国立国際医療研究センター研究所 糖尿病情報センター医療政策研究室

〒162-8655 東京都新宿区戸山 1-21-1

03(3202)7181

(4)研究協力者 川村顕

研究者番号：10422198

早稲田大学 政治経済学術院

03 (3203) 6150

〒169-8050 東京都新宿区西早稲田 1-6-1