

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

報告書

平成 23 年 3 月

まえがき

平成 22 年国民生活基礎調査の企画案を審議する統計委員会人口・社会統計部会において「近年、非標本誤差を解析し、集計値を補正する理論の研究が進んできていることから、それらの利用可能性に関する検討も併せて行う必要がある」との課題を私から提示したところであります。

本研究はその課題に答えるため、「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」を設置し、石井 太（国立社会保障・人口問題研究所国際関係部第三室長）、稲葉 由之（慶應義塾大学経済学部教授）、西郷 浩（早稲田大学政治経済学術院教授）、樋田 勉（群馬大学社会情報学部准教授）の各先生をはじめ、オブザーバーとして伊藤 伸介（明海大学経済学部講師）先生にご参画いただきました。研究会では各先生方から多くの提案や知見を頂戴し、研究を推進できましたことに厚く御礼申し上げます。また、資料編に掲載している補正結果の算出に関しては、厚生労働省大臣官房統計情報部企画課審査解析室に負うところが大きく、ここに改めて深く感謝致します。

今回の研究で引き続き検討する必要があるとされた事項について、23 年度以降整理検討し、本研究をさらに推進していく次第です。

平成23年3月

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
座 長 岩 崎 学

【研究会委員名簿】

座長

岩崎 学 成蹊大学理工学部教授

委員

石井 太 国立社会保障・人口問題研究所国際関係部第三室長

稲葉 由之 慶應義塾大学経済学部教授

西郷 浩 早稲田大学政治経済学術院教授

樋田 勉 群馬大学社会情報学部准教授

オブザーバー

伊藤 伸介 明海大学経済学部講師

(五十音順、敬称略)

【研究会開催日程】

第1回 平成22年4月16日(金) 15:00～17:00

於：中央合同庁舎5号館21階2111室内会議室

第2回 平成22年6月25日(金) 16:00～18:00

於：中央合同庁舎5号館21階2111室内会議室

第3回 平成22年10月21日(木) 15:00～17:00

於：中央合同庁舎5号館21階2111室内会議室

第4回 平成22年12月10日(金) 15:30～17:30

於：中央合同庁舎5号館21階2111室内会議室

第5回 平成23年3月3日(木) 15:30～17:30

於：中央合同庁舎5号館21階2111室内会議室

目 次

1	はじめに.....	3
	(1) 国民生活基礎調査の概要.....	3
	(2) 統計委員会における標本設計・推計方法等に関する議論と本研究会の検討.....	3
2	標本設計.....	4
	(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.....	4
	(2) 集落抽出について.....	4
	(3) 代替サンプリングについて.....	7
	(4) 国勢調査地区の問題点について.....	7
	(5) 母集団の推定方法について.....	8
3	調査実施上の留意点.....	9
	(1) 記入者負担.....	9
	(2) 調査員負担.....	10
	(3) 調査費用.....	11
4	調査結果の母集団推定における質の向上.....	11
	(1) 非標本誤差の縮小.....	11
	(2) 施設入所者の扱い.....	11
5	一部不詳データの補整の考え方と試算結果.....	12
	(1) 所得票無回答世帯の総所得金額.....	12
	(2) 可処分所得不詳世帯の可処分所得.....	12
	(3) 主症状・主傷病不詳世帯の主症状・主傷病.....	13
6	全部不詳データ（無回答世帯）の補整の考え方と試算結果.....	13
7	まとめ.....	14
	【資料】.....	15

1 はじめに

(1) 国民生活基礎調査の概要

国民生活基礎調査は、保健、医療、福祉、年金、所得等の国民生活の基礎的事項を調査し、厚生労働省の所掌事務に関する政策の企画及び立案に必要な基礎資料を得るとともに、各種調査の調査客体を抽出するための親標本を設定することを目的としている。

国民生活基礎調査は、厚生行政基礎調査、国民健康調査、国民生活実態調査及び保健衛生基礎調査の4調査を統合して、昭和61年を初年とし、3年ごとに大規模な調査を実施し、その中間の各年は、簡易な調査を実施している。

(2) 統計委員会における標本設計・推計方法等に関する議論と本研究会の検討

現行の母集団推定の方法では、調査票回収率の偏りにより国勢調査との間で世帯構造別の世帯数に乖離があるが、調査票回収率の偏りに関する正確な情報を得るために直ちに採用できる効果的な手段がない現状では、母集団推定の方法の工夫によって世帯数の乖離を縮小することは当面困難であると認められたところである。このため、平成22年の大規模調査において、まず調査票回収率の向上に努めるべきとされている。

今後の課題として、調査票回収率の向上策の効果を検証する必要があるとされ、具体的には、単に回収率の増減をみるだけでなく、国勢調査も平成22年に実施されることから、両者の差異を検証する必要があるとされている。

また、近年、非標本誤差を解析し、集計値を補正する理論の研究が進んできていることから、それらの利用可能性に関する検討を行う必要があるとされたところである。本研究会ではこの検討を行うことを第1の目的としている。この補正は一部の調査票情報が得られているときに、完全に回答のあった調査票情報を利用して無回答の調査票情報を補正するものである。なお、国勢調査との乖離の主因となっている、全ての調査票情報が無回答の場合にはこの研究は利用できないが、国民生活基礎調査にとっては大きな問題であるため、本研究会ではこの対策も取り上げて検討している。

最後に、調査票回収率の向上策や近年の研究の利用可能性に関する検討が思わしい成果を上げない場合には、平成25年に実施する調査の企画に当たり、調査方法等について、見直しを検討する必要があるとされている。

2 標本設計

(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

昭和61年から3年ごとに実施されている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するものとして設計している。また、大規模年の間の年である中間年については、世帯票、所得票を全国表章するものとして設計している。

(2) 集落抽出について

① 調査方法について

国民生活基礎調査の標本設計では、発足当初から集落抽出とよばれる手法を用いている。すなわち、集落として50世帯程度からなる国勢調査地区を用い、国勢調査地区をランダムサンプリングにより抽出し、その国勢調査地区内の世帯を悉皆で調査するという手法である。集落抽出法を用いている調査としては、他には患者調査などがあげられる。患者調査では、集落として病院や診療所を用いている。

また、国民生活基礎調査では集落抽出に加えて、2段抽出という手法を用いている。1段目の抽出は世帯票と健康票の調査対象の抽出であり、2段目の抽出は所得票、貯蓄票及び介護票の調査対象の抽出である。所得票と貯蓄票では、一世帯当たりの調査員の負担が大きいことから、世帯票の調査対象となった国勢調査地区を地理的に20～30世帯程度に細分化して単位区を作成し、全単位区から調査対象となる単位区を抽出して、その単位区内の世帯を悉皆で調査している。また、介護票は世帯票の調査対象となった国勢調査地区から調査対象となる地区を抽出して、その地区内の要介護者・要支援者のいる世帯を調査の対象としている。なお、所得票・貯蓄票と介護票の調査対象地区は重ならないようにしている。

② 世帯のランダムサンプリングについて

国民生活基礎調査で国勢調査地区内の世帯を悉皆で調査するのではなくランダムサンプリングを行い調査することが考えられるが、このためには、国勢調査地区内の世帯名簿が無い場合、新たに世帯名簿を作成する必要がある。しかしながら、ランダムサン

リングのために名簿を作成するとなると、調査員の負担が大きく、実査を行うのと同様の困難さがある。また、仮に名簿が作成できたとして、ランダムサンプリングを行うこととした場合、抽出の具体的な方法も検討する必要がある。なお、調査する国勢調査地区数は増加することになるため、調査員の増加や実査の費用の増加なども考えられる。

抽出の具体的な方法については、仮に名簿作成者が抽出の作業も行うこととした場合、一定のルールを決めて抽出を行ってもらうことになるが、そのルールを前提に調査しやすい世帯が抽出されるように名簿を並べ替えて作成することが可能であることから、より厳密にランダムサンプリングを行い無作為性を確保するとすれば、名簿作成者と抽出を行う者を分ける必要がある。具体的には、全国の集落内の名簿を一旦厚生労働省に報告し、厚生労働省で抽出作業をすることになる。この手法では、調査員の拘束期間がそれだけ延び、ひいては調査員の確保の困難さが増大するとともに、費用が余計にかかることになる。一方、次善の方法として、患者調査で行っている方法を参考に、世帯主の生年月日に着目しその日付により抽出世帯を決めて調査するという手法も考えることができる。しかしながら、患者調査では、皆保険制度により患者の生年月日は正確に把握できるが、国民生活基礎調査では世帯主の生年月日を正確に把握できる保証はない。

いずれにしろ、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ほか、推計結果についての推計精度の評価もできないことになる。

③ 他調査の調査手法の導入の可能性について

国民生活基礎調査においては世帯のランダムサンプリングを行っていないが、他の世帯を対象とした統計調査でランダムサンプリングを実施している例があり、なぜ国民生活基礎調査でランダムサンプリングを行わないのかと疑問をもたれている。他調査の調査費用などの実態を把握しているわけではないが、現行の国民生活基礎調査の調査方法に比べて、調査費用の増大や調査員の長期間の拘束など招くと考えられることから、単純に、他調査の手法をそのまま導入できるわけではないと考えられる。

また、他調査で回収率を向上させるため代替サンプリングを行っている例があり、国民生活基礎調査においても回収率を向上させるため導入できないかと考えられる。ここでは名簿に限ってみても、国民生活基礎調査では国勢調査地区内を悉皆で調査しているため、代替の世帯は他の国勢調査地区から選ぶ必要がある。このためには新たな国勢調

查地区の世帯名簿作成が必要となる。また、代替の世帯を選ぶ基準をどうするか、また、その基準で選ぶための名簿が作成できるかという問題がある。ランダムサンプリングの場合以上に詳細な世帯名簿の作成が必要となるため、現行の調査期間、調査費用、調査員数などでは対応できない。その他、代替サンプリングの問題点は後で検討する。

④ 世帯名簿作成の困難さについて

国民生活基礎調査では集落抽出により調査地区内を悉皆で調査しているが、以前、調査地区内について調査に無回答の世帯も含めた名簿の作成を行った。この名簿は、世帯員数等の情報に限った基本的なものである。

具体的には、無回答の世帯について、住民基本台帳の閲覧や近隣世帯からの聞き取りなどにより作成しようとした。ところが、住民基本台帳の閲覧が個人情報保護の制約で難しかったり、閲覧できても、住所や世帯単位となっておらず世帯員全員を見つけるのが容易ではないという問題があった。また、近隣世帯からの聞き取りも、近所づきあいが希薄となってきており、必ずしも有効な回答が得られなかった。このように、調査員に負担をお願いして名簿を作成しようとしたがうまくいかなかった。

世帯名簿の作成は、新たに一つの調査を行うことに匹敵するほど困難であり、限られた予算と人員では、他の統計調査のようなランダムサンプリングはできないと考えている。

⑤ 集落抽出のメリットとデメリットについて

最後に集落抽出のメリットとデメリットについて整理することにする。

まず、デメリットであるが、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ということである。このため、国民生活基礎調査では、標本誤差を推計し、目標精度を達成していることを確認している。また、近年は調査環境の悪化により、回答率の低い国勢調査地区が増えてきており、このような調査地区ではご近所が協力しないから自分も協力しないといった対応が増えてきているのではないかという見方もある。なお、回答率の低いような調査地区は近所のつながりも希薄ではないかという考え方もあることから、どこまでデメリットとなっているかは分からない。

一方、メリットであるが、最大のメリットは調査が効率的に行えるというものである。調査員調査で、限られた時間と費用でできるだけ高い有効回答を得るためには、調査対

象が地理的にまとまっていることは効果的である。次に、名簿作成の負担が小さいこともあげられる。ランダムサンプリングはきちんとした名簿があることが大前提であるが、集落抽出の際には悉皆で抽出するので世帯名簿は必要ない。また、実査において隣近所が協力してくれているという理由で調査の協力を消極的な世帯に対しても調査の依頼がしやすいという面もある。なお、集落抽出法についても標本調査理論に基づく推計精度の評価が行える。

集落抽出は、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) 代替サンプリングについて

近年、調査環境は悪化しており、不在や拒否などにより調査に協力してもらえない世帯が増加している。このため、他の統計調査では代替サンプリングという手法がとられているものもある。この代替サンプリングでは、調査できない世帯と似た世帯をサンプリングして調査対象とする必要がある。しかしながら、国民生活基礎調査では前述のようにそもそも悉皆調査なので代替する世帯が同じ調査地区内に存在しない。また、何らかの方法で代替する世帯の候補を用意しても、その中から選ぶための名簿を作成することは容易ではないと考えられる。

また、代替世帯を選ぶ基準をどうするかという問題があり、例えば出現頻度の少ない父子世帯が無回答であった場合どのような世帯を選ぶべきかなど、世帯調査である国民生活基礎調査では、この基準をより慎重に検討する必要がある。いずれにしろ、代替サンプリングは、答えてくれやすい世帯に回答が偏る危険があることに留意する必要がある。

なお、所得票については、代替サンプリングを導入することも考えられるが、この点については後で検討する。

(4) 国勢調査地区の問題点について

国民生活基礎調査の調査地区は国勢調査地区からランダムサンプリングを行っているが、国勢調査地区は国勢調査が行われたときに更新されるため、5年間同じ情報を使う

こととなる。なお、国勢調査が行われたときから実際に使えるようになるまでのタイムラグもあるため、最大7年前の国勢調査地区情報を使うことになる。この情報の劣化の影響も考慮する必要がある。

これまでの国民生活基礎調査の調査結果では、国勢調査地区情報が更新される前後の年で極端な結果の差は観察されていないので、影響は少ないとみることはできる。しかしながら、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ではないかと考えられる。

(5) 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量とする比推定、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ている。この問題の対応策については後で検討するが、現在の推定方法でどのような影響が出ているのかを簡単に検証することとする。

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身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。

3 調査実施上の留意点

(1) 記入者負担

国民生活基礎調査は、調査項目が非常に多く記入者負担が大きいことから、標本設計を考える際にも配慮できるか検討する。記入者負担が大きいことによりサンプリングのゆがみが大きくなっているとすれば、その緩和策について検討する必要がある。

標本設計で対応できる範囲は限られているが、例えば、世帯票の全項目について現行の精度、すなわち、都道府県・指定都市別に精度を確保する必要があるのか再検討する余地はあると考えられる。全部無回答データは後で述べるように事後的に統計数理の手法では対応は困難であるが、基本的な情報の回答が得られれば、補整の余地がでてくる。世帯票の質問項目はかなり多く記入者負担が大きく回収率に影響している可能性があるため、例えば、世帯票の質問項目を都道府県別表章をするものと全国規模で表章できればよいものに整理ができれば、世帯票を2つに分割し、全国表章の項目については標本数が少なくてもいいので別の調査票にして調査対象世帯を少なくすることが考えられる。ただし、世帯票という最も基本的な調査票について調査方法を変更した場合、調査結果の継続性や後続調査への影響も考慮する必要がある。また、所得票は世帯票とのクロス集計が重要なことにも留意する必要がある。

また、現在は、介護票を導入したときの統計審議会の審議により、世帯票調査世帯の一部について調査している所得票・貯蓄票・介護票について、記入者負担が過重なものにならないようにするため、所得票・貯蓄票と介護票で対象となる国勢調査地区が重ならないようにしている。

調査地区数については、世帯票・健康票は都道府県・指定都市別表章を行うことから、都道府県別、指定都市別に一定数確保するように配分している。なお、人口規模が大きな地域、具体的には東京都と大阪府については、有限母集団修正の影響も考慮して、調査地区数に上積みを行っている。一方、所得票・貯蓄票及び介護票は標本設計上は全国

一本の表章であるため、本来は、都道府県・指定都市別には、人口比例若しくは国勢調査地区数比例とするのが望ましい。しかしながら、世帯票調査世帯の一部を調査対象としていることの制約から、人口比例では、規模の大きな都道府県・指定都市において必要な調査地区数が確保できない。このため、介護票先取りで都道府県別に同地区数調査し、残りからなるべく規模に比例したものとなるように所得票・貯蓄票の調査地区を選んでいる。結果として、記入者負担や調査員負担の重い所得票・貯蓄票が、都市部で少なめに、地方で多めに調査されている。この調査地区数のゆがみが調査結果の精度にも影響を及ぼしていると考えられる。そこで、介護票と所得票・貯蓄票の重複排除をやめ、調査地区単位では、介護票と所得票・貯蓄票の重複を認めるように改めることを考慮してはどうかと思われる。ただし、世帯単位で介護票と所得票・貯蓄票の両者の調査対象となる世帯の記入者負担には配慮が必要である。

記入者負担の軽減策を検討すると、現在、高齢化が進んでいるとはいえ、介護票の出現率は低くなっている。一方、所得票・貯蓄票は国勢調査地区内を全て調査するのではなく、いくつかの単位区に分割し、そのうちの一つの単位区を調査している。そこで、世帯単位で介護票と所得票・貯蓄票の両方の調査対象となった場合は、所得票・貯蓄票の調査対象世帯から外し、代替サンプルとして、同じ国勢調査区内の別の単位区から新たに介護票の調査対象となっていない世帯を抽出する方法が考えられる。既に世帯票と介護票の調査が行われており、世帯名簿は世帯票が回答のあった世帯については正しく記載できると考えられるので、この世帯名簿の活用により抽出は可能である。なお、要介護者のいる世帯の所得分布及び貯蓄分布が要介護者のいない世帯のものと差がある場合には代替サンプルを別集計してその影響をみるなど、代替サンプリングの影響について分析する必要がある。

別の対策としては、調査項目数を見直し、介護票と所得票・貯蓄票の両方の調査対象となった場合の記入者負担を軽減することが考えられる。

いずれにしろ、記入者負担の問題は重要であるが、調査設計を変更する場合は、テスト調査を実施して調査結果への影響を分析することや後続調査への影響を分析する等、慎重に検討する必要がある。

(2) 調査員負担

国民生活基礎調査では、集落抽出を行っており調査員負担の軽減に寄与している。し

かしながら、個人情報保護意識の高まり、共働きによる昼間の不在、オートロックマンションの増加などにより調査が難しくなっている。このような調査環境の悪化や質問項目の難しさにより調査員の確保も難しくなっている。

(3) 調査費用

国民生活基礎調査の標本設計を考える際には、調査費用も無視出来ない。現在は、集落抽出というかなり効率的な手法を用いており、調査費用を増大させず、標本設計を変えるのはかなり困難と考えられる。予算制約も年々厳しくなっており、調査費用を無視した標本設計の議論はできないことにも留意する必要がある。

4 調査結果の母集団推定における質の向上

(1) 非標本誤差の縮小

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。

(2) 施設入所者の扱い

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。

5 一部不詳データの補整の考え方と試算結果

一部不詳データの補整については、全項目について統一的な方法を適用するのは難しい。不詳データの性質を考慮しながら、個別に推定方法を検討する必要がある。

なお、全項目について一部不詳データの補整を行うのは項目数が多いため困難であることや、例えば、所得について所得の内訳ごとの補整を行うこととした場合、所得の内訳ごとの補整結果の合計が総所得の補整結果とは必ずしも一致しないと考えられることなど、一部不詳データの補整にはここで検討している事項以外にも多くの問題があるため、今回検討した手法がそのまま適用できるかを含め、慎重に検討する必要がある。

(1) 所得票無回答世帯の総所得金額

所得票調査対象世帯のうち所得票無回答世帯について世帯票の回答は得られていることから、世帯票情報を用いて総所得金額を推定できないか検討した。なお、世帯票に所得票調査対象世帯かどうかの情報がないため、国勢調査地区及び単位区についての情報により代替している。

世帯票情報は豊富にあり、その中から所得と関係がありそうな項目をできるだけ抜き出し、世帯票の回答のあった全世帯について傾向スコアと呼ばれる方法で数値化した。この数値をもとに、所得票の無回答世帯に対して数値に近い所得票の回答世帯を対応させた。最後に、対応した所得票回答世帯の世帯票情報と総所得金額の関係についてモデルを作成し、所得票無回答世帯の世帯票情報から総所得金額を推定した。

傾向スコアにより対応させる方法とモデルの作成方法により総所得の推定を何通りか試みたが、それぞれの所得分布はかなり違いがみられており、どの手法を用いるのが最も適当か等、この手法についての妥当性の評価は十分には行えなかったが、現在研究が進められている分野であり、新たな研究の成果の適用を試みるなど、引き続き検討する必要があると考えられる。

(2) 可処分所得不詳世帯の可処分所得

所得票回答世帯のうち可処分所得不詳世帯について総所得金額の回答は得られていることから、可処分所得金額を推定できないか検討した。

可処分所得が分かっている世帯について、総所得金額と可処分所得金額との関係のみ

ると強い相関がみられた。したがって、回帰分析により総所得金額から可処分所得金額を推定した結果、可処分所得金額の分布は可処分所得不詳世帯を含む場合と含まない場合で違いがみられた。

しかしながら、可処分所得の補整を検討する際には、そもそも所得票の回答率が低いことなど、所得分布に影響を与える他の要因があることや、所得分布を補整することとした場合には他の統計への影響も大きいことなど、多くの問題があるため、引き続き検討する必要がある。

(3) 主症状・主傷病不詳世帯の主症状・主傷病

健康票回答世帯のうち、主症状・主傷病不詳世帯について、気になる症状等の回答状況から主症状・主傷病を推定できないか検討した。

症状・傷病の選択の多いものから5項目を選び、その5項目の選択状況と主症状・主傷病の関係をパターン化できないか検討したが、症状・傷病の選択状況は多様であり、十分なパターン化にはならなかったため、引き続き検討する必要がある。

さらに、症状・傷病を1つしか選んでいないため、自動的に主症状・主傷病が決まるケースや、ほとんど全部の項目を選択しているケースなど一定数以上を選択しているものについての扱いについて、この手法の妥当性の評価と合わせて更に検討する必要がある。

6 全部不詳データ（無回答世帯）の補整の考え方と試算結果

全部不詳データの補整については、無回答世帯の情報が何もないことから、補整は困難である。したがって、回答世帯から母集団推定する際に、無回答世帯があることを考慮して何らかの対応をすることを検討した。具体的には、①同じ国勢調査地区内では世帯の性質が似ていると仮定、②世帯構造、世帯主の年齢別分布が国勢調査と同じと仮定、③無回答世帯の割合が、所得票の無回答世帯の割合と同じと仮定した3通りについて検討した。しかし、どれも一長一短あり、補整結果が補整しない場合より良くなったかどうかを含め、有効性が判断できなかった。平成22年については、国民生活基礎調査と国勢調査が同じ年で実施されているので、このデータを用いて、さらに検証する必要がある。

なお、世帯構造、世帯主の年齢別分布が国勢調査と同じと仮定する方法は、国民生活基礎調査と国勢調査が同じ年に実施されている場合は適用できるが、異なる年の場合、国勢調査結果をそのままでは適用できないことに注意する必要がある。

7 まとめ

標本設計、調査実施上の留意点及び母集団推定の方法については、記入者負担、費用負担及び調査員負担などの考慮すべき問題が多くあり、現行の方法は合理性があると考えられる。

しかしながら、改善の余地はあると考えられることから、厚生労働省において検討が求められる。

不詳データの補整については、全部不詳データの補整は難しいが、一部不詳データの補整については研究の余地があることが確認できた。したがって、一部不詳データの補整については、単年のデータだけでなく、いくつかの年について試算を行い、その有効性を確認していく必要がある。

なお、一部不詳データの補整について、調査結果として作成する統計表にどのように反映するかについては、厚生労働省で検討すべきものと考えている。

【資料】

1	一部不詳データの補整結果	
1-1	所得票無回答世帯の総所得金額 世帯票情報を用いた傾向スコアによる総所得金額の推定について.....	17
1-2	可処分所得不詳世帯の可処分所得 総所得金額を用いた回帰分析による可処分所得金額の推定について.....	45
1-3	主症状・主傷病不詳世帯の主症状・主傷病 性別・年齢階級別症状・傷病の選択状況を用いた 主症状・主傷病の推定について.....	50
2	全部不詳データの補整結果	
2-0	はじめに.....	80
2-1	世帯票・健康票の推定について.....	80
2-2	所得票・貯蓄票の推定について.....	91