

分担研究報告書

都道府県別指標の精度とサンプリング・統計解析手法

分担研究者 横山徹爾 国立保健医療科学院技術評価部 研究動向分析室長

研究要旨

メタボリックシンドローム関連の指標等、生活習慣病関連リスクを把握するために都道府県レベルで調査を行い経年比較や相互比較を行う際には、十分な推定精度を確保する必要がある。特に集団において栄養素等摂取量や歩行数などの個人内変動の大きな量的変数の長期間の平均的な値(習慣的摂取量等)の分布を把握するためには複数日調査を行う必要があるが、その際に複数日調査の対象をどのように選んだらよいか、モンテカルロ・シミュレーションによって検討した。①全地区の全世帯に複数日調査を行う、②全地区で複数日調査を行うが各地区では無作為抽出した一部の世帯だけで複数日調査を行う、③無作為抽出した一部の地区で全世帯に複数日調査を行う、の3パターンについて比較したところ、①が困難な場合には、③よりも②の方が推定精度が高いことが示された。この結果を考慮して都道府県民健康・栄養調査における調査設計を行うことが望まれる。

A. 研究目的

メタボリックシンドローム関連の指標を中心とした生活習慣病関連リスクを把握するための都道府県レベルでの調査としては、国民健康・栄養調査に上乗せの形で行われるいわゆる都道府県民健康・栄養調査等がある。標本誤差の大きな調査データを用いて経年変化および都道府県の相互比較を行うことは誤った解釈をすることにもなりかねないため、その推定精度については十分に検討する必要がある。集団の平均値や個人内変動の少ない比率の推定精度についてはこれまで検討されてきたが^{1), 2)}、個人内変動の大きな変数の集団レベルでの分布の推定に関する検討はなされていない。特に、国民健康・栄養調査と同様、都道府県民健康・栄養調査はほとんどの場合、食事調査が1日のみであるが、集団の野菜摂取量、脂肪エネルギー比率、他の栄養素等摂取量の分布を把握し、不足者・過剰者の割合を算出するためには1日のみの調査では不十分であることが指摘されている³⁾。同様の問題は歩行数のように個人内変動が大きな量的変数一般についても起こりえる。そこで、複数日、少なくとも2日間の食事調査等を行い、統計学的な手法を用いることで習慣摂取量等の集

団における分布を把握する必要があるが³⁾、複数日調査を行う対象をどのようにして選び、サンプルサイズをどの程度にすればよいのかはよく分かっていない。そこで本研究では、今後の調査設計に役立たせるために、実際の複数日の食事調査データに基づくコンピュータシミュレーションによって、複数日調査の対象の抽出方法およびサンプルサイズと習慣的摂取量の推定精度との関係を明らかにする。

一方、複数日調査の結果から習慣的摂取量の集団における分布を計算するためにはやや複雑な統計学的計算が要求されるため、一般の保健医療従事者には容易ではない。そこで、簡単な操作で複数日調査結果から習慣的摂取量等の分布を推定可能な日本語で操作できるアプリケーション・ソフトウェアを開発する。

B. 研究方法

都道府県民健康・栄養調査において、栄養素等摂取量のように日々の個人内変動の大きな量的変数の長期間の平均的な値(習慣的摂取量等)の分布を推定するために、どのように複数日調査を行うのが効率的であるかを考える。ここでは特に、メタボリックシンドローム関連の指標

の一つである脂肪エネルギー比率について検討する。複数日調査の日数は2日または3日とし、都道府県健康・栄養調査マニュアル³⁾に示されたように調査単位区数は50単位区とする。複数日調査を行う対象としては、

- ① 全地区の全世帯に複数日調査を行う。
- ② 全地区で複数日調査を行うが、各地区では無作為抽出した一部の世帯だけで複数日調査を行う。
- ③ 無作為抽出した一部の地区で、全世帯に複数日調査を行う。

等が考えられる。このうち、①が最も推定精度は高いと考えられるが調査の負担がかなり大きい、それにより協力率が低下するのは望ましいことではなく、②または③も検討すべきであろう。その場合、例えば、

②' 各単位区から1世帯ずつ、計50世帯

③' 5単位区の全世帯、約50世帯強

とすると、いずれも複数日調査の世帯数は50世帯程度ではほぼ同じであるが、クラスター抽出の場合には両者の推定精度はかなり異なる可能性がある。そこで、これらの方法と①の全地区の全世帯に複数日調査を行った場合とで、どの程度推定精度が異なるかを比較することにする。

また、同様にその2倍ずつで、

②" 各単位区から2世帯ずつ、計100世帯

③" 10単位区の全世帯、約100世帯強

の場合についても検討する。

推定精度の計算のためには、主任研究者の吉池らが全国の14地区で実施した国民健康・栄養調査方式による1年間12回の繰り返し調査データを元にして、以下の手順でモンテカルロ・シミュレーションを行った(図1)。

(1) 吉池らの実際の食事調査データを用い、単位区内で世帯の復元抽出を繰り返して仮想的な単位区を3000個作成し、これを「仮想母集団」とする。その際、単位区あたりの世帯数は平成16年国民・健康栄養調査のものとほぼ一致するようにした。仮想母集団において米国 National Research Council が提唱した方法³⁾(以下、ANOVA法)により習慣摂取量の分布を計算し、これを

「真の値」とする。

(2) 仮想母集団から50単位区を復元抽出し、これを「仮想県民健康・栄養調査」とする。①、②、③の方法で複数日調査を実施する世帯を決め、ANOVA法で習慣摂取量の分布を推定する。ただし、一部の世帯のみ複数日調査を行って個人間変動 σ_b^2 および個人内変動 σ_w^2 を推定するためには、次式を用いた。ここで、個体の効果の平均平方 s_b^2 、その自由度 d_b 、誤差の平均平方 s_w^2 、全自由度 d_t である。

$$\hat{\sigma}_b^2 = (s_b^2 - s_w^2) d_b / (d_t - 1), \quad \hat{\sigma}_w^2 = s_w^2$$

(3) (2)を2000回繰り返し、(1)の真の値と(2)の推定値の分布とを比較する。

C. 研究結果

表1は、複数日調査の対象の選び方によって、習慣的摂取量の脂肪エネルギー比率の分布の推定精度がどのように異なるかをモンテカルロ・シミュレーションで評価した結果の要約である。脂肪エネルギー比率(%)の個人間標準偏差の真の値は、現実に近いと思われる4.04に設定してある。上段の値は仮想県民・健康栄養調査を2000回繰り返した時に推定された「脂肪エネルギー比率(%)の個人間標準偏差」の平均とその標準偏差(=標準誤差)、誤差率(標準誤差÷平均)、中央値である。複数日調査の対象の選び方にかかわらず、平均は真の値にほぼ一致するが、標準誤差あるいは誤差率は、複数日調査を「①全単位区の全世帯」、「②' 全単位区2世帯ずつ」、「②' 全単位区1世帯ずつ」、「③' 10単位区の全世帯」、「③' 5単位区の全世帯」に行った場合の順に小さく、推定精度が高いと考えられる。②' と③'、あるいは②" と③" は複数日調査の対象世帯数の合計がほぼ同一(③の方が少し多い)であるが、全単位区から1または2世帯ずつ調査をした②'、②"の方が推定精度がはるかに良いことがわかる。②' は③"と比べて複数日調査対象の世帯数の合計は約半数であるにもかかわらず、推定精度は高い。複数日調査日数

を2日間と3日間で比べると3日間調査の方が推定精度が高く、3日間調査を全単位区から2世帯ずつ行った場合の推定精度は、2日間調査を全単位区の全世帯に行った場合と比べて大きくは劣らない。

表1の中段と下段は同様にして推定された脂肪エネルギー比率 $\geq 25\%$ および $\geq 30\%$ の割合である。やはり、推定値の平均は真の値にほぼ一致するが、その標準誤差は個人間標準偏差の推定精度がそのまま反映されたものとなっている。

図2は同様にして推定値の頻度分布を度数折れ線で表したものである。いずれも真の値を中心として左右にばらつくが、③' 5単位区の全世帯と③" 10単位区の全世帯の場合にそのバラツキが大きく、計算不能(割合=0)となることもかなりある。②' と②" のように全単位区から1~2世帯ずつ選んだ方が望ましいことがわかる。脂肪エネルギー比率 $\geq 30\%$ の割合のように真の値(5.3%)が小さい場合には、推定値のバラツキは大きくなりやすいようである。

D. 考察

脂肪エネルギー比率や各栄養素、野菜摂取量、歩数など、日々の個人間変動が大きい指標について、都道府県のような集団レベルでその頻度分布を把握して評価を行うためには、複数日調査を行って習慣的な値の頻度分布を把握する必要がある³。一方で、複数日調査を行うことは、調査の負担が大きくなるという問題点もある。全単位区の全世帯で複数日(2~3日)調査を行うことが高い推定精度を得るためには望ましいが、それが困難で一部の世帯にのみ複数日調査を行うとした場合、本研究のシミュレーション結果から、複数日調査を行う総世帯数が同じならば、一部の単位区の全世帯で複数日調査を行うのではなく、全ての単位区で一部の世帯のみに複数日調査を行った方がよいことが確認された。都道府県健康・栄養調査で複数日調査を導入する場合には、本研究結果をふまえて、望まれる推定精度と実現可能性とを考慮しつつ、調査計画を立てるとよいだろう。

一方、習慣摂取量の分布の推定を行うためのコンピュータ・ソフトウェアとしては米国のアイオワ州立大学で開発されたSIDE(SAS-IML上で動作)およびC-SIDE(X Windows上で動作)が有料で提供されているが⁴、英語版であることと、SAS-IMLまたはX Windows serverが必要とされることから、今のところ利用は必ずしも容易ではない。日本でも習慣摂取量等の分布推定を行い、食事摂取基準を用いて集団の評価を行っていくためには誰でも容易に使える簡便なPC用ソフトウェアが必要であろう。今年度は、その試作を行い、CD-ROMとして添付した。

E. 結論

都道府県民健康・栄養調査で、栄養素等の習慣的摂取量の集団における頻度分布を把握するために、調査対象のうち一部の世帯でのみ複数日調査を行う場合には、一部の単位区の全世帯で複数日調査を行うよりも、全単位区で一部の世帯に複数日調査を行った方がよい。

F. 健康危険情報

この研究において健康危険情報に該当するものはなかった。

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 参考文献

- 1). 横山徹爾. 標本抽出方法及びデータ解析手法の検討. 厚生労働科学研究費補助金健康科学総合研究事業「国民健康・栄養調査における各種指標の設定及び精度の向上に関する研究(主任研究者:吉池信男)」総括研究報告書, 2006年3月31日.
- 2). 都道府県健康・栄養調査マニュアル. 厚生労働省健康局総務課生活習慣病対策室, 平成18年6月.

- 3). Subcommittee on Criteria for Dietary Evaluation, Coordinating Committee on Evaluation of Food Consumption Surveys, Food and Nutrition Board, National Research Council 1986. Nutrient Adequacy Assessment Using Food Consumption Surveys, National Academies Press. Washington D.C.
- 4). <http://www.statlab.iastate.edu/survey/software/side.html>

1. 知的財産権の出願・登録状況

この研究において、知的財産権に該当するものはなかった。

図1. 都道府県民健康・栄養調査で複数日調査を行い習慣摂取量等の集団における分布を評価した場合の推定精度を評価するためのモンテカルロ・シミュレーションの手順

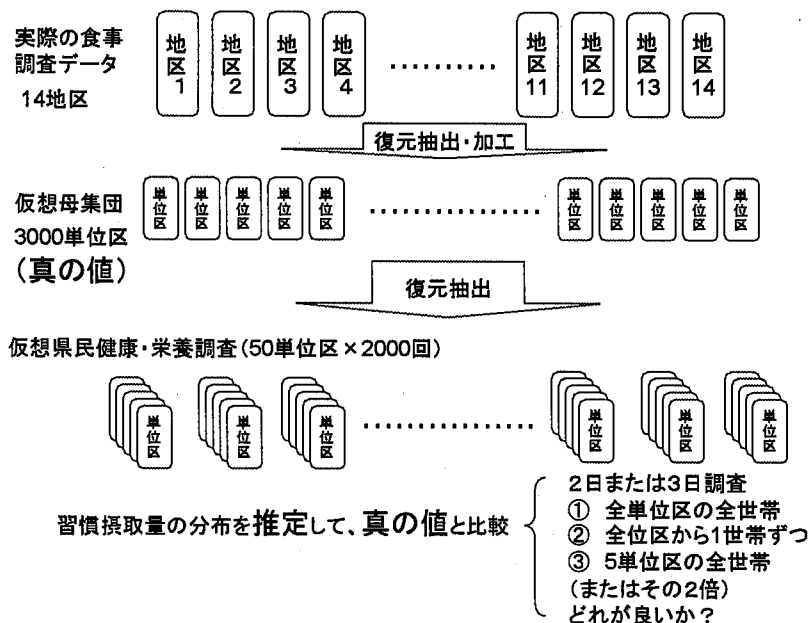


表1. 複数日調査の対象の選び方と、脂肪エネルギー比率(習慣的摂取量)の集団における分布の推定精度

複数日調査の対象※	2日間調査				3日間調査			
	平均	標準 誤差	誤差率	中央値	平均	標準 誤差	誤差率	中央値
脂肪エネルギー比率(%)の個人間標準偏差								
①全単位区の全世帯	4.01	0.28	7%	4.01	4.02	0.22	6%	4.02
②'全単位区1世帯ずつ	3.90	1.06	27%	4.11	3.96	0.73	18%	4.04
②"全単位区2世帯ずつ	3.97	0.69	17%	4.07	4.00	0.48	12%	4.03
③'5単位区の全世帯	3.85	1.63	42%	4.29	3.93	1.44	37%	4.32
③"10単位区の全世帯	3.80	1.30	34%	4.07	3.89	1.06	27%	4.09
真の値	4.04			4.04	4.04			4.04
脂肪エネルギー比率≥25%の割合								
①全単位区の全世帯	32.7%	3.5%	10.6%	32.7%	32.7%	3.2%	9.8%	32.6%
②'全単位区1世帯ずつ	31.1%	7.6%	24.3%	32.7%	32.0%	5.0%	15.8%	32.6%
②"全単位区2世帯ずつ	32.2%	4.9%	15.4%	32.7%	32.4%	3.8%	11.9%	32.6%
③'5単位区の全世帯	29.4%	11.5%	39.0%	33.1%	30.2%	10.2%	33.7%	33.2%
③"10単位区の全世帯	30.1%	9.5%	31.6%	32.6%	31.0%	7.7%	25.0%	32.6%
真の値	32.4%			32.4%	32.4%			32.4%
脂肪エネルギー比率≥30%の割合								
①全単位区の全世帯	5.3%	1.6%	29.9%	5.3%	5.4%	1.4%	26.2%	5.3%
②'全単位区1世帯ずつ	5.4%	3.3%	59.8%	5.7%	5.3%	2.7%	50.9%	5.4%
②"全単位区2世帯ずつ	5.4%	2.6%	49.1%	5.5%	5.3%	2.1%	40.4%	5.3%
③'5単位区の全世帯	6.1%	4.1%	67.8%	6.6%	6.1%	3.7%	61.5%	6.6%
③"10単位区の全世帯	5.4%	3.6%	65.5%	5.6%	5.4%	3.2%	58.9%	5.7%
真の値	5.3%			5.3%	5.3%			5.3%

※全調査単位区数は50単位区。複数日調査の対象以外の世帯は全て1日調査を行う。

図2. 全50単位区のうち複数日調査を行う世帯の選び方で比較した、習慣摂取量(脂肪エネルギー比率)の推定精度

