出生日を用いた標本抽出法についての一考察

高田（タカダ） 崇司（タカシ）*1　石井（イシイ） 太（フトシ）*2

Ⅰ はじめに

厚生労働省では，医療施設や社会福祉施設など厚生労働行政に関係する様々な施設に対する統計調査を行っている。施設の基本的な情報については全数調査が行われることが多いが，施設の利用者などについては，全数調査ではなく，対象から一定の大きさの標本を抽出して行う標本調査が多い。この標本抽出に当たっては，まず（一定数の）施設を標本抽出した後，客体となった施設における利用者などからさらに標本抽出して調査を行う，二段抽出法がよく用いられる。このとき，施設の標本抽出は全数調査により作成された名簿があるため，これを用いて厚生労働省側で標本抽出を行うことが可能であるが，一般的に施設の利用者などは調査時点での対象者が事前に把握できないため，各施設において標本抽出を行う必要が生じる。これを行うためには，最も簡易な系統抽出法の場合でも，利用者等の名簿を整備して一定間隔で客体を抽出するなどの手間が必要になるとともに，実務的に複雑な作業を行うことから生じるミスなどによる標本の無作為性のクオリティ低下が起きる危険性がないとはいえない。

そこで，厚生労働省が実施する標本調査では，施設において利用者などを標本抽出する際，「出生日が奇数の利用者のみを客体とする」など出生日の特性を利用して標本抽出を行うという標本抽出法が採られているものがいくつかある（患者調査，社会福祉施設等調査，介護サービス施設・事業所調査，地域児童福祉事業等調査など）。このような方法を採ることにより，施設などの現場でも比較的容易に標本抽出を行うことが可能になるとともに，（後述するように，一定の条件の下で）標本の無作為性についても一定のクオリティが担保されることとなる。

ところで，標本調査には抽出された標本が全体とは異なることから生じる標本誤差があり，この標本誤差を一定の精度に管理する標本設計が必須のものとなる[1)]。「統計行政の新たな展開方向」（平成15年６月各府省統計主管部局長等会議申合せ）[2)]の中でも，指定統計については達成誤差などの誤差情報を提供していくこととされたほか，既に情報提供している統計調査についても「その内容の充実を図ることとし，承認統計や届出統計についても指定統計に準じて情報提供を図ること」とされており，すべての官庁統計について，標本調査における誤差情報提供の一層の充実は，まさに必須の重要課題である。

さて，出生日を利用した標本抽出法の理論的整理を試みようとすると次のような問題があることに気がつく。すなわち，ある調査日における利用者の出生日は，母集団において既に確定しているのであるから，施設を抽出すると同時に客体となる利用者も決定しており，利用者を標本抽出することによる確率的な変動はない。したがって，通常，誤差情報として提供を行っているsampling designによる標本誤差は，二段

＊１厚生労働省労働基準局勤労者生活部勤労者生活課（前同省大臣官房統計情報部企画課審査解析室総合解析係）
＊２国立社会保障・人口問題研究所企画部第四室長（前厚生労働省大臣官房統計情報部企画課審査解析室長補佐）

目の標本抽出については考えられないのではないかという問題である。この点については，平成17年患者調査（指定統計第66号）計画案の審議の場においても，「標本設計の面で，最初から生年月日の末尾でもって配り分けられるべき調査票というのが分かれてしまっているという形になっています。だから，ランダムな過程が入っていないので，(中略)たとえ全数調査をしたとしても，簡単な調査票を配った方の人については詳しい情報はわからないという形になっているわけです。(中略)ただ，恐らく調査されている項目と生年月日との間にはあまり関係はないであろうという大きな前提条件があって，その条件の下では，このように調査したとしても標本誤差の評価というのが可能になって，多分，そういう整理になると思われます」との問題提起がされている[3)]。このように，出生日を利用した標本抽出は，通常の標本抽出とは理論的に異なった側面をもっていると考えられるが，この場合の推定量やその精度に関し，標本調査論における理論的な位置づけと，これら厚生労働省の実際の調査を直接的に関連づけて整理を行った論文はあまり多くない。本稿は，厚生労働省で実際に行われている調査に近い例を用いて，出生日を利用した標本抽出法の理論的な位置づけの整理を試みるとともに，具体的な数値シミュレーションによる評価を行ったものである。

Ⅱ 方　　法

出生日を利用した標本抽出法の理論的整理に当たり，次のような例を用いて考える。施設の利用者に対して，あるサービスに関する１年の延べ利用回数を調査することを目的とした標本調査を行うこととする。調査対象となる施設は全体で50施設であり，各施設には表Ⅰで示されるような利用者数があったとする。また，表Ⅰには各利用者の１年におけるあるサービスの延べ利用回数が各利用者の属性値として示されている。これから母集団全体の総延べ利用回数は28,606回となるが，標本調査を行うことによりこの総延べ利用回数を推定する問題を考える。

標本抽出と総延べ利用回数の推定は，実際の厚生統計にあわせ，次のように行うこととする。

1)　まず調査客体とする施設を単純無作為抽出により標本抽出する。

2)　抽出された施設の利用者のうち，一定の確率π(以下，単純無作為抽出の場合と同様，「第２次抽出率」という）に応じて出生日が一定の基準を満たす者を標本とし，その延べ利用回数を調査して推定を行う（例えば，出生日が奇数日であるものを抽出する場合はπが約1/2と考えられる)。

一般的な二段抽出法では，1)のプロセスに引き続き，あらかじめ定められた抽出率で各施設から利用者を抽出し，それらを標本として調査を行うわけであるが，この出生日を利用した標本抽出法が一般的な二段抽出法と異なるのは，利用者の出生日が一定の基準を満たすかどうかは母集団において既に確定しており，一般的な二段抽出法のように利用者の抽出という確率的プロセスが一見組み込まれていないようにみられる点である。

このような標本抽出を理論的にとらえるためには，各利用者の出生日が一定の基準を満たすかどうかは確率πで定まるという確率的プロセスに従っているものとし，それらは施設の抽出とは独立であると仮定する。この場合，施設の抽出を行うときの母集団は一般的な標本抽出法で考えるような確定的なものではなく，利用者の出生日により変動する確率的なものであるということになる。

さて，推定式などを定式化するため，次のような記号を導入する。

$(i,\ j)$：施設iにおける利用者j
　$(i=1,\cdots,M\ ;\ j=1,\cdots,N_i)$

X_{ij}：母集団における利用者$(i,\ j)$の延べ利用回数

$$Y_{ij}=\begin{cases}1 & \text{：母集団における利用者}(i,j)\text{の出生日が基準を満たすかどうか}\\ 0 & \end{cases}$$

一段目の抽出では，M(＝50)施設からm施設を単純無作為抽出法により抽出する。このとき，標本として抽出された施設の中で，$y_{ij}=1$である

利用者に対して調査を行うこととする。各施設では，N_i人の利用者のうち，

$$n_i=\sum_{j=1}^{Ni} y_{ij}$$

を満たすn_i人ずつの利用者に対して調査を行うこととなるが，これらn_i人に対してあらためてindexを振り，抽出された利用者の述べ利用回数を，$x_{ij}(i=1,\cdots,m\ ;\ j=1,\cdots,n_i)$と表すこととする。母集団の総延べ利用回数は，

$$T_X=\sum_{i=1}^{M}\sum_{j=1}^{Ni} X_{ij}=\sum_{i=1}^{M} T_{Xi}$$

であるから，これを標本から推定することが筆者らの命題となる。さて，出生日と施設の抽出は独立と仮定したことから，まず抽出された施設に注目し，施設内の利用者の出生日が確率的に変動すると考えて施設ごとの総延べ利用回数を推定し，その後にこれを用いて施設全体での総延べ利用回数を推定すると考えてもよい。そこで，抽出されたある施設における二段目の抽

表1　シミュレーションに用いた母集団

施設番号	利用者数	各利用者の1年におけるあるサービスの延べ利用回数																											
	(人)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	11	37	48	18	36	29	23	48	45	30	29	14																		
2	15	47	12	41	30	10	20	19	22	29	29	10	49	39	11	27														
3	27	36	20	18	12	37	40	23	14	28	40	24	13	34	26	30	36	14	26	18	43	13	22	22	21	29	11	34		
4	23	13	18	37	27	31	28	29	36	21	21	13	46	40	27	11	49	43	27	36	35	36	44	46						
5	12	27	42	10	42	49	39	36	41	28	23	17	34																	
6	30	16	45	42	21	36	14	22	25	41	38	40	30	20	45	33	15	23	21	10	30	36	12	27	11	37	25	31	40	45	11
7	12	12	11	36	19	25	37	14	43	47	18	25	40																	
8	29	13	19	13	31	20	25	48	22	17	18	42	43	31	32	42	10	39	48	43	28	11	29	21	24	43	38	46	39	21	
9	21	12	14	36	12	18	25	12	32	46	47	30	12	12	40	26	27	45	40	25	47	28									
10	10	18	45	19	29	29	13	42	26	49	13																			
11	25	30	29	25	18	25	28	11	38	36	21	37	37	21	14	27	14	22	31	47	22	40	20	45	33	42				
12	16	34	35	16	48	41	19	16	42	10	34	16	36	12	48	42	40													
13	14	10	34	34	31	23	39	16	22	13	41	31	46	25	41															
14	21	45	46	23	16	24	49	45	32	20	22	17	16	23	32	10	13	34	47	35	13	11								
15	16	28	43	32	44	33	10	36	44	49	46	43	10	39	19	28	29													
16	23	29	40	44	35	16	30	29	30	40	23	21	26	26	37	32	41	38	33	43	30	22	44	10						
17	15	12	38	42	46	21	26	42	48	22	15	26	12	34	31	31														
18	17	47	27	46	13	28	28	22	37	27	35	47	35	27	24	35	17	18												
19	23	46	37	16	47	48	13	34	11	34	34	44	12	27	33	41	49	47	36	22	32	13	42	17						
20	27	13	23	42	10	22	39	37	30	37	18	34	49	42	23	47	17	14	17	31	17	40	26	28	41	10	39	36		
21	11	49	28	31	12	37	47	38	49	40	22	45																		
22	16	36	23	43	39	14	22	28	37	13	10	33	41	34	25	13	42													
23	21	23	14	25	36	32	17	12	14	17	39	25	30	14	46	26	13	38	19	13	20	43								
24	10	15	40	22	39	36	45	18	27	29	46																			
25	20	45	23	18	32	29	11	47	21	12	25	44	44	30	42	29	28	35	49	39	42									
26	10	32	29	32	35	27	10	36	25	20	38																			
27	11	31	16	48	36	38	47	49	39	20	11	38																		
28	29	41	34	39	30	42	30	25	42	34	23	26	27	36	31	49	26	13	31	30	36	13	32	39	22	26	12	46	47	23
29	11	17	18	22	32	29	28	22	38	11	13	45																		
30	17	41	33	23	36	12	16	38	29	14	42	36	20	17	33	31	17	41												
31	27	43	45	14	28	39	38	16	44	11	25	35	18	12	29	15	36	10	48	10	23	27	39	47	19	10	19	41		
32	21	38	35	30	46	47	32	20	23	40	49	28	31	44	42	49	22	46	36	28	45	24								
33	20	16	38	21	14	25	47	11	34	38	36	12	21	40	30	39	49	40	28	23	47									
34	10	46	25	38	37	22	33	14	31	43	48																			
35	18	24	13	31	48	19	12	12	20	14	26	25	23	43	33	10	28	47	15											
36	29	48	45	47	34	33	23	49	29	12	36	42	37	35	44	45	48	38	15	36	44	29	33	25	41	34	36	46	20	24
37	18	20	42	31	38	47	45	27	10	20	34	28	35	33	18	11	12	12	37											
38	20	15	17	36	23	16	36	43	16	45	46	11	11	48	45	43	36	22	31	10	49									
39	15	13	19	11	45	39	30	41	11	19	17	29	14	41	29	46														
40	21	11	33	37	45	40	45	41	45	11	23	33	13	39	35	48	31	49	18	21	34	13								
41	29	28	20	35	15	48	10	11	26	13	16	22	32	42	31	37	25	46	41	34	37	12	31	47	22	36	30	46	38	29
42	19	17	43	29	43	26	33	20	44	32	34	44	15	39	21	37	30	19	13	25										
43	21	15	28	34	39	41	39	19	37	32	39	32	47	31	48	36	25	48	29	49	17	33								
44	26	45	15	39	16	39	17	17	23	44	12	11	38	22	21	49	16	18	47	16	28	16	43	26	10	37	49			
45	21	12	10	49	21	38	41	11	30	23	20	38	28	44	46	27	28	32	41	45	12	18								
46	11	44	10	47	48	31	22	18	48	34	23	38																		
47	19	42	17	30	45	33	15	10	34	47	34	16	29	28	46	48	18	22	32	37										
48	24	22	17	30	16	16	26	48	38	11	22	48	49	39	44	32	31	17	35	20	43	25	42	25	36					
49	17	38	41	10	28	19	31	15	22	29	39	42	37	28	42	44	46	39												
50	26	21	38	36	48	32	24	33	33	33	44	32	39	20	38	37	30	34	44	10	26	29	37	18	48	33	24			

出を考えてみよう。通常，二段目の抽出を単純無作為抽出で行う際に，施設全体の総延べ利用回数の推定量は，各施設においてN_i人からn_i人が抽出される場合，

$$\hat{T}_{Xi}^{(1)}=\frac{N_i}{n_i}\sum_{j=1}^{n_i}x_{ij}$$

という推定量（以下「推定量(1)」）を用いる。ところが，出生日を利用した抽出においては抽出される人数n_iは確率変数であり，各利用者が出生日に応じて確率πで抽出される。このような標本抽出法をBernoulli抽出と呼ぶが，この場合，もう１つの推定量の考え方として，各個体が抽出される確率の逆数を乗じて推定を行うものがある。すなわち，

$$\hat{T}_{Xi}^{(2)}=\frac{1}{\pi}\sum_{j=1}^{n_i}x_{ij}$$

を当該施設の総延べ利用回数の推定量（以下「推定量(2)」）とする考え方である。これは一般にHorvitz-Thompson推定量と呼ばれる不偏推定量であるが，事後的に定まったN_i, n_iを用いて単純無作為抽出を行ったときと形式的に同じ形の推定量である$\hat{T}_{Xi}^{(1)}$の方がよい推定量となっている。これは，Särndalら[4]が示しているように，各推定量の分散が一定の前提の下で近似的に，

$$V_{BE}(\hat{T}_{Xi}^{(1)})\simeq N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\left(1+\frac{1}{n_i}\right)\sigma_{Xi}^2$$

$$V_{BE}(\hat{T}_{Xi}^{(2)})\simeq N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\left\{1-\frac{1}{N_i}+\left(\frac{\bar{x}_i^2}{\sigma_{Xi}^2}\right)\right\}\sigma_{Xi}^2$$

（ただし，$n_i=\pi N_i$）

と表され，N_i，n_iがある程度大きい場合には

$1-\frac{1}{N_i}\simeq 1$，$1+\frac{1}{n_i}\simeq 1$と考えられるため，

$V_{BE}(\hat{T}_{Xi}^{(1)})<V_{BE}(\hat{T}_{Xi}^{(2)})$となることによる。したがって，このような標本設計の場合，実際の厚生統計においても単純無作為抽出を行った場合と形式的に同じ形の推定量(1)が用いられているものの，これは単純無作為抽出とは理論的な位置づけはやや異なるものなのである。

では，このような標本抽出法は，実際の標本設計でよく考えられているように$n_i=\pi N_i$として客体となる利用者数を固定した単純無作為抽出法とどの程度の違いがあるものなのだろうか。この場合の推定量の分散は，単純無作為抽出法の公式から，

$$V_{SR}(\hat{T}_{Xi}^{(1)})\simeq N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\sigma_{Xi}^2$$

となることから，単純無作為抽出の場合の分散に対するBernoulli抽出の場合の分散の比は，各施設において客体として抽出される利用者数の期待値n_iを用いて，$1+\frac{1}{n_i}$となっていることがわかる。したがって，各施設における利用者数が一定規模以上の大きさがあり，かつ，出生日による抽出確率がある程度大きく各施設の客体数に一定の大きさが確保されていれば，推定量の精度はほぼ同じと考えてよい。これは第２次抽出に係る分散であるから，第１次抽出に係る分散が大きい場合には両者の差はより小さいものとなる。しかしながら，記入者の負担軽減などの観点から１施設当たりの客体数を減らすことが行われることがあり，実際に平成17年患者調査では，一部の施設において第１次抽出を悉皆にするとともに第２次抽出率を下げるという標本設計の改定が行われた。同調査では，病床数の多い，１施設当たり利用者数が比較的大きい施設に対して改定を行ったことから，影響は大きくないものと推察されるが，１施設当たり利用者数の規模が小さいような調査対象について同様に第1次抽出率を上げ，第２次抽出率を下げた場合，推定量の精度に影響を及ぼす可能性がある。そこで，後述のⅢにおいて，数値シミュレーションに基づいてこの影響の大きさを評価してみることとする。

そして，両者の違いについてさらに注意すべき重要な点がある。実際の標本調査においては母集団の特性値を知ることができないため，標本誤差の評価は，実際の調査において抽出された標本から母集団の分散を推定することによって行われる。報告書に掲載される達成精度の評価結果や，新たな調査の標本設計を行う場合に用いられるのはこの推定値であるため，分散の推定精度が低くなるとこれらに影響を及ぼす可能性がある。Bernoulli抽出による場合でも，母

集団の分散の推定値は形式的に単純無作為抽出法と同じ形の推定量により不偏推定量が得られる。しかしながら，この分散の推定量は標本を用いた推定であり，抽出された標本に応じて変動するが，Bernoulli抽出による場合，この分散の推定精度にも影響が出ることがある。これは，利用者数が固定されている単純無作為抽出に比べ，Bernoulli抽出では利用者数が確率変数となっていることから分散の推定にも影響が及ぶことによるものであるが，この影響についてもⅢにおいて数値シミュレーションで評価することとする。

Ⅲ　結果と考察

(Ⅰ)　総延べ利用回数推定量の比較

Ⅱで述べた方法に基づき，表１の母集団から標本抽出のシミュレーションを10,000回行って総延べ利用回数を推定し，推定量の分散・標準誤差(率)を求めたものが表２である（ただし，各施設において客体が０または１の場合は，推定量と推定分散が算定できないため除外している)。それぞれ，Ⅱで示した理論的な近似式による値も併せて示している。

まず，ケース①「第１次抽出率　100%，第２次抽出率　1/2」の場合を見てみよう。シミュレーション結果による分散を比較してみると，Bernoulli抽出の場合，推定量⑵の分散は推定量⑴の約７倍となっており，圧倒的に推定量⑴の精度が良いことがわかる。ケース②「第１次抽出率60%，第２次抽出率　1/2」の場合では約２倍までその比は縮まるもののやはり推定量⑴の精度が良い。このように，Bernoulli抽出の下で考えても形式的に単純無作為抽出と同じ形をした推定量⑴を使うべきであることがわかる。しかしながらこのことは，この標本設計を $n_i = \pi N_i$ の単純無作為抽出とみなしてよいと考えることと必ずしも同じではない。そこで，次に，推定量⑴を用いる場合，$n_i = \pi N_i$ の単純無作為抽出と精度がどの程度違うかを比較してみよう。ケース①では，Bernoulli抽出の場合の分散は単純無作為抽出の場合に比べて１割程度(９%)大きくなっていることがわかる。一方，ケース②では，第１次抽出に係る分散があるため，両者の分散の大きさはほぼ同程度となっている。次に，表の下側にある第２次抽出率を1/3に下げたケース③，④を見てみると，ケース③「第１次抽出率100%，第２次抽出率　1/3」の場合では，Bernoulli抽出の場合の分散は単純無作為抽出の場合の分散に比べて15%程度大きく，ケース④「第１次抽出率　60%，第２次抽出率　1/3」の場合でも４%程度大きくなっていることがわかる。

表２　推定量の分散・標準誤差(率)のシミュレーション結果

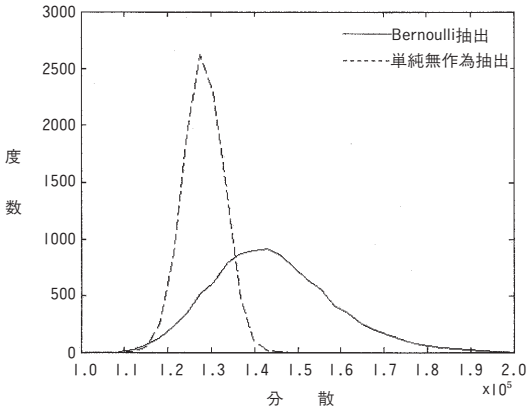
		ケース①(第1次抽出率 100%, 第2次抽出率 1/2)			ケース②(第1次抽出率 60%, 第2次抽出率 1/2)		
		分散	標準誤差	標準誤差率(%)	分散	標準誤差	標準誤差率(%)
Bernoulli抽出	推定量(1)	141 223	376	1.31	1 386 969	1 178	4.12
	(理論値)	(141 990)	(377)	(1.32)	(1 401 017)	(1 184)	(4.14)
	推定量(2)	971 812	986	3.45	2 803 106	1 674	5.85
	(理論値)	(985 714)	(993)	(3.47)	(2 807 223)	(1 675)	(5.86)
単純無作為抽出	推定量(1)	129 581	360	1.26	1 381 808	1 176	4.11
	(理論値)	(128 224)	(358)	(1.25)	(1 378 074)	(1 174)	(4.10)
		ケース③(第1次抽出率 100%, 第2次抽出率 1/3)			ケース④(第1次抽出率 60%, 第2次抽出率 1/3)		
		分散	標準誤差	標準誤差率(%)	分散	標準誤差	標準誤差率(%)
Bernoulli抽出	推定量(1)	299 201	547	1.91	1 650 530	1 285	4.49
	(理論値)	(297 577)	(546)	(1.91)	(1 660 328)	(1 289)	(4.50)
単純無作為抽出	推定量(1)	260 129	510	1.78	1 582 582	1 258	4.40
	(理論値)	(253 046)	(503)	(1.76)	(1 586 109)	(1 259)	(4.40)

このように，第１次抽出率が低く，第２次抽出率が高いケースではBernoulli抽出を行っても，標本設計上，$n_i=\pi N_i$ の単純無作為抽出とほぼ同様であると考えることができるが，第１次抽出率が悉皆になっているなど高い場合や，第１次抽出率がある程度低くても１施設当たりの利用者数の規模が小さい，あるいは第２次抽出率が低いなどにより，各施設の客体となる利用者数が小さい場合には実際に抽出された標本の大きさが $n_i=\pi N_i$ からずれることが多くなるために，推定量の精度が低下することがあるので注意が必要となる。

（２） 推定分散の精度の比較

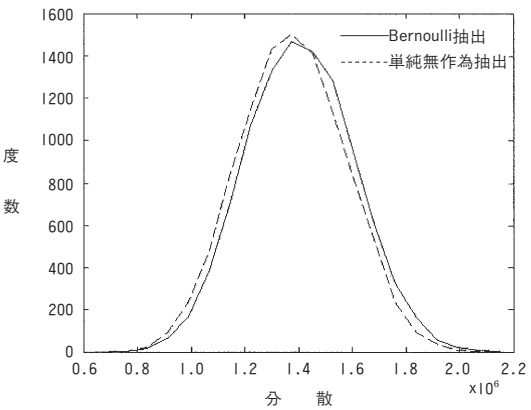
図１～４は，表２で行った４つのケースに対して，標本から推定される推定量(1)を用いた場合の分散の分布のシミュレーション結果を示したものである。これを見ると，ケース①，③の第１次抽出が悉皆のケースでは，Bernoulli抽出による分散の推定値の分布の方が，単純無作為抽出のものに比べ大きく広がっていることがわかる。ケース②，④の第１次抽出率を60％に下げたケースでは，ケース①，③ほどの違いではないが，やはり第２次抽出率の低いケース④でやや分散の推定精度が低下していることがわかる。このように，Bernoulli抽出を用いた場合には，$n_i=\pi N_i$ の単純無作為抽出の場合と比べて，推定量の精度の低下だけではなく，分散の推定精度が低下することにも注意が必要となる。特に，第１次抽出を悉皆で行う場合には $n_i=\pi N_i$ の単純無作為抽出法の場合に比べて分散の推定精度が低くなっていることから，達成精度評価さ

図１　推定分散の分布（ケース①）



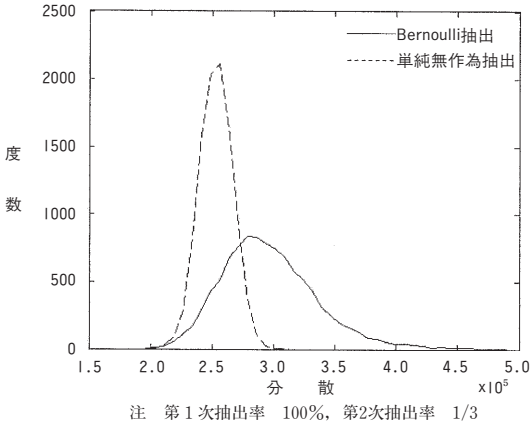
注　第１次抽出率　100％，第2次抽出率　1/2

図２　推定分散の分布（ケース②）



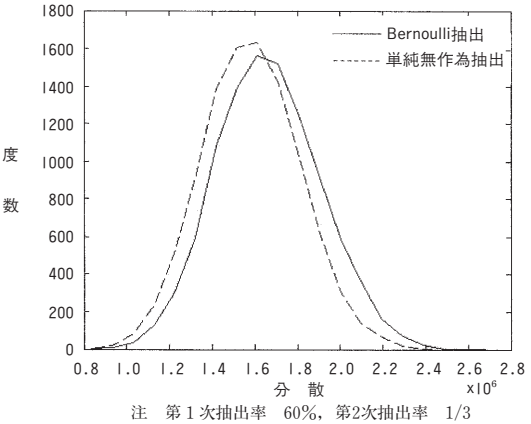
注　第１次抽出率　60％，第2次抽出率　1/2

図３　推定分散の分布（ケース③）



注　第１次抽出率　100％，第2次抽出率　1/3

図４　推定分散の分布（ケース④）



注　第１次抽出率　60％，第2次抽出率　1/3

れた標準誤差(率)や，これを利用して標本設計を行う場合にはこの点に十分留意する必要がある。

Ⅳ おわりに

出生日を利用した標本抽出法は，実務的な簡便性などもあり，厚生統計で幅広く用いられてきた。そして，通常，$n_i=\pi N_i$ の単純無作為抽出を行うのとほぼ同等の精度を得られるという認識で標本設計などが考えられることが多かった。本稿で示したように，各施設において一定規模の客体数が確保されているときにはこれらの事実は理論的にも妥当である一方，第１次抽出率が高く，各施設の客体数が小さい場合には $n_i=\pi N_i$ の単純無作為抽出法に比べ精度が低下し，さらに達成精度評価などに用いられる分散の推定精度が低下することから，その利用に当たっては十分な注意が必要であることがわかった。近年，統計調査に対する被調査者の負担軽減が求められてきているが，その対応策の１つとして，標本誤差に関する情報を活用して標本設計に工夫をすることが考えられる。このようなものとして，平成17年患者調査で行われたように，対象施設を増やす一方で１施設当たりの標本抽出率を下げるような標本設計の変更を考えなければならないケースは今後もあり得よう。しかしながら，各施設の利用者規模などが小さい調査対象に対してこのような標本設計を適用する場合には，本稿で考察を行った推定量の精度や分散の推定精度の問題も考慮しつつ，十分な検討を行うことが必要と言える。

謝辞

本研究に関連し，平成17年患者調査の標本設計について貴重なご助言，ご協力を賜った社会保障審議会統計分科会委員の西郷浩早稲田大学教授に対し，感謝の意を表します。

文　献

1）村山令二, 鈴木健二, 石井太, 他. よくわかる標本調査法. 東京：厚生統計協会, 2004；32-139.
2）総務省統計局統計基準部. 統計行政の新たな展開方向. 東京：全国統計協会連合会, 2004.
3）第７回社会保障審議会統計分科会(平成16年10月14日(木)). 厚生労働省ホームページ(http://www.mhlw.go.jp/shingi/2004/10/s1014-14. html).
4）Särndal C, Swenson B, Wretman J. Model Assisted Survey Sampling. New York：Springer, 1992.