

表8(つづき) 脱落の有無により変数がとる値及び第1回調査との有意差の有無

集計対象	第1回		第2回		第3回	
	回答者	脱落者	回答者	脱落者	回答者	脱落者
除いた対象	なし					
父の育児(入浴させる) 4段階(1~4)	1.81	1.80	1.80	1.80*	1.80	1.80*
父の家事(部屋等の掃除をする) 4段階(1~4)	3.06	3.06	3.06	3.05	3.06	3.06
父の家事(日常の買い物をする) 4段階(1~4)	2.57	2.57	2.57	2.58	2.57	2.57
子育てで意識(よい言葉をかせる)	18.8	18.9	18.9	18.6	18.8	18.8*
子を持ってよかったこと(身近な人が喜んでくれた)	78.1	78.4	78.4	78.2	78.5	78.1
子を持ってよかったこと(上の子に弟・妹ができた)	46.1	46.2	46.2	47.1***	46.2	46.2
子を持って負担に思うこと(子育てによる自分の視界が広がった)	43.9	44.0	44.0	43.6	44.0	43.2*
子を持って負担に思うこと(子育てで自分の疲れが大きい)	39.5	39.8	39.8	39.6	39.8	39.5
子を持って負担に思うこと(子育てで出費がかさむ)	34.7	34.4	34.4	34.7	34.2*	34.7
子を持って負担に思うこと(自分の自由な時間が持たない)	55.2	55.6	55.6	55.4	55.7*	55.8**
子育ての不安や悩みを相談する人	99.0	99.1	99.1	99.1	99.1*	99.2*
子育ての相談相手(配偶者)	81.5	82.4***	82.4***	82.4***	82.5***	82.7***
子育ての相談相手(自分の両親)	72.3	72.6	72.6	72.4	72.8*	72.9**
子育ての相談相手(配偶者の両親)	30.3	30.5	30.5	30.8*	30.7*	30.8*
子育ての相談相手(友人・知人)	70.5	70.9	70.9	70.9	71.0*	71.1**
子育ての相談相手(保健師)	14.2	14.4	14.4	14.2	14.6*	14.4
出産1年前の父の職の有無	98.3	98.5*	98.5*	98.6***	98.5**	98.6***
月齢6ヶ月時の父の職の有無	25.3	25.1	25.1	25.9**	25.3	26.7***
月齢6ヶ月時の母の職の有無	98.3	98.4	98.4	98.5*	98.5*	98.5***
母の労働時間 5段階(1~5)	1.27	1.26	1.26	1.28	1.27	1.28***
母の通勤時間 5段階(1~5)	1.91	1.90	1.90	1.89	1.90	1.88*
父の労働時間 5段階(1~5)	4.02	4.03*	4.03*	4.02	4.03**	4.03
父の通勤時間 5段階(1~5)	2.51	2.51	2.51	2.51	2.51	2.50*
母の就労収入(有無)	50.7	50.9	50.9	51.1	51.0	51.5**
母の就労収入(金額:万円)	99.4	101.1*	101.1*	101.2*	101.6**	103.0***
父の就労収入(有無)	98.8	99.0**	99.0**	99.0**	99.0**	99.0***
父の就労収入(金額:万円)	445.5	451.1***	451.1***	447.7	450.0***	448.9*
その他の収入(有無)	26.2	26.4	26.4	26.6	26.3	26.9**
父母の就労収入(有無)	99.4	99.5*	99.5*	99.5*	99.5*	99.5**
父母の就労収入(金額:万円)	546.1	553.3***	553.3***	550.2*	558.1***	553.1***
父母の就労収入+その他の収入(有無)	99.6	99.7	99.7	99.7*	99.7	99.7*
父母の就労収入+その他の収入(金額:万円)	557.3	564.6***	564.6***	561.4*	567.3***	564.6***
父母の就労収入+その他の収入(子育て費用の割合)	80.5	78.3***	78.3***	78.8*	77.4***	77.1***
1か月の子育て費用(万円)	4.1	4.0	4.0	4.0	4.0*	3.9***
保育料(有無)	5.8	5.6	5.6	5.8	5.6	5.8
母の1日の喫煙本数	2.01	1.8***	1.8***	1.9***	1.8***	1.8***
父の1日の喫煙本数	11.81	11.6***	11.6***	11.7	11.5***	11.7*

表 8 付表 理想のサンプルとの間に有意差がない項目

性別が男、出生年月が 2001 年 1 月、子の体重、子の身長、単胎・多胎の別、ふだんの保育者が保育ママやベビーシッター、母の父親と同居、母が子供を寝かしつける、母が散歩などで子供を屋外に連れていく、母が部屋等の掃除をする、母が洗濯をする、母がゴミを出す、父が子供の食事の世話をする、父が子供のおむつを取り換える、父が子供を寝かしつける、父が家の中で子供の相手をする、父が散歩などで子供を屋外に連れていく、父が食事をつくる、父が食事の後片づけをする、父が洗濯をする、父がゴミを出す、子育てで意識して行っていることの有無、子育てで意識してよく話しかける、子育てで意識してよくだっこする、子育てで意識して外気浴をさせる、子育てで意識して子どもの生活リズムをくずさない、子どもをもってよかったと思うことの有無、子どもをもってよかったこと：家庭が明るくなった、子どもをもってよかったこと：生活にはりあいができた、子どもをもってよかったこと：子育てを通じて自分の友人が増えた、子供を持って負担に思うことの有無、子どもをもって負担に思うこと：夫婦で楽しむ時間がない、子どもをもって負担に思うこと：仕事が十分にできない、子どもをもって負担に思うこと：子育てが大変なことを身近な人が理解してくれない、子どもをもって負担に思うこと：子どもが病気がちである、子育てでの不安や悩みの有無、親戚が子育ての相談相手である、医師が子育ての相談相手である、助産師・看護師が子育ての相談相手である、保育士が子育ての相談相手である、カウンセラー（電話相談を含む）が子育ての相談相手である、育児サークルの仲間が子育ての相談相手である、その他の収入の金額、父の家事・育児の総合点数、父の育児の点数、父の家事の点数、保育料の金額、出産 1 年前の母の就労状況、母の育児休業取得期間、妊娠日数

強い方向に偏っている。また、父母の就労収入が多く、父母それぞれの喫煙本数は少ない方向に歪んでいる。また、今回の出産に関係した引越・増築をした率は低くなっている。

第 2 回には有意差が無かったが第 3 回で有意差が出たものには、母方の祖母との同居率・保育士等の利用・保育料が小さい方向に偏っていること、子育てで出費がかさむとの回答が少なめに、自由な時間が持てないとの回答が多めになっていること、子育ての相談相手の多きに差が出てきたことがある。また第 2 回よりも第 3 回で歪みが強くなったものは、出産をきっかけとした引越・一年前に父が無職・父が同居していないが少なめに、母乳のみで育てたケースが多めになったことである。

第 2 回の時点で理想サンプルと実際のサンプルに有意差があった項目は、126 項目中 29 項目、第 3 回の時点で有意差があった項目は、126 項目中 41 項目である。サンプルが多いので信頼区間が狭くなることを考慮したとしても、回を追うごとに、歪みが多くの項目にでてくること、また上述のように歪みが強まる項目が出て来ることが分かる。

### 3-5. 復活・移動への対処をしなかった場合のサンプル特性—理想のサンプルとの比較

第 1 回調査に回答した全サンプル、すなわち理想のサンプルと、第 3 回調査まで 3 回とも回答した（つまり復活者は除く）サンプルとで諸変数の値を比較した（表 8 「第 3 回回

答者（復活除く）」の欄）。

第3回の実際のサンプル（つまり復活者が含まれている）と理想のサンプルを比べた場合と有意差の傾向が違うものを挙げてみる。兄弟姉妹数が少なめに、核家族世帯が多めに、母乳のみを授乳したケースがより多めに出ている。子供を持って身近な人が喜んでくれた率も高めだが子育てによる身体の疲れを負担とする率も高く、自分の自由な時間が持てない率もいっそう高めに歪んでいる。父の労働時間と母の就労収入が多めになり、子育て費用・保育料は低めに出ている。

第3回調査までに移動しなかったサンプルについても、第1回調査の諸変数の値を理想のサンプルと比較してみた（表8「第2回非移動者」「第3回非移動者」の欄）。もし移動者を追跡しなかった場合に起こったはずのサンプルの歪みについて述べる。

まず、実際の回答では理想サンプルと比べて歪みが見られないが、移動者が脱落していたら歪んでいた項目としては、母の出産児数・兄弟姉妹数・同居者人数が多い方に歪むこと、三世帯世帯・父方の祖父母との同居・ふだんの保育者が親と祖父母である率が高めになり、核家族世帯が少なめになること、13大都市出身者が少なめになって他地域が多くなること、父と母の通勤時間が短めになることが挙げられる。逆に、実際の回答では理想サンプルと比べて歪みが見られるが、移動者が脱落していたら歪みがなくなっていた項目には、子育てによる出費や自由な時間が持てないことの負担感・父の労働時間、歪みがなくなりはしないが小さくなっていた項目には、母乳で育てる率・父の収入・喫煙本数がある。

実際の回答の歪みと同じ傾向で、歪みをいっそう強める結果となる項目を挙げてみる。これらの項目は逆に言えば、実際には移動者を捕捉したために、歪みの拡大を防ぐことができている項目と言える。父母の年齢は、移動者が脱落していたら、理想のサンプルよりも、また実際のサンプルよりも高い方に歪む。父母とも日本人である確率・父との同居率・配偶者が子育ての相談者になる確率・出産一年前の父の無職率・母の就労収入もより高めになり、ひとり親世帯はより少なめに出る。

21世紀出生時調査は、脱落者に復活を促すことで、また移動者の捕捉に努めることで、これらのような歪みを防いでいると言える。

なお、理想のサンプルと差がなかった項目（表8付表）は、子どもの身長体重、父の育児・家事、子育てで気をつけている内容、子供を持ってよかったことと負担に思うことの内容、子育ての相談相手に多い。

#### 4. まとめ

本報告では、第3回までの21世紀出生児縦断調査における脱落・居住地移動・（脱落からの）復活サンプルの発生状況と、それぞれのサンプルの特徴、回による違い、脱落によって生じるサンプルの歪み、復活の促進や移動者捕捉を行わなかった場合に起こっていたはずのサンプルの歪みを検討した。

脱落によって過小に扱われるようになってしまっている層には、父母のどちらかが外国人である子、ひとり親世帯、父の関わりや母の育児への関わりが弱いケース、収入が少ないケースが含まれる。いずれも、政策的な支援が必要な可能性がある層であり、本調査ではこれらの層が過小評価された数値になっていることは、結果の読み取りと政策への適応の際に留意すべきであろう。

移動者を追ったことでは、項目によっては歪みの悪化を防いでいるものと、独自の傾向を見せた項目がある。それは移動を自己申告する人が、移動による脱落者と重なる側面と、脱落者とむしろ反対で調査に積極的な面とを持つことに由来すると考えられる。

それでは、移動による脱落を含めて、脱落を未然に防ぐには、実際にどのような方法があるのか。

第1節で触れた「消費生活に関するパネル調査」では、転居はがき送付の依頼、年賀状・前年度の調査結果の概要・調査の結果を報じた新聞雑誌記事集の送付、調査回収時の謝礼の配布を行っている。また、第10年度の調査時には、調査対象者と調査員にプレゼントを渡したという(村上、2003a:30;村上、2003b:120)。東京都老人総合研究所の「高齢者の生活と健康に関する長期縦断調査」では調査員の教育、調査員への手土産、対象者が高齢のため回答できない場合は親しい人による代行調査、調査結果報告、謝礼の支払いを行っている。早稲田大学人間総合研究センターのパネル調査「からだ・こころ・つながりの発達研究」では、調査報告や年賀状・暑中見舞い・災害時の見舞い状の送付、シンポジウムへの招待、回答者の知人を事務局で把握しておく方法をとっている。ニッセイ基礎研究所の「中高年パネル調査」では、謝金の額を上げたり、郡部では図書券・商品券の利用が困難であるために現金に変更した結果、回収率が上昇したという(以上、永井他、2002)。

海外ではどのような工夫があるのか。家計経済研究所が諸外国のパネル調査実施機関にインタビューをしたところによると、礼状を送ったり、対象者に調査結果を簡単に報告することで調査の意義を理解してもらうことは、調査協力につながるとされる。毎年同じ調査員が調査に行くことで、対象者との関係ができたり、在宅時間の予想が容易になったりする効果があるという。拒否しようとする対象者にはベテラン調査員が説得にあたる、対象者と親しい人の連絡先を控えて対象者の転居先が不明になることを防止するなどの工夫もある。また、多忙で拒否する場合には、短い調査のみを行うという方法もある(村上、2003b:120)。

アメリカでは、謝礼が重要になる場合が報告されている。調査にかかる時間が2倍になったら謝礼を2倍にしたり、なかなか調査に応じない対象者には謝礼を一定上限まで値上げしていくことを調査員に許容している調査もある。他国の例では、謝金ではなく謝品を配布する場合もあるという(永井他、2001:64)。

Farringtonらは、1960年代に8歳の時点で第1回の調査をし、32歳まで追跡した調査の様子を紹介している(D.Farrington et al., 1990)。専任スタッフが、選挙人名簿や電話帳のチェックを初め、近隣・親族への問合せ、犯罪記録の閲覧や住宅局等の助けまで得て、

表 9 実査時に捕捉に留意すべき属性

脱落防止に留意すべき属性	父母の年齢が若い、父母のどちらかが外国人、ひとり親世帯（特に父が同居していないケース）、父が無職、父母とも収入が少ない、母が育児・家事に関わりにくい、父母の喫煙が多い、妊娠・出産に伴い引越・増築があった
脱落後の復活を特に促すべき属性	父母の年齢が若い、父母のどちらかが外国人、ひとり親世帯（特に父が同居していないケース）、保育所等を利用、父母とも収入が少ない

時には数ヶ月かけて居所を突きとめたというこの調査は、約 95%の回答率を誇った。ただし、そこまでの追跡には、時間的・費用的な問題と、現在では情報公開上の制度的・倫理的な問題が出てくるだろう。

21 世紀出生児調査の場合は誕生日カードを送付している。そして、調査票を発送する際及び誕生日カードを送付する際に、住所等が変わった場合の連絡用はがきを同封し、対象者に転居の際の連絡を促している。このような方法に加えて、国による調査であることを活かして、予め回答者の了承を得て、転居の自己申告をしなかった場合も追跡できる何らかの方法が取れたならば、調査の精度が飛躍的に高まるだろう。

なお、本報告の結果から、特に捕捉に留意すべきと考えられる層を表 9 にまとめた。今後の実査で調査員に周知するなどして、脱落防止の一助としてもらえれば幸いである。

偏った脱落を防ぐためにも、脱落数自体を減らすためにも、21 世紀出生児調査で引き続き、またその他のパネル調査において、転居者の追跡において一層の工夫を期待したい。

### 【参考文献】

D.Farrington et al. (1990) 'Minimizing attrition in longitudinal research', in D. Magnusson and L. R. Bergman eds. *Data quality in longitudinal research*, 122-147.

金子隆一・福田節也(2005)「21 世紀出生児縦断調査における脱落要因の分析」『パネル調査（縦断調査）のデータマネジメント方策及び分析に関する総合的システムの開発研究』厚生労働科学研究費平成 16 年度報告書。

金子隆一・西野淑美 (2005)「21 世紀出生児縦断調査における居住地移動サンプルの考察」『パネル調査（縦断調査）のデータマネジメント方策及び分析に関する総合的システムの開発研究』厚生労働科学研究費平成 16 年度報告書。

村上あかね(2003a)「消費生活に関するパネル調査」の概要」『人口学研究』32:25-31。

村上あかね(2003b)「なぜ脱落したのか：「消費生活に関するパネル調査」における脱落理由の分析」『家計・仕事・暮らしと女性の現在：消費生活に関するパネル調査(第10年度)』，家計経済研究所，115-122。

永井暁子，久木元真吾，溝口由己(2002)「「パネル調査・カンファレンス」レポート」『季刊家計経済研究』54:70-81。

永井暁子，溝口由己，樋口美雄，チャールズ・ユウジ・ホリオカ(2001)「世界のパネル調査：ミシガン・パネル調査会議から」『季刊家計経済研究』50:63-68。

杉澤秀博他(2000)「全国高齢者に対する12年間の縦断調査の脱落者・継続回答者の特性」『日本公衆衛生雑誌』47(4):337-349。

## 付 録

### 21 世紀出生児縦断調査における脱落・居住地移動・復活サンプルの分析 分析方法マニュアル編

筆者は、「パネル調査（縦断調査）のデータマネジメント方策及び分析に関する総合的システムの開発研究」の平成 17 年度報告書論文「21 世紀出生児縦断調査における脱落・居住地移動・復活サンプルの分析」を執筆するに当たって、SPSSver13.0 の Base 及び Regression を用いて分析した。

以下、「市町村合併への対応」「分散分析および多重比較」「ロジスティック回帰分析」「母平均の差の検定・1 サンプルの検定」について、具体的な分析手順を、マニュアルの形にまとめた。

#### 1. 市町村合併への対応

分析に入る前に、居住地移動サンプル（各調査の間に、住所地の自治体コードが変わっているサンプル、ただし市町村合併によるコード変更を除く。詳しくは報告書論文を参照）を特定するために、住所地の自治体コードが違うサンプル内で、市町村合併によってコードが変わっている場合を識別する作業が必要となった。

具体的には、2001 年 1 月（1 月出生児の出生票届出時点＝この時点の住所地が第 1 回調査の送り先）から 2003 年 2 月（7 月出生児の第 2 回目調査時点）までに自治体コードに変更があった自治体、2003 年 3 月から 2004 年 2 月（7 月出生児の第 3 回目調査時点）までに自治体コードに変更があった自治体をリスト化し、各ケースの移動の有無を変数化する際にそれらを識別するシンタックスを作成し、「合併による自治体名変更の可能性」とのカテゴリをあてがった。実際には、合併をする自治体間を合併前に移動している可能性もあるが、その点は識別不能であるため、移動ありかなしかの二項に分類する際は、「合併による自治体名変更の可能性」に該当するケースは「移動なし」に分類している。

#### 2. 分散分析および多重比較

報告書論文では、第 3 回調査時点での脱落者・移動者・非移動者について、カテゴリ別分散分析およびシェフェの多重比較を行い、それぞれの特性を把握した。具体的には、次のステップで行った。

- ①変数のチェックと作成：第 1 回調査の項目のほぼすべてに近い 120 項目（従属変数となる）に対して、量的変数、順序変数、または 01 の変数になっているかを確認する。無回答などは欠損値に指定しておく。また、独立変数である、分散分析のグループ（この場合は脱落者・移動者・非移動者の分類）を規定する変数を作成する。

②分散分析及び多重比較を実行する。

【操作】SPSS のメニューの「分析」→「平均の比較」→「一元配置分散分析」で、オプションの「その後の検定」で「Scheffe」にチェックを入れる。120 の従属変数は「従属変数リスト」に、独立変数であるグループを規定する変数を「因子」に入れる。

③多重比較の結果、変数ごとにサブセット間の有意差（5%水準）の状況が表に示される。報告書論文では、その状況をパターンごとに一覧表にまとめた。

### 3. ロジスティック回帰分析

報告書論文では、第3回調査時点での継続回答者と脱落者、移動者と移動非経験者について、それぞれロジスティック回帰分析を行い、他の項目を調整した上でも影響の大きい要因を検討した。具体的には、次のステップで行った。

#### ①従属変数の作成

第3回までの継続回答者を0、脱落者を1とした変数（attrtn01）を作成した。また、第3回までに移動を経験していない人を0、移動を経験した人を1とした変数（move01）を作成した。

#### ②独立変数の選定

a)第1回調査の項目のほぼすべてに近い116項目のうち、8項目は欠損が多いため除外した。

b)残りの108項目に対し、多重共線性をチェックした。SPSSではロジスティック回帰分析の出力結果には共線性の情報が含まれない。よって、これら108項目を独立変数とし、上記①で作成した変数を従属変数として通常の回帰分析を行い、共線性の情報を得た。

【操作】SPSS のメニューの「分析」→「回帰」→「線型」で、オプションの「統計」で「共線性の診断」にチェックを入れる。

※ロジスティック分析ではダミーコード化の機能があるが、回帰分析の場合は事前にダミーコード化が必要。

c)回帰分析の出力結果の「係数」の中で、「共線性の統計量」の「許容度」が0に近づく変数が多くなるにつれて、分析結果は多重共線性のために不安定になる。よって、許容



度が 0.1 以下等であるものは除外し、また項目が重複しているもの（同居者変数と世帯構成変数、ふだんの保育者変数とふだんの保育者の組み合わせ変数、収入の有無と収入の金額など）は一方のみを採用することとした。最終的に 70 項目をロジスティック回帰に投入することとした。

③上記①の従属変数と、上記②の独立変数で、ロジスティック回帰分析を行った。

【操作】SPSS のメニューの「分析」→「回帰」→「二項ロジスティック」で、「従属変数」に①の変数を入れる。独立変数は「共変量」に入れる。

※ダミー変数化する必要があるカテゴリー変数は、オプションの「カテゴリ」で「カテゴリ共変量」に指定する。それぞれの変数について、参照カテゴリを指定できる。

④各変数の有意確率と EXP(B)（係数 B の指数値）を示した。

「Exp(B)」は、独立変数が 1 単位増加した時のオッズの増減量を表す。例えば Exp(B) が 2 ならば、オッズは 2 倍に増えるということである。

#### 4. 母平均の差の t 検定

報告書論文では、以下の組み合わせについて、連続変数の母平均の差を t 検定で比較した。

- ・第 2 回調査で脱落したサンプルと、第 3 回調査で初めて脱落したサンプル
- ・第 1 回調査から第 2 回調査の間に移動したサンプルと、第 2 回調査から第 3 回の調査の間に移動したサンプル
- ・第 2 回で脱落したサンプルのうち、第 3 回に復活したサンプルと第 3 回も脱落したままのサンプル

具体的手順は以下のとおりである。

①「独立したサンプルの T 検定」を実行する。

【操作】SPSS のメニューの「分析」→「平均の比較」→「独立したサンプルの T 検定」を行う。比較する項目（変数）を「検定変数」に、グループ（上記の各組み合わせ）を規定する変数を「グループ化変数」に入れる。

②「等分散性のための Levene の検定」の出力を見て、「等分散性を仮定する」の有意確率が低くて仮定が棄却される（すなわち、等分散性が仮定できない）場合は、「等分散性を

仮定しない」の行の t 検定結果を読む。「等分散性を仮定する」の有意確率が高くて仮定が棄却されない（すなわち、等分散性が仮定できる）場合は、「等分散性を仮定する」の行の t 検定結果を読む。その結果を、報告書論文では一覧表にまとめた。

## 5. 母平均の差の $\chi^2$ 検定

報告書論文では、4 と同じサンプルの組み合わせについて、離散変量の母比率の差を  $\chi^2$  検定で比較した。

具体的手順は以下のとおりである。

① 比較に使用するサンプルのみを分析ケースに指定しておく。

【操作】メニューの「データ」→「ケースの選択」→「IF 条件が満たされるケース」でフィルタリング指定するか、上記のサンプル組合せがカテゴリ化されている変数で、当該サンプルの値以外を欠損値に指定しておく。

② 「クロス集計表」を作成し、 $\chi^2$  検定を実行する。

【操作】SPSS のメニューの「分析」→「記述統計」→「クロス集計表」を行う。オプションの「統計」で「カイ二乗」にチェックを入れる。グループ（上記の各組み合わせ）を別に、比較する項目（変数）を列に入れる（行列が入れ替わっても検定は可能）。

## 6. 1 サンプルの検定

報告書論文では、以下の組み合わせについて、第 1 回調査に回答した全サンプル（理想のサンプル）を母集団として想定し、1 サンプルによる検定を行った。離散変量は  $\chi^2$  検定で、連続変量は t 検定で、1 サンプルによる検定を行った。

- ・ 第 1 回調査に回答した全サンプルと、第 3 回調査に回答したサンプルとで第 1 回調査の諸変数の値を比較し、脱落によるサンプルの歪みが生じているかを確認した。
- ・ 第 1 回調査に回答した全サンプルと、第 3 回調査まで 3 回とも回答した（つまり復活者は除く）サンプルとで諸変数の値を比較し、もし復活者を受け入れなかった場合に起こったはずのサンプルの歪みを確認した。
- ・ 第 1 回調査に回答した全サンプルと、第 3 回調査までに移動しなかったサンプルとで第 1 回調査の諸変数の値を比較し、もし移動者を追跡しなかった場合に起こったはずのサン

プルの歪みを確認した。

具体的手順は以下のとおりである。

- ① 理想サンプルに対して比較するサンプル(例:第3回調査に回答したサンプル)のみを、分析ケースに指定しておく。

【操作】メニューの「データ」→「ケースの選択」→「IF 条件が満たされるケース」でフィルタリング指定する。

- ②連続変数の場合：当該変数の理想サンプルの平均値と、分析ケースに指定した比較サンプルとで「1 サンプルの T 検定」を実行する。

【操作】SPSS のメニューの「分析」→「平均の比較」→「1 サンプルの T 検定」を行う。当該変数を「検定変数」に、理想サンプル (= 第 1 回調査に回答した全サンプル) の平均値を「検定値」に入れる。

※変数によって理想サンプルの平均値が違うため、この分析は各変数ごとに 1 回 1 回実行する必要がある。コピー&ペーストを活用して、シンタックスを書く方が省力できる。

- ③離散変数の場合

a) 各変数について理想サンプルの度数分布表を予め出力しておく。

b) 比較するサンプルを分析ケースに指定した状態で、1 サンプルのカイ二乗検定 (ノンパラメトリック検定) を実行する。

【操作】SPSS のメニューの「分析」→「ノンパラメトリック検定」→「カイ 2 乗」を行う。当該変数を「検定変数リスト」に、理想サンプル (= 第 1 回調査に回答した全サンプル) の度数分布を、カテゴリーに割り振られている値の昇順に「期待度数」に入れる。

※カテゴリーの値の昇順に「期待度数」に入れるとは、例えば男=1 で 3000 ケース、女=2 で 7000 ケースならば、1→2 の順で、「3000 7000」と入れることを意味する。

※変数によって理想サンプルの度数分布が違うため、この分析は各変数ごとに 1 回 1 回実行する必要がある。コピー&ペーストを活用して、シンタックスを書く方が省力できる。

- ④それぞれ有意差を読み取り、報告書論文では一覧表にまとめた。

**捕捉： 欠損値分析 (Missing Value Analysis)**

SPSS にはオプションで「欠損値分析」という機能がある。これを用いて、ある変数の欠損値が、別の変数での欠損値の有無に関連しているか、またある特性をもったケースにおいて系統的に発生していないか等をチェックできる。

しかし、メモリを多く使用するため、限られた数の変数しか投入できなかったこと、また欠損ケースが 5% を超す変数は少なく、深刻な影響は少ないと考えられたため、優先順位を下げ、本報告書論文では扱わなかった。

## 1 1 不詳回答の発生状況とその処理に関する検討

小山 泰代

### はじめに

統計調査、とりわけ回答者自身が記入する形式の調査では、回答がなかった、適切な回答ではなかった等の理由により、不詳となる回答が発生する。不詳回答は、データクリーニングにおいては、有効回答を確保するため、他の項目との関係によってあるカテゴリーを割り当てるなどの方策がとられることが多い。パネル調査の場合には、時系列で遡ることによって不詳を補完する方法も考えられる。官庁統計として、パネル調査として、不詳回答にどのように対処するかは、極めてベーシックな課題である。そこで本研究では、不詳回答がどのように発生しているのかを観察し、また、それが厚生労働省パネル調査でどのように処理されているのかを確認し、不詳回答の処理方法について考察を加える。

以上の目的のもとに、本研究では以下の分析を行う。すなわち、一般の統計調査やパネル形式のデータにおける不詳処理の事例の検討と整理、厚労省パネル調査における不詳回答の発生状況の確認、不詳の発生に関連する要因の検討、そして、不詳回答の処理方法（現行と今後の方向）について検討する。

### 1. 21世紀出生児縦断調査における不詳の状況

21世紀出生児縦断調査第1回および第2回における不詳回答の発生の状況を表1（章末）に示す。

ここでは、データ管理の観点から、整理されたデータにおいて「不詳」というコード（ラベル）を与えられたものを「不詳」としてカウントしている。また、当初想定されたコード以外のコードを与えられたものを別途カウントし、図表中の「その他」欄に併記している。本データでは、変数によっては不詳とラベリングされたものにいわゆる非該当が含まれており、それによって不詳回答の割合が非常に高くなっているものがある。

本調査の不詳調査の発生状況を見ると、同居家族の状況や保育者、収入、学歴など、分析の際に説明変数となるような基本属性における不詳の発生が目立つ。また、一般に言われるように、子どもをもってよかったと思うことや子どもをもって負担に思うことといった意識に関する質問でも不詳回答の割合が高いことが見てとれる。社会調査では、基本属性を示す変数は分析上の重要性が高いゆえに、これらの変数についてはできるだけ不詳を減らすようななんらかの手段を講じることが多い。パネル調査では、その特徴である異なる時点における回答によって補完することも可能な場合がある。とりわけ基本属性に関する変数は客観的な指標であることが多いため、時間的前後関係からの推察によって補完しようとする傾向が強い。しかし、逆にそうした遡及的な補完データが、その後の回答と矛盾する場合も考えられ、重要な変数であればこそ、不詳回答の補完には一層の慎重を期する必要がある。

データ管理の面からは、不詳と非該当とは区別されることが望ましく、今後長期に渡る

データ整備と将来的な分析の精度確保のためにも、ドキュメント上での明確な定義と区別が求められる。同時に、不詳を何らかの方法で保管した場合には、その記録が後に重要となる。不詳コードを割り当てられたものだけでなく、図表1においてその他とされたものについても同様の配慮が必要である。パネル調査に限らず、一般の統計調査においてどのような不詳回答の処理がなされているのか、その事例を整理・検討し、本調査におけるそれと照らして、不詳回答の処理ないし補完について、現在どのような方策がとられているのか、今後の方策としてどのような可能性があるのかを検討していく。

## 2. 不詳発生 の 検討

不詳回答は、ふだん意識していないことを尋ねられたり、質問内容が難しかったり、回答の選択肢の定義や区分が曖昧で判断に困ったりする場合に発生しやすい。また、容易な質問ばかりであっても、質問項目が多すぎて最後まで回答するのに大きな負担を伴う場合にも不詳は発生する。不詳の発生について、質問の性格、あるいは回答者の属性によってなんらかの傾向があることが明らかになれば、不詳割合を低下させるヒントを見いだすこともできよう。この分析においては、以下のような視点が考えられる。

(1) 質問の性格(質問内容、位置、回答形式等): 一般に、政治的なことなど身近な生活には直接関係しないような事項、実態よりも意識に関する事項、現在よりも過去に関する事項あるいは過去から現在に至る事項等については不詳回答が多いとされる。出生児調査は、育児という日常生活に密着したテーマの調査であり、その点では不詳回答の発生しにくい調査であるといえる。逆に、育児に関する調査は場合によっては家庭内の非常にプライベートな領域にも触れるため、回答拒否の意味での不詳回答が発生する可能性を含んでいる。本調査では質問の性格によって意味のある差異が見受けられるか。不詳の発生した事項間にはどのような関係があるのか。回答形式による不詳発生の特徴は見られるか。特に不詳の発生しやすい項目については、質問文や選択肢の状況も観察・検討する必要がある。

21世紀出生児縦断調査は2001年1月10日～17日生まれの子ども(回収数23,421。以下、1月出生児)と同年7月10日から17日生まれの子ども(回収数23,586。以下、7月出生児)を対象として継続的に実施される。第1回調査は調査対象となる出生児が月齢6ヶ月時点での調査で、平成13(2001)年度に実施された。調査日は、1月出生児については平成13(2001)年8月1日、7月出生児については平成14(2002)年2月1日である。同一の調査票を用いており、異なるのは調査日のみということになる。

この第1回調査の1月出生児と7月出生児の調査結果(データ)を見ても、不詳等の出現状況にはぶれがあることが分かる。

- ・ 父母の就業状況: 1月出生児ではその他の出現率が高い。一方7月出生児ではその他はほとんど出現していない。父母の企業規模が不詳コードあるいはその他となる割合が高いが、父親において大きな差が見られる。父親の労働時間についても7月出生児のその他出現率が1月出生児に比して高い。
- ・ 育児・家事の分担: いずれにおいてもほとんどがその他のコードとして現れているが、その頻度は7月出生児がきわめて高い。また、7月出生児ではその他のほかに不詳コードの現れているものがある(部屋の掃除(母)、買い物(母)、食事の世話(父))。

- ・住宅：出現頻度はほぼ同じであるものの、1月出生児ではその他、7月出生児では不詳コードとして出現している。
- ・父母の喫煙状況：いずれにおいてもその他として現れているが、出現頻度には大きな差があり、7月出生児の方が総じて高い。
- ・日頃意識して行っていること：複数回答の間である。1月出生児と7月出生児の不詳コードとその他の分布が大きく異なっている。1月出生児ではコードの昇順にしたがい累積的に分布しているが、7月出生児ではそういった傾向は見られない。
- ・子どもをもってよかったと思うこと：1月出生児ではコードの昇順にしたがい累積的に分布しているが、7月出生児ではそういった傾向は見られない。
- ・子どもをもって負担に思うこと：1月出生児ではコードの昇順にしたがい累積的に分布しているが、7月出生児ではそういった傾向は見られない。また、1月出生児にはその他として現れるものはみられないが、7月出生児ではそれがある。
- ・子育ての不満や悩みの相談相手：1月出生児ではコードの昇順にしたがい累積的に分布しているが、7月出生児ではそういった傾向は見られない。また、不満や悩みの有無について、7月出生児では不詳コードやその他の出現率が非常に高い。
- ・母乳の状況：回答するのに抵抗感が強い設問は質問紙の後半に置かれることがおおいが、本調査でも最後の部分に置かれている。不詳コード、その他とも、7月出生児のほうが総じて高い。

一方、対象出生児が1歳6ヶ月時点で行われた第2回調査（1月出生児：平成14（2002）年8月1日実施、7月出生児：平成15（2003）年2月1日実施）では、出現の状況や頻度分布に大きな差は見られない。

これらは、実際の回答における不詳やその他というよりも、コーディングなどのデータクリーニングにおける操作の結果によるところが大きいものと思われる。調査の最終的な結果は1月出生児と7月出生児とを合わせた形で集計されたものが公表されるので、コーディングの統一作業が重要となる。

（2）回答者の属性：一般に社会調査では、回答者の性別や年齢層によって不詳回答の発生状況がある程度規定されるといわれる。本調査では回答者の性別や年齢層は比較的限定的と考えられるが、どのような属性の回答者がどのような事項で不詳回答を示しているのか。どのような回答者が不詳回答を行いやすいのか。逆に、不詳のない回答者とはどのような属性をもっているのか。

ここで、21世紀出生児縦断調査の第1回と第2回のデータから、個々の回答者の回答状況をみる。使用するデータは第2回調査実施後に作成された確定データで、1月出生児、7月出生児が統合されており、サンプル数は47,015となっている。また、回答者個人のマッチングにより、2回分の調査データが連結されている。今回は、確定データのデータセットから、回答者の基本的属性や社会経済的属性を中心とする主要な変数（85変数）を抽出した。

回答者ごとの不詳数は、図1の通りである。抽出された85変数について不詳のない者は25,323で全体の53.9%、不詳のある者は21,692で46.1%と、不詳のない者のほうがやや多い結果となっている。最大の不詳数は29（1サンプル）で、つまり対象となる変数の3分の1近くが不詳となっていることになる。不詳をもつ者の多くは1、2項目にとどまるが、例えば10以上の不詳をもつ者は286で、不詳をもつ者の1.3%を占める。なお、章末

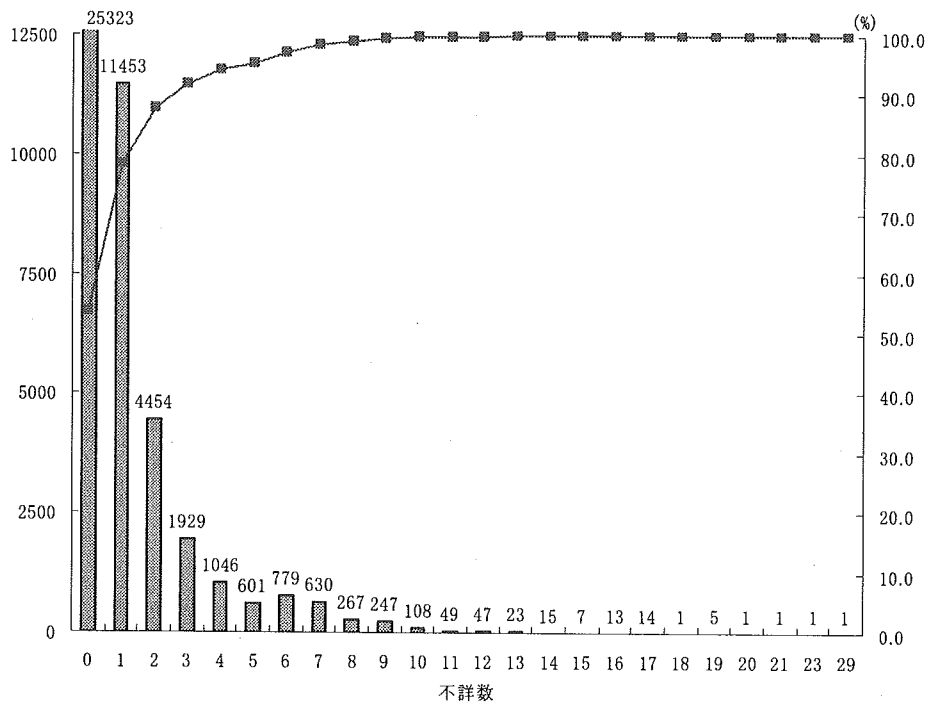


図1 不詳数の分布

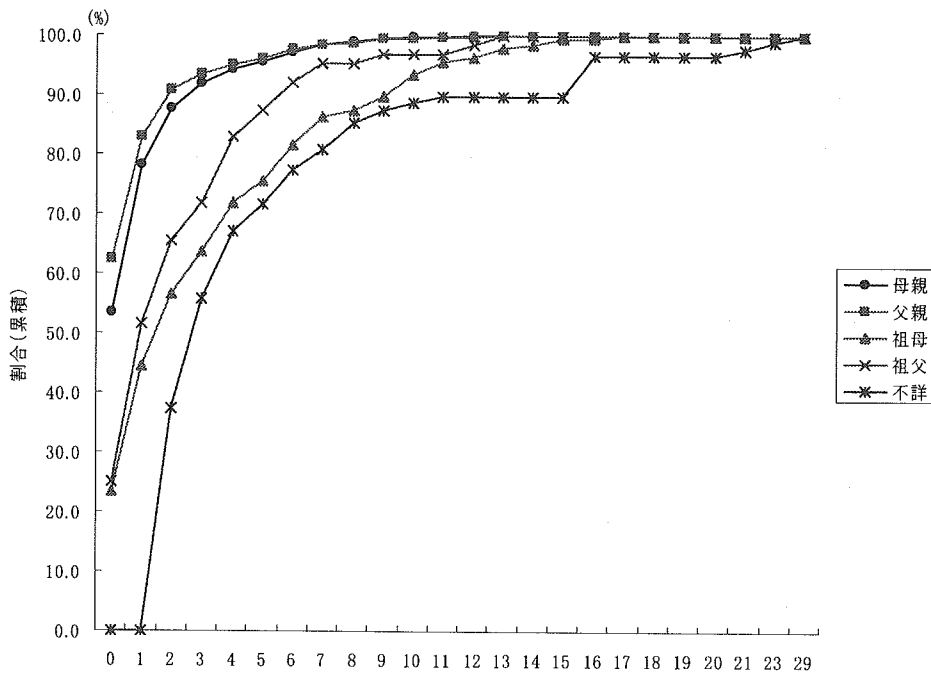


図2 回答者の続柄別不詳数の分布(累積割合)

には15以上の不詳をもつサンプル(44/不詳をもつ者の0.2%)のリストを添えた(表2)。

これを回答者の続柄別にみたものを図2に示す。回答者の続柄は第1回調査時点のもので、母親、父親、祖母、祖父とした。また、回答者の続柄不詳についても合わせて示してある。なお、回答者の続柄の分布は、母親(43,211/全体の91.9%)、父親(5,571/11.8%)、



祖母（184/0.4%）、祖父（64/0.1%）で母親が大多数を占めている。また、回答者の続柄不詳は88で、全体の0.2%にあたる。

これをみると、母親・父親に対して祖母・祖父は相対的に不詳の発生する割合が高いことが分かる。その背景としては、回答者が高齢であるということのほか、調査の内容が祖父や祖母にとっては孫や子ども夫婦のことについてであるため、把握できていない事柄が多いということが考えられよう。図2からはまた、父親よりも母親、祖父よりも祖母に不詳数の多い者が多いことがみてとれる。一般には子どもに関する調査は母親が回答するケースが多いと考えられるが、こうした調査に回答する父親や祖父は家庭の事柄により積極的に関わっているといえるかもしれない。しかしまたその逆も考えられ、回答の内容（質）については吟味が必要な場合もあるだろう。また、回答者の続柄不詳のケースでは、不詳の発生率が高いことが分かる。実際不詳数が20以上の4ケースのうち3ケースは回答者の続柄が不詳である。回答者の続柄は非常に基本的な項目であり、意図的に秘匿する場合もないとはいえないが、それが不詳になるようなケースでは、調査に対する態度そのものが消極的であるか、調査票の設計が理解されにくいものであったといった点も要因として考えられる。これらは、「誰が」「誰のことについて」回答するのかが明確に伝わるような工夫を加えることによって回避できる場合があるかもしれない。

なお、第1回と第2回の回答者を比較すると（表3）、第1回で母親が回答していたもののうち、第2回も母親が回答したものは39,039(96.6%)、他の者に変更されたものは1,287(3.2%)、回答者の続柄不詳となったものは96(0.2%)となっている。同様に第1回に父親が回答したものについては、1,937(37.4%)、1,287(62.2%)、21(0.4%)である。父親が回答するケースはそのとき限りのものであるという可能性がうかがえる。また、回答者の変更によって回答内容にぶれが生じる可能性も当然考えられよう。

調査では、就業状況や収入など母親と父親のそれぞれのことを尋ねている設問がある。これについて、回答者の続柄によって、自分のことと配偶者のことに関する回答の状況に差があるかをみたのが表4である。この表からは、回答者が母親であっても父親であって

表3 回答者の続柄の比較

第1回 (サンプル数)		第2回		
		第1回と同じ	変更	不詳
母親	40,422	39,039	1,287	96
父親	5,185	1,937	3,227	21
	(%)			
母親	100.0	96.6	3.2	0.2
父親	100.0	37.4	62.2	0.4

表4 回答者の続柄の比較 (単位:%)

回答者 質問の対象	母親		父親	
	自分のこと	配偶者のこと	自分のこと	配偶者のこと
出産前の就業	0.8	1.2	1.0	1.3
出産前の企業規模	1.6	4.4	1.4	1.5
H13年の就業	1.1	1.8	1.3	1.4
H13年の企業規模	0.3	2.8	1.1	0.2
労働時間	1.0	4.1	2.2	1.3
通勤時間	1.5	6.0	4.2	1.6

も、おおむね自分のことよりも相手のことについての質問に不詳が発生する割合が高いことが分かる。とくに、母親が回答する場合、配偶者すなわち「父親」の労働時間や通勤時間など具体的なことは不詳となる割合が高い。一方、父親が回答する場合には、配偶者すなわち「母親」の企業規模や労働時間、通勤時間などでは自身についてよりも不詳の割合が低くなっている。母親のほうが総体的に常勤で働いている割合が低いためと考えられよう。

(3) 時点による発生状況：たとえば、ある回答者について、ある時点では不詳でなかった項目がその後の調査で不詳になるといったケースが考えられる。対象児の成長につれて育児の性格も変化し、育児の各領域に対する関心の濃淡が変化することによって不詳の発生箇所も変化するということも考えられよう。

21世紀出生児縦断調査の1月出生児について、第1回(平成13(2001)年8月1日実施)と第2回(平成14(2002)年8月1日実施)をみてる。前者は出生児の月齢6ヶ月、後者は1歳6ヶ月である。なお、両調査間では調査項目に若干の変更がある。

- ・ 父母の就業状況：第2回では父母の就業状況についてのみ質問があり、不詳コードとその他コードが現れている。第1回では現れていなかった不詳コードが出現している。
- ・ 育児・家事の分担：第2回では第1回と同様にその他のコードとして現れており、その頻度も同水準である。
- ・ 住宅：第2回では質問項目なし。
- ・ 父母の喫煙状況：第2回では質問項目なし。
- ・ 日頃意識して行っていること：第2回では質問項目なし。
- ・ 子どもをもってよかったと思うこと：不詳コードの頻度分布は第1回、第2回とも類似している。第2回では一部のその他コードの頻度が第1回よりも小さい。
- ・ 子どもをもって負担に思うこと：不詳コードの頻度分布は第1回、第2回とも類似している。
- ・ 子育ての不満や悩みの相談相手：不詳コードの頻度分布は第1回、第2回ともおおむね類似している。第2回ではその他コードの出現は極端に減っている。
- ・ 母乳の状況：第2回では質問項目なし。

これらは、実際の回答における不詳やその他というよりも、コーディングなどのデータクリーニングにおける操作の結果によるところが大きいものと思われる。調査の最終的な結果は1月出生児と7月出生児とを合わせた形で集計されたものが公表されるので、コーディングの統一作業が重要となる。

図3として先に抽出した85変数のそれぞれについて不詳の割合を示す。なお、図中B265以降の変数は第2回調査による項目である。

調査間の比較に先立って、目立つものをみると、父親の育児休暇に関する項目(B157, B158)が挙げられる。この背景としては、男性が育児休暇を取得するという点について、制度として認識されていないということが考えられるのではないだろうか。制度として定められていても実際に男性の育児休暇の取得率はきわめて低く、それに対する質問は回答者にとっては現実感に乏しく、回答に迷ったという可能性は否定できない。パネル調査に限らず、統計調査の難しい点の表れであるといえるだろう。このほか、先に見た就業や、収入に関する項目で不詳の割合が比較的高い。とくに収入については第2回調

査での不詳の割合が大きく伸びている。その背景としては、プライバシーに対する意識が年々厳しくなっていることのほかに、毎年同じことを尋ねられるのが面倒だということも考えられる。収入だけでなく、1年単位で大きな変化のない事項について毎年尋ねられるのを煩わしいと感じる回答者もいるだろう。それはパネル調査の難しさといえる。子育てにかかる費用と収入とのバランスは少子化対策や子育て関連施策で大きな関心を寄せられるところであるだけに、こうした点の検討を行わなければならない。数量的な裏付けはないが、収入の金額などを回答するのをためらう理由として、調査員など他人に知られるのが嫌だという心情的なもののほか、回答内容を税金関係の確認等に利用されるのではないかという不安があると聞く。個人情報保護の立場は当然遵守しつつ、調査データの目的外の利用は決して行われないことを効果的にアピールする必要もあるかもしれない。

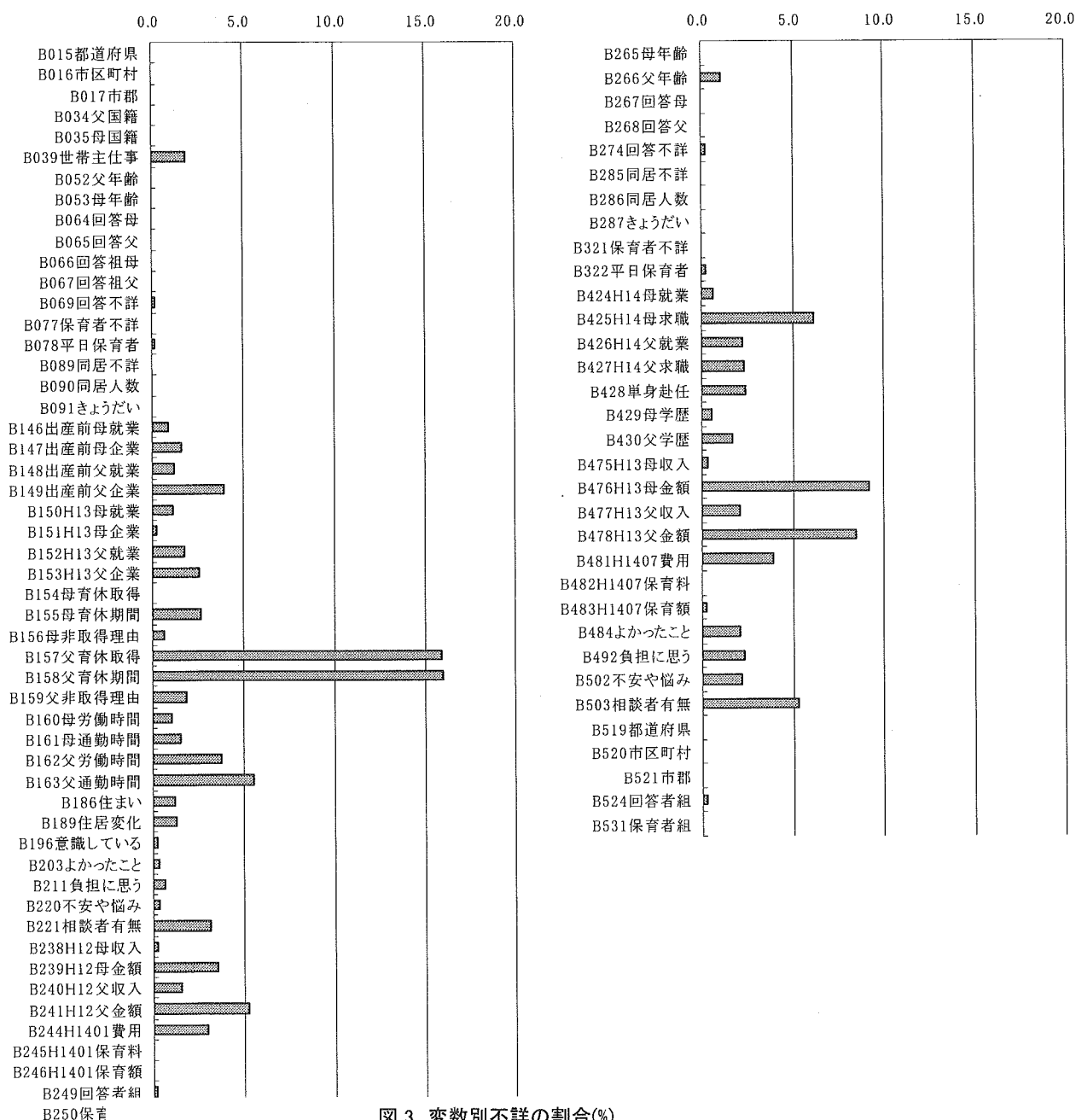


図3 変数別不詳の割合(%)

表5 第1回と第2回との回答の比較(収入・費用)

母親の収入の有無		第2回		第2回 (%)	
第1回	非不詳 不詳	非不詳	不詳	非不詳	不詳
第1回	非不詳 不詳	43,823 103	43,725 76	98 73.8	0.2 26.2
母親の収入金額					
第1回	非不詳 不詳	11,046 505	10,196 296	850 209	92.3 58.6
父親の収入の有無					
第1回	非不詳 不詳	第2回		第2回 (%)	
第1回	非不詳 不詳	43,354 572	42,889 134	465 438	98.9 23.4
父親の収入金額					
第1回	非不詳 不詳	40,589 2,044	37,722 1,291	2,867 753	92.9 63.2
子育て費用					
第1回	非不詳 不詳	第2回		第2回 (%)	
第1回	非不詳 不詳	42,726 1,200	41,281 938	1,445 262	96.6 78.2
保育料金額					
第1回	非不詳 不詳	1,925 0	1,904 0	21 0	98.9 -

収入や子育て費用について、回答者ごとに第1回と第2回での回答状況を比較したものを表5にまとめた。まず母親の収入の有無については、第1回では103(0.2%)だった不詳は第2回では125(0.3%)と若干割合が増えている。これに対して収入の額については、先に見たように505(4.0%)から1,059(9.2%)と倍増している。収入の額では、第1回、第2回とも不詳のものは209(全体の1.8%)にすぎないが、非不詳だったものから第2回に不詳に転じたものが850(第1回非不詳の7.7%)と大きな値を示している。逆に第1回の不詳から第2回の非不詳へ転じたものも296(第1回不詳の58.6%)存在しており、不詳から非不詳へ、非不詳から不詳へとどちらの推移も起こってはいるが、後者の動きが非常に大きいため結果として不詳の増加となっていることがわかる。父親についても同様で、収入の額では不詳から非不詳への動きが1,291(第1回不詳の63.2%)、非不詳から不詳への動きが2,867(第1回非不詳の7.1%)で、結果として不詳が2,044(4.8%)から3,620(8.5%)と大きく増加した。また、子育て費用についても、不詳から非不詳へ938(第1回不詳の78.2%)、非不詳から不詳へ1,445(第1回非不詳の3.4%)で、結果として不詳が1,200(2.7%)から1,707(3.9%)と大きくなっているものの、収入ほどの大きな伸びではない。金銭に関する設問でも、収入に比べ支出については回答に対する抵抗が小さいといえる。

最後に、意識の面について、2回の調査間での回答状況の違いをみってみる。表6に、子どもをもってよかったと思うこと、負担に思うこと、不安や悩み、相談相手の有無についての比較をまとめた。まず、よかったと思うこと、負担に思うこと、不安や悩みの3項目では、第1回、第2回とも不詳のもの割合が収入関係の項目に比べて非常に低いことがみてとれる。非不詳から不詳への動きも2%程度である。第1回調査のサンプル47,015の