

特集：『第5回全国家庭動向調査（2013年）』の個票データを
利用した実証的研究（その1）

第5回全国家庭動向調査の無回答の発生状況ならびに 平成25年国民生活基礎調査（世帯票） の個票データとのマッチングに関する検討

山内昌和・菅 桂太・菊池 潤

本稿は、第5回全国家庭動向調査（以下、第5回調査とする。他の調査回についても同様の表記とする）の無回答の発生状況と、平成25年国民生活基礎調査（世帯票）の個票データとのマッチングについて検討するものである。

まず、第5回調査の無回答のうち欠票、すなわち標本のうち無効となったものについて検討した。標本に占める欠票の割合を意味する欠票率は、第4回調査までは増加傾向にあったが、第5回調査では第4回調査とほぼ同じ36.1%であった。この欠票率を未配布率、未回収率、無効率に分解すると、未回収率は低下していたが、未配布率と無効率は上昇していた。未配布と未回収の理由としては、いずれも標本となった世帯に接触できなかったことを意味する不在（全国家庭動向調査では調査員が調査票の配布と回収を担う）が多く、第4回調査より高い割合を占めていた。

次に、第5回調査の無回答のうち不詳、すなわち集計の際にいわゆる「不詳」や「無回答」等として処理されるものについて検討した。変数別およびケース（集計表作成に利用した客体のこと）別に分析したところ、第4回調査に比べ、不詳割合の低い変数やケースが増えた。また、データクリーニングによる調査結果への影響を検討したところ、変数別およびケース別のいずれにおいてもデータクリーニングが不詳割合の減少に寄与していることが分かった。

続いて、データマッチングについては、第5回調査の地区番号、単位区番号、世帯番号、妻の出生年月の4つの識別子を用いて平成25年国民生活基礎調査（世帯票）との照合を試みた。その結果マッチング可能と判断できたデータは、第5回調査のうち93.3%、平成25年国民生活基礎調査（世帯票）のうち61.7%であった。

以上を踏まえ、最後に第5回調査のデータ利用上の留意点や今後の調査の課題について考察した。

I. はじめに

全国家庭動向調査は、国立社会保障・人口問題研究所（以下、社人研とする）が5年に1度実施する反復横断調査であり、1993年の第1回調査以降、1998年、2003年、2008年、2013年の5回にわたって実施されてきた。この調査の目的は、家庭内における出産・子育て、老親の扶養・介護をはじめとする家庭機能の実態とその変化を捉えることである。その成果は、報告書（厚生省人口問題研究所 1995、国立社会保障・人口問題研究所 2000、2007、2011a、2011b、2015a、2015b）や雑誌「人口問題研究」の特集号¹⁾等で公表され

1) 全国家庭動向調査に関係する「人口問題研究」の特集号は次の通りである。第54巻3号、第57巻2号、第57巻3号、第63巻4号、第66巻4号、第67巻1号、第68巻1号、第72巻1号。

てきた。

本稿は、第5回全国家庭動向調査（以下、原則として第5回調査とする。なお、他の調査回についても同様の表記とする）の分析や結果の解釈に必要な基礎的情報の提供と、今後の調査実施に向けた課題の明確化を目的として、第5回調査の無回答の発生状況ならびに平成25年国民生活基礎調査（世帯票）の個票データとのマッチングについて検討するものである。

全国家庭動向調査の無回答の発生に関する検討は、前述の報告書を除き山内（2012a, b）が行った。山内（2012a）では第4回調査における無回答の発生状況と標本属性の関係について分析し、山内（2012b）では第1回調査から第4回調査で集計対象となった有配偶女性の年齢や仕事の有無、学歴といった基本属性について国勢調査の結果と比較して、無回答の集計結果への影響を検討した。本稿では、山内（2012a）の分析を参考にしつつ、第5回調査の無回答の発生状況について第4回調査の結果と比較検討する。本作業は、無回答の発生状況を把握することに加え、後述する第5回調査で導入した変更点の影響を評価するために実施するものである。

国民生活基礎調査（世帯票）の個票データとのマッチングは第2回調査の際に実施され、その結果は報告書にまとめられている（国立社会保障・人口問題研究所 2000, pp.119-120）。第2回調査では有効標本である世帯を構成する全世帯員についてデータマッチングが行われたが、第5回調査については結婚経験のある女性のみを対象としてデータマッチングを行う²⁾。同作業は、データマッチングを前提とした調査設計の可能性と、国民生活基礎調査に含まれる結婚経験のある女性を全国家庭動向調査がどの程度補足できたのかを評価するために実施する。

以下、Ⅱで調査員調査による大規模な標本調査を念頭に置きつつ日本で行われた既存研究について概観し、Ⅲで第5回全国家庭動向調査の調査デザインについて、第4回調査からの変更点を含めて整理する。続くⅣとⅤで無回答を構成する欠票と不詳について、それぞれ第4回調査と比較しながらその発生状況を明らかにする。さらにⅥでは、平成25年国民生活基礎調査（世帯票）の個票データとのマッチング結果を整理し、Ⅶで全体をまとめる。

Ⅱ. 既存研究

1. 無回答

無回答は社会調査では必ずと言ってよいほど発生する。全国家庭動向調査を含む標本調査の場合、無回答（survey nonresponse）は欠票（unit nonresponse）と不詳（item nonresponse）に区別される（Groves et al. 2001）。このうち欠票とは、標本のうち無効

2) 全国家庭動向調査では、被験者の負担軽減の立場から、第4回調査以降、世帯員全員の続柄や性別、出生年月、従業上の地位といった情報を調査しない形式に改めた。そのため、第2回調査と同じようなデータマッチングを実施することは困難である。

となったものである。これには、調査票への記入状況が悪かった標本だけでなく、調査協力を得られなかった標本も含まれる。それに対して不詳とは、調査票への回答内容が未回答や誤答³⁾などのために集計表で「不詳」や「無回答」等として処理されるものことである。

欠票と不詳はいずれも調査結果の非標本誤差の要因の1つであり、その発生状況を分析することは調査結果を適切に理解するために必要である。そのため、標本調査では欠票や不詳に関する何らかの情報提供がなされることが一般的である。なかには欠票や不詳に関する分析が比較的多くなされている標本調査も存在する。その例として、日本版総合社会調査⁴⁾(朝倉 2006, 2007; 善教 2011; 保田 2005, 2008; 保田・宍戸・岩井 2008; 埴淵ほか2012など)や慶応義塾家計パネル調査(宮内・McKenzie・木村 2005, 2006; 木村 2005, 2006; 直井 2007, 2008, 2009など)を挙げることができる。

欠票や不詳に関する分析は、単にその発生状況を示すにとどまらず、発生を抑制するための知見を得たり、調査結果への影響を評価するといった目的にも資するものである。前者の例としては、欠票の軽減を目指して調査員の行動と欠票との関係进行分析した保田・宍戸・岩井(2008)、質問紙調査とWeb調査の併用によって欠票が減るのかどうかを検討した山本・直井(2011)、欠票や不詳の分析を通じて調査票の改善について論じた山内(2012a)などがある。後者の例としては、調査結果の示す時系列変化に対する欠票増加の影響を論じた佐藤・武下(2009)、欠票を多数含む場合に調査結果から妥当な母集団の推定値を算出する方法を論じた星野(2010)や土屋(2010)、新田(2004)、欠票となった標本の回答を推定する方法を論じた土屋(2005)、不詳の補正方法と分析結果への影響を論じた保田(2000a, 2000b, 2005)などがある。

2. データマッチング

本稿でいうデータマッチングとは、同一標本に対してなされた複数の異なる統計調査の個票情報を共通の識別子を利用し、個票レベルで連結する完全照合(exact matching)のことである。完全照合には他に、同一母集団を対象とした2つの全数調査の個票情報を共通の識別子を利用して連結するものがある。また、完全照合以外のデータマッチングには、統計的照合(statistical matching)と呼ばれるものがある。これは、同一標本或いは同一母集団を対象としてなされた複数の異なる統計調査の結果を共通の識別子がないにもかかわらず標本属性などの情報を利用して連結する場合や、対象の異なる統計調査をあたかも同一対象の統計調査であるかのようにみなして連結するものである。

完全照合によるデータマッチングの研究成果には次のようなものがある。美添・荒木(2000)は、総務省の家計調査と同調査の一部標本を対象として実施される貯蓄動向調査とのデータマッチングを行って、収入、貯蓄、消費支出に関する分析を行った。橋本ほか

3) 例えば、1~4の数値を記入すべきところにそれら以外の数値や文字が記入されているものがある。こうした場合、データクリーニングの過程で不詳に置きかえるといった処理がなされる。

4) JGSS (Japanese General Social Surveys の略称) と称されることも多い。

(2001)は、旧厚生省所管の統計調査についてデータマッチングの実施可能性を整理し、国民生活基礎調査とその後継調査である国民栄養調査の個票データがどの程度マッチングできるのかを検討した。川戸ほか(2003)は、国民生活基礎調査とその後継調査である国民栄養調査とのデータマッチングを行って、生活習慣と自覚症状との関連を分析した。石井(2013)は、国民生活基礎調査とその後継調査である出生動向基本調査とのデータマッチングを行って、子供数と健康に関する諸指標との関連を分析した。

Ⅲ. 第5回全国家庭動向調査の調査デザイン

全国家庭動向調査は、社人研が毎年実施する「社会保障・人口問題基本調査」の中の1つであり⁵⁾、国民生活基礎調査の体系の中で実施されている。図1は平成25年国民生活基礎調査の体系を示したものである。平成25年国民生活基礎調査では、平成22年国勢調査による調査地区のうち、後置番号が1と8の地区から5,530地区を標本抽出し、この地区に居住する全世帯の名簿が作成され、世帯票・健康票の調査が実施される。このように抽出された世帯の標本は図の下側に示された様々な後続調査の親標本となっており、後続調査である第5回調査はここからさらに標本抽出された地区の全世帯に対して実施される。石井(2013)が指摘するように、後続調査は国民生活基礎調査(世帯票)のための親標本を第一相標本とする二相抽出法により標本抽出されており、両調査の結果は原則として完全照合によるデータマッチングが可能である⁶⁾。

第5回調査の実査では、地方自治体や保健所の協力のもと、地方自治体によって選ばれた統計調査員が配票・回収する訪問留置法で実施された。調査票の配布に際しては、原則として標本となった世帯の世帯員に面会し、配布するよう依頼している。調査実施日は2013年7月1日であり、おおむね6月下旬から7月上旬が実査の期間となる。

全国家庭動向調査が想定する回答者は、結婚経験のある女性⁷⁾である。そのため各世帯に配布された調査票への回答は、結婚経験のある女性がいる場合には当該女性⁸⁾が、結婚経験のある女性がいなかった場合には世帯主が回答する⁹⁾。世帯主が回答した調査票も結

5)「社会保障・人口問題基本調査」には、世帯動態調査、出生動向基本調査、人口移動調査、生活と支え合いに関する調査が含まれる。

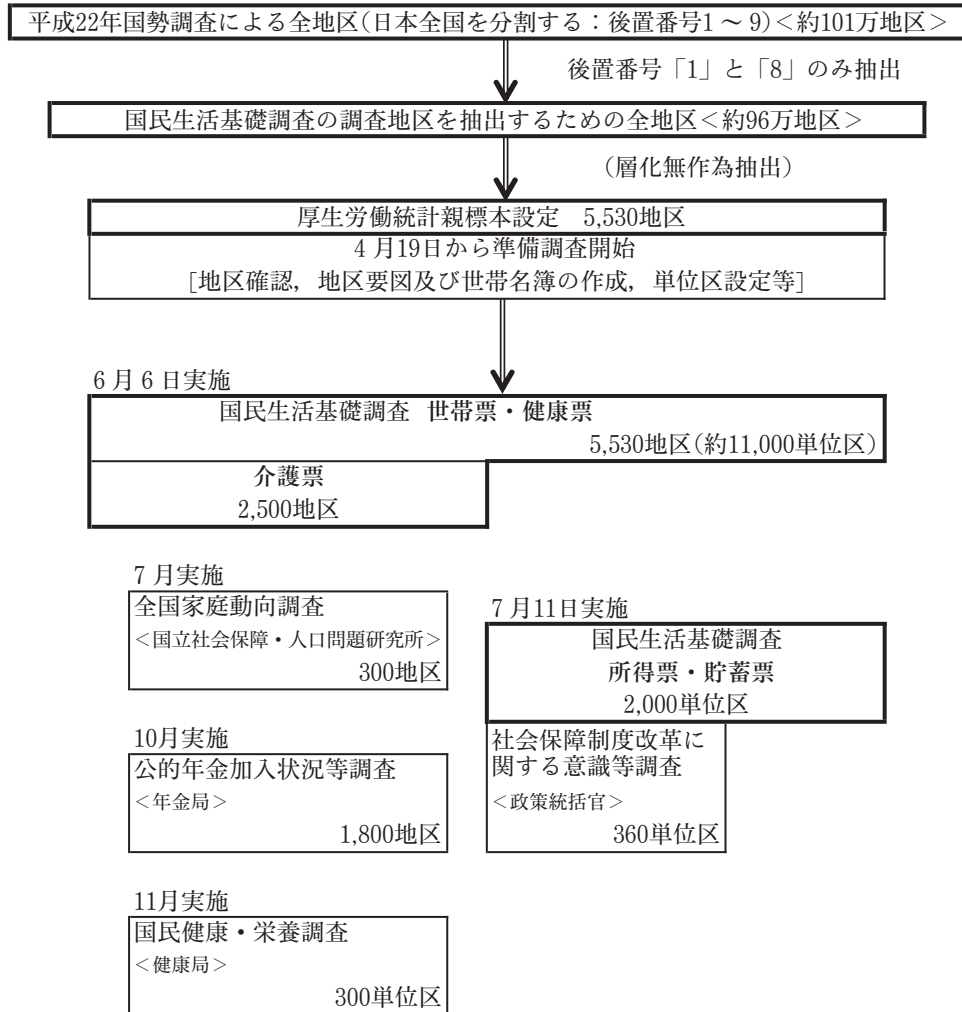
6)ただし、国民生活基礎調査の実施(6月6日)から全国家庭動向調査の実施(7月1日)の間に世帯や世帯員の転出入、世帯の統廃合が起こることや、調査対象者名簿の作成などの対象者の選定に係る問題、世帯員の識別子として用いる性・出生年月の不詳や誤回答、パンチングエラーなどの存在により、完全照合によるデータマッチングには困難な面がある。参考までに、2000年10月1日に実施された国勢調査とその2ヶ月半後の2000年12月15日に実施された事後調査を比較すると、国勢調査の集計対象世帯数57,523のうち事後調査時点で同じ調査区に居住していたのは93.95%であった(松村, 2003)。

7)結婚経験のある女性には配偶関係が有配偶と離死別の場合がある。公表段階では有配偶女性のケースについて集計・分析した結果を報告し(鈴木ほか 2014)、報告書(国立社会保障・人口問題研究所 2015a)を刊行した後、離死別女性のケースについて集計・分析した報告書を作成した(国立社会保障・人口問題研究所 2015b)。なお、第3回調査までは有配偶女性のみを調査対象としていた。

8)結婚経験のある女性が複数含まれる場合、もっとも若い結婚経験のある女性が回答する。

9)結婚経験のある女性がいなかった世帯で世帯主が調査票に記入する場合、回答が必要な設問は世帯の規模と構成に関するもののみである。

図1 平成25年国民生活基礎調査の体系



婚経験のある女性が回答したものと同様に回収されて後述のデータ入力やデータクリーニングが行われるが、集計・分析の段階で除外される。

調査に使用する調査票は1種類のみである¹⁰⁾。調査票は両面印刷されたA4用紙で20枚から成り、このうち設問に関する部分は16枚分、設問は問1から問30まであり、変数¹¹⁾の数にして613である。主要な調査項目は従来同様、①夫婦の人口学的・社会経済的的属性、②両親・子どもに関する事項、③出産・育児、扶養・介護に関する事項、④日常生活でのサポート資源に関する事項、⑤夫の家事・育児に関する事項、⑥夫婦関係に関する事項、

10) 調査票については社人研 HP (<http://www.ipss.go.jp/tosho/questionnaire.html>) または報告書(2015a, pp.155-174)を参照のこと。

11) ここでの変数とは、個票データを構成する項目のことである。例えば、結婚経験のある女性の結婚開始年月に関する問1(4)(ア)に関しては、個票データでは「元号」、「年」、「月」という項目で構成されており、変数の数としては3ということになる。

⑦子どもや家族に関する考え方（意識）に関する事項、⑧資産の継承に関する事項、である。なお、回答が必要な項目は、回答者自身の配偶関係や就業履歴、子の有無や年齢、親の生死、家族介護の状況によって異なる。

調査票回収後は、国立社会保障・人口問題研究所で回収票の点検作業を行い、データの入力とデータクリーニング、集計に必要な変数の作成、集計表作成の順に作業を実施する。データクリーニングには、値範囲が不適切な場合や回答内容に論理エラーがある場合に回答を修正する作業のほか、記入内容に不備が多いものを無効票¹²⁾として除外する作業も含まれる。

第5回調査では、第4回調査の結果を踏まえて主に下記6点を変更した。最初の3点は実査に関わることで、残る3点は調査票に関わることである。

1点目は世帯名簿の記入欄を増やしたことである。世帯名簿とは、国民生活基礎調査の標本となった世帯のリストを全国家庭動向調査実施のために整理するためのもので、調査員が作成する¹³⁾。この世帯名簿は、国民生活基礎調査実施後の転出入の有無や、調査票の配布・回収時の特記事項を記録するためのものである。従来の世帯名簿は、これら情報を記入する欄が1箇所であった。第5回調査では、この欄を2つに分け、調査票の配布と回収のそれぞれの時点での特記事項を別々に記入できるようにした。

2点目は、調査員が記入する調査票の未配布や未回収の理由を標準化したことである。従来から調査票を配布または回収できなかった場合、その理由を世帯名簿に記すように指示していた。しかし、記入された理由は実態に応じて多様な表現が用いられており、担当した調査員による差も大きかった。第5回調査では、未配布の理由として「不在」「長期不在」「拒否」「拒否（病気）」「言語」「その他」の6区分、未回収の理由として「不在」「拒否」「その他」の3区分に整理し、未配布や未回収の場合はいずれかの理由を記すように指示した（国立社会保障・人口問題研究所 2015a, pp.188-190）。

3点目は標本となった世帯の全世帯員が不在時の調査票の配布方法である。第3回調査までは世帯人員数や世帯構成によって調査対象世帯に配布する調査票の部数が異なっていたが、第4回調査以降は世帯人員数や世帯構成にかかわらず配布数は1世帯につき1部となったため、全世帯員が不在であっても調査票の配布が可能となった。第4回調査では、全世帯員が不在の場合には郵便受け等に調査票を含む関係書類一式の配布を認めた（国立社会保障・人口問題研究所 2011, pp.105）。しかし、実際には全世帯員が不在の場合に配布しなかった例も少なからずみられたことを踏まえ、第5回調査では、標本となった世帯の世帯員に面会してから調査票を含む関係書類を直接配布するように指示した（国立社会保障・人口問題研究所 2015a, pp.192）。

4点目は調査票の形式的な変更である。第4回調査まではA3タテで見開き8ページの調査票を使用していたが、第5回調査ではA4横で見開き20ページとした。この変更に関

12) 原則として、世帯構成あるいは記入者となった結婚経験のある女性（または世帯主）の出生年月の回答内容に不備がある場合を無効票とした。

13) 第4回調査と第5回調査の世帯名簿の詳細は報告書（国立社会保障・人口問題研究所 2011, pp.98-104；国立社会保障・人口問題研究所 2015a, pp.183-191）を参照のこと。

わせて、表紙の裏とその次の頁に調査の説明（目的や対象者、調査の仕組み、個人情報取り扱い等）を記し、裏表紙は白紙として調査票の記入内容が見えにくくするようにした。こうした変更は、主として被験者への丁寧な説明と記入のしやすさに配慮したためであり、調査員に対して配布・回収時の調査票の取り扱いを容易にする意図もあった。

5点目は調査票の質問内容の変更である。変更点は細かい点を含めれば幾つかあるが、ここでは特に重要な就業履歴について言及しておきたい。全国家庭動向調査では、結婚経験のある女性の就業と結婚・子育ての関連性を把握する試みがなされてきたが、これまでは就業の開始時期や継続期間を十分に捕捉できていなかった。第5回調査では、初職、結婚を決めた時点、第1子の妊娠が判明した時点、介護の開始時点、調査時点での就業に注目し、その時々での就業の開始時期や継続期間、雇用形態などを尋ねる質問を追加した。

6点目は、回答者の状況によって回答する必要がない設問については、調査票にその旨を明示した。これは主に回答者が離死別女性の場合に当てはまるもので、例えば離死別女性の回答者に対してはかつての配偶者に関する設問への回答が最低限となるようにした。これに伴い、回答する必要がない設問への回答は非該当として処理するようになった¹⁴⁾。

IV. 欠票の発生状況

1. 発生段階別にみた欠票

欠票は以下のように調査実施過程の3つの段階で生じる。①調査拒否等により調査票を配布できない、②調査拒否等により配布した調査票を回収できない、③回収した調査票の記入内容に不備が多い。この①～③を通じた欠票に関する指標として欠票率、各段階で発生する欠票に関する指標として、未配布率、未回収率、無効率を考える。欠票率とは標本（標本となった世帯に調査票を1部ずつ配布する全国家庭動向調査では、配布されるべき調査票数は標本数と一致する）に対する①～③のいずれかの段階で欠票となった調査票数の比であり、未配布率とは標本に対する配布できなかった調査票数の比、未回収率とは配布できた調査票数に対する回収できなかった調査票数の比、無効率とは回収できた調査票数に対する無効となった調査票数の比、を指す（山内 2012a）。未配布率と未回収率は、本来は保田（2008）のいう接触不能と回収不能を表すものであるが、第5回調査では標本となった世帯の全世帯員が不在でも調査票を配布した事例が少なからずみられたため¹⁵⁾、それらとは異なる指標として理解すべきものである。

表1は、有効、無効、未回収、未配布の分布ならびに欠票率、未配布率、未回収率、無効率を、標本の属性との関連で示したものである。標本の属性については、山内（2012a）と同様に世帯人員数¹⁶⁾と居住地域、居住する市区町村の人口規模を用いた。

第5回調査全体でみると、欠票率は36.1%で第4回調査とほぼ同じ値であり、第2回調

14) 第4回調査では該当する設問の結果は他の設問の結果と同様に処理した。

15) 標本となった世帯の世帯員に面会してから調査票を含む関係書類を直接配布するように指示したが、世帯名簿から判断する限り、一部には全世帯員が不在の場合であっても配布された形跡がみられた。

16) ここでの世帯人員数は世帯名簿に調査員が記載した値である。

表1 欠票と標本属性

標本属性	第5回調査										第4回調査				
	総標本数 a=b+c+d+e	有効 b	無効 c	未回収 d	未配布 e	欠票率 (c+d+e)/a	未配布率 e/a	未回収率 d/(a-e)	無効率 c/(a-e-d)	総標本数	欠票率	未配布率	未回収率	無効率	
総数	15,081	9,632	1,548	1,109	2,792	36.1	18.5	9.0	13.8	15,644	36.0	16.8	15.1	9.4	
世帯人員数	1人	3,705	2,328	434	357	586	37.2	15.8	11.4	15.7	4,206	46.9	22.3	25.9	7.6
	2人	4,001	2,803	534	231	433	29.9	10.8	6.5	16.0	4,051	29.9	11.8	8.9	12.7
	3人	2,556	1,890	278	142	246	26.1	9.6	6.1	12.8	2,827	24.4	10.3	7.2	9.2
	4人	1,928	1,513	156	106	153	21.5	7.9	6.0	9.3	2,106	20.1	8.3	7.0	6.2
	5人以上	1,304	1,044	135	36	89	19.9	6.8	3.0	11.5	1,385	19.6	6.4	6.4	8.3
	不明	1,587	54	11	237	1,285	96.6	81.0	78.5	16.9	1,069	99.7	61.6	97.8	66.7
居住地域	三大都市圏	7,693	4,504	671	738	1,780	41.5	23.1	12.5	13.0	7,898	42.7	22.6	18.7	9.0
	非三大都市圏	7,388	5,128	877	371	1,012	30.6	13.7	5.8	14.6	7,746	29.2	10.9	12.0	9.7
居住する市区町村の人口規模	1万人未満	467	357	91	9	10	23.6	2.1	2.0	20.3	789	20.7	6.5	3.5	12.1
	1～5万人	1,937	1,402	303	20	212	27.6	10.9	1.2	17.8	1,451	25.2	8.8	8.2	10.7
	5～10万人	2,212	1,561	251	109	291	29.4	13.2	5.7	13.9	2,004	33.2	17.5	9.0	11.0
	10～30万人	3,311	2,176	298	323	514	34.3	15.5	11.5	12.0	4,003	34.7	16.1	14.4	9.1
	30～50万人	2,311	1,460	195	237	419	36.8	18.1	12.5	11.8	2,505	33.3	13.6	15.1	9.0
50万人以上	4,843	2,676	410	411	1,346	44.7	27.8	11.8	13.3	4,892	45.4	22.9	23.2	7.8	

資料：第4回調査は山内（2012a）

注1）世帯人員数は世帯名簿の値。

注2）居住地域のうち三大都市圏は埼玉県、千葉県、東京都、神奈川県、愛知県、京都府、大阪府、兵庫県とし、他の道県を非三大都市圏とした。

注3）居住する市区町村の人口規模は調査時点の値である。

査から第4回調査にかけてみられた欠票率上昇（山内，2012a）は頭打ちとなった。欠票率の内訳を見ると、未配布率は18.5%、未回収率は9.0%、無効率は13.8%であった。第4回調査と比較すると、未配布率と無効率は上昇したのに対して未回収率は低下した。

第5回調査について欠票率等と標本の属性との関連をみていく。あらかじめ指摘しておきたいのは、世帯人員数が不明の世帯数が第4回調査の1,069から1,587へ518（32.6%）増加したことである。世帯人員数は、調査員が標本となった世帯の世帯員に聞き取りを行って記入する項目である。したがって、世帯人員数が不明の世帯数の増加は標本となった世帯への面会が困難になっていることを反映したものと考えられる。

世帯人員数と欠票率等との関連をみると、世帯人員数が不明の場合に欠票率等は軒並み高い値であり、それ以外では世帯員数が少ないほど欠票率等の値は高くなりやすい。三大都市圏と非三大都市圏を比較すると、前者の方が欠票率、未配布率、未回収率は高く、無効率は低い。居住する市区町村の人口規模についてみると、人口規模が大きい方が欠票率、未配布率、未回収率は高く、無効率はおおむね低い傾向にある。こうした傾向は第4回調査でもおおむね同様であった。一方、第4回調査と比べて第5回調査では世帯人員数が1の世帯で欠票率、未配布率、未回収率の値が低下しているのが目立つ。これについては、1人世帯の協力が得られやすくなった可能性は否定できないものの、本来は1人世帯であるはずの世帯において世帯人員数が不明となったケースが大幅に増えたことが影響している可能性も考えられる。

2. 配布または回収できなかった理由

表2は、実査の過程で発生した未配布または未回収の理由を標本の属性との関連で示したものである。ここに示した理由は調査員が世帯名簿に記入したものを再集計したものであり、調査員によって記入状況が異なることに留意する必要がある。理由のうち、拒否はⅢで述べた「拒否」「拒否（病気）」に相当するもの、不在はⅢで述べた「不在」「長期不在」に相当するもの、その他はⅢで述べた「言語」「その他」に相当するもの、不明は未

表2 未配布と未回収の理由

未配布の理由											
標本属性		第5回調査					第4回調査				
		標本数	構成(%)				標本数	構成(%)			
			拒否	不在	その他	不詳		拒否	不在	その他	不詳
総数		2,792	30.1	49.9	10.4	9.6	2,634	47.6	31.5	15.6	5.2
世帯人員数	1人	586	31.1	55.6	7.7	5.6	940	35.2	43.7	18.3	2.8
	2人	433	51.7	33.5	8.1	6.7	480	65.6	14.6	18.1	1.7
	3人	246	58.1	32.5	4.5	4.9	292	65.8	17.8	12.3	4.1
	4人	153	54.2	35.3	5.2	5.2	175	70.3	10.9	16.6	2.3
	5人以上	89	60.7	25.8	6.7	6.7	88	75.0	13.6	9.1	2.3
	不明	1,285	12.1	59.6	14.3	14.0	659	34.4	40.5	12.0	13.1
居住地域	三大都市圏	1,780	28.8	54.7	8.5	8.0	1,786	45.2	35.4	14.4	5.0
	非三大都市圏	1,012	32.5	41.6	13.5	12.4	848	52.7	23.5	18.0	5.8
居住する市区町村の人口規模	1万人未満	10	10.0	70.0	0.0	20.0	51	64.7	33.3	2.0	0.0
	1～5万人	212	30.2	41.5	25.9	2.4	127	40.9	34.6	24.4	0.0
	5～10万人	291	30.6	34.7	3.1	31.6	351	45.9	31.3	15.4	7.4
	10～30万人	514	32.1	49.4	8.4	10.1	646	50.8	30.3	15.8	3.1
	30～50万人	419	40.6	46.3	6.7	6.4	341	54.5	37.5	2.6	5.3
	50万人以上	1,346	26.2	55.7	11.4	6.7	1,118	44.2	30.1	19.1	6.6

未回収の理由											
標本属性		第5回調査					第4回調査				
		標本数	構成(%)				標本数	構成(%)			
			拒否	不在	その他	不詳		拒否	不在	その他	不詳
総数		1,109	23.4	52.8	6.0	17.9	1,967	33.4	48.2	15.7	2.6
世帯人員数	1人	357	16.5	58.5	6.4	18.5	847	25.1	51.1	21.3	2.5
	2人	231	38.1	37.7	4.8	19.5	318	50.6	30.8	13.5	5.0
	3人	142	29.6	47.9	4.9	17.6	182	63.7	23.1	10.4	2.7
	4人	106	33.0	40.6	4.7	21.7	136	60.3	22.8	14.7	2.2
	5人以上	36	44.4	36.1	5.6	13.9	83	65.1	24.1	8.4	2.4
	不明	237	8.4	69.6	7.6	14.3	401	7.7	81.0	10.0	1.2
居住地域	三大都市圏	738	23.3	59.6	6.9	10.2	1,140	33.2	41.0	21.8	4.0
	非三大都市圏	371	23.7	39.1	4.0	33.2	827	33.7	58.3	7.3	0.7
居住する市区町村の人口規模	1万人未満	9	0.0	66.7	0.0	33.3	26	76.9	11.5	11.5	0.0
	1～5万人	20	50.0	40.0	5.0	5.0	108	43.5	46.3	2.8	7.4
	5～10万人	109	40.4	35.8	7.3	16.5	148	43.2	45.3	11.5	0.0
	10～30万人	323	22.3	38.7	5.9	33.1	482	34.4	32.0	30.7	2.9
	30～50万人	237	14.3	72.2	4.6	8.9	327	40.7	39.1	13.1	7.0
	50万人以上	411	24.3	57.4	6.6	11.7	876	25.9	62.4	10.8	0.8

資料：第4回調査は山内（2012a）

注）標本属性は表1と同じ。

配布・未回収の理由が明記されていないものである。

未配布の理由について、全体をみると、不在がもっとも多い49.9%、次いで拒否が30.1%、その他が10.4%、不明が9.6%の順であった。第4回調査と比較すると、不在の割合が大きく上昇する一方で拒否の割合が大きく低下した。これはオートロックマンションの増加等で標本となった世帯へ接触することが難しくなったことや、前述のように第5回調査では直接会って配布するよう指示したことの影響とみられる。

第5回調査について未配布の理由と標本の属性との関連をみると、世帯人員数が1や不明、三大都市圏、50万人以上の場合に不在の割合が高い傾向がみられる。また、緩やかながら、世帯人員数が多いほど不在ではなく拒否の割合が高くなる傾向はみられた。こうした傾向は第4回調査と比較的類似する。

一方、未回収の理由について、全体をみると、不在がもっとも多い52.8%、次いで拒否が23.4%、その他が6.0%、不明が17.9%の順であった。第4回調査と比較すると、未配布と同様に、不在の割合が上昇する一方で拒否の割合が低下した。

第5回調査について未回収の理由と標本の属性との関連をみると、世帯人員数が1や不明、三大都市圏、30万人以上の場合に不在の割合が高い傾向がみられる。また、世帯人員数が多いと拒否の割合が高い傾向がみられる。第4回調査と比較すると、水準は異なるものの、未回収の理由と標本の属性との関連の仕方は比較的似ている。

V. 不詳の発生状況

不詳の発生状況に関する分析は、集計表作成に利用した客体を対象とし、配偶関係の違いを考慮して有配偶女性と離死別女性に分けて分析する。分析では変数別とケース（集計表作成に利用した客体のこと）別に不詳の発生状況を整理するとともに、データクリーニングによって回答に対してどの程度の修正が加えられたのかを示す。

1. 変数別にみた不詳の発生状況

変数別の不詳割合を整理したのが表3である。ここでの不詳割合は変数ごとに算出したもので、全ケースに占める不詳となったケースの割合である（非該当¹⁷⁾のケースを除く¹⁸⁾。回答者が有配偶女性の場合、不詳割合が10%未満となるのは613ある変数のうちの26.8%、10~20%が33.6%、20~30%が15.0%であり、合わせて30%未満は75.4%であった。これを第4回調査の68.9%と比べると、第5回調査の方が相対的に不詳割合の低い変数が多い。ただし、不詳割合が50%を超える変数、すなわち当該変数への回答が半分以上不詳である変数が全体の15.5%で、第4回調査の7.2%を上回っていることから、不詳割合の高い変数も増えている。

17) 回答する必要が無い変数については非該当として処理される。例えば、子のいない有配偶女性の場合、子の属性に関する設問や育児に関する設問への回答は非該当となる。

18) 全てのケースが非該当になってしまう変数については不詳割合を0とした。

表3 不詳割合別にみた変数の分布

不詳割合	第5回調査				第4回調査			
	有配偶女性		離死別女性		有配偶女性		離死別女性	
	度数	構成(%)	度数	構成(%)	度数	構成(%)	度数	構成(%)
10%未満	164	26.8	203	33.1	178	30.4	55	9.4
10～20%	206	33.6	104	17.0	125	21.4	127	21.7
20～30%	92	15.0	92	15.0	100	17.1	65	11.1
30～40%	45	7.3	67	10.9	89	15.2	86	14.7
40～50%	11	1.8	33	5.4	51	8.7	45	7.7
50～60%	32	5.2	28	4.6	14	2.4	55	9.4
60～70%	15	2.4	10	1.6	11	1.9	26	4.4
70～80%	14	2.3	35	5.7	13	2.2	10	1.7
80～90%	14	2.3	18	2.9	3	0.5	16	2.7
90%以上	20	3.3	23	3.8	1	0.2	100	17.1
計	613	100	613	100	585	100	585	100

資料：第4回調査は山内（2012a）

不詳割合が50%以上の変数をみると（参考表1）、有配偶女性の場合は問9、問13、問14、問15に集中していた。問9は第1子の妊娠が分かったときと末子が小学校に入学したときの就業に関する設問であり、とくに不詳割合が高かったのは末子が小学校に入学したときの就業の継続状況と仕事を辞めている場合の再就業に関する設問である。問13は親の介護認定の現状と変化についての設問で、不詳割合が高かったのは具体的な要介護度や介護認定を受けた時期、病院や施設に入院・入所中の場合の入院・入所以前の居住状態についての設問である。問14は家族の介護の状況と就業との関係に関する設問であり、設問全体で不詳率が高かった。問15は出産や育児などの日常生活で手助けを頼む相手を複数選択する設問であり、不詳割合が高かったのは2人目以降の相手に関する回答である。これら不詳割合の高い変数のうち、問9、問13、問14は今回新たに作成した設問であり、問15に関しては第4回調査でも同様の傾向がみられた設問である。なお、これらの設問には非該当が多数存在する。

一方、離死別女性の場合、有配偶女性に比べて総じて不詳割合の高い変数が多い（表3）。不詳割合が30%未満の変数は65.1%であるのに対して、50%以上は18.6%となっている。第4回調査と比較すると、第5回調査では不詳割合10%未満の変数割合が大幅に上昇したのに対し、90%以上の変数割合が大幅に低下した。これは、Ⅲの最後に触れたように、離死別女性が回答する必要のない設問であることを調査票に明示することで回答を非該当として処理したためである。

不詳割合が50%以上の変数をみると（参考表2）、上述の有配偶女性と共通する変数の他に、問6、問10、問12、問16、問30に不詳割合の高いものがみられた。問6はもっとも最近の結婚に関する設問であり、不詳割合が高かったのは結婚することを決めた時の就業を辞めた場合の再就業に関する設問である。問10は18歳以上の子との関係についての設問であり、不詳割合が高かったのはもっとも世話をしているお子さんかどうかを尋ねた設問

である。問12は親との関係についての設問であり、不詳割合が高かったのは自分の父親との関係についての設問である。問16は出産・育児や介護をする上での不安や苦勞に関する設問で、全般に不詳割合が高かった。問30は夫婦の収入に関する設問で、夫の収入の不詳割合が高い¹⁹⁾。

表4は設問ごとに不詳割合の分布を示したものである。有配偶女性の場合、先に挙げた問9、問13、問14、問15の他に、問10、問11、問12、問16、問27も平均値や最大値が大きい傾向にある。これらは18歳以上の子や両親との関係、出産・育児や介護をする上での不

表4 設問別にみた変数別不詳割合の分布

設問	設問の概要	変数の数	有配偶女性			離死別女性		
			最小値	平均値	最大値	最小値	平均値	最大値
問1	世帯に関する事項	26	0.0	0.0	0.0	0.0	0.0	0.0
問2	配偶関係に関する事項	5	0.0	9.2	19.0	0.0	13.2	26.6
問3	夫婦の基本属性に関する事項	20	0.0	2.4	4.3	0.0	17.2	46.0
問4	夫婦の現在の仕事に関する事項	30	2.2	13.7	22.6	0.0	7.2	19.8
問5	はじめての仕事(妻)に関する事項	10	5.3	13.8	25.1	11.1	28.2	47.3
問6	現在の結婚と結婚が決まったときの仕事(妻)に関する事項	16	2.3	11.1	29.4	11.3	26.8	51.9
問7	子どもの数に関する事項	3	0.7	4.7	12.6	1.8	8.7	21.2
問8	子の基本属性に関する事項	24	2.9	9.4	24.6	6.2	18.1	41.8
問9	第1子の妊娠がわかったときと末子が小学校入学時の仕事(妻)に関する事項	24	3.9	38.0	71.7	9.7	52.0	79.8
問10	18歳以上の子との関係に関する事項	108	4.1	12.5	41.0	9.6	19.7	55.8
問11	親の基本属性に関する事項	24	8.6	22.5	37.8	0.0	14.1	40.9
問12	親との関係に関する事項	92	15.9	25.4	36.3	0.0	21.6	51.1
問13	親の健康状態に関する事項	44	15.6	73.1	98.3	0.0	40.1	97.5
問14	介護に関する事項	46	11.3	66.3	93.7	20.5	81.3	96.7
問15	出産・育児等のサポートネットワークに関する事項	28	18.3	38.6	68.5	34.0	57.5	85.7
問16	出産・子育て、介護の不安・苦勞	4	25.4	36.3	46.3	40.6	51.5	60.9
問17	家族の果たすはたらきに関する事項	2	11.7	12.2	12.7	26.7	27.9	29.1
問18	家族・子どもに関する意識	12	9.0	9.8	10.5	22.5	23.2	24.5
問19	家族の範囲に関する事項	14	7.2	14.9	19.0	26.5	31.2	35.0
問20	家族に対する考え方に関する事項	7	8.1	9.2	10.5	22.4	24.1	25.8
問21	生活満足度に関する事項	4	6.6	8.3	12.8	0.0	19.8	32.9
問22	夫婦の家事に関する事項	18	7.5	10.5	13.9	0.0	0.0	0.0
問23	夫婦の育児に関する事項	21	13.6	17.9	22.5	0.0	0.0	0.0
問24	夫婦の勢力関係に関する事項	4	9.5	11.6	17.6	0.0	0.0	0.0
問25	夫婦のコミュニケーションに関する事項	7	6.3	7.1	8.1	0.0	0.0	0.0
問26	夫から妻への情緒支援に関する事項	6	7.1	7.3	7.8	0.0	0.0	0.0
問27	母親の家事支援に関する事項	2	19.5	21.2	23.0	0.0	22.6	45.1
問28	理想の子ども数に関する事項	1	7.7	7.7	7.7	21.0	21.0	21.0
問29	資産の保有と継承に関する事項	9	6.0	14.3	17.5	16.2	25.5	30.3
問30	夫婦の収入に関する事項	2	13.9	14.4	15.0	20.4	59.3	98.3

19) 離死別女性の夫の収入は、非該当として処理すべきものである。

安や苦勞，母親の家事支援を尋ねるものである。一方，離死別女性の場合，設問全体が非該当の場合を除けば，不詳割合の高低はおおむね有配偶女性と類似するが，総じて不詳割合の平均値や最大値は有配偶女性に比べて大きい傾向にある。第4回調査との設問ごとの比較は，設問の内容や順番が変わっているために困難であるが，同じ設問であっても離死別女性の不詳割合が有配偶女性より高くなる傾向については第4回調査と同様である。

2. ケース別にみた不詳の発生状況

ケース別の不詳割合を整理したのが表5である。ここでの不詳割合はケース別に算出したもので，全変数に占める不詳となった変数の割合である（非該当²⁰⁾の変数を除く）。有配偶女性の場合，不詳割合10%未満が55.8%，10～20%が16.8%，20～30%が9.1%であり，これらを合わせると81.8%となる。第4回調査では，それぞれ44.4%，22.1%，11.1%，77.5%であり，第5回調査の方が総じて不詳割合の低いケースが増えている。

離死別女性の場合，有配偶女性に比べて不詳割合の高いケースの割合が高いものの，やはり第4回調査に比べて不詳割合の低いケースの割合が増えた。例えば第5回調査では10%未満が52.3%，10～20%が17.3%，20～30%が9.5%でこれらを合計すると79.1%となるのに対し，第4回調査ではそれぞれ29.8%，22.1%，11.8%，63.7%であった。

表6はケース別の不詳割合と属性の関係をみたものである。有配偶女性の場合，50歳代以降は年齢が上がると不詳割合の平均が高く，就学年数の長さとは逆相関の関係がみられる。調査時の仕事の有無との関係では仕事ありの方が，末子年齢との関係では末子12歳未満の方がそれぞれ不詳割合は低い。離死別女性の場合も，属性と不詳割合との関連は有配偶女性と同様である。

表5 不詳割合別にみたケースの分布

不詳割合	第5回調査				第4回調査			
	有配偶女性		離死別女性		有配偶女性		離死別女性	
	度数	構成(%)	度数	構成(%)	度数	構成(%)	度数	構成(%)
10%未満	3,577	55.8	466	52.3	3,047	44.4	356	29.8
10～20%	1,078	16.8	256	17.3	1,518	22.1	264	22.1
20～30%	586	9.1	149	9.5	761	11.1	141	11.8
30～40%	385	6.0	108	6.4	423	6.2	117	9.8
40～50%	260	4.1	97	4.6	333	4.8	84	7.0
50～60%	179	2.8	81	3.4	281	4.1	70	5.9
60～70%	136	2.1	53	2.4	179	2.6	56	4.7
70～80%	113	1.8	37	1.9	167	2.4	44	3.7
80～90%	64	1.0	46	1.4	144	2.1	56	4.7
90%以上	31	0.5	25	0.7	17	0.2	7	0.6
計	6,409	100	1,318	100	6,870	100	1,195	100

資料：第4回調査は山内（2012a）

20) 回答する必要が無い変数については非該当として処理される。例えば，子のいない有配偶女性の場合，子の属性に関する設問や育児に関する設問への回答は非該当となる。

表6 ケース別にみた不詳割合の属性別分布

属性	第5回調査						第4回調査						
	有配偶			離死別(夫関連除く)			有配偶			離死別(夫関連除く)			
	度数	平均値	標準偏差	度数	平均値	標準偏差	度数	平均値	標準偏差	度数	平均値	標準偏差	
年齢	29歳以下	226	10.3	14.2	13	18.7	17.5	330	14.9	17.2	15	15.2	15.5
	30-39歳	971	9.5	13.9	111	13.9	14.1	1,369	12.1	15.8	106	16.2	17.4
	40-49歳	1,404	10.0	13.9	133	16.6	19.7	1,401	13.7	15.9	147	15.4	15.4
	50-59歳	1,406	13.2	16.9	178	16.0	18.0	1,603	16.4	16.7	210	19.9	19.4
	60-69歳	1,478	19.8	20.0	271	21.1	20.7	1,366	26.7	21.6	265	24.8	20.6
	70歳以上	924	31.8	25.7	612	36.0	26.6	801	42.2	25.2	452	40.2	26.0
学歴	中学	846	30.7	25.2	389	34.1	24.9	967	35.2	24.7	327	36.6	25.5
	高校	2,619	15.0	17.6	540	21.9	21.3	2,959	19.4	19.5	513	25.7	22.8
	専門	762	14.2	17.4	112	21.5	23.6	695	16.5	18.5	119	21.5	20.2
	高専・短大	1,331	11.3	16.0	156	17.9	17.8	1,284	13.3	15.2	122	17.9	17.4
	大学・院	713	8.5	12.7	70	14.4	18.9	825	13.0	16.1	66	12.9	12.7
仕事の有無	仕事あり	3,523	13.7	17.5	550	17.6	18.8	3,741	17.2	18.8	583	20.4	20.0
	仕事なし	2,745	17.3	20.0	717	30.6	25.6	2,898	22.0	21.8	532	32.6	24.9
末子	末子12歳以下	1,467	8.8	12.8	112	15.0	15.9	1,783	12.1	14.9	110	15.4	15.7
	末子12歳以上	4,089	16.2	18.7	977	22.7	20.8	3,808	20.4	19.4	810	25.1	21.5
	子どもなし	587	17.7	19.2	109	32.6	24.7	636	19.3	22.6	107	34.9	28.6
計	6,409	16.0	19.6	1,318	26.3	24.6	6,870	20.0	20.9	1,195	27.7	24.1	

資料：第4回調査は山内（2012a）

第4回調査と比較すると、有配偶女性の場合はいずれの属性でみても不詳割合の平均値は低下している。低下幅が特に大きいのは年齢の70歳以上や60～69歳である。離死別女性の場合、一部例外はみられるものの、総じて第5回調査の方が不詳割合の平均値は低い傾向にあるが、有配偶女性に比べて差は小さい傾向にある。

3. データクリーニングと不詳の発生

Ⅲで述べたように、データクリーニングの過程で回答を修正することがある²¹⁾。データクリーニングでは回答が適切な値範囲にあるのかどうか、論理的な整合がないかどうかを確認することが主目的であり、問1から順に検討を進め、問題があれば他の設問への回答も含めて記入内容を確認するなどして修正する。ただし、論理的な整合性を確認することが困難な変数については値範囲の確認のみとなる²²⁾。

データクリーニングによる回答の修正には次の3タイプがある。①記入された値を他の値に置き換える（置換）、②記入された値を不詳にする（不詳代入）、③無記入のものに何らかの値を代入する（値代入）。このうち①は不詳に直接影響しないが、②は不詳を増や

21) データクリーニングの作業内容は事前に文書化されている。同文書は原則として非公開だが、データクリーニングを外部の事業者と共同で実施するため、業者選定の際の仕様書には同文書が含まれている。ただし、データクリーニングの過程で当初予定したものは一部異なる作業内容となることもある。

22) このような設問には、家族に関する考え方を尋ねた問18などがある。

表7 変数別，ケース別にみたデータクリーニングによって値を変更した割合

割合	変数別の集計				ケース別の集計			
	有配偶女性		離死別女性		有配偶女性		離死別女性	
	度数	割合(%)	度数	割合(%)	度数	割合(%)	度数	割合(%)
0～2.5%未満	545	88.9	548	89.4	6,038	94.2	1,237	93.9
2.5～5%未満	37	6.0	31	5.1	339	5.3	75	5.7
5～7.5%未満	16	2.6	18	2.9	31	0.5	5	0.4
7.5～10%未満	5	0.8	7	1.1	1	0.0	1	0.1
10%以上	10	1.6	9	1.5	0	0.0	0	0.0
計	613	100	613	100	6,409	100	1,318	100

注1) 変数別の集計は，全ケースに占めるデータクリーニングで値が変更されたケースの割合を各変数について算出し，整理したもの。

注2) ケース別の集計は，全変数に占めるデータクリーニングで値が変更された変数の割合を各ケースについて算出し，整理したもの。

表8 データクリーニングで値を変更した割合が5%以上の変数

有配偶女性					離死別女性						
設問番号	変数記号	データクリーニングで値を変更した割合				設問番号	変数記号	データクリーニングで値を変更した割合			
		計	置換	不詳代入	値代入			計	置換	不詳代入	値代入
問1(3)	Q010301	27.5	7.0	0.0	20.5	問1(3)	Q010301	35.1	3.5	0.0	31.6
問1(3)	Q010302	20.3	5.1	0.0	15.2	問3(2)	Q030205H	24.3	14.3	0.0	9.9
問4(5)	Q040501H	15.5	0.0	0.2	15.2	問1(3)	Q010302	20.6	1.4	0.0	19.2
問4(8)	Q040802H	15.0	14.8	0.3	0.0	問2(2)	Q0202	20.2	0.4	0.0	19.8
問4(8)	Q040802W	13.7	13.6	0.0	0.0	問3(2)	Q030205W	14.3	4.1	0.0	10.2
問5(3)	Q050301	13.3	1.5	1.4	10.5	問5(3)	Q050301	12.4	1.9	2.3	8.2
問6(7)	Q060701	11.3	1.8	1.3	8.3	問5(2)	Q050201	12.1	0.7	2.7	8.7
問4(5)	Q040501W	10.3	0.0	0.1	10.2	問10(11)	Q1011C1	11.5	2.0	2.1	7.4
問10(11)	Q1011C1	10.2	4.0	2.1	4.2	問4(5)	Q040501W	10.2	0.0	0.0	10.2
問3(2)	Q030205W	10.0	2.0	0.0	8.0	問4(8)	Q040802W	9.4	9.4	0.0	0.0
問3(2)	Q030205H	9.9	1.8	0.0	8.1	問6(7)	Q060701	9.2	0.9	1.7	6.6
問5(2)	Q050201	9.1	0.9	1.7	6.4	問9(4)	Q090401	9.0	3.6	3.4	2.0
問9(4)	Q090401	8.6	3.8	2.9	1.9	問1(2)	Q010204a	8.6	0.2	0.0	8.4
問11(4)	Q110401HM	8.6	0.0	0.0	8.5	問11(4)	Q110401WM	8.1	0.0	0.0	8.1
問11(4)	Q110401HF	8.1	0.0	0.0	8.1	問7(1)	Q070101	7.9	0.0	0.0	7.9
問5(2)	Q050202	6.8	4.3	1.9	0.6	問9(3)	Q090301	7.5	0.5	1.2	5.8
問8(2)	Q080201C2	6.5	0.3	0.0	6.2	問9(4)	Q090403	7.3	1.6	3.2	2.5
問9(1)	Q090101	6.4	5.0	1.3	0.2	問6(6)	Q060601	7.1	0.5	2.8	3.7
問11(4)	Q110401WM	6.4	0.0	0.0	6.4	問8(2)	Q080201C2	6.8	0.2	0.0	6.7
問4(8)	Q040801H	6.4	1.6	1.3	3.5	問11(4)	Q110401WF	6.8	0.0	0.0	6.8
問9(4)	Q090403	6.3	1.5	2.5	2.3	問5(2)	Q050202	6.6	2.7	2.9	1.0
問9(3)	Q090301	6.2	0.5	0.6	5.1	問9(1)	Q090101	6.6	4.6	1.5	0.5
問7(1)	Q070101	6.0	0.0	0.0	6.0	問9(2)	Q090201	6.3	1.4	1.5	3.4
問6(5)	Q060503	5.8	2.7	2.0	1.1	問9(5)	Q090501	6.2	0.3	3.3	2.7
問9(2)	Q090201	5.7	1.2	1.1	3.5	問9(4)	Q090404	6.0	0.8	2.5	2.7
問9(4)	Q090404	5.5	0.9	2.3	2.3	問1(1)	Q0101	6.0	2.6	0.0	3.4
問6(6)	Q060601	5.5	0.6	1.9	3.0	問8(2)	Q080201C1	5.7	0.5	0.0	5.2
問9(5)	Q090501	5.1	0.5	2.8	1.9	問1(2)	Q010205a	5.6	0.6	0.0	5.0
問9(1)	Q090103	5.0	2.5	1.2	1.2	問8(3)	Q0803C1	5.5	1.4	0.0	4.2
問10(11)	Q1011C2	5.0	1.1	2.3	1.6	問6(5)	Q060503	5.4	1.8	2.7	0.9
問11(4)	Q110401WF	5.0	0.0	0.0	5.0	問10(11)	Q1011C2	5.4	0.5	2.1	2.7
						問2(1)	Q0201	5.3	2.7	0.0	2.6
						問9(6)	Q090601	5.2	0.4	2.4	2.4
						問5(1)	Q050103	5.1	1.5	2.3	1.3

注1) 全ケースに占めるデータクリーニングで値が変更されたケースの割合を各変数について算出し，整理したもの。

し、③は不詳を減らす。表7、表8、表9は、データクリーニングによって回答の修正がどの程度あるのか、不詳をどの程度増減させたのかを整理したものである。

表7（左欄）には変数別に全ケースに占めるデータクリーニングで値が変更されたケースの割合を算出し、その分布を示した。全体としてみれば、一部の変数を除いてデータクリーニングによる回答の修正はほとんどない。有配偶女性の場合、613変数のうちの88.9%を占める545変数が回答を修正した割合は2.5%未満、5.0%以上となるのは31変数であった。離死別女性の場合もほぼ同じであり、613変数のうちの89.4%を占める548変数が回答を修正した割合が2.5%未満、5.0%以上となるのは34変数であった。

表8は全ケースに占めるデータクリーニングで値が変更されたケースの割合が5%を超える変数について①～③の修正処理の内訳を示したものである。回答を修正した割合が相対的に高い変数として、世帯構成のうち結婚経験のある人数に関する変数（問1(3)）、きょうだい数に関する変数（問3(2)）、通勤時間と帰宅時間に関する変数（問4(5)(8)）、妻の就業履歴に関する変数（問5、問6、問7、問9）、子の性別に関する変数（問8(2)）、もっとも世話した18歳以上の子に関する変数（問10(11)）、両親の生死に関する変数（問11(4)）があり、これらは有配偶女性と離死別女性に共通していた。また離死別女性では、世帯員数と世帯構成に関する変数（問1(1)(2)）、配偶関係に関する変数（問2(1)(2)）も回答を修正した割合が5%を超えた。

これら変数について、上記①～③の内訳をみると、多くの変数で③の値代入の割合が高いのに対し、①の置換の割合が高いのは一部の変数に限られ、②の不詳代入はほとんどの変数で割合が低かった。

表9 データクリーニングで値を変更した割合が5%以上のケース

配偶関係	ID	データクリーニングで値を変更した割合			
		計	置換	不詳代入	値代入
有配偶女性	1	8.6	0.2	7.8	0.7
	2	6.9	0.2	6.4	0.3
	3	6.9	0.7	3.3	2.9
	4	6.7	1.8	0.3	4.6
	5	6.4	2.4	0.0	3.9
	6	6.4	1.0	3.9	1.5
	7	6.2	0.3	5.2	0.7
	8	5.9	0.2	4.2	1.5
	9	5.9	0.5	2.9	2.4
	10	5.9	1.5	1.5	2.9
	11	5.9	0.3	5.4	0.2
	12	5.9	0.0	2.9	2.9
	13	5.9	0.3	5.1	0.5
	14	5.7	0.0	5.4	0.3
	15	5.7	1.6	0.7	3.4
	16	5.7	0.0	4.6	1.1
	17	5.7	1.5	1.6	2.6
	18	5.5	1.6	0.3	3.6
	19	5.5	1.0	3.3	1.3
	20	5.5	0.2	1.6	3.8
	21	5.4	0.8	3.1	1.5
	22	5.4	0.8	0.7	3.9
	23	5.4	1.1	0.2	4.1
	24	5.4	0.0	4.9	0.5
	25	5.4	1.5	0.2	3.8
	26	5.2	0.8	4.1	0.3
	27	5.2	0.0	4.9	0.3
	28	5.1	0.3	0.2	4.6
	29	5.1	1.1	0.0	3.9
	30	5.1	0.0	4.6	0.5
	31	5.1	0.2	0.0	4.9
	32	5.1	1.1	2.8	1.1
離死別女性	1	7.8	3.1	0.2	4.6
	2	6.2	0.7	4.7	0.8
	3	5.5	3.9	0.0	1.6
	4	5.4	0.3	3.8	1.3
	5	5.2	0.5	2.1	2.6
	6	5.1	0.2	4.1	0.8

注) 全変数に占めるデータクリーニングで値が変更された変数の割合を各ケースについて算出し、整理したもの。

②と③の割合を変数ごとに比較すると、ほとんどの変数で③が②を上回り、②が③を上回る場合でも両者の差は1ポイント程度である。したがって、データクリーニングは、全体として不詳を増やすのではなくむしろ減らしていたと言える。

表7（右欄）にはケース別に全変数に占めるデータクリーニングで回答を修正した変数の割合を算出し、その分布を示した。全体としてみれば、一部のケースを除いてデータクリーニングで回答を修正したことの影響はほとんどない。有配偶女性の場合、6,409ケースのうちの94.2%を占める6,038ケースで回答を修正した割合が2.5%未満、5.0%以上となるのは32ケースである。離死別女性の場合も同様で、1,318ケースのうちの93.9%を占める1,237ケースで回答を修正した変数の割合が2.5%未満、5.0%以上となるのは6ケースである。

表9は全変数に占めるデータクリーニングで回答を修正した変数の割合が5%を超えるケースについて、上記①～③の割合を示したものである。同表によればケースによって②の不詳代入の割合が高いものと③の値代入が高いものに分かれる傾向にあり、①は総じて低い。こうした結果は、記入内容に論理的不整合が多くなるのは特定のケースに限られること、本来記入すべき個所であっても回答者にとって自明である場合に設問への回答を省略することが多いのも特定のケースに限られることを反映したものと考えられる。

VI. 国民生活基礎調査とのデータマッチング

データマッチングは、第5回調査を基準として、平成25年国民生活基礎調査（世帯票）の個票を連結させることを目的として実施した。データマッチングの識別子となる変数は地区番号、単位区番号、世帯番号及び妻の出生年月とした。

データマッチングは3段階に分けて行った。第1段階は、全ての標本を含む第5回調査の世帯名簿と平成25年国民生活基礎調査（世帯票）の個票のデータマッチングであり、地区番号、単位区番号、世帯番号を識別子とした。ここでの目的は、第5回調査の世帯名簿に割り当てられた地区番号、単位区番号、世帯番号が平成25年国民生活基礎調査（世帯票）との間で矛盾が無いかどうかを確認することである²³⁾。その結果、300ある地区番号は完全に照合できたが、4つの地区で単位区に不整合がみられた。このうち3地区では、第5回調査の実査の際に平成25年国民生活基礎調査とは異なる単位区番号と世帯番号で世帯名簿が作成されたと考えられるものであった。残る1地区については、抽出されたものではない世帯が調査された可能性が高いことがわかった。世帯番号についても確認したところ、第5回調査の世帯名簿に掲載された15,605ケースのうち、地区番号、単位区番号、世帯番号が全て一致したのが11,695ケース、単位区番号と世帯番号のいずれかが一致しなかった

23) 地区番号、単位区番号、世帯番号は両調査で原則として同じものが利用される。しかし、調査時期が異なるために転出入等が起こることから、両調査に回答した世帯であっても世帯番号が一致しないことが少なくない。単位区番号についても、調査員の変更や転出入等の調査実施上の事情により両調査で一致しない可能性がある。地区番号については両調査で必ず一致するはずのもので、不一致があったとすれば標本とは異なる対象を調査したことを意味する。

のが3,910ケースであった。

第2段階は、第5回調査のうち結婚経験のある女性のいる世帯の個票を基準として、平成25年国民生活基礎調査（世帯票）の個票を照合させるもので、地区番号、単位区番号、世帯番号、妻の出生年月の4つの識別子を用いて照合した。第5回調査の7,727ケースのうち、7,482ケースは地区番号、単位区番号、世帯番号が全て一致した。この7,482ケースのうち、妻の出生年月と共通する出生年月の女性が含まれる平成25年国民生活基礎調査（世帯票）の個票は7,043ケースであった。ひとまずこの7,043ケースはデータマッチングが可能と判断した。

第3段階は、第2段階でデータマッチングできなかった684ケースを対象に、識別子を地区番号と妻の出生年月に減らしてデータマッチングを行い、新たにデータマッチングできた場合に単位区番号や世帯番号の修正可能性を検証した。その結果、新たに170ケースのデータマッチングが可能となり、第5回調査の7,727ケースのうち最終的に7,213ケースをデータマッチング可能と判断した²⁴⁾。

データマッチングの結果を整理したのが図2である。第5回調査のうち結婚経験のある女性がいるケースからみた場合にデータマッチングできたのは93.3%の7,213ケースであった。データマッチングできなかったのは514ケースで、これらは2つに分けられる。1つ目は、平成25年国民生活基礎調査（世帯票）の個票に該当する世帯が存在しないと考えられる個票であり、176ケースあった。2つ目は、平成25年国民生活基礎調査の個票に該当する世帯は存在するが、出生年月の一致する女性が存在しない個票であり、338ケースあった。

一方、平成25年国民生活基礎調査（世帯票）からみた場合にデータマッチングできたのは61.7%の7,213ケースであった。データマッチングできなかったのは4,598ケースで、こ

図2 データマッチングの結果

第5回全国家庭動向調査 （結婚経験のある女性のいる 世帯）の個票 7,727	データマッ チング不可 ①	データマッ チング不可 ②	データマッチング可能 7,213 有配偶女性の個票：5,998 離死別女性の個票：1,215			
	176	338				
平成25年国民生活基礎調査 （世帯票）の個票 11,811			データマッチング可能 7,213	データマッ チング不可 ③ 338	データマッチング 不可 ④ 2,311	データマッチング 不可 ⑤ 1,949

注1) 図中の数値はケースの数を表すが、枠の大きさはそれを正確に反映したものではない。

注2) データマッチング不可①は、平成25年国民生活基礎調査（世帯票）の個票に該当する世帯が存在しないと考えられる個票のこと。

注3) データマッチング不可②は、平成25年国民生活基礎調査の個票に該当する世帯は存在するが、出生年月の一致する女性が存在しない個票のこと。

注4) データマッチング不可③は、第5回全国家庭動向調査の個票に該当する世帯は存在するが、出生年月の一致する女性が存在しない個票のこと。

注5) データマッチング不可④は、第5回全国家庭動向調査の個票に該当する世帯が存在しないと考えられる個票のこと。

注6) データマッチング不可⑤は、結婚経験のある女性のない世帯の個票のこと。

24) 以上の処理に伴い、第5回調査の世帯名簿のうち330ケースについて単位区番号や世帯番号の修正を行った。

れらは3つに分けられる。1つ目は、結婚経験のある女性のいない世帯の個票であり、1,949ケースあった。2つ目は、第5回調査の個票に該当する世帯が存在しないと考えられる個票であり、2,311ケースあった。3つ目は、第5回調査の個票に該当する世帯は存在するが、出生年月の一致する女性が存在しない個票であり、338ケースあった²⁵⁾。このうち2つ目の2,311ケースは、平成25年国民生活基礎調査（世帯票）で捕捉できたにもかかわらず、第5回調査では捕捉できなかったケースと考えられる。

最後にデータマッチング可能と判断した7,213ケースについて世帯人員数ならびに女性の配偶関係の整合性を確認した。その結果、世帯人員数と配偶関係まで一致したのは6,608ケースであった。したがって、世帯人員数と配偶関係まで考慮すると、データマッチングできたのは第5回調査からみた場合は85.5%、平成25年国民生活基礎調査（世帯票）からみた場合は55.9%となる。

Ⅶ. おわりに

本稿では、第5回全国家庭動向調査の無回答の発生状況と、平成25年国民生活基礎調査（世帯票）の個票データとのマッチングについて検討してきた。

まず、第5回調査の無回答のうち欠票、すなわち標本のうち無効となったものについて検討した。標本に占める欠票の割合を意味する欠票率は、第4回調査までは増加傾向にあったが、第5回調査では第4回調査とほぼ同じ36.1%であった。この欠票率を未配布率、未回収率、無効率に分解すると、未回収率は低下していたが、未配布率と無効率は上昇していた。未配布と未回収の理由としては、いずれも標本となった世帯に接触できなかったことを意味する不在によるものが多く、第4回調査より高い割合を占めていた。

次に、第5回全国家庭動向調査の無回答のうち不詳、すなわち集計の際にいわゆる「不詳」や「無回答」等として処理されるものについて検討した。変数別およびケース別に分析したところ、第4回調査に比べ、不詳割合の低い変数やケースが増えた。また、データクリーニングによる調査結果への影響を検討したところ、変数別およびケース別のいずれにおいてもデータクリーニングで回答が修正される割合は小さく、データクリーニングが不詳割合の減少に寄与していることが分かった。

続いて、データマッチングについては、第5回調査を基準に、地区番号、単位区番号、世帯番号、妻の出生年月の4つの識別子を用いて平成25年国民生活基礎調査（世帯票）との照合を試みた。その結果マッチング可能と判断できたデータは、第5回調査のうち結婚経験のある女性のいるケースからみて93.3%、平成25年国民生活基礎調査（世帯票）のうち61.7%であった。

以上を踏まえ、最後に、第5回調査のデータ利用に関する留意事項や今後の調査の課題に関する4点について述べたい。

25) 平成25年国民生活基礎調査（世帯票）には結婚経験のある女性のいない世帯も含まれる。

1点目は、第5回調査のデータ利用に際しての注意点についてである。欠票率は、前回調査とほぼ同じであったとはいえ、それ以前の調査に比べれば依然として高い水準にある。また、今回新たに追加した就業履歴の設問において特に不詳が多く見受けられるなど、不詳割合の高い変数やケースが存在する。こうした点は、第5回調査のデータ利用に際して特に注意を払う必要がある。

2点目は、第5回調査で導入した調査法の変更点に関する評価についてである。欠票率が前回調査とほぼ同じであったことや、不詳割合が変数別・ケース別ともに総じて低下していたことを考えると、調査票をA4サイズに変更して調査に関する丁寧な説明書きを加えたり記入不要の箇所を明示したりといった調査票の変更により一定の効果がみられた可能性がある。その一方で、調査対象世帯の世帯員に直接調査票を含む関連資料一式を配布するように調査員に指示したことが欠票率を高めた可能性もあることから、今後は郵便受け等を利用した配布を認めるなどといった検討を進める必要があるだろう。

3点目は、データマッチングを前提とした調査設計の可能性と限界についてである。データマッチングは、被験者への負担軽減を図りつつ様々な情報を効率的に取得できる利点がある反面、多くの作業時間を要する上にデータマッチングできない個票が発生してしまう問題も抱えている。今回の場合、2カ月程の作業時間を投じながら第5回調査の1割近くの間票についてはデータマッチングできなかった。第5回調査の統計表作成に必須の変数が国民生活基礎調査（世帯票）の設問に必ずしも含まれているわけではないこと、両調査に共通する設問が極力生じないような調査票の見直しを第4回調査の時点で既に行っていること等を考えると、情報損失をできるだけ減らしつつ迅速に調査結果を公表するという立場から言えば、データマッチングを前提として調査を設計することは適切な判断とは言い難い。

4点目は、調査実施上の今後の課題についてである。調査実施上の最大の課題は、調査対象者からできるだけ多くの協力を得ること、その上で調査票に可能な限り正確に回答してもらうようにすること、それを通じて調査結果の精度や信頼性を高めていくことにある。そのため、調査票配布時に標本となった世帯の全世帯員が不在だった場合の対応をはじめとする実査のあり方や、不詳の多い変数については調査票の設計を再度見直すといった検討を継続する必要がある。

参考表 1 不詳割合が50%以上の変数（有配偶女性）

設問	変数記号	不詳割合	設問	変数記号	不詳割合	設問	変数記号	不詳割合
問 9	Q090203	52.2		Q140201	53.4	問15	Q150204	62.4
	Q090207	54.3		Q140202	53.4		Q150303	54.0
	Q090404	51.8		Q140203	53.4		Q150304	67.8
	Q090502	58.8		Q140204	53.4		Q150702	68.5
	Q090503	64.4		Q140205	53.4		Q150802	64.6
	Q090601	55.8		Q140206	53.4	Q150902	58.1	
	Q090602	68.8		Q140207	53.4			
	Q090603	71.7		Q140208	53.4			
問13	Q130202HF	88.9	問14	Q140209	53.4			
	Q130202HM	72.0		Q140210	53.4			
	Q130202WF	79.3		Q140301	55.8			
	Q130202WM	62.0		Q140302	63.5			
	Q130301HF	86.4		Q140401	55.9			
	Q130301HM	68.9		Q140402	65.6			
	Q130301WF	77.6		Q140501	56.7			
	Q130301WM	57.3		Q140502	65.4			
	Q130302HF	97.4		Q140601	54.9			
	Q130302HM	92.3		Q140602	69.8			
	Q130302WF	94.9		Q140603	71.8			
	Q130302WM	87.1		Q140604	70.6			
	Q130303HF	91.4		Q140605	74.2			
	Q130303HM	78.1		Q140701	68.7			
	Q130303WF	85.1		Q140702	79.2			
	Q130303WM	69.1		Q140703	83.7			
	Q130401HF	87.2		Q140704	88.2			
	Q130401HM	69.9		Q140705	89.9			
	Q130401WF	77.9		Q140706	90.3			
	Q130401WM	59.5		Q140707	93.7			
	Q130402HF	94.9		Q140801	74.2			
	Q130402HM	84.8		Q140802	92.9			
	Q130402WF	90.9		Q140803	93.4			
	Q130402WM	77.6		Q1409	92.4			
	Q130403HF	98.3		Q141001	80.0			
	Q130403HM	93.3		Q141002	79.7			
	Q130403WF	96.9		Q141003	80.4			
	Q130403WM	90.4		Q141004	80.4			
	Q130404HF	96.5		Q141101	59.4			
	Q130404HM	90.9		Q141102	59.4			
	Q130404WF	94.1		Q141103	59.4			
	Q130404WM	85.0		Q141104	59.4			
Q1305HF	94.7	Q141105	59.4					
Q1305HM	84.1	Q141106	59.4					
Q1305WF	90.1	Q141107	59.4					
Q1305WM	79.2	Q141108	59.4					

注) 不詳割合の単位は%。

参考表2 不詳割合が50%以上の変数（離死別女性）

設問	変数記号	不詳割合	設問	変数記号	不詳割合	設問	変数記号	不詳割合
問6	Q060703	51.9		Q140201	75.1		Q150103	58.5
	Q090104	51.8		Q140202	75.1		Q150104	67.2
	Q090202	75.1		Q140203	75.1		Q150202	56.9
	Q090203	79.8		Q140204	75.1		Q150203	73.5
	Q090205	53.6		Q140205	75.1		Q150204	82.9
	Q090206	56.7		Q140206	75.1		Q150302	57.4
	Q090207	71.8		Q140207	75.1		Q150303	76.6
	Q090302	51.8		Q140208	75.1		Q150304	85.7
	問9	Q090303		62.8	Q140209		75.1	問15
Q090403		57.0	Q140210	75.1	Q150502	67.9		
Q090404		64.9	Q140301	77.0	Q150602	65.7		
Q090502		59.4	Q140302	83.9	Q150701	52.1		
Q090503		68.2	Q140401	77.0	Q150702	81.7		
Q090601		56.3	Q140402	82.0	Q150802	77.2		
Q090602		69.4	Q140501	77.0	Q150902	73.3		
Q090603		75.0	Q140502	81.4	Q151002	70.2		
問10		Q1011C2	50.9	Q140601	76.5	Q151102	64.3	
	Q1011C3	55.8	Q140602	83.3	Q1602	51.6		
問12	Q1206WF	50.9	問14	Q140603	83.6	問16	Q1603	53.0
	Q120701WF	51.1		Q140604	83.6		Q1604	60.9
	Q120702WF	51.1		Q140605	85.3	問30	Q30H	98.3
	Q120703WF	51.1		Q140701	82.1			
	Q120704WF	51.1		Q140702	91.7			
	Q120705WF	51.1		Q140703	94.6			
	Q120706WF	51.1		Q140704	92.2			
	Q120707WF	51.1		Q140705	92.5			
	Q120708WF	51.1		Q140706	94.5			
Q120709WF	51.1	Q140707	95.8					
Q1208WF	50.2	Q140801	89.3					
問13	Q130201WF	50.0	Q140802	96.4				
	Q130202WF	88.8	Q140803	96.7				
	Q130202WM	74.8	Q1409	96.4				
	Q130301WF	88.8	Q141001	90.4				
	Q130301WM	72.4	Q141002	90.4				
	Q130302WF	96.7	Q141003	90.4				
	Q130302WM	94.8	Q141004	90.7				
	Q130303WF	92.0	Q141101	77.3				
	Q130303WM	80.3	Q141102	77.3				
	Q130401WF	88.8	Q141103	77.3				
	Q130401WM	73.4	Q141104	77.3				
	Q130402WF	94.3	Q141105	77.3				
	Q130402WM	86.9	Q141106	77.3				
	Q130403WF	97.5	Q141107	77.3				
	Q130403WM	94.2	Q141108	77.3				
Q130404WF	96.7	Q141109	77.3					
Q130404WM	92.6							
Q1305WF	96.5							
Q1305WM	86.8							

注) 不詳割合の単位は%.

付記

本研究は、国立社会保障・人口問題研究所「全国家庭動向調査プロジェクト」の研究成果であり、本稿で使用した「国民生活基礎調査」,「全国家庭動向調査」に関する分析結果には、統計法第32条の規定に基づき、調査票情報を二次利用したものが含まれている。

文献

- 朝倉真粧美 (2006) 「無回答の発生要因」『社会学研究科年報』13号, pp.41-54.
- 朝倉真粧美 (2007) 「訪問調査によるデータの偏り—住居形態が及ぼす影響」『社会学研究科年報』14号, pp.69-81.
- Groves, Robert M., Dillman, Don A., Eltinge, John L. and Little, Roderic J. A. (eds.), (2001) *Survey Nonresponse*, J. Wiley, New York.
- 埴淵知哉・中谷友樹・村中亮夫・花岡和聖 (2012) 「社会調査における回収率の地域差とその規定要因：個人および地域特性を考慮したマルチレベル分析」『地理学評論』85巻5号, pp.447-467.
- 橋本修二・川戸美由紀・松村康弘・小栗重統・岡山明・中村好一・柳川洋 (2001) 「保健統計におけるレコードリンケージの実施可能性」『厚生指標』第48巻第11号, pp.1-5.
- 星野崇宏 (2010) 「調査不能がある場合の標本調査におけるセミパラメトリック推定と感度分析：日本人の国民性調査データへの適用」『統計数理』第58巻第1号, pp.3-23.
- 石井太 (2013) 「出生動向基本調査と国民生活基礎調査とのデータマッチングを用いた子ども数の分析」『人口問題研究』第69巻第2号, pp.53-73.
- 川戸美由紀・橋本修二・松村康弘・小栗重統・岡山明・中村好一・柳川洋 (2003) 「国民生活基礎調査と国民栄養調査のレコードリンケージに基づく自覚症状と生活習慣の関連」『厚生指標』第50巻第13号, pp.8-13.
- 木村正一 (2005) 「2004年慶応義塾家計パネル調査の標本特性」(樋口美雄・慶應義塾大学経商連携21世紀COE『日本の家計行動のダイナミズム [I] —慶応義塾家計パネル調査の特性と居住・就業・賃金分析』慶応義塾大学出版会) pp.13-41.
- 木村正一 (2006) 「KHPSにおける回答率の変化とその影響」(樋口美雄・慶應義塾大学経商連携21世紀COE『日本の家計行動のダイナミズム [II] —税制改革と家計の対応』慶応義塾大学出版会) pp.53-71.
- 国立社会保障・人口問題研究所 (2000) 『現代日本の家族変動—第2回全国家庭動向調査 (1998年社会保障・人口問題基本調査)—』調査研究報告資料第15号
- 国立社会保障・人口問題研究所 (2007) 『現代日本の家族変動—第3回全国家庭動向調査 (2003年社会保障・人口問題基本調査)—』調査研究報告資料第22号
- 国立社会保障・人口問題研究所 (2011a) 『現代日本の家族変動—第4回全国家庭動向調査 (2008年社会保障・人口問題基本調査)—』調査研究報告資料第27号
- 国立社会保障・人口問題研究所 (2011b) 『全国家庭動向調査関連資料—第1回～第4回調査のコード表ならびに第4回調査の集計表 (離死別サンプル)—』所内研究報告第35号
- 国立社会保障・人口問題研究所 (2015a) 『現代日本の家族変動—第5回全国家庭動向調査 (2013年社会保障・人口問題基本調査)—』調査研究報告資料第33号
- 国立社会保障・人口問題研究所 (2015b) 『現代日本の家族変動 (離死別編)—第5回全国家庭動向調査 (2013年社会保障・人口問題基本調査)—』所内研究報告第58号
- 厚生省人口問題研究所 (1995) 『1993 (平成5) 年 第1回全国家庭動向調査—現代日本の家族に関する意識と実態—』調査研究報告資料第9号
- 松村迪雄 (2003) 「2000年国勢調査事後調査について」『人口学研究』第33号, pp.139-143.
- 宮内環・McKenzie, C. R.・木村正一 (2005) 「回答行動の分析—調査受託と拒否の選択行動—」(樋口美雄・慶應義塾大学経商連携21世紀COE『日本の家計行動のダイナミズム [I] —慶応義塾家計パネル調査の特性と居住・就業・賃金分析』慶応義塾大学出版会) pp.43-91.
- 宮内環・McKenzie, C. R.・木村正一 (2006) 「パネルデータ継続と回答行動の分析」(樋口美雄・慶應義塾大学経商連携21世紀COE『日本の家計行動のダイナミズム [II] —税制改革と家計の対応』慶応義塾大学出版

- 会) pp.9-52.
- 直井道生 (2007) 「家計の住居移動行動とサンプル脱落問題」(樋口美雄・瀬古美喜・慶應義塾大学経商連携21世紀 COE 『日本の家計行動のダイナミズム [Ⅲ] —経済格差変動の実態・要因・影響』慶應義塾大学出版会) pp.77-98.
- 直井道生 (2008) 「KHPS2007新規対象サンプルの標本特性」(樋口美雄・瀬古美喜・慶應義塾大学経商連携21世紀 COE 『日本の家計行動のダイナミズム [Ⅳ] —制度政策の変更と就業行動』慶應義塾大学出版会) pp.9-35.
- 直井道生 (2009) 「項目回答率とパネル調査回答継続率—KHPS 新規追加サンプルを用いた検証」(樋口美雄・瀬古美喜・照山博司・慶應—京大グローバル COE 『日本の家計行動のダイナミズム [Ⅴ] —労働市場の高質化と就業行動』慶應義塾大学出版会) pp.11-36.
- 新田功 (2004) 「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」『厚生指標』第51巻第11号, pp.24-28.
- 佐藤朋彦・武下朋広 (2009) 家計消費状況調査における調査票回収率の低下による調査結果への影響について—統計調査の民間委託から得られたインプリケーション—. 統計研究彙報第66号, pp.77-99.
- 土屋隆裕 (2005) 「調査不能者の特性に関する一考察—「日本人の国民性第11次全国調査」への協力理由に関する事後調査から—」『統計数理』第53巻第1号, pp.35-56.
- 土屋隆裕 (2010) 「調査への指向性変数を用いた調査不能バイアスの二段補正—「日本人の国民性第12次全国調査」への適用—」『統計数理』第58巻第1号, pp.25-38.
- 山本耕資・直井道生 (2011) 「Web 調査モードの追加は回収率を上昇させるのか—統制群がない場合に選択肢追加の効果を評価する」(樋口美雄・宮内環・C. R. McKenzie・慶應義塾大学パネルデータ設計・解析センター 『パネルデータによる政策評価分析 [2] 教育・健康と貧困のダイナミズム—所得格差に与える税社会保障制度の効果』慶應義塾大学出版会) pp.19-45.
- 保田時男 (2000a) 「クロス集計表における欠損データの分析—学歴移動表を例として—」『理論と方法』15巻1号, pp.165-180.
- 保田時男 (2000b) 「欠損データの分析がもたらす新たな知見—1995年 SSM 調査の学歴移動表分析を例として—」『大阪大学教育学年報』第5号, pp.139-152.
- 保田時男 (2005) 「JGSS データによる父学歴の欠損メカニズムの分析—「わからない」と「無回答」の違い—」『JGSS 研究論文集』4, pp.243-256.
- 保田時男 (2008) 「低下する回収率と回収不能の要因」(谷岡一郎・仁田道夫・岩井紀子 『日本人の意識と行動—日本版総合的社会調査 JGSS による分析—』東京大学出版会) pp.447-458.
- 保田時男・穴戸邦章・岩井紀子 (2008) 「大規模調査の回収率改善のための調査員の行動把握—JGSS における訪問記録の分析から—」『理論と方法』23巻2号, pp.129-136.
- 山内昌和 (2012a) 「第4回全国家庭動向調査の無回答に関する検討」『人口問題研究』第68巻第1号, pp.70-89.
- 山内昌和 (2012b) 「政府統計調査の標本特性の検討—全国家庭動向調査を例に」『新情報』第100号, pp.20-28.
- 美添泰人・荒木万寿夫 (2000) 「完全照合」(松田芳郎・伴金美・美添泰人編 『講座ミクロ統計分析 2: ミクロ統計の集計解析と技法』日本評論社) pp.239-250.
- 善教将大 (2011) 「回答拒否者の論理—JGSS を用いた一般の信頼感と「協力の程度」の分析—」『日本版総合的社会調査共同研究拠点 研究論文集』11, pp.259-271.

An Assessment of Data Quality of *The 5th National Survey on Family: Evaluation of Survey Nonresponses and Data Matching with the Data Set of the 2013 Comprehensive Survey of Living Conditions*

Masakazu YAMAUCHI, Keita SUGA and Jun KIKUCHI

This study has two objectives. First, we assess the data quality of *The 5th National Survey on Family* (NSF), which was conducted by the National Institute of Population and Social Security Research in 2013. Second, we examine the practical feasibility of combining two micro-data sets in a micro-level analysis. For the first objective, we analyze both unit and item survey nonresponses in the NSF. For the second, we perform a micro-level match-merging of the NSF to the 2013 *Comprehensive Survey of Living Conditions* (CSLC), conducted by the Ministry of Health, Labour and Welfare.

A unit nonresponse refers to the complete absence of a sampled unit's response to a survey. In the study, it was found that the overall rate of unit nonresponses in NSFs increased from 24.3% in the 2nd round (1998) to 36.0% in the 4th (2008), and the rate was 36.1% in the 5th NSF (2013). An examination of the data collection process revealed that the contact rate in the 5th NSF worsened from the 4th round, but the response rate in the 5th NSF improved.

Here, the contact rate refers to the proportion of sampled units who could be successfully contacted by field workers, and the response rate refers to the proportion of contacted persons who finally submitted completely filled questionnaires. The reason for unsuccessful contacts was that field workers could not deliver or collect questionnaires due to the lack of effective contacts, and we found that the rate of unsuccessful contacts in both the stages (delivery and collection) increased from the 4th round.

An item nonresponse refers to the absence of valid answers from the respondents. In the study, a higher number of item nonresponses was found in the 5th NSF compared to the 4th. Moreover, item nonresponse rates decreased in terms of both variables and cases in the 5th NSF micro-data set. However, we found that the data cleaning procedure in data processing, which identified inconsistent or illogical answers to determine whether a sensible correction was attainable, generally improved item response rates among both items and responded units.

For the second objective, we note that, in principle, the NSF micro-data can be merged on a one-to-one basis to cases in CSLC, because a common sampling frame is used for both. In practice, however, the exact matching cannot be realized. Indeed, we found that the proportions of matched cases were 93.3% of the 5th NSF units and 61.7% of the 2013 CSLC units.

On the basis of these results, we encourage those using the 5th NSF data to exercise caution and be aware of the limitations. Further, we point out desirable revisions to the survey design, especially in the construction of the questionnaire, for the next round of NSF, which is scheduled to be conducted in 2018.