

特集：第8回世帯動態調査

世帯動態調査における非標本誤差の動向：
50歳未満離家経験者は減少しているのか？

菅 桂 太

世帯動態調査の回収率低下による非標本誤差の拡大が調査結果に及ぼす影響の評価を行う。調査実施年次を挟む2時点の国勢調査の調査区における世帯（主）属性別世帯数の集計表を用いて、国勢調査の調査区という小地域を単位とした世帯数の推計を行い、標本フレームの属性別世帯数を推計した。そして、ポワソン回帰モデルを用いて有効回収率に及ぼす世帯（主）属性の影響を推定し、属性別の世帯回収率ウェイトとその信頼区間を構築した。非標本誤差の影響を評価する調査項目として、実地調査において捕捉が困難な単独世帯の回収状況悪化の影響を顕著に受ける50歳未満の離家経験割合について分析した。

分析の結果、2014～2019年にかけての20歳代の離家経験割合の低下には回収率の低下幅が拡大したことが大きく影響していることが明らかになった。また、20歳代の信頼区間幅は2014年から2019年にかけて顕著に拡大しており、少子化による世代規模の縮小も調査精度に影響を及ぼしていることが示唆された。

キーワード：統計調査の精度、回収率、非標本誤差、離家

I. 課題

本稿では統計行政上の必要性和実証的な意義というやや趣の異なる課題について、世帯動態調査の個票データの再集計を通じて取り組むものである。具体的には、標本抽出元である国勢調査の調査区についての調査票情報の二次利用によって得た統計表を用いて、世帯動態調査の調査対象世帯（以降、「標本フレーム」と呼ぶ）の世帯（主）属性（以降、とくに誤解のない限り「世帯属性」には世帯主属性を含む）別の世帯数を推計し、国勢調査を客体抽出に用いる標本調査の非標本誤差が調査結果に及ぼす影響を評価するための枠組みを提示するとともに、精度改善のための可能性を検討する。調査結果への実証的な影響の評価として、統計調査の実施が困難とされる若年単独世帯の回収状況に推定値が顕著に左右されると考えられる50歳未満の離家経験割合を取り上げる。分析の結果を先取りすれば、国立社会保障・人口問題研究所（2022）が示した2014年から2019年の20歳代を中心とした離家経験割合の低下には、回収状況の悪化が大きく影響していることが明らかになった。

わが国の統計制度は改革の最中にある。1947年に策定された統計法が全面改定され、

2007年5月に公布、2009年4月に全面施行された（西村他 2020：第1～2章）。この新統計法に基づき、2007年10月に内閣府に統計委員会が設置（2016年に総務省に移管）されるとともに、公的統計の整備に関する総合的かつ計画的な推進を図るための「公的統計の整備に関する基本的な計画」（第Ⅰ期2009～2013年度、第Ⅱ期2014～2017年度）が実施されてきた。さらに、2018年には統計委員会に勧告機能を付与する等の措置を講じる統計法の一部改正が行われ、また第Ⅲ期基本計画（2018～2022年度）が策定されている。このようななかで、2018年末から2019年初にかけて毎月勤労統計における不適切処理事案が発覚し、同統計の2004～2018年の遡及推計に基づくと雇用保険等の過少給付が延べ2,000万人に総額約795億円生じる他、追加給付に必要なプログラム改修などの関連事務費に約195億円を要することが明らかになった（西村他 2020：図表4-12、日本統計学会 2019：p.1）。毎月勤労統計の不適切処理事案は、公的統計が社会インフラの重要な部分となっており（西村他 2020：p.614/4317）、公的統計が国の政策の企画立案の根拠となるだけでなく、国民が国の運営の実情を知り、政策を評価し、意思決定に利用するために不可欠の社会情報基盤であることを国民に広く認識させた（日本統計学会 2019：p.1）とされる。2019年1月から統計委員会点検検証部会において政府統計全体の再点検が必要となり、統計委員会建議として勧告された（総務省統計委員会 2019）。また、内閣総理大臣が招集した統計改革推進会議のなかに設置された行政新生部会においても、今後の再発防止及び品質向上のための具体策が策定され（統計改革推進会議統計行政新生部会 2019）、第Ⅲ期基本計画を改定して盛り込まれることになった（総務省 2020）。第8回世帯動態調査は、このような調査環境の2019年7月1日に実施されたものである。

毎月勤労統計における不適切処理事案を発端とした改革において進められている今後の再発防止及び品質向上のための取り組みのなかで、基本となる考えは統計作成プロセスの過程を改善することで品質を向上させるという総合的品質管理である。具体的にはPDCA（Plan-Do-Check-Act）のサイクルを回すことによる不断の品質改善が求められており（総務省統計委員会 2019：p.2-3、統計改革推進会議統計行政新生部会 2019：p.16）、オートロックマンションの増加や一時的に居住している学生の増加、過疎化の進展等による調査環境の悪化といった社会・経済状況の変化に対応できるよう回収率等の調査精度について点検・評価を行って調査計画を改善するというPDCAサイクルの確立が求められている。統計調査の精度検証は人口統計学において最重要課題の一つであり、我が国においても2000年頃からのプライバシー意識の高まりによる影響を危惧する研究（たとえば、菅 2007）や、世帯動態調査に関連するものとしてその主要な利活用目的である世帯推計への影響の検証が行われてきた（たとえば、鈴木 2014）。世帯動態調査は国民生活基礎調査の後続調査として実施されているが、親調査の国民生活基礎調査においても標本設計や非標本誤差、実施方法についての一連の研究が、厚生労働省において実施されている（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 2011、国民生活基礎調査の非標本誤差の縮小に向けた研究会 2018（以降、「2018年国生研究会」）、国民生活基礎調査の改善に関するワーキンググループ 2021（以降、「2021年国生研究会」））。なかでも非標本誤

差の影響を検証した「2018年国生研究会」では、2005年国勢調査の調査区を標本フレームとする2010年国民生活基礎調査の2010年国勢調査における調査対象調査区を特定し、国民生活基礎調査における20～30歳代単独世帯の国勢調査と比べた捕捉率が45%を下回ることを明らかにしている。第Ⅱ章でみる通り、2009年調査から最新の2019年調査にかけて世帯動態調査の回収率は急激に悪化しており、その調査結果への影響の評価が求められているというのが行政上の必要性である。

調査結果への実証的な影響の評価として取り上げる若い年齢層における離家は世帯形成の主要な要素である。世帯動態調査は、世帯の形成・拡大・縮小・解体といった世帯の変化に関する動態データを継続的に収集する唯一の大規模標本調査として¹⁾、国立社会保障・人口問題研究所が「日本の世帯数の将来推計（全国推計）」を実施するために不可欠な配偶関係別・家族類型別世帯内地位に関する推移確率行列を得るために設計され、用いられてきた。世帯数の将来推計結果は、国民生活基礎調査の母集団推定に用いることが検討されている（「2021年国生研究会」）。世帯動態調査の精度が厚生労働行政施策全般の企画及び立案に影響を及ぼす可能性が生じており、精度検証は喫緊の課題と言える。

実証的には、我が国のみならず米国や韓国においても、離家は結婚を促進し、その効果は20歳代前半などの若い年齢ほど強い可能性がある（菅 2009, 스가・조（菅・チョ）2017）。結婚は、伝統的な人口学的分析において出生の最重要な近接要因であるが（Bongaarts 1978）、婚姻と出生が分かちがたく結びついた我が国では、少子化の要因としての結婚の動向の的確な把握が極めて重要であり²⁾、したがって、20歳代などの若い年齢層における離家の動向の正確な把握は重要な政策的含意を有する。

一方で、「2018年国生研究会」が指摘するように20歳代の単独世帯は実地調査が最も困難な集団である。若い年齢層の単独世帯は、ほぼ全員が親元から離家したことがあるため、その捕捉率の低下は調査結果から得られる離家経験割合の低下に直結する。このような世帯動態調査の調査状況が離家経験割合の観測値を過小に評価させる可能性は十分に認識され、1994年（第3回）調査から2004年（第5回）までのそれぞれの調査結果を分析した一連の研究では男女・年齢別に国勢調査の単独世帯割合を復元するウェイトを用いて集計がなされてきた（鈴木 1997, 2003, 2007）。鈴木の一連の研究と比べて本研究は、調査実施年月を挟む2時点の国勢調査の個票データの二次利用によって得た基本単位区（調査区）別の世帯属性別世帯数についての集計表を用いて、標本フレームの世帯数を推計する手法の精緻化を図り、さらに世帯が居住する住宅の種類や世帯主の配偶関係といった属性を考慮することで非標本誤差の補正精度の改善を目指すものである。本稿が提示する標本フレー

1) 総務省が実施する労働力調査を用いて世帯動態の分析を試みた例（山口 2011）があるが、そこでの対象は1年間同じ居所に居住した世帯・世帯員に限定され、結婚・離婚や親世帯からの独立といった世帯間の異動を伴うライフイベントの実態把握は困難である。

2) 日本の出生数に占める嫡出生の割合は、1947年96.2%から1955年98.3%へと急速に上昇し、1970年代は99%を上回る水準であった。2000年代以後においても約98%を維持しており、法的な婚姻をしていないカップルから子が生まれることはほとんどない。岩澤（2008）によれば、日本の出生率が人口置換水準を継続的に下回り続け始めた頃から歴史上最低水準を記録した時期（1975～2005年）における合計出生率の低下の77.7%が初婚（晩婚化・非婚化）の影響による。

ムの世帯数推計手法は、国立社会保障・人口問題研究所が実施する「社会保障・人口問題基本調査」など、国勢調査の調査区を抽出元に利用し、世帯名簿を作成する等によって調査時点の調査対象世帯数が把握されている調査に容易かつ概ね機械的な適用が可能であり、標本抽出の方法が集落抽出で、抽出元の国勢調査から3年以上が経過した時点で実施しようとしている場合に効果的な手法になっている³⁾。

続くⅡ章では世帯動態調査の調査実施状況を振り返り、2014年から2019年にかけて回収率は急速に悪化し、とくに単独世帯の回収率の悪化が著しく、離家経験割合観測値を過小にする影響が大きくなった可能性を指摘する。Ⅲ章で分析の枠組みを概観し、Ⅳ章で分析結果を示す。Ⅴ章でまとめる。なお、本稿は他の調査に分析手法を機械的に応用できることを狙っており、技術的な手法の詳細を含む。技術的側面に関心のない読者には、Ⅲ章の表3まで読み進めていただいた後、Ⅳ章分析結果に進むことを勧めたい。

Ⅱ. 世帯動態調査における非標本誤差の動向

2000年頃から個人情報保護意識の高まりなどを背景に、公的統計においても調査の実施が困難になっている。世帯動態調査と、親調査として約1ヶ月前に実施されている国民生活基礎調査の調査対象世帯数と有効回収世帯数、並びに有効回収率の推移を表1にみた。世帯動態調査の回収状況は国民生活基礎調査より約10%ポイント低いが、回収状況の変化は非常に似通っており、2000年代前半と2010年代を通じて回収率は急速に悪化したことがわかる。

一般に、都市部の若年男性の捕捉が困難である（回収率は低い）（たとえば、増淵他2012）ことはよく知られるが、回収率の推定には標本フレームの世帯属性等についての詳細かつ精確な情報を要するため、容易ではない。冒頭で指摘した「2018年国生研究会」も都市部の若年・単独世帯の捕捉率が低いことの検討を行うことを主要な目的としており、国勢調査を用いて世帯属性別に標本フレームの推定を行っている。世帯動態調査は国民生活基礎調査と標本フレームを共有していることもあり、本稿の分析の基本的な考え方は「2018年国生研究会」に酷似する。そこで、国勢調査との比較を通じた捕捉状況の検証方法について少し詳しく見ておきたい。

3) 従来から国立社会保障・人口問題研究所（社人研）では、以下のような考えによって、これまでのところ標本抽出率をとくに考慮しない集計を行っており、本稿の集計においても標本抽出率は調整しない。社人研の調査では実査を行う調査区を、国勢調査の調査区数に比例して政令指定都市と都道府県別その他地域に配分している。国勢調査の調査区は調査区あたりの世帯数が概ね50となるよう設定されていることから、調査区数に比例させた無作為抽出は、概ね人口規模に比例配分させた無作為抽出となり、標本の単純平均により母平均のバイアスのない推定を行うことができる（たとえば、石井2004）。しかしながら、標本調査の目的は本来的には母集団推定にあり、標本設計の誤差と非標本誤差（標本欠落及び変数値の欠損）を考慮する必要があることから、両面から推定精度を向上させるための検討が国民生活基礎調査においても長らくなされているという指摘を石井太教授（慶應義塾大学経済学部）から受けた。集落抽出法は調査の費用や負担を抑えることができる方法ではあるが、一般には母集団推定の精度は無作為抽出より低くなり、世帯動態調査のように抽出数が300（調査区）という限られた規模の場合の精度は今後慎重に検証を行う必要があるだろう。

表1 調査対象世帯数と有効回収世帯数、有効回収率、国民生活基礎調査における捕捉率

調査実施年次 (調査回)	調査対象 世帯数 ^(注1)	有効回収 世帯数 ^(注2)	有効回収率 ^(注3) (再掲)単独	
A. 世帯動態調査				
1999 (第4回)	15,903	12,434	78.2	—
2004 (第5回)	15,575	10,711	68.8	—
2009 (第6回)	15,666	11,355	72.5	53.4
2014 (第7回)	16,349	11,053	67.6	51.7
2019 (第8回)	16,402	10,213	62.3	45.0
B. 国民生活基礎調査				
1999	56,813	49,403	87.0	—
2004	276,682	220,836	79.8	—
2009	56,882	46,528	81.8	—
2010	289,363	228,864	79.1	—
2014	59,475	46,804	78.7	—
2019	301,334	217,179	72.1	—
C. 国民生活基礎調査における捕捉率				
2010			79.1	61.5

出典：パネル A の世帯動態調査は、国立社会保障・人口問題研究所（2001，2007，2011，2016，2012）。パネル B の国民生活基礎調査は、厚生労働省大臣官房統計情報部（2001，2006，2010，2012，2015）及び厚生労働省政策統括官（統計・情報政策担当）（2020）。パネル C の国民生活基礎調査における捕捉率は国民生活基礎調査の非標本誤差の縮小に向けた研究会（2018：p.6）に掲載された「表. 世帯構造・世帯主の性・世帯主の年齢階級別にみた捕捉率」による。（注1）パネル A の世帯動態調査では、調査区世帯名簿の記載を精査したことにより出典に示した資料と数字が異なる場合がある。（注2）パネル A の世帯動態調査のうち、国立社会保障・人口問題研究所（2016）の記載は回収調査票数であり、回収世帯数ではない。（注3）パネル A とパネル B は有効回収率、パネル C は捕捉率である。パネル A の世帯動態調査における単独世帯回収率はⅢ章3節（1）項で作成したデータセット（後出表2の世帯属性のうち「基本属性」の組み合わせを単位とするもの）を用いて筆者推計。

「2018年国生研究会」では2010年国民生活基礎調査を検証の対象とし、同調査の標本フレーム（抽出元）である2005年国勢調査の調査対象調査区について、基本単位区番号と調査区番号をキーとする突合によって2010年国勢調査区を特定している。その結果、調査対象となった5,510国勢調査区のうち、3,826（69.4%）の2010年国勢調査区が特定され、概ね同時期の国勢調査における世帯主の男女・年齢別にみた世帯の家族類型や住宅の種類といった属性別の世帯数との比較を可能にしている。そして、このようにして得た国勢調査の世帯数に対する国民生活基礎調査の比を捕捉率と呼び、評価尺度としている。表1パネルCには、「2018年国生研究会」報告書から、集計対象全体の捕捉率と単独世帯の捕捉率を引用した。同時に、本稿の分析の結果（Ⅳ章2節）から直接比較可能と考えられる単独世帯の回収確率を示した。

表1によれば、2010年国民生活基礎調査における単独世帯の捕捉率は61.5%であり、同時期の世帯動態調査を比べると、全世帯の回収率と同様に約10%ポイント世帯動態調査の単独世帯回収確率は低いと見られる。しかしながら、単独世帯回収確率の時系列変化のパターンは全世帯とは異なっている。世帯動態調査の全世帯を対象とする有効回収率は2009年以後、5年毎に約5%ポイント低下した。単独世帯の回収確率の低下は2009～2014年は約2%ポイントと相対的に良好であったのに対し、2014～2019年に7%ポイント近く低下し、急速に悪化した。世帯属性別にみた回収確率と分析対象とする変数が統計的に独立

(回収・非回収別の離家経験割合が同じ)ならば、非回収は母集団平均の推定に影響しない。また、繰り返し横断面調査は前回調査からの変化の大きさを測るという時系列変化の実態把握を目的としており、精度に変化がないならば時系列比較についての問題は少ないのかもしれない。しかしながら、単独世帯は概ね全員が離家経験者であると考えられるため、単独世帯の回収状況の悪化は、悪化した分だけ調査結果の離家経験者を確実に減らすことになる。国民生活基礎調査のみならず世帯動態調査においても従来から単独世帯の調査実施が他の集団と比べて困難で、離家経験割合が過小であることは指摘されてきた(鈴木 1997, 2003, 2007)。本稿は、さらに、2014~2019年の回収状況の変化が、離家経験割合観測値をより過小にしたという側面を強調する。

III. 分析手法

本稿の分析においては、回収確率は世帯属性に依存するが、これら世帯属性別にみた有効回収データの離家行動は非回収者の離家行動と変わらないこと(無作為の標本欠損 missing at random, Little and Rubin 2020: 1.3節, Lohr 2022: 8.4節)を、最終的に仮定する。目的は、離家行動に影響し離家経験割合が異なると考えられる世帯属性(これらの世帯属性別にみた個人の離家経験と回収確率は統計的に独立)の別に有効回収確率を推定し、回収確率で割り戻す(あるいは回収確率の逆数をウェイトとして離家経験有無についての二項変数を平均する, Horvitz-Thompson 推定量, Little and Rubin 2020: 3.3節, Lohr 2022: 8.4-8.5節)ことで男女・年齢別離家経験割合を標本フレームにおける水準に還元し、非標本誤差(バイアス)を除去した趨勢を観察することである。具体的には、表2の世帯属性の別に有効回収確率をモデル推定する。

表2 t年世帯動態調査の有効回収確率の推定に用いる世帯(主)属性

世帯(主)属性	変数符号	基本属性
世帯主の男女・年齢5歳階級 ^(注1)	男女別, 25歳未満, 25~29歳, ..., 80~84歳, 85歳以上	○
t年常住地(調査地)の地域 ^(注2)	東京特別区, 政令指定都市, その他・東日本, その他・西日本	○
世帯の家族類型	単独, 夫婦のみ, 夫婦と子, ひとり親と子, その他	○
住宅の種類と所有関係 ^(注3)	一戸建ての持ち家, 共同住宅及び公営/公団の賃貸, その他(民間借家等)	-
世帯主の5年前(t-5年)常住地	現住所(t年調査地), 別住所	-
世帯主の配偶関係	未婚, 有配偶, 死別, 離別, 不詳	-

出典: 筆者作成。(注1) 年齢は出生年に基づき、世帯動態調査実施t年の誕生日を迎えた後のもの。(注2) その他・東日本とは東京特別区と政令指定都市を除く地域のうち中部地方以東の地域、その他・西日本は近畿地方以西の地域を指す。ここでは中部地方の西端は福井県・岐阜県・三重県、近畿地方の東端は京都府・滋賀県・奈良県・和歌山県とした。データセットは(基本単位区×)調査区別に作成し、推定の直前に集計した。(注3) 国勢調査では、住居の種類・住宅の所有関係(8区分)と住宅の建て方(4区分)を再区分した。「共同住宅及び公営/公団の賃貸」とは、都道府県・市区町村営もしくは都市再生機構・公社等の賃貸住宅と、持ち家のうち一戸建て以外の建て方(アパート・マンションなどの共同住宅、長屋建・テラスハウス、その他)を含む。「その他(民間借家等)」には、民間の賃貸住宅、社宅・公務員住宅などの給与住宅、間借り、会社等の独身寮・寄宿舎のほか、施設等の世帯を含む。世帯動態調査では前者は持ち家(マンションなどの共同住宅)と公社・公団・公営などの賃貸住宅、後者は民間借家または賃貸住宅、社宅・公務員住宅などの給与住宅及びその他から成る。

このような目的のためには標本フレームから抽出された調査対象についての、できる限り正確な表2に掲げた世帯の属性別世帯数が必要になるが、全国をカバーする調査時点(t年)の世帯リストは存在しない。世帯動態調査では集落抽出法を用いた標本抽出が行われていること、標本抽出元となる国勢調査から約4年後に調査が実施されていることを鑑み、t-4年とt+1年の2時点の国勢調査の個票データを二次利用することによって得た統計表を用いて標本フレームの世帯を推計する。推計されるのは標本フレームの世帯主リストではなく、調査区別世帯属性別世帯数(集計(カウント)データ)である。そのため、標本フレーム世帯数を推計した後に回収確率を推定する際には、有効回収世帯数を被説明変数とするポワソン回帰モデル(カウントデータに関するもっとも単純なモデル)を用いる。そして、ポワソン回帰モデルを用いた回収確率予測値の95%信頼区間を推定し、回収確率推定値の不確かさが男女・年齢別離家経験割合に及ぼす影響をみる。

分析対象とする期間は、表1でみた回収率が顕著に低下した2009年(第6回)から2019年(第8回)とした。回収率の変化よりも水準自体が調査結果に及ぼす影響に着目するならば、より長期(2000年より前)に対象期間を広げることが望ましい。本稿の分析手法では、2000年以後の国勢調査について公表されている「国勢調査小地域(町丁・字等)境界(都道府県別)」⁴⁾及び大規模調査年の国勢調査にのみ調査項目に含まれる「5年前常住地」を用いるため、2004年(第5回)以前の調査を対象とすることが困難であった。

分析の手順は表3の通りである。最終的な男女・年齢別離家経験割合についてのウェイト有無・種別集計結果は、世帯(主)属性別にみた所属世帯員の離家経験割合と当該属性の世帯回収確率についての複雑な関係の影響を受けることになる。ある世帯属性の回収確率を推定するためには、当該属性の標本フレーム世帯数の推計が必要になるが、調査区単位という小地域の世帯推計を詳細な属性別に行おうとすればするほど、国勢調査もしくは世帯動態調査(のいずれか)における誤回答やプロセッシングエラーといった有限の予算や時間・人員等の資源を用いて実施する実地調査では不可避な誤差(Groves et. al 2011: 2章)の影響が大きくなる。そのため、最低限必要な世帯属性(表2の「基本属性」)のみを用いた分析を表3の手順で一通り試みた後に、他の世帯属性を追加する詳細なデータセットを表3の手順で再構築した。繰り返しとなるが、表3の分析過程で構築するデータセットは、標本フレーム世帯数を推計する世帯属性や有効回収確率を推定する世帯属性によって定義される(調査区別の世帯属性の組み合わせをデータ・レコードとする)集計(カウント)データであることに留意されたい⁵⁾。

4) 政府統計の総合窓口(e-Stat)内の統計地理情報システムからダウンロード可能である(2022年10月28日アクセス:<https://www.e-stat.go.jp/gis/statmap-search?page=1&type=2&aggregateUnitForBoundary=A&toukeiCode=00200521>)。

5) たとえば、2019年調査についての巻末表3のモデル2を推定するためのデータを例に挙げると、特別区・政令都市以外の東日本において5年以上前から調査時の一戸建ての持ち家に居住する夫婦のみ世帯の70~74歳有配偶男性の世帯主が、標本フレームに230人いるという構造のデータを構築し分析に用いる。この場合、有効回収が202世帯からあり、有効回収率は約87.8%となる。

表3 標本フレーム世帯数推計と世帯回収確率推定並びに離家経験割合の集計の手順

番号	主な内容
1	t年世帯動態調査の調査対象となったt-4年国勢調査区に対応するt+1年国勢調査における調査区を特定する。
2	手順1で特定された世帯動態調査の調査区別に、t-4年及びt+1年国勢調査の世帯数を世帯（主）属性別に集計し、世帯（主）属性をキーとして突合する。
3	世帯動態調査の調査区別、世帯（主）属性別の世帯数をt-4～t+1年に線形補間し、t年世帯数推計値（初期値）を得る。
4	世帯動態調査の調査区世帯名簿による調査区別総世帯数に対し、t年世帯数推計値（初期値）を比例的に補正することで、標本フレームの調査地区別にt年世帯（主）属性別世帯数を得る。
5	世帯動態調査の有効回収数を調査区別世帯（主）属性別に集計し、同じ属性別に集計した手順4の推計世帯数に突合する。
6	手順5で作成したデータを用いて、有効回収世帯数を被説明変数（event）、標本フレームのt年世帯（主）属性別世帯数をリスク人年（exposure）とするポワソン回帰モデルを用いて、世帯（主）属性が世帯回収確率に及ぼす影響を推定する。モデル係数についての推定結果を用いて、予測値とその95%信頼区間を計算する。
7	予測回収確率を世帯（主）属性をキーとして世帯動態調査の個人単位個票データに突合する。予測回収確率の逆数をウェイトとして、男女年齢別に世帯員の離家経験割合を集計する。同時に、予測回収確率の95%信頼区間から男女年齢別離家経験割合の上限と下限を算出する。

出典：筆者作成。（注）5年前常住地を用いる場合、手順3で転出入はt-1.5年に起こると仮定する。

以下では、まず世帯動態調査の標本抽出と実施法について簡単に振り返る。その後、国勢調査を用いた標本フレーム世帯数推計（表3の1～4）、回収確率の推定と離家経験割合の集計（表3の5～7）、それぞれの方法について節を改めて述べる。

1. 世帯動態調査の標本抽出及び実施方法

世帯動態調査は、同年の国民生活基礎調査（世帯票（大規模調査年のみ健康票を同時に実施））の対象となった国勢調査区のなかから300調査区を無作為に抽出し、抽出された国勢調査区に居住するすべての世帯を実査の対象とする集落抽出法によって実施されている。国民生活基礎調査は3年毎に大規模調査を実施しており、第8回世帯動態調査を実施した2019年は大規模調査を行う年次であった。大規模調査年次においては約5,500、その他の簡易調査年次においては約1,100の（直近の国勢調査において設定された）国勢調査の調査区を抽出する。層化の基準は、都道府県（政令指定都市がある場合は除く地域）・政令指定都市及び調査区の基本属性（「産業・従業上の地位別就業者数等による調査区分類」、総務省統計局 2022）である。大規模調査年次においては都道府県別結果の表章を行うため各都道府県・政令指定都市に一定数を配分して層化無作為抽出、簡易調査年次においては各層の調査区数に比例配分して層化無作為抽出している。ただし、国民生活基礎調査が調査対象として選定するのは国勢調査の調査区（無人調査区を除く）に付された後置番号が1（一般調査区）と8（おおむね50人以上の単身者が居住している寄宿舍・寮等がある区域）のみであり、後置番号が2～7の社会施設や入院病床数200以上の病院等のある特別調査区や水面調査区は、抽出の際にあらかじめ除外されている。国勢調査の調査区の抽出には、基本属性分類符号が付与された調査区名簿が必要になるが、国勢調査実施の翌年までは利用できないため、最大で6年以上前の調査区名簿から調査対象の抽出が行われて

いるようである。世帯動態調査では4年前の国勢調査の調査区から（無作為抽出された国民生活基礎調査の対象調査区を抽出元として）都道府県・政令指定都市別国勢調査の調査区数に比例配分するよう300調査区の無作為抽出が行われている。

国民生活基礎調査の実査にあたり、まず調査員が対象となった国勢調査区の調査区世帯名簿を4～5月頃に作成する。対象に選ばれた国勢調査区内のすべての世帯について、調査員による訪問留め置き調査が6月上旬に実施される（2019年調査では郵送回収が一部地域に導入され、世帯動態調査にも導入された）。世帯動態調査は、国民生活基礎調査のために作成した調査区世帯名簿の複製を基として、6月中の世帯異動（転出入、単独世帯主の死亡、世帯主交代等）を修正し、調査期日（7月1日）の対象調査区内の世帯数を把握したうえで、すべての世帯を調査している。

2. 国勢調査を用いた標本フレーム世帯数についての小地域推計

まず、 t 年世帯動態調査の対象として抽出された $t-4$ 年国勢調査の調査区について、 $t+1$ 年国勢調査における同じ地域を特定する。その上で、調査時点を挟む2時点の国勢調査の調査対象地区について集計された世帯数を用いて、表4の世帯属性別に標本フレーム世帯数を推計する。なお、表3の手順1にあたる $t+1$ 年国勢調査の調査区の中から $t-4$ 年国勢調査の調査対象地区を特定する方法については、紙幅の関係で「Web付論. 国勢調査における小地域の区分と $t-4$ 年調査対象調査区に対応する $t+1$ 年国勢調査の調査区の特定」とし、『人口問題研究』Webページからダウンロードできるようにした。以下では国勢調査における最小の地域単位である基本単位区番号（9桁）と調査区番号（6桁）のうち主番号（4桁）の組み合わせから成る地域を「基本単位区×調査区」と略記する。また、とくに誤解のない限り主番号（4桁）の示す領域を調査区と呼ぶ。調査対象となった（ $t-4$ 年国勢調査の）300調査区のうち、対応する $t+1$ 年調査区を特定できなかった2009年（第6回）調査の5調査区、2014年（第7回）の5調査区、2019年（第8回）の6調査区

表4 標本フレーム世帯数を推計する（ t 年世帯動態調査の調査対象となった $t-4$ 年国勢調査区と $t+1$ 年国勢調査区の突合に用いる）世帯（主）属性

世帯（主）属性 ^(注1)	変数符号	基本属性
基本単位区番号及び調査区番号 ^(注2)	$t-4$ 年調査区番号に対応する $t+1$ 年基本単位区と調査区番号を別途特定	○
世帯の種類	一般世帯，1人世帯（独身寮入居者を含む），寮・寄宿舎の学生・生徒，病院・療養所の入院者，社会施設の入所者，その他	○
世帯主の男女・出生年	男女別，出生年（1902，1903，…，2003）	○
世帯の家族類型	単独，夫婦のみ，夫婦と子，ひとり親と子，その他	○
住宅の種類と所有関係	一戸建ての持ち家，共同住宅及び公営/公団の賃貸，その他（民間借家等）	—
$t+1$ 年調査における世帯主の5年前($t-4$ 年)常住地	現住所（ $t+1$ 年調査地），別住所	—
世帯主の配偶関係	未婚，有配偶，死別，離別，不詳	—

出典：筆者作成。（注1）有効回収確率の推定に用いる世帯（主）属性（表2）と共通の変数の定義は同一である。（注2）Web付論参照。

を除外する。詳細は Web 付論を参照されたい。

t 年世帯動態調査の対象となった t-4 年国勢調査の基本単位区×調査区と、Web 付論の方法で特定された t+1 年国勢調査の基本単位区×調査区について、表 4 の世帯属性別世帯数を国勢調査の調査票情報の二次利用によって得た。これらを基本単位区×調査区別に、表 4 の世帯属性をキーとして突合した。突合に先立って、t+1 年国勢調査における世帯主の 5 年前(t-4 年)常住地を用いる場合には、5 年前常住地「不詳」の世帯主を同じ調査区内で似通った属性の世帯主に最近傍突合して補綴した⁶⁾。出生年と配偶関係（回収確率モデル推定に用いる場合のみ）の「不詳」については、「不詳」のまま国勢調査間の突合作業を行った。

t-4 年基本単位区×調査区に常住していた世帯主が移動も死亡もせず世帯異動もなかったなら t+1 年にも同じ基本単位区×調査区に常住し t-4~t+1 年国勢調査の突合ができるはずだが、世帯移動、世帯の形成や解体、世帯主変更等による世帯主属性の変化が起これると、t-4 年もしくは t+1 年に常住者のいた世帯属性において他方で常住者なしになって突合できない。t-4 年もしくは t 年に当該の基本単位区×調査区に居住していたにも関わらず、突合できなかった世帯属性は他方の世帯数をゼロとした。その上で、t-4~t+1 年の変化を基本単位区×調査区別、世帯属性別に線形補間して、t 年世帯数を推計する。すなわち、基本単位区×調査区が $j = \{1, \dots, J\}$ 、世帯属性が $k = \{1, \dots, K\}$ の t 年 7 月 1 日推計世帯数（初期値）を $N^c(j, k)$ 、t-4 年 10 月 1 日世帯数を $N^b(j, k)$ 、t+1 年 10 月 1 日世帯数を $N^e(j, k)$ と書くと、 $N^c(j, k) = (N^b(j, k) \times 15 + N^e(j, k) \times 45) / 60$ とする。

この推計では、ある基本単位区×調査区において t-4 年に常住していた世帯属性の世帯が t+1 年まで移動も死亡もせず世帯異動もなかったなら、 $N^b(j, k) = N^e(j, k) = N^c(j, k)$ であり、世帯数は変化しない。一方、t-4 年に常住していた世帯属性の世帯が t+1 年までに転出もしくは世帯解体や世帯主属性の変更でゼロになる場合、 $(N^b(j, k) \times 15 + 0 \times 45) / 60 = N^b(j, k) / 4 = N^c(j, k)$ とする⁷⁾。これは、転出等の世帯異動の発生確率が t-4~t+1 年の間を通じ一定で、t-4~t 年発生率を 3/4 と仮定していることになる。逆に、t-4 年に常住していなかった世帯属性の世帯が t+1 年までに転入や世帯主属性の変更もしくは新世帯形成が生じた場合も同様に、 $(0 \times 15 + N^e(j, k) \times 45) / 60 = N^e(j, k) \times 3/4 = N^c(j, k)$ とする。

6) 同じ調査区内で 5 年前(t-4年)常住地「不詳」の世帯主と、マハラノビス距離が最も近い（5 年前常住地が既知の）世帯主を突合せ、後者の t-4 年常住地で補綴した。マハラノビス距離は多変数間の分散共分散構造を考慮して縮約した 2 つの変数値（の組合せ）の間の距離を表すものであり、世帯人員、世帯の家族類型、世帯の種類（表 4 参照）、世帯主の男女・出生年、配偶関係、最後に行った学校の種類、従業上の地位が似通った世帯主と突合せた。実装には統計解析ソフトウェア Stata の ultimatch (Doherr 2021) を用いた。

7) たとえば、2015 年国勢調査のある基本単位区×調査区に 1947 年生まれ有配偶・男性が世帯主の夫婦のみ世帯しか常住していなかったとしよう。この世帯主が 2015~2020 年に妻を亡くし死別の単独世帯主になって 2020 年国勢調査において同じ住所に常住しており、他に転入がなかったとする。この基本単位区×調査区の 2019 年世帯数推計値は 1947 年生まれ有配偶・男性が世帯主の夫婦のみ世帯が 1/4 世帯、1947 年生まれ死別の男性単独世帯で 5 年前常住地が「現住所」の世帯が 3/4 世帯となる。これは、2015~2019 年の 1947 年生まれ男性世帯主が移動も死亡もせず有配偶・夫婦のみから死別・単独へ推移する確率を 3/4 と仮定・推定することになる。

ただし、 $t+1$ 年調査における世帯主の5年前常住地を世帯属性として用いる場合に、転出・死亡等の世帯異動と転入の両者を線型補間すると、 $t+1$ 年調査において5年前常住地が「現住所」の世帯主と「現住所」以外の世帯主を区別しない場合と同様の推計になるため⁸⁾、5年前常住地という追加的な情報の効果的な利用とは言えない。そのため、 $t+1$ 年調査における世帯主の5年前常住地を用いる場合の転入・転出等の世帯異動は、 $t-4\sim t+1$ 年の期間の中央（したがって世帯動態調査の実施より前）に起こると仮定する。期間の中央ですべての異動が発生するというのは、線型補間と同様に、考える中で最も単純な想定である。具体的には、 $t+1$ 年国勢調査において5年前常住地が「現住所」の世帯属性別世帯数と突合できなかった $t-4$ 年国勢調査の世帯は $t-4\sim t-1.5$ 年に転出等の世帯異動が生じたと推定されるため、 t 年世帯数はゼロと仮定した。 $t+1$ 年国勢調査において5年前常住地が「現住所」以外の世帯は t 年までに転入すると仮定した。5年前常住地が「現住所」の世帯については、住所が同じであっても属性が変化する可能性はあるものの、 t 年においても $t+1$ 年と同じ状態であると仮定した。

世帯動態調査では、調査対象となった国勢調査区の t 年7月1日現在の世帯数について調査区世帯名簿を作成して調査している（Ⅲ章1節）。 $t-4\sim t+1$ 年の線型補間または転入・転出等の世帯異動が期間の中央で起こると仮定することによって得た粗い推計初期値を、世帯名簿の総世帯数に対し調査区別に一律に補正し、最終的な標本フレーム世帯数推計値とする。ただし、この補正を行う前に世帯主の出生年「不詳」の世帯は標本フレームから除外し、補正を通じて、世帯主の出生年「不詳」を、出生年既知の世帯属性分布に比例的にあん分する⁹⁾。また、2009年（第6回）調査において、調査員調査が不能により世帯名簿を作成することができなかった1調査区をこの補正の段階で分析から除外した。

調査区が $i = \{1, \dots, 300\}$ の調査区世帯名簿の総世帯数を $N^{NSHC}(i)$ 、基本単位区 \times 調査区が $j = \{1, \dots, J\}$ 、世帯属性の組み合わせが $k = \{1, \dots, K\}$ の t 年7月1日推計世帯数（補正済み）を $N^a(j, k)$ で表すと、 $N^a(j, k) = N^{NSHC}(i) \times N^c(j, k) / \sum_{j=1}^J \{I(j|i) \sum_{k=1}^K (1 - ma(k)) N^c(j, k)\}$ とする。ここで、 $I(j|i)$ は基本単位区 \times 調査区 j が調査区 i に含まれる時に1をとる指標関数、 $ma(k)$ は世帯属性 k のなかの出生年が「不詳」の場合に1をとる指標関数である。

8) $t+1$ 年10月1日世帯数 $N^e(j, k)$ のうち、5年前常住地が「現住所」で世帯異動のなかった世帯主を $N_{stay}^e(j, k)$ 、5年前の常住地が「現住所」以外（転入）の世帯主を $N_{in}^e(j, k)$ と書くと、転出等の世帯異動があった世帯主は $N^b(j, k) - N_{stay}^e(j, k)$ と推定される。転入・転出等を線形補間して t 年7月1日世帯数を推計すると、 $N^c(j, k) = N_{stay}^e(j, k) + (N^b(j, k) - N_{stay}^e(j, k)) / 4 + N_{in}^e(j, k) \times 3 / 4 = \{N^b(j, k) + (N_{stay}^e(j, k) + N_{in}^e(j, k)) \times 3\} / 4 = (N^b(j, k) + N^e(j, k) \times 3) / 4$ となる。

9) このように世帯主の出生年「不詳」をあん分するのは以下3つの理由による。最初の2つの理由から分析結果への影響は限定的と考えられる。第1に、5年前常住地などに比べると年齢「不詳」の発生は限定的である。第2に、国勢調査における年齢「不詳」は「聞き取り調査世帯」（回収が得られない場合、調査員が近隣の世帯から男女別世帯員数のみを聞き取って世帯名簿に回答することになっている）の可能性が高く、ほとんどの属性が「不詳」になる。そのため、これらの世帯を含む地域別の世帯属性分布は、除外する場合とほとんど変わらない。第3に、世帯動態調査では世帯主の出生年不詳は無効票としており、世帯主の出生年が「不詳」の有効票は存在しない。したがって、有効回収率をモデル推定する際（確率ゼロの事象は予測できないため）、年齢「不詳」の係数は識別できない。

表5 世帯動態調査の調査年次（t年）別 国勢調査による調査対象地区^(注)の総世帯数（t-4年及びt+1年）、世帯動態調査の調査区世帯名簿^(注)の総世帯数と世帯数推計値^(注)（t年）

		2009(第6回)	2014(第7回)	2019(第8回)
A	t-4年国勢調査における世帯動態調査対象調査区の総世帯数	15,582	15,877	16,043
B	t年世帯数推計値（補正前）	15,896	16,519	16,550
C	t年世帯動態調査の調査区世帯名簿	15,504	16,104	16,140
D	t+1年国勢調査において特定された調査対象調査区の総世帯数	16,001	16,733	16,719
E	D-A	419	856	676
F	E/A(%)	2.7	5.4	4.2
G	B-C	392	415	410
H	G/C(%)	2.5	2.6	2.5

出典：筆者算出。（注）世帯動態調査の調査対象となった各回の300調査区のうち、調査員による世帯名簿の作成が不能だった2009年（第6回）の1調査区と、t-4年調査対象地区に対応するt+1年調査区を特定することができなかった2009年（第6回）の5調査区、2014年（第7回）の5調査区、2019年（第8回）の6調査区を除く。

表5に標本フレーム世帯数推計の結果を示した。t-4年及びt+1年国勢調査におけるt年世帯動態調査の調査対象となった調査区の総世帯数は、標本フレーム世帯数推計の最初の手順（表3の手順1，Web付論2節）において特定されたものである。国勢調査の一般世帯総数の増加率は2005～2010年5.7%，2010～2015年2.9%，2015～2020年4.4%とややぶれのある推移をしてきた。同じ期間の世帯動態調査の調査対象調査区について総世帯増加率（表5F行）をみると、2015～2020年4.2%は同水準だが、2005～2010年は一般世帯総数の増加率より3.0%ポイント低く、2010～2015年は2.5%ポイント高い。一方、推計総世帯数と世帯名簿の総世帯数の差率（表5H行）は2009～2019年の間ほとんど変化しておらず、推計世帯数が世帯名簿を2.5～2.6%上回る。この差率の水準を評価するため、調査区世帯名簿の調査区別総世帯数に基づき母集団総世帯数を推定し、その標準誤差率を計算すると¹⁰⁾、2009年は2.2%，2014年と2019年は2.6%であった。したがって、総世帯数でみる限り、Web付論2節のt+1年調査対象調査区の機械的な特定やt-4～t+1年線型補間による粗い推計の誤差はサンプリングエラーで生じる範囲内の水準であり、推計世帯数は標本フレーム世帯数と一定の整合性があると言えるだろう。

3. ポワソン回帰モデルを用いた世帯動態調査における世帯（主）属性別 有効回収率の推定と非標準誤差を考慮した男女・年齢別離家経験割合の集計

(1) 有効回収率を推定するためのデータセットの作成

基本単位区×調査区別に表4の世帯属性別に推計した標本フレーム世帯数を、調査区別に表2の世帯属性別に集計する。そして、世帯動態調査の有効回収世帯数を、調査区別に表2の世帯属性別に集計し、突合する。突合に先立って、世帯動態調査の住宅の種類と所

10) 標準誤差率 SEr は次の式で算出した（Lohr 2022：5.2.3項）。t年世帯動態調査の対象調査区の抽出元であるt-4年国勢調査に設定された調査区総数を M 、世帯動態調査 調査区世帯名簿における調査区 i の総世帯数を $N^{NSHC}(i)$ 、と書くと、 $SEr = SE/\hat{N}$ 。ここで、 $\hat{N} = (M/300)\sum_i N^{NSHC}(i)$ 、 $SE = M\left\{(1-300/M)(1/300)\sum_i (N^{NSHC}(i) - \hat{N}/M)^2\right\}^{1/2}$ である。

有関係「不詳」と世帯主の5年前常住地「不詳」を用いる場合には、同じ調査区内で似通った属性の世帯に最近傍突合を行って補綴した¹¹⁾。

調査区別に表2「基本属性」の世帯属性をキーとして標本フレーム推計世帯数と有効回収数を突合したところ、有効回収数のうち約95%の突合に成功した(表6C行)。突合に用いる世帯属性を増やすと、国勢調査データからは特定が困難な転出入のタイミングや国勢調査もしくは世帯動態調査への誤記入等の誤差の影響の増大を指数関数的に受ける¹²⁾。そのため、突合状況は悪化し、すべての変数を用いる場合には4分の1近くの有効回収世帯を突合できなかった。このような調査区別世帯属性別に標本フレーム推計世帯数と有効回収数が突合できないケースは、国勢調査もしくは世帯動態調査への誤回答等による無作為なものを中心とみられるものの、捕捉が困難な(回収確率が低い)属性の世帯に偏っている可能性もある。その場合、突合できなかった世帯属性の世帯を回収確率推定から除外すると男女・年齢別離家経験割合の集計対象も限定され、バイアスを生じる。非標本誤差の検討のためには、できる限りすべての有効回収票を分析対象とするべきであり、突合に失敗した有効回収世帯は、表2の世帯属性のマハラノビス距離を用いて同じ調査区内で似通った属性の世帯に最近傍突合させた¹³⁾。

なお、表6には、すべての有効回収票を(最近傍)突合させた後の標本フレーム推計世帯数の突合状況も示した。すべての世帯属性を用いて突合を行ったときに有効回収と突合した標本フレーム推計世帯総数は、2009～2014年で僅かに有効回収総数を下回る。標本フレームの真の世帯リストから抽出を行って、各抽出世帯の回収・非回収を観測できるなら抽出世帯数が回収世帯数を下回ることはないため、標本フレーム世帯数が有効回収数を下

11) 国勢調査の5年前常住地「不詳」の補綴で行ったのと同様に、「不詳」の世帯主をマハラノビス距離でみて変数値が既知の最も似通った世帯主に最近傍突合させた。まず、住宅の種類と所有関係についての補綴を実施した。マハラノビス距離の算出に用いた変数は、世帯人員、世帯の家族類型、世帯主の男女・出生年、配偶関係、最後に行った学校の種類、離家経験の有無、世帯主経歴(現世帯主は5年前も世帯主であった不変型(現世帯員の中に5年前世帯主がいない場合)と合併型(世帯員に5年前世帯主がいる場合)、現世帯主が5年前に非世帯主であった場合の交代型(交代時に前世帯主が同居・死亡)と発生型(交代時に前世帯主が別居)、不詳)である。その上で、5年前常住地「不詳」について、世帯の家族類型、世帯主の男女・出生年、配偶関係、最後に行った学校の種類、最後の結婚・死離別が5年前より前か否か、世帯主経歴を用いてマハラノビス距離を算出し、調査区内の住宅の種類と所有関係(一戸建ての持ち家か否か)が同じ世帯主の中で最近傍突合を行った。

12) 表2の「基本属性」に加えて世帯属性を1つずつ追加したときの有効回収世帯に占める突合に成功した世帯割合の変化は、住宅の種類と所有関係の場合には-4.3～-3.5%ポイント、世帯主の配偶関係は-4.5～-3.1%ポイントであったのに対し、5年前常住地は-14.5～-11.6%ポイントとやや大きく悪化した。本稿では、 $t-4 \sim t+1$ 年国勢調査間の t 年世帯数推計において5年前常住地を用いる場合には、単純に世帯動態調査の実施前(期間の中央)に転入や転出等の世帯異動が起こると仮定したが、実際には $t \sim t+1$ 年に転出入する世帯や(5年前常住地が「現住所」であっても)世帯異動による世帯属性の変化が生じていることの影響があると見られる。5年前常住地の効果的な利用を通じた小地域世帯推計の精度向上は今後の主要な課題である。

13) 調査区別にみた表2の世帯属性(データレコードの単位)のほか、突合状態別世帯数(突合不能であった属性の標本フレーム世帯数推計値(N^F)及び有効回収世帯数(N^{ER}))、突合に成功した属性の標本フレーム世帯数推計値の有効回収世帯数からの差(N^{Dif})を加えた。①国勢調査及び世帯動態調査が完全で、②両調査に誤記入等の誤差がなく、③標本フレーム世帯数の推計に誤差がないならば、 N^{Dif} と N^F は世帯動態調査の非回収世帯数に対応する。標本フレーム推計世帯数に突合できない有効回収世帯数(N^{ER})が存在することは、①～③の誤差が生じていることを意味する。 N^F と N^{Dif} の世帯属性が同じならばどちらに突合しても回収確率推定の結果は変わらない。 N^F と N^{Dif} の世帯属性が異なるとき、 N^{ER} を世帯属性が似通った方に突合させるのだが、同程度に似通っているときには非回収世帯数(N^F もしくは N^{Dif})と N^{ER} が近いとき、標本フレームと世帯動態調査の整合性は高いと考える。

回るのは推計誤差ということになる。ただし、ここで推計された世帯数について世帯属性別にみた世帯数が2未満の推計世帯総数に占める割合は、調査年次や推計に用いる世帯属性によらず8割を超えており、標本フレームの個々の世帯（の有無）についての推計に近いものになっている。推計世帯数の有効回収数からの差がマイナスの世帯属性の組み合わせは（標本フレームの世帯属性の組み合わせのうち）約3割弱だが、このうちマイナスの差の絶対値が0.5未満は約6割、1未満は8～9割と、非常に小さくなっている（結果表は割愛）。標本フレーム世帯数の推計誤差ではあるが、むしろ有効回収数が自然数でしか観測できないことによる観測誤差の影響とも言えよう。

t年世帯数推計の最初の手順（Web付論2節）において基本単位区を複数に分割して調査区を設定する場合（Web付論表1の類型B）や調査区が基本単位区境界をまたぐ場合（Web付論表1の類型D）のt-4年基本単位区×調査区に対応するt+1年調査区を特定できていない可能性を指摘した。t-4～t+1年調査区の間で推計された世帯数と世帯動態調査の調査区世帯名簿の世帯数の差は3%未満でサンプリングエラーの範囲内であったこと（Ⅲ章2節）、有効回収世帯数のうち表4の「基本属性」で推計された標本フレームに存在しない世帯属性の世帯は5%未満（表6）というように整合性が高いことを踏まえると、推計世帯数に誤差はあるものの、全般的にはt-4年調査対象地区をt+1年国勢調査において概ね特定することができていると考えてよいだろう。これらの結果は、t-4年調査対象調査区に対応するt+1年調査区を特定するのが困難な場合（Web付論表1の類型BやD）で誤ったt+1年調査区を特定していたとしても、同一の基本単位区内にある隣接調査区を特定しており隣接調査区には似通った世帯が多いためと考えられるが、回収確率を推定するための標本フレーム世帯数の推計という目的は果たしていることを示唆する。

表6 調査区^(注1)別世帯（主）属性別の標本フレーム推計世帯数と世帯動態調査有効回収数の突合結果

	基本属性			全属性		
	2009(第6回)	2014(第7回)	2019(第8回)	2009(第6回)	2014(第7回)	2019(第8回)
A 有効回収数 総数	11,223	10,892	10,099	11,223	10,892	10,099
B 厳格な突合(exact match)成功	10,618	10,288	9,569	8,339	7,776	7,604
C B/A (%)	94.6	94.5	94.8	74.3	71.4	75.3
D 標本フレーム推計世帯数 総数	15,504	16,104	16,140	15,504	16,104	16,140
E 有効回収数と突合 ^(注2)	11,492	11,440	11,220	10,769	10,688	10,376
F E/D (%)	74.1	71.0	69.5	69.5	66.4	64.3
G 有効回収率 (A/D) (%)	72.4	67.6	62.6	72.4	67.6	62.6

出典：筆者算出。（注1）世帯動態調査の調査対象となった各年次の300調査区のうち、調査員による世帯名簿の作成が不能だった2009年（第6回）の1調査区と、t-4年調査対象地区に対応するt+1年調査区を特定することができなかった2009年（第6回）の5調査区、2014年（第7回）の5調査区、2019年（第8回）の6調査区を除く。（注2）厳格な突合ができなかった有効回収世帯を最近傍突合せた後の最終的な状態において、有効回収のある世帯（主）属性について標本フレーム推計世帯数を合計したもの。

(2) ポワソン回帰モデルを用いた有効回収確率の推定と男女・年齢別離家経験割合の集計

前項（Ⅲ章3節(1)項）で作成した世帯属性別標本フレームの推計世帯数と世帯動態調

査の有効回収数を含むデータは調査区別であるが、推定に用いる地域は大都市（特別区、政令指定都市）と非大都市（東日本か西日本の別）という大区分¹⁴⁾とし、推計世帯数と有効回収数を地域別世帯属性別に集計する。その上で、有効回収数を被説明変数、推計世帯数をリスク人年とするポワソン回帰モデルを用いて、有効回収確率に及ぼす世帯属性の影響を推定する。ポワソン回帰モデルはカウントデータの分析に用いる最も単純で広く用いられるモデルである（たとえば、Cameron and Trivedi 2013：3章）。

まず、表2の「基本属性」のみを用いた推定を行い、その他の世帯属性を1つずつ追加してモデル全体の説明力の変化や係数推定値の変化等の影響を検討してモデル選択を行う。その後、モデル係数推定値を用いて、予測回収確率を算出し、その逆数（標本フレーム世帯数÷予測有効回収数）によって世帯有効回収確率ウェイトを作成する。同時に、予測回収確率の95%信頼区間を計算し、回収確率ウェイトの上限と下限を得る。これらの世帯（主）属性単位のウェイトを世帯動態調査の個人単位個票データに世帯（主）属性をキーとして突合し、世帯（主）属性別世帯の所属世帯員全員に適用して、男女・年齢5歳階級別に離家経験割合を18歳以上50歳未満について集計する。離家経験割合は、調査票における「(18歳以上の方について) 親と離れた暮らした経験についておたずねします。この方は、親と離れて別の世帯で暮らしたことがありますか。」に対する回答が「親と離れ別の世帯に住んだことがある」である場合に1、「生まれてからずっと親と同居」の場合に0をとる二項変数についての不詳を除く平均である。

なお、回収確率推定値の統計学的な不確かさによって離家経験割合が確率的に取り得る信頼区間は、集計対象個人の離家経験有無を用いて離家経験割合が最大・最小になるウェイトを作成した。すなわち、離家経験割合が最大になるウェイトは、離家経験者の回収確率ウェイトが信頼区間上限（回収確率は信頼区間の下限）、かつ離家未経験者の回収確率ウェイトが信頼区間の下限（回収確率は上限）の場合とした。離家経験割合が最小になるウェイトも同様とした。

IV. 分析結果

1. ポワソン回帰モデルを用いた有効回収確率の推定

世帯有効回収確率モデル（共変量の）選択に関し、表7に共変量を1つずつ用いて計算した赤池情報量基準（AIC）を示した¹⁵⁾。一般に、異なったデータ間でAICを比較することはできないため、表2の「基本属性」のみを含むデータとすべての共変量を含むデータを用いる場合の2通り（それぞれ表7の「モデル1」と「モデル2」）に示した。

表7（2～8行目）の共変量を1つだけ用いるAICを比較すると、モデル1のデータ

14) 大都市地域は単独世帯が多くかつ調査が困難とされる。その他の地域については、総じて東日本よりも西日本の離家経験タイミングは早い。

15) この結果表については、国勢調査をはじめとする統計調査において、特定の変数の情報量の豊富さや有用性、利活用の状況は当該項目についての調査を実施する際のコストパフォーマンスを測る重要な指標であるという指摘を津谷典子教授（慶應義塾大学）から受けた。

では世帯の家族類型，モデル2のデータでは世帯主の配偶関係のAICがすべての調査年次で最も小さく，予測パフォーマンスがよい．次善の変数は，モデル1のデータでは世帯主の年齢5歳階級，モデル2のデータでは調査年次によって若干の違いはあるが家族類型と住宅の種類と所有関係が優れていた．「基本属性」に共変量を追加する場合においても（表7の9～11行目），世帯主の配偶関係の予測パフォーマンスがよい．

ただし，世帯主の配偶関係については，配偶関係「不詳」の回収確率（の低さ）を説明する効果が卓越している．他の共変量の「不詳」はデータ作成過程で補綴しており，記入状況の悪さが世帯主の配偶関係「不詳」に集約される形になっている．そこで世帯主の配偶関係「不詳」の世帯を除外して推定を行ってAICを再計算したところ，モデル2のデータにおいて共変量を1つだけ用いる場合には，2009年は住宅の種類と所有関係，次いで家族類型，2014年は世帯主の5年前常住地，次いで住宅の種類と所有関係，2019年は住宅の種類と所有関係，次いで年齢5歳階級が優れていた（結果表は割愛）．「基本属性」に共変量を追加する場合には，2009年と2014年は世帯主の5年前常住地，2019年は住宅の種類と所有関係の予測パフォーマンスがよくなっていた．この結果は，表2の世帯属性のうち，世帯を区別して，標本調査の欠票（有効回収確率）の違いを予測する情報量が多いのは「基本属性」のなかでは世帯の家族類型，すべての世帯属性のなかでは住宅の種類と所有関係である場合が多いことを示唆する．ただし，調査年次によって各共変量の予測パフォーマンスは若干異なること，単変量の係数推定値と多変量を用いる係数に不整合はなかったため，「基本属性」のみを用いるモデル1と全共変量を同時に用いるモデル2を，すべての調査年次について推定し，モデル間の違いと離家経験割合への影響を検討した．

表7 有効回収確率に関するポワソン回帰モデルに用いる 世帯（主）属性の組み合わせ別にみた赤池情報量基準

行番号	共変量有無 ^(注1)							赤池情報量基準 (AIC) ^(注2)					
	男女	年齢階級	地域	家族類型	住宅	5年前常住地	配偶関係	2009(第6回)		2014(第7回)		2019(第8回)	
								モデル1	モデル2	モデル1	モデル2	モデル1	モデル2
1	○	○	○	○				2,344.3	8,045.1	2,377.5	8,467.3	2,390.7	8,264.2
2	○							3,073.3	8,774.1	3,227.9	9,317.7	3,512.9	9,386.3
3		○						2,898.4	8,599.2	2,864.2	8,954.0	3,030.3	8,903.7
4			○					3,053.6	8,754.4	3,238.2	9,327.9	3,499.7	9,373.2
5				○				2,653.8	8,354.5	2,850.9	8,940.7	2,935.7	8,809.1
6					○			-	8,358.6	-	8,857.4	-	8,691.5
7						○		-	8,547.6	-	9,004.5	-	9,079.1
8							○	-	8,185.4	-	8,561.4	-	8,103.0
9	○	○	○	○	○			-	7,978.7	-	8,403.1	-	8,165.7
10	○	○	○	○		○		-	7,969.1	-	8,387.2	-	8,215.4
11	○	○	○	○			○	-	7,879.7	-	8,180.2	-	7,696.6
12	○	○	○	○	○	○	○	-	7,767.2	-	8,071.1	-	7,587.9

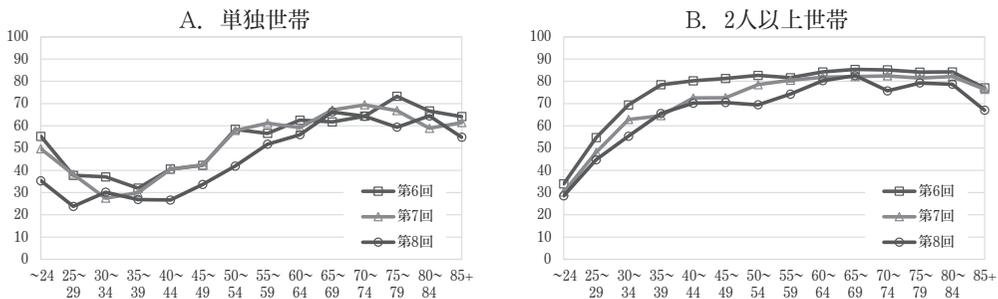
資料：t年世帯動態調査，並びにt-4～t+1年国勢調査に基づく世帯動態調査の調査対象調査区の世帯数推計値を用いて筆者算出．（注1）表2の世帯（主）属性をモデル推定に含めるか否かを示す．（注2）AICはデータに対するモデル予測のパフォーマンスを測る指標である（小さい方が優れている）．共変量有無に「○」のある共変量を含めたモデルのAICを示す．一般に，異なるデータでの比較はできないため，巻末表2のモデル1と2の推定に用いるデータそれぞれで算出した．

有効回収確率に関するポワソン回帰モデルの係数推定値の指数値（発生率比）は巻末表2、記述統計量を巻末表3に示した。発生率とはリスク人年（ここでは標本フレーム推計世帯数）あたりの発生率（すなわち有効回収確率）であり、定数項の発生率比はすべての共変量がゼロのときの発生率（ベース・ライン発生率）、各共変量の発生率比はリスク人年あたりの発生率のベース・ライン発生率に対する比を示している。発生率比の大きさには調査年次間で若干の差があるものの、回収確率への各共変量の影響の方向性と統計的な有意性は調査年次間で一貫している。

具体的には、50歳未満（とくに20歳代）、東京特別区に居住、単独世帯、共同住宅や民間借家等に居住、5年前常住地が「現住所」以外の（転入した）世帯主の回収確率は低い。逆に、大都市以外に居住、夫婦と子世帯、（有配偶と比べて）死別の世帯主の回収確率は統計的に有意に高い。とくに、（東京特別区居住者と比べて）非大都市居住者の回収確率は1.5倍以上、（年齢やその他共変量を一定にし、単独世帯と比べた）夫婦のみ世帯や夫婦と子世帯の回収確率は約1.5倍以上になっていた。逆に、2019年調査（モデル1）では、世帯主年齢40歳代前半以下の年齢層の回収確率は65～69歳と比べ25%以上低く、なかでも25～29歳の回収確率は半分未満であった。これら年齢や東京特別区、単独世帯の回収確率ウェイトは大きくなり、調査結果への影響も大きくなる。なお、ポワソン回帰モデル推定結果（巻末表2）によれば、（夫婦と子世帯と比べた相対的な）単独世帯の回収確率は2009～2014年についても低く、2019年に突然調査が困難になったわけではない。

2. 単独世帯の年齢別 予測有効回収確率

図1には、巻末表2のモデル1を用いて計算した単独世帯か否かの別に世帯主の年齢別予測回収確率を示した。40歳代以下の単独世帯の回収率は50%を下回っており、2019年調査の50歳代以下で2009～2014年と比べて低下していることが目立つ。50歳以下の単独世帯の回収率は、30歳代前半を除けば、2009～2014年ではほとんど変化がなかった。しかし、2014年から2019年にかけて20歳代では14%ポイント以上低下しており、最も低い25～29歳では23.8%になった。40歳代前半や50歳代前半においてもそれぞれ14%ポイントと16%ポイント低下している。



資料：巻末表2（モデル1）の係数推定値を用いて筆者算出。

図1 単独世帯か否かの別にみた世帯主の年齢別有効回収確率

また、2人以上世帯においても、世帯主の年齢が30歳代後半を中心に2009年から2014年にかけて（35～39歳で14%ポイント）低下、50歳代前半において2014年から2019年にかけて回収率は約9%ポイント低下した。離家未経験の20歳代以下の子は50歳代前半の世帯主の世帯に同居していることも多く、世帯主が50歳代前半の2人以上世帯の回収状況の悪化は観察される離家経験割合を引き上げる可能性がある。これは若年単独世帯の回収率の低下が離家経験割合観測値に及ぼす影響を打ち消す方向に作用するものではあるが、回収数の低下は標準誤差を増大させ、調査結果の精度を下げる可能性がある。

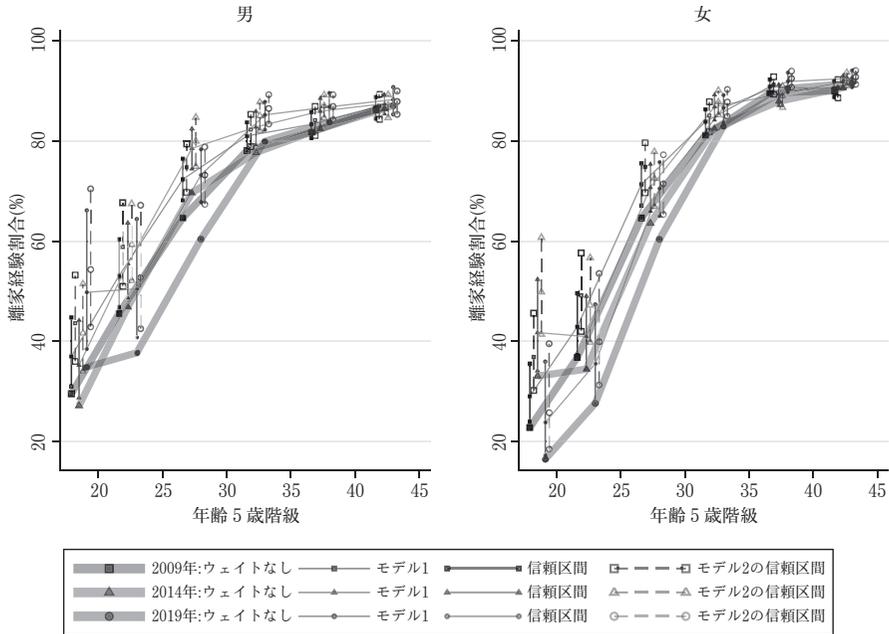
回収率の水準が低くても悪化していなければ、同等のバイアスを及ぼし、一定の留保の上で同等の質の時系列比較が可能かもしれない。しかし、単独世帯回収率の低下は調査結果から観測される離家経験割合の低下に直結するため時系列比較を困難にする。

3. 有効回収確率の離家経験割合への影響

世帯回収確率に関するポワソン回帰モデル（巻末表2）の種類（モデル1とモデル2）別の予測回収確率ウェイトを用いて集計した男女・年齢別離家経験割合を巻末表1とともに図2に示した。離家経験割合の信頼区間はⅢ章3節（2）項で説明した方法で算出した。

男女・年齢別離家経験割合を2009～2019年の調査年次間、ウェイト有無・種類の別に比較検討した結果から、少なくとも以下の4点を指摘したい。まず、国立社会保障・人口問題研究所（2022）で示したように世帯回収確率ウェイトを用いない単純集計では、2014～2019年にかけて男性20歳代と女性の20歳代前半で、離家経験割合は大きく低下した。この低下が大きかった男女年齢層で、回収確率ウェイトを用いる場合の（ウェイトなしに対する）離家経験割合の変化が著しい。世帯属性として表2の「基本属性」を回収確率の予測に用いるモデル1の推定結果によるウェイトを用いて集計した離家経験割合をウェイトなしの割合と比較した変化幅は、男性20歳代で2014年調査の約9%ポイントに対して2019年調査は約13%ポイント上昇、女性20歳代でも2014年調査の約7%ポイントに対し2019年調査は約8～10%ポイント上昇し、非標本誤差の拡大が2019年調査で観察された離家経験割合を過小にしており、その影響は2014年調査と比べても2019年調査で拡大した可能性を示唆する。

第二に、回収確率を予測するモデルの違い（表2「基本属性」以外の共変量を用いるか否か）は、調査年次と男女年齢によって多少の違いはあるものの、総じてウェイトを用いるか否かの差よりも小さい。回収確率モデル係数推定値の統計学的な不確かさに基づく信頼区間には一定の幅があり、信頼区間を考慮すると、ほとんどの調査年次と男女・年齢において「基本属性」のみを用いるモデル1と全変数を用いるモデル2の信頼区間は相互に重なっている。一方、男女ともすべての調査年次において20歳代（及び2014年以後の30歳代前半）についてのウェイトを用いない離家経験割合はモデル予測の信頼区間を下回っている。さらに、2019年調査の特徴は、信頼区間の下限を下回る幅が拡大したことにある。たとえば、20歳代後半男性におけるウェイトなし離家経験割合のモデル1の信頼区間下限からの差をみると、2009年は約3%ポイント低く、2014年は約5%ポイントだったが、



出典：巻末表1。(注)世帯員の年齢は調査時点のもの。95%信頼区間は回収確率に関するポワソン回帰モデルに基づき推定されたウェイトの統計学的な信頼区間によって離家経験割合が確率的に取り得る上限と下限を指す。

図2 回収確率ウェイトの種類別 男女・年齢階級別 離家経験割合(%)とその信頼区間

2019年は7%ポイント以上に悪化した。回収確率ウェイトの信頼区間の下限を下回るということは、これらの年齢層では回収状況の悪化が統計的に有意に離家経験割合の集計に影響を及ぼしていることを意味する。

第三に、このようにモデル予測の信頼区間は場合によっては10%ポイントを超える幅があるため、特定の男女年齢階級における時系列変化について、信頼区間が相互に分離するといった明瞭な変化を見出すことはできない。時系列比較を通じて何らかの結論を導くためには、慎重な検討が求められる。

第四に、とくに20歳代以下の若い年齢層において、回収確率モデル予測の不確実性に起因する離家経験割合の信頼区間は2014年から2019年にかけて顕著に広がっている。たとえば、男性20歳代の信頼区間幅は2009～2014年は14～15%ポイントだったが、2019年は24～28%ポイントへと倍増に近い拡大をした。この背景には回収状況の悪化による回収数の低下とともに、推定モデル上はリスク人年と呼ばれる標本フレーム世帯数の減少、すなわち20歳代人口の減少(少子化によるコーホートサイズの縮小)の影響があると推察される。

V. 結語

本稿では、世帯動態調査における回収状況の悪化が非標本誤差の拡大を通じて調査結果に及ぼした影響を検証するために、調査実施年を挟む2時点の国勢調査の個票データの二

次利用によって得た集計表を用いて世帯動態調査の調査対象となった標本フレームの世帯属性別世帯数を推計し、ポワソン回帰モデルを用いて世帯属性別の有効回収確率を推定することで世帯回収確率ウェイトを作成した（世帯属性には世帯主の属性を含む）。実地調査において捕捉が困難な単独世帯の回収状況の影響を顕著に受ける50歳未満の男女・年齢別離家経験割合に着目し、2009年（第6回）調査から2019年（第8回）の推移について世帯回収確率ウェイト有無・種類別に再集計を行って、再集計結果の比較検討を行った。その結果、2014年から2019年にかけての20歳代の離家経験割合の低下には回収率の低下幅が拡大したことが大きく影響していることが明らかになった。

今後の課題として、本稿では世帯動態調査の調査事項のうち50歳未満の離家経験割合への影響のみを取り上げたが、より総合的な調査結果への影響の検証が望ましい。一方、分析結果から20歳代で信頼区間が顕著に拡大しており、その背後には回収率の低下とともにコーホートサイズの縮小の影響もあることが示唆された。回収率向上のための努力を重ねることはもちろん必要であろうが、抽出数や抽出方法といった調査フレームの再検討も重要であろう。

分析手法の課題として、本稿のように非標本誤差に対し、回収確率を推定して調査結果の補正に利用するというアプローチでは、標本フレーム世帯数の（推計）精度が結果を大きく左右する。本稿の標本フレーム世帯数推計は、調査区別に世帯属性別の世帯数を期首・期末国勢調査間で線型補間する、もしくは転入・転出等の世帯異動が期間の中央で起こることを仮定するという単純なものであったが、国勢調査及び世帯動態調査の5年前常住地に関する結果をより効果的に活用して小地域世帯数推計の精度を向上することは今後の主要な課題である。具体的には、 $t+1$ 年調査における5年前($t-4$ 年)常住地が「現住所」の場合、 $t-4$ 年国勢調査の同一地点に当該世帯を見つけられるはずであり、変化しない世帯属性別にみた $t-4\sim t+1$ 年世帯数を用いて、 $t-4\sim t+1$ 年に転出もしくは消滅した世帯数を期首・期末の差によって推定することができるはずである。しかし、実際には $t-4$ 年もしくは $t+1$ 年国勢調査への回答における種々の「不詳」及び「不詳」の補綴方法や誤記入等による誤差のため、5年前常住地が「現住所」の期末世帯数が期首を上回る場合が散見され、別の世帯属性に移動していない世帯主がいると想定される。転出入した世帯を特定することができれば、地域や世帯属性に応じて転入率・転出率を可変的に設定する等の方法によって t 年世帯数の推計精度を向上できる可能性がある¹⁶⁾。抜本的な改善方法としては、

16) たとえば、 $t-4\sim t+1$ 年の間に共同住宅が新設されたことによって国勢調査間の世帯数が急増した場合で、世帯動態調査の結果から t 年までに新築・入居があったことが明らかとなるとき、 $t-4\sim t+1$ 年の転入や世帯異動が期間平均等に発生していると考えて $t-4\sim t$ 年発生率を $3/4$ とするのは過小であろう。転出入率のピークが18～22歳頃にあり、加齢によって顕著に低下するならば、5年転出入率のうち4年間の発生率は18歳から各歳でみて加齢により低下するだろう。また、学生寮等の位置する（移動性行が高い世帯が集住する）調査区のように $t-4$ 年世帯のほぼ全員が転出し、 $t+1$ 年までに同様の属性の世帯が転入している場合、出生年ではなくピリオドの年齢別に $t-4\sim t+1$ 年推計を実施するという考え方によって t 年の属性別世帯数の推計精度を向上できる可能性がある。さらに、世帯動態調査は世帯主・非世帯主の地位に関する情報を豊富に含むが十分に活用していない。転出もしくは死亡した人口を精確に全国で特定できるならば、 $t-4$ 年転出者は $t+1$ 年国勢調査における別の地域に常住する可能性が高く、死亡については人口動態調査で補足されるため、これら統計との整合性の検証を通じて転出と死亡を分解し、小地域人口・世帯分析を飛躍的に前進させる可能性がある。このような可能性の検証を進めるには、調査対象地区が確実に特定できていることが前提となる。

たとえば国勢調査の調査区地図を用いることができれば、 $t-4$ 年調査対象調査区に対応する $t+1$ 年の調査区を機械的かつ確実に特定することができる。そうなれば、分析を行おうとする $t-4$ 年国勢調査の調査区に $t+1$ 年調査の5年前常住地「現住所」の世帯が捕捉されている蓋然性が高まり、 $t-4\sim t+1$ 年国勢調査間の突合において世帯属性別にみた5年前常住地が「現住所」の($t+1$ 年)世帯数は $t-4$ 年世帯数を上回るという制約を課すといった5年前常住地の効率的な利用方法等の検討を深めて、精度向上を図ることができると期待される。

このように本稿で提案した標本フレーム世帯数の推計には、調査実施年を挟む2時点のうち期末年の国勢調査における調査区の特定方法や5年前常住地に関する調査結果のより効果的な利用等の改善の余地がある。課題はあるものの、本稿における回収確率ウェイトの作成方法は、国勢調査の調査区を標本フレーム(抽出元)に用いる他の統計調査に容易に応用が可能であり、とくに国立社会保障・人口問題研究所が実施する「社会保障・人口問題基本調査」等の国民生活基礎調査の後続調査に効果的な適用が可能である。毎月勤労統計調査の不適切処理事案を発端とした政府統計に対する国民の不信が高まる中での統計委員会や統計改革推進会議統計新生部会での検討の結果、公的統計作成プロセスの中でのPDCA(Plan-Do-Check-Act)サイクルを回すことによる作成過程の改善を通じた統計品質の不断の向上が求められている。世帯動態調査だけでなく他の統計調査においても本研究と同様の検証が進められ、非標本誤差の調査結果への影響についての知見が蓄積されることを期待したい。

謝辞

本研究は、「世帯動態調査」および厚生労働行政推進調査事業補助金(政策科学総合研究事業(政策科学推進研究事業))「長期的人口減少と大国際人口移動時代における将来人口・世帯推計の方法論的発展と応用に関する研究」(研究代表者小池司朗, 課題番号20AA2007)による研究成果の一部である。本稿における「世帯動態調査」に関する分析には、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。また、本研究で利用した「国勢調査」の集計表には、厚生労働行政推進調査事業補助金事業(研究代表者:小池司朗, 課題番号:20AA2007)のために、統計法第33条に基づき調査票情報を二次利用したものが含まれる。津谷典子教授(慶應義塾大学)、石井太教授(慶應義塾大学経済学部)と可部繁三郎氏(日本経済新聞)から草稿執筆段階で受けた本質的な指摘は極めて重要だった。関野秀峰氏と亀本薫氏(総務省統計局)から国勢調査の実施方法等を含む調査全般について詳細な教えを受けた。また、社人研の菊池潤室長からは5つの「社会保障・人口問題基本調査」のサンプリングに関する資料、中川雅貴室長からは「人口移動調査」の標本誤差についての資料の提供を受けた。記し感謝申し上げたい。言うまでもなく、残された誤謬は筆者の責に帰す。

参照文献

- 石井太（2004）「標本設計 理論編」，財団法人厚生統計協会編『よくわかる標本設計法—厚生統計で学ぶ標本設計の理論と実践—』，第2部，財団法人厚生統計協会。
- 岩澤美帆（2008）「初婚・離婚の動向と出生率への影響」『人口問題研究』第64巻第4号，pp.19-34。
- 厚生労働省大臣官房統計情報部（2001）『平成11年 国民生活基礎調査』，厚生労働省大臣官房統計情報部，2001年3月。
- 厚生労働省大臣官房統計情報部（2006）『平成16年 国民生活基礎調査 第1巻 解説編』，厚生労働省大臣官房統計情報部，2006年3月。
- 厚生労働省大臣官房統計情報部（2010）『平成21年 国民生活基礎調査』，厚生労働省大臣官房統計情報部，2010年12月。
- 厚生労働省大臣官房統計情報部（2012）『平成22年 国民生活基礎調査 第1巻 結果の概要 全国編（世帯、所得・貯蓄）』，厚生労働省大臣官房統計情報部，2012年3月。
- 厚生労働省大臣官房統計情報部（2015）『2014（平成26）年 国民生活基礎調査』，厚生労働省大臣官房統計情報部，2015年12月。
- 厚生労働省政策統括官（統計・情報政策担当）（2020）『2019（令和元）年 国民生活基礎調査』，厚生労働省政策統括官（統計・情報政策担当），2020年12月。
- 国民生活基礎調査の改善に関するワーキンググループ（2021）「国民生活基礎調査の改善に関するワーキンググループ 報告書（令和3年3月）」，（2022年11月19日アクセス：https://www.mhlw.go.jp/stf/newpage_18134.html）
- 国民生活基礎調査の非標本誤差の縮小に向けた研究会（2018）「国民生活基礎調査の非標本誤差の縮小に向けた研究会 報告書（平成30年3月）」，（2022年11月19日アクセス：<https://www.mhlw.go.jp/stf/shingi2/0000204385.html>）
- 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（2011）「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）」，（2022年11月19日アクセス：<https://www.mhlw.go.jp/toukei/list/20-21ronbun.html>）
- 国立社会保障・人口問題研究所（2001）『第4回世帯動態調査（1999年社会保障・人口問題基本調査）現代日本の世帯変動』，調査研究報告資料第16号，（2023年1月30日アクセス：<https://www.ipss.go.jp/syoushika/bunken/data/pdf/127224.pdf>）
- 国立社会保障・人口問題研究所（2007）『第5回世帯動態調査（2004年社会保障・人口問題基本調査）現代日本の世帯変動』，調査研究報告資料第21号，（2023年1月30日アクセス：<https://www.ipss.go.jp/syoushika/bunken/data/pdf/132274.pdf>）
- 国立社会保障・人口問題研究所（2011）『第6回世帯動態調査（2009年社会保障・人口問題基本調査）現代日本の世帯変動』，調査研究報告資料第28号，（2023年1月30日アクセス：<https://www.ipss.go.jp/syoushika/bunken/data/pdf/206714.pdf>）
- 国立社会保障・人口問題研究所（2016）『第7回世帯動態調査（2014年社会保障・人口問題基本調査）現代日本の世帯変動』，調査研究報告資料第34号，（2023年1月30日アクセス：<https://www.ipss.go.jp/ps-dotai/j/DOI7/houkoku/Hhoukoku.pdf>）
- 国立社会保障・人口問題研究所（2022）『第8回世帯動態調査（2019年社会保障・人口問題基本調査）現代日本の世帯変動』，調査研究報告資料第39号，（2022年6月20日アクセス：<https://www.ipss.go.jp/ps-dotai/j/DOI8/houkoku/Hhoukoku.pdf>）
- 総務省（2020）「公的統計の整備に関する基本的な計画（令和2年6月2日）」，（2023年1月30日アクセス：https://www.soumu.go.jp/main_content/000690298.pdf）
- 総務省統計局（2022）『令和2年国勢調査 調査区関係資料利用の手引き』，総務省統計局。
- 菅桂太（2007）「近年の「国勢調査」日本人人口の精度に関する一考察」『人口学研究』第41号，pp.61-73。
- 菅桂太（2009）「離家とパートナー形成形成タイミングの日米比較」『人口問題研究』第65巻第3号，pp.40-57。

- 스가케이타・조성호 (2017) 「한국과 일본 청년층의 경제적 자립과 가족형성 간의 관계」 조성호 편저 『청년층의 경제적 자립과 가족형성에 관한 한일 비교분석』, 한국보건사회연구원, 제4장 제2절, pp.127-167. (2023年1月30日アクセス：<https://www.kihasa.re.kr/publish/report/view?searchText=%EC%A1%B0%EC%84%B1%ED%98%B8&page=1&type=all&seq=27828>) 菅桂太・チョソンホ (2017) 「地域差を考慮した若年者の自立と家族形成：日韓比較」, チョソンホ編著 『韓国と日本における青年層の経済的自立と家族形成間の関係』, 韓国保健社会研究院, 第4章第2節, pp.127-167.
- 鈴木透 (1997) 「世帯形成の生命表分析」 『人口問題研究』 第53巻第2号, pp.18-30.
- 鈴木透 (2003) 「離家の動向・性差・決定因」 『人口問題研究』 第59巻第4号, pp.1-18.
- 鈴木透 (2007) 「世帯形成の動向」 『人口問題研究』 第63巻第4号, pp.1-13.
- 鈴木透 (2014) 「全国世帯推計の方法論的諸問題」 『人口問題研究』 第70巻第2号, pp.81-96.
- 統計改革推進会議 統計行政新生部会 (2019) 『統計行政の新生に向けて～将来にわたって高い品質の統計を提供するために～』. (2023年1月30日アクセス：https://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/20191224_shinsei_honbun.pdf)
- 日本統計学会 (2019) 『公的統計に関する臨時委員会報告書—第1部毎月勤労統計調査の不正をめぐる事案に関する見解』. (2023年1月30日アクセス：https://www.jss.gr.jp/wp-content/uploads/kouteki_toukei_report_main.pdf)
- 西村清彦・山澤成康・肥後雅博 (2020) 『統計 危機と改革 システム劣化からの復活』 (Kindle版), 日本経済新聞社.
- 増淵知哉・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』 第85巻第5号, pp.447-467.
- 山口幸三 (2011) 『現代日本の世帯構造と就業構造の変動解析—公的統計のミクロ統計解説序説』 財団法人日本統計協会.
- Bongaarts, John (1978) "A Framework for Analyzing the Proximate Determinants of Fertility," *Population and Development Review*, Vol.4, No.1 (Mar., 1978), pp.105-132.
- Cameron, A. Colin and Pravin K. Trivedi (2013) *Regression Analysis of Count Data, Second Edition*, Cambridge: Cambridge University Press.
- Doherr, Thorsten (2021) "ultimatch: counterfactual matching," Online. <Accessed on 2023/1/13 at <https://github.com/ThorstenDoherr/ultimatch> >
- Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer and Roger Tourangeau (2011) *Survey Methodology, Second Edition*, Wiley Series in Survey Methodology, Hoboken, NJ: Wiley.
- Little, Roderick J. A. and Donald B. Rubin (2020) *Statistical Analysis with Missing Data, Third Edition*, Wiley Series in Probability and Statistics, Hoboken, NJ: John Wiley & Sons, Inc.
- Lohr, Sharon L. (2022) *Sampling Design and Analysis, Third Edition*, Boca Raton, FL: CRC Press.

巻末表 1 男女・年齢5歳階級別 有効回収世帯員数と標本フレーム世帯員数（推計値）、世帯回収率ウエイトの有無・種類別 離家経験割合（%）とその信頼区間

A. 男

	有効回収世帯員数	標本フレーム世帯員数(推計値)	離家経験割合(%)								
			ウエイトなし	モデル1				モデル2			
				点推定値	95%信頼区間		点推定値	95%信頼区間			
					下限	上限		下限	上限		
2009年(第6回)											
50歳未満総数	5,458	7,293	73.9	76.8	73.2	80.2	78.2	74.0	82.1		
18～19歳	281	374	29.6	37.0	31.0	44.8	43.6	36.0	53.3		
20～24歳	668	903	45.7	53.1	46.9	60.5	59.0	51.1	67.8		
25～29歳	679	1,000	64.7	72.3	68.1	76.5	74.7	69.8	79.4		
30～34歳	915	1,264	78.1	81.0	78.0	83.8	82.3	78.9	85.3		
35～39歳	1,069	1,405	81.9	83.4	80.7	85.8	84.2	81.2	87.0		
40～44歳	925	1,190	86.3	86.7	84.5	88.7	87.0	84.4	89.3		
45～49歳	921	1,157	87.8	88.2	86.1	90.1	88.5	86.4	90.3		
2014年(第7回)											
50歳未満総数	4,795	7,255	75.3	79.4	75.9	82.8	80.7	76.9	84.2		
18～19歳	245	356	27.3	35.4	28.9	44.4	41.8	34.2	51.7		
20～24歳	546	812	46.9	55.5	48.7	63.7	59.3	52.1	67.5		
25～29歳	613	1,051	69.8	78.7	74.6	82.6	80.1	75.1	84.8		
30～34歳	675	1,108	77.7	82.8	79.3	86.0	85.1	82.1	87.8		
35～39歳	850	1,297	82.5	86.0	83.2	88.6	87.1	84.7	89.3		
40～44歳	978	1,385	86.6	87.6	85.6	89.4	87.3	84.8	89.4		
45～49歳	888	1,244	87.7	88.3	86.4	89.9	88.7	86.7	90.5		
2019年(第8回)											
50歳未満総数	4,342	7,044	74.4	78.9	74.4	83.4	79.2	74.9	83.6		
18～19歳	245	444	35.2	50.2	38.9	66.5	54.7	43.3	70.8		
20～24歳	546	859	37.8	50.5	40.8	64.4	52.7	42.6	67.2		
25～29歳	613	898	60.7	73.5	68.2	78.6	73.5	67.7	79.1		
30～34歳	675	919	79.9	85.2	82.2	87.8	86.5	83.4	89.2		
35～39歳	850	1,159	83.8	86.8	83.6	89.6	86.9	84.3	89.2		
40～44歳	978	1,300	87.1	88.3	85.3	90.8	87.9	85.3	90.0		
45～49歳	888	1,465	87.9	88.8	86.6	90.7	88.3	86.1	90.2		

B. 女

	有効回収世帯員数	標本フレーム世帯員数(推計値)	離家経験割合(%)								
			ウエイトなし	モデル1				モデル2			
				点推定値	95%信頼区間		点推定値	95%信頼区間			
					下限	上限		下限	上限		
2009年(第6回)											
50歳未満総数	5,573	7,378	76.3	78.9	75.6	81.9	80.2	76.4	83.7		
18～19歳	284	367	22.8	29.0	24.0	35.6	36.9	30.2	45.6		
20～24歳	617	804	36.8	42.9	37.4	49.6	49.3	42.0	57.7		
25～29歳	776	1,121	64.8	71.6	67.3	75.7	75.0	69.9	79.9		
30～34歳	921	1,286	81.3	83.9	81.2	86.4	85.1	81.9	87.9		
35～39歳	1,073	1,400	89.5	90.8	89.2	92.2	91.2	89.3	92.8		
40～44歳	999	1,275	90.0	90.5	88.8	91.9	90.6	88.6	92.3		
45～49歳	903	1,124	91.8	92.1	90.7	93.3	92.0	90.4	93.4		
2014年(第7回)											
50歳未満総数	4,890	7,140	76.6	79.8	76.5	83.0	81.1	77.4	84.4		
18～19歳	297	438	33.1	41.8	34.2	52.4	49.9	41.4	60.8		
20～24歳	544	752	34.6	41.1	35.0	49.2	47.4	39.9	56.8		
25～29歳	608	937	63.5	70.6	65.9	75.3	72.5	66.8	77.8		
30～34歳	668	1,063	82.6	86.9	84.3	89.3	87.9	85.3	90.2		
35～39歳	889	1,332	87.9	89.3	87.1	91.3	89.0	86.7	91.0		
40～44歳	979	1,373	90.6	91.8	90.3	93.0	92.4	91.0	93.7		
45～49歳	905	1,243	93.2	93.6	92.5	94.5	93.6	92.3	94.6		
2019年(第8回)											
50歳未満総数	4,363	6,874	76.5	79.3	75.1	83.2	80.3	76.3	84.1		
18～19歳	255	460	16.1	23.6	17.1	35.7	25.6	18.3	39.4		
20～24歳	477	768	27.5	35.5	27.9	47.3	39.8	31.2	53.5		
25～29歳	434	763	60.4	70.6	65.1	75.8	71.5	65.4	77.3		
30～34歳	560	935	83.0	86.6	83.7	89.2	87.8	84.9	90.3		
35～39歳	708	1,099	90.5	91.8	89.7	93.6	92.5	90.7	93.9		
40～44歳	941	1,393	91.6	92.6	90.8	94.1	92.9	91.4	94.1		
45～49歳	988	1,457	91.0	91.6	89.8	93.1	91.9	90.4	93.2		

資料：巻末表 2 の回収率に関するポワソン回帰モデル推定値による世帯回収率ウエイトを用いて、世帯動態調査を再集計。ただし、世帯動態調査の対象となった調査区のうち、調査員調査不能により世帯名簿の作成ができなかった2009年(第6回)調査の1調査区と、t-4年国勢調査の調査区に対応するt+1年調査区を特定できなかった2009年(第6回)の5調査区(162世帯の有効回収134票)、2014年(第7回)の5調査区(245世帯の有効回収162票)、2019年(第8回)の6調査区(262世帯の有効回収114票)を除く。世帯主が18歳未満の世帯と性別か年齢が不詳の世帯主及び世帯員を集計対象から除く。有効回収世帯員数及び標本フレーム世帯員数(推計値)には、離家経験不詳を含む。(注)世帯員の年齢は世帯動態調査の調査時点のもの。95%信頼区間とは、回収率に関するポワソン回帰モデルに基づき推定されたウエイトの統計学的な信頼区間によって離家経験割合が率的に取り得る上限と下限を指す。

巻末表2 回収確率に関するポワソン回帰モデルの発生率比 (Incidence Rate Ratio)^(注1) 推定結果

t年世帯動態調査 調査時の世帯(主)属性	2009年(第6回)				2014年(第7回)				2019年(第8回)			
	モデル1		モデル2		モデル1		モデル2		モデル1		モデル2	
	IRR	P値	IRR	P値	IRR	P値	IRR	P値	IRR	P値	IRR	P値
世帯主の男女・年齢												
男 ^(*)	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
女	0.962	0.301	0.927	0.041	0.978	0.578	0.878	0.002	1.040	0.375	0.934	0.154
25歳未満	0.872	0.305	0.924	0.561	0.760	0.090	0.853	0.265	0.581	0.018	0.688	0.102
25～29	0.653	0.000	0.750	0.001	0.611	0.000	0.719	0.002	0.460	0.000	0.592	0.000
30～34	0.758	0.000	0.848	0.008	0.628	0.000	0.696	0.000	0.590	0.000	0.701	0.000
35～39	0.832	0.001	0.914	0.092	0.685	0.000	0.745	0.000	0.674	0.000	0.795	0.000
40～44	0.893	0.026	0.902	0.023	0.804	0.000	0.840	0.000	0.729	0.000	0.805	0.000
45～49	0.909	0.067	0.908	0.011	0.824	0.000	0.865	0.000	0.769	0.000	0.857	0.000
50～54	0.962	0.281	0.954	0.255	0.934	0.114	0.935	0.087	0.795	0.000	0.852	0.000
55～59	0.942	0.106	0.958	0.319	0.961	0.366	0.988	0.783	0.880	0.010	0.954	0.267
60～64	0.997	0.940	1.023	0.530	0.968	0.402	0.976	0.527	0.950	0.296	0.972	0.523
65～69 ^(*)	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
70～74	1.016	0.683	0.951	0.208	1.016	0.723	0.951	0.176	0.933	0.182	0.933	0.031
75～79	1.056	0.253	0.990	0.787	1.004	0.915	0.955	0.241	0.956	0.359	0.943	0.091
80～84	1.047	0.386	1.000	0.997	0.984	0.721	0.943	0.222	0.984	0.786	0.962	0.366
85+	1.002	0.970	1.105	0.108	0.973	0.595	1.061	0.259	0.860	0.009	0.929	0.148
t年常住地の地域												
東京特別区 ^(*)	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
政令指定都市	1.206	0.000	1.147	0.004	1.211	0.000	1.206	0.000	1.361	0.000	1.240	0.000
その他 東日本	1.385	0.000	1.246	0.000	1.358	0.000	1.245	0.000	1.561	0.000	1.326	0.000
その他 西日本	1.420	0.000	1.262	0.000	1.327	0.000	1.242	0.000	1.545	0.000	1.328	0.000
世帯の家族類型												
単独世帯 ^(*)	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
夫婦のみ	1.342	0.000	1.282	0.000	1.351	0.000	1.346	0.000	1.486	0.000	1.387	0.000
夫婦と子	1.532	0.000	1.503	0.000	1.481	0.000	1.492	0.000	1.632	0.000	1.484	0.000
ひとり親と子	1.120	0.003	1.030	0.450	1.145	0.000	1.004	0.908	1.187	0.000	1.051	0.187
その他	1.400	0.000	1.263	0.000	1.287	0.000	1.122	0.008	1.395	0.000	1.203	0.000
住宅の種類と所有関係												
一戸建ての持ち家 ^(*)			1.000	-			1.000	-			1.000	-
共同住宅 及び 公営/ 公団賃貸			0.874	0.000			0.905	0.000			0.897	0.000
その他(民間借家等)			0.875	0.000			0.883	0.000			0.803	0.000
世帯主のt-5年常住地												
現住所(t年調査地) ^(*)			1.000	-			1.000	-			1.000	-
別住所			0.789	0.000			0.783	0.000			0.815	0.000
世帯主の配偶関係												
未婚			0.987	0.800			1.044	0.406			1.049	0.330
有配偶 ^(*)			1.000	-			1.000	-			1.000	-
死別			1.142	0.003			1.284	0.000			1.224	0.000
離別			1.044	0.391			1.132	0.018			1.124	0.027
不詳			0.296	0.000			0.274	0.000			0.151	0.000
定数項	0.459	0.000	0.575	0.000	0.479	0.000	0.567	0.000	0.387	0.000	0.515	0.000
対数 リスク人年 (オフセット)	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
対数尤度	-1,169.3		-3,854.6		-1,197.5		-4,006.6		-1,195.9		-3,764.9	
N	504		2,888		505		3,026		509		2,956	
総リスク人年	15,504		15,504		16,104		16,104		16,140		16,140	

資料：t年世帯動態調査，並びにt-4～t+1年国勢調査に基づく世帯動態調査の調査対象調査区の世帯数（推計値）を用いて筆者算出。ただし，世帯動態調査の対象となった調査区のうち，調査員調査不能により世帯名簿の作成ができなかった2009年(第6回)調査の1調査区と，t-4年国勢調査の調査区に対応するt+1年調査区を特定できなかった2009年(第6回)の5調査区(162世帯の有効回収134票)，2014年(第7回)の5調査区(245世帯の有効回収162票)，2019年(第8回)の6調査区(262世帯の有効回収114票)を除く。

(*)レファレンス・カテゴリー。(注1)発生率はリスク人年（ここでは調査対象世帯数）あたりの発生率（すなわち有効回収確率）を指す。定数項の発生率比はすべての共変量がゼロのときの発生率（ベース・ライン発生率），各共変量の発生率比はリスク人年あたりの発生率のベース・ライン発生率に対する比を示す。P値は頑健標準誤差推定値による。

巻末表 3 回収確率に関するポワソン回帰モデルに用いる変数の平均と標準偏差

t年世帯動態調査 調査時の世帯(主)属性 ^(注1)	2009年(第6回)				2014年(第7回)				2019年(第8回)			
	モデル1		モデル2		モデル1		モデル2		モデル1		モデル2	
	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差
有効回収数	22.3	38.3	3.9	13.7	21.6	35.4	3.6	12.1	19.8	35.6	3.4	12.4
世帯主の男女・年齢												
男 ^(†)	0.548	-	0.564	-	0.545	-	0.552	-	0.532	-	0.547	-
女	0.452	0.498	0.436	0.496	0.455	0.499	0.448	0.497	0.468	0.499	0.453	0.498
25歳未満	0.056	0.229	0.032	0.177	0.059	0.237	0.030	0.170	0.065	0.246	0.029	0.168
25～29	0.067	0.251	0.052	0.223	0.073	0.261	0.045	0.208	0.069	0.253	0.047	0.211
30～34	0.077	0.267	0.064	0.246	0.073	0.261	0.061	0.240	0.073	0.260	0.057	0.232
35～39	0.077	0.267	0.078	0.268	0.077	0.267	0.074	0.262	0.077	0.266	0.070	0.255
40～44	0.077	0.267	0.086	0.281	0.077	0.267	0.087	0.282	0.073	0.260	0.081	0.273
45～49	0.077	0.267	0.093	0.291	0.079	0.270	0.089	0.285	0.079	0.269	0.099	0.299
50～54	0.075	0.264	0.092	0.290	0.075	0.264	0.091	0.287	0.077	0.266	0.096	0.295
55～59	0.073	0.261	0.092	0.289	0.075	0.264	0.089	0.285	0.075	0.263	0.095	0.294
60～64	0.075	0.264	0.108	0.310	0.073	0.261	0.096	0.295	0.075	0.263	0.082	0.274
65～69 ^(†)	0.069	-	0.085	-	0.075	-	0.088	-	0.071	-	0.083	-
70～74	0.071	0.258	0.071	0.257	0.069	0.254	0.080	0.271	0.073	0.260	0.080	0.272
75～79	0.067	0.251	0.060	0.237	0.065	0.247	0.067	0.250	0.067	0.250	0.068	0.252
80～84	0.067	0.251	0.048	0.214	0.063	0.244	0.057	0.231	0.067	0.250	0.061	0.239
85+	0.067	0.251	0.037	0.190	0.063	0.244	0.046	0.209	0.063	0.243	0.052	0.222
t年常住地の地域												
東京特別区 ^(†)	0.236	-	0.171	-	0.236	-	0.170	-	0.224	-	0.158	-
政令指定都市	0.248	0.432	0.234	0.423	0.251	0.434	0.245	0.430	0.253	0.435	0.231	0.422
その他 東日本	0.260	0.439	0.320	0.466	0.253	0.435	0.303	0.460	0.261	0.440	0.332	0.471
その他 西日本	0.256	0.437	0.276	0.447	0.259	0.439	0.282	0.450	0.261	0.440	0.279	0.449
世帯の家族類型												
単独世帯 ^(†)	0.222	-	0.424	-	0.222	-	0.441	-	0.220	-	0.454	-
夫婦のみ	0.183	0.387	0.101	0.302	0.178	0.383	0.102	0.302	0.183	0.387	0.103	0.303
夫婦と子	0.163	0.369	0.107	0.310	0.166	0.373	0.104	0.305	0.181	0.385	0.106	0.308
ひとり親と子	0.212	0.409	0.220	0.414	0.214	0.410	0.221	0.415	0.202	0.402	0.212	0.409
その他	0.220	0.415	0.147	0.354	0.220	0.415	0.133	0.339	0.214	0.411	0.125	0.331
住宅の種類と所有関係												
一戸建ての持ち家 ^(†)			0.337	-			0.309	-			0.339	-
共同住宅 及び 公営/公団賃貸			0.263	0.440			0.295	0.456			0.280	0.449
その他(民間借家等)			0.400	0.490			0.396	0.489			0.382	0.486
世帯主のt-5年常住地												
現住所(t年調査地) ^(†)			0.663	-			0.655	-			0.679	-
別住所			0.337	0.473			0.345	0.475			0.321	0.467
世帯主の配偶関係												
未婚			0.221	0.415			0.222	0.415			0.210	0.408
有配偶 ^(†)			0.353	-			0.343	-			0.341	-
死別			0.146	0.353			0.144	0.351			0.122	0.327
離別			0.185	0.388			0.187	0.390			0.167	0.373
不詳			0.095	0.294			0.104	0.306			0.159	0.366
N	504	2,888	505	3,026	509	2,956						
総リスク人年 ^(注2)	15,504	15,504	16,104	16,104	16,140	16,140						

資料：t年世帯動態調査，並びにt-4～t+1年国勢調査に基づく世帯動態調査の調査対象調査区の世帯数（推計値）を用いて筆者算出。ただし，世帯動態調査の対象となった調査区のうち，調査員調査不能により世帯名簿の作成ができなかった2009年(第6回)調査の1調査区と，t-4年国勢調査の調査区に対応するt+1年調査区を特定できなかった2009年(第6回)の5調査区(162世帯の有効回収134票)，2014年(第7回)の5調査区(245世帯の有効回収162票)，2019年(第8回)の6調査区(262世帯の有効回収114票)を除く。

(†)レファレンス・カテゴリー。(注1)有効回収数は被説明変数，その他の共変量はすべて二項変数。ケース数Nは4個（もしくは7個）共変量の組み合わせ（理論上は約560通り，もしくは約16,800通り中）の観測数である。(注2)総リスク人年は世帯動態調査の調査区世帯名簿の世帯数に対応する。

Patterns of Unit Nonresponse Biases in the National Survey on Household Changes: Did the Japanese Youths under the Ages of the Fifty Leaving the Parental Home Decrease in the Late 2010s?

SUGA Keita

This study examines and assesses the accuracy and precision of the National Survey on Household Changes in the face of increasing non-sampling errors due to substantial declines in response rates over the past decades. In general, a detailed assessment of a unit nonresponse in household surveys is cumbersome because we need a complete list of household heads by their characteristics at the time of the survey. For this purpose, we derive small-area household estimates for the enumeration area of the survey corresponding to that in the Population Census of Japan (the size of households is about fifty in each area) by interpolating changes in the number of households by their characteristics such as migration status, marital status, age and sex of heads, and housing type. We then estimate the Poisson regression models for the nonresponse probability of the household head, with using the household estimates for the sampling frame as the exposure. We obtain predictions and their confidence intervals for the inverse probability weights.

As an empirical application, we compare the rate of the population who ever left the parental home before the age of fifty between with and without the nonresponse weight of several models. The rate suffers severely from the nonresponse of lone households whose coverage is often among the worst in a sample survey.

The results show that the expansion of nonresponse from 2014 to 2019 critically caused the decline in the rate of the population who ever left the parental home in the age twenties. Moreover, the confidence intervals of the age twenties significantly widened in the same period. The precision and accuracy of a survey are increasingly dependent on the shrinkage of the cohort size due to fertility decline.

Keywords: Precision and accuracy of a survey, response rate, non-sampling errors, leaving the parental home