

## パネル調査からの対象の脱落について ——生存時間解析を用いた分析

坂口 尚文

(公益財団法人 家計経済研究所 次席研究員)

### 1. はじめに

本稿は、生存時間解析の枠組みを用いて、「消費生活に関するパネル調査」(以下、JPSC)における対象者の脱落の傾向を評価したものである。生存時間解析は、特定の処置や環境の効果を、興味あるイベントの発生までの時間(time-to-event)を比較することで評価する統計手法である。例えば医学の臨床分野等では、この手法を用いて、ある治療法を処置した群と処置していない群との間に、患者の死亡までの時間、あるいは寛解までの時間に違いがあるかどうかで、その治療法の効果を判定することが多い。しかし、あらゆる分野において、実際に観測されるデータの多くは、観察期間中に対象が追跡不能となる、あるいは実験の打ち切り等によりイベント発生までの正確な時間を計測できないケースを含んでいる。そのため、イベントまでの発生時間を単純に比較することはできない。このような観測データの不完全性を克服するため、生存時間解析ではイベントの発生時間について背後に何らかの関数を想定したアプローチを行う。

本稿では、対象のパネル調査からの脱落をイベントの発生とみなし、調査に参加した回数(期間)を生存時間として解析する。目的は、対象のどのような属性が、脱落までの期間を早めるか明らかにすることである。属性間で脱落するスピードに違いが認められる場合、残存サンプルにおける属性の構成比が、初回抽出時のサンプル構成比と異なることが推し計れる。今回、主として着目する

属性は、対象者の初回調査参加時の年齢と対象者が属するコーホート間の相違である。対象者の初回参加時の年齢は、調査期間を通じて不変であり、同一コーホート内の年齢構成比のずれを指し示す指標となる。コーホート間の相違は、異なるコーホートのデータを結合して、そのまま集計を行うことが妥当であるかを指し示すことになる。

パネル調査からの脱落は、調査に初めて参加した時の年齢や出生年といった、同一対象については時間(=調査回)に関して不変の属性だけではなく、調査時点の状況に依存する要因も大きく関わっていることが予想される。本稿では、時間依存の変数として、対象者の1日の生活基礎時間が脱落に与えている影響を推計する。生活基礎時間は、睡眠や食事、身の回りのことに費やす時間であるが、JPSCの調査票上は、実質的には24時間から労働時間や家事・育児時間等を除いたものである。そのため、この生活基礎時間を回答者が持つ時間的余裕の代理指標とみなすことも可能である。対象者の時間的余裕と脱落の関係は、サンプル保持の観点から、調査票の設計や調査遂行の面において重要な情報となる。

本稿の構成は以下の通りである。第2節では、分析に使用するJPSCデータのサブセットの説明と生存時間解析を行う際の追跡期間の条件設定について述べる。第3節では、今回の分析に用いる共変量についての記述を行う。第4節は使用するモデルの説明を、第5節では推計結果について述べる。第6節において、結果の要約と今後の課題を示して本稿を閉じることにする。

図表-1 JPSCを構成する追跡コーホートと基本情報

	調査参加時年齢	出生年	サンプルサイズ (初回抽出時)	調査開始年	調査回数	第21回調査時 回答者数
コーホート A	24～34	10/1/58～9/30/69	1,500	1993	21	732
コーホート B	24～27	10/1/69～9/30/73	500	1997	17	231
コーホート C	24～29	10/1/73～9/30/79	836	2003	11	478
コーホート D	24～28	10/1/79～9/30/84	636	2008	6	451
コーホート E	24～28	10/1/84～9/30/89	648	2013	1	648

## 2. 使用データの説明と追跡期間の設定

分析には、JPSCの第1回調査（1993年）から第21回調査（2013年）のデータを用いる。JPSCでは、ほぼ5年おきに後生の人口学的コーホートを調査に追加しており、現在5つのコーホートを追跡集団として有している。各コーホートは、出生年の早い順にA～Eのアルファベットを付け識別している。コーホートの情報については、図表-1にまとめている。コーホートAだけが初回調査時に30歳代前半までの対象を含み、年齢幅も11歳と他のコーホートの2倍以上になっている。同様にサンプルサイズも2倍以上の値で、全データを用いて分析する際、コーホートAの動向が推計値に与える影響力は大きい。他のコーホートと条件を近づけるために、コーホートAについては調査参加時30歳未満の対象に限定して分析する。第21回調査で追加したコーホートEは、本稿執筆時に初回調査しか経験していないため、使用データ上で脱落は発生していない。本稿ではA～Dのコーホートに属する対象者を分析に用いる。

脱落までの追跡期間の設定は以下である。始点となる追跡開始時点は、各対象の初回調査参加回に設定する。初回の調査参加回（年）は、図表-1に示すように対象が属するコーホート単位で異なる。終点とするイベント発生の時点は、当該対象の観測ができなかった調査回に設定する。この設定の下で、生存時間は各対象が何回の調査を経て脱落したかということになる。観測の打ち切りは、初回の調査参加回から6回分の調査、年の単位では5年間とする。最も早生のコーホートAに属する参加者では第1回調査から参加しているため、

調査21回分の観測データが得られる。一方で、第16回調査から参加した後生のコーホートDでは6回分の観測データしか得られない。いわゆる組み入れ遅延（late entry）の問題である。今回はコーホートDに合わせて、各コーホートとも初回参加時から6回分のデータを使用し、追跡期間を統一する。各コーホートとも全期間のデータを用いる際の対処法の模索は、今後の課題としたい。

イベントの発生を調査からの脱落自体に設定しているため、打ち切りおよびイベント発生前の追跡不能は観測されない。そのため、脱落した対象については、イベント生起までの生存期間を正確に計測できる。なお、JPSCはすべてのコーホートを毎年10月に一斉に調査している。そのため、観測の時間単位は年であり、時間間隔はコーホート間で共通している。

設定に関して1つの留意点を述べておく。一度調査から脱落しても、回をあけて調査に復活する対象の存在である。結婚や出産などのライフイベントが調査時期に重なり、当該回のみ調査に参加しない事例が、調査復活の多くを占める。結婚、出産は対象のライフステージに依存する事象である。調査への復活は、対象が若い、すなわち調査への参加回数が少ない時期に多く観測され、調査回数を経るにつれ減少していく傾向にある。今回は、この調査復活の事象について、休止した回を脱落とみなす一方、復活後のデータについては当該年度まで脱落がなかったものとして扱っている。なお復活のケース数は、各コーホートとも調査回単位でみて、多い場合でもおおむね10ケース程度である。

図表-2 コーホート別、回収率の推移

調査回	年	コーホート A		コーホート B		コーホート C		コーホート D	
		観測数	回収率	観測数	回収率	観測数	回収率	観測数	回収率
1	1993	1,500	41.4% <sup>1)</sup>	—	—	—	—	—	—
2	1994	1,422	94.8%	—	—	—	—	—	—
3	1995	1,342	94.4%	—	—	—	—	—	—
4	1996	1,298	96.7%	—	—	—	—	—	—
5	1997	1,255	96.7%	500	34.3% <sup>1)</sup>	—	—	—	—
6	1998	1,196	95.3%	442	88.4%	—	—	—	—
7	1999	1,137	95.1%	412	93.2%	—	—	—	—
8	2000	1,102	96.9%	386	93.7%	—	—	—	—
9	2001	1,059	96.1%	366	94.8%	—	—	—	—
10	2002	1,032	97.5%	344	94.0%	—	—	—	—
11	2003	980	95.0%	323	93.9%	836	28.4% <sup>1)</sup>	—	—
12	2004	944	96.3%	312	96.6%	724	86.6%	—	—
13	2005	904	95.8%	292	93.6%	674	93.1%	—	—
14	2006	875	96.8%	278	95.2%	621	92.1%	—	—
15	2007	847	96.8%	271	97.5%	588	94.7%	—	—
16	2008	828	97.8%	261	96.3%	560	95.2%	636	30.5% <sup>1)</sup>
17	2009	799	96.5%	255	97.7%	541	96.6%	573	90.1%
18	2010	778	97.4%	246	96.5%	522	96.5%	543	94.8%
19	2011	765	98.3%	243	98.8%	507	97.1%	509	93.7%
20	2012	748	97.8%	233	95.9%	491	96.8%	483	94.9%
21	2013	732	97.9%	231	99.1%	478	97.4%	451	93.4%

注: 1) 調査応諾者数を事前に送付した調査依頼状で割った値

### 3. 使用する共変量

主として着目する属性は、対象者が調査に初めて参加した時の年齢と対象者が属するコーホートについてである。これら2点の属性に着目する理由は、調査経過後の残存サンプルとサンプル抽出時とのずれを把握する指標となるためである。JPSCのサンプル抽出は、都市規模で層化した後、母集団の年齢と配偶状態に応じた比例抽出を行っている。対象者の配偶状態は調査経過に伴い変化するが、調査参加時の年齢は不変である。調査参加時の年齢が調査からの脱落に影響を与えている場合、調査が継続するにつれ同一抽出集団（コーホート）内での年齢構成にひずみが生じていることになる。初回抽出の年齢はおおむね20歳代半ばから後半の年齢である。5年の調査を経た場合を考えると、最も若い対象についてはその20歳代後半を、同一コーホート内で年齢の高い対象者についてはその30歳代前半を主に調査したことになる。そのため、中長期的スパンでみて、参加時点での年齢が低いほど脱落している傾向が強い

場合は、20歳代での調査継続が相対的に難しいことを示唆する。逆に年齢が高いほど脱落の傾向が強い場合は、むしろ30歳代以降で調査継続を阻む要素が多いことになる。

コーホート間の脱落の相違は、図表-2に示したように、各回の回収率レベルでも違いが見てとれる。概して、早生コーホートの方が脱落率は低い傾向にある。分析では、各コーホートの年齢幅の違いをコントロールした上での効果を測ることになる。対象者が属するコーホート間で脱落に与える影響の違いが認められる場合は、異なるコーホートのデータを結合して集計する際には注意する必要がある。仮に後生コーホートが脱落の傾向が高い場合、全コーホートのデータを用いて20歳代と30歳代の集計値を算出すると、30歳代に占める後生コーホートの割合は相対的に低くなる。初回参加時年齢を調整したうえで、属するコーホートが脱落に与える影響は、主として出生年代と抽出時点が異なることの相違を反映したものである。出生年代による社会調査に対する意識の違い、あるいは各コーホートの初回調査からの6回目までの調査期間の調

査環境を反映しているとも解釈できる。JPSCでは、サンプルのリフレッシュメントなどにより同一出生年代の対象を異時点間で抽出することはしていないため、属するコーホートの効果を出生年代と抽出時点の要素に識別することは容易ではない<sup>1)</sup>。

本稿では、これら2点の属性に加えて、対象の1日の生活基礎時間と脱落の関係についても、その影響を把握することにする。生活基礎時間は、主として睡眠や食事、入浴など身の回りのことに費やす時間である。JPSCでは1日の活動時間を24時間になるように尋ねており、「通勤・通学」、「勉強」、「仕事」、「家事・育児」、「趣味・娯楽・交際」に割り振れない／割り振られていない時間は、その他の時間として基礎時間に計上するように指示している。そのため、実質的に24時間から労働時間や家事・育児時間等を除いたものとなっており、対象が持つ時間的余裕の代理指標とみなすことも可能である。生活基礎時間の長さが脱落に影響を与えている場合、単に対象の忙しさが脱落に影響を及ぼしているという事実だけでなく、調査票の分量や煩雑さなど調査設計、遂行面での改良が脱落の水準に影響を及ぼす可能性を含意する点において有用な情報となる。

生活基礎時間は、その値が各回の調査を通じて変化する点において、対象の調査参加時年齢や属するコーホートとは性質が異なる。追跡調査が中長期にわたる場合、対象の属性および対象をとりまく環境の変化が、調査からの脱落を誘発する可能性がある。むしろ、調査開始から長い期間を経た時点で脱落するケースでは、これまで参加してきた調査時と異なる状況が対象の脱落に寄与したと考える方が自然ともいえる。ただし、個人の中長期の脱落過程を描写するモデルにおいては、JPSCの調査項目から用いることが可能な変数は自ずと制限される。

例えば坂本（2006）では、JPSCのコーホートAとBに関して、年収の低い層、あるいは年収低下を経験した層での脱落確率が高いことが示されている。ただ、今回の分析枠組みの下で対象の収入を分析に導入することは難しい。主な理由は、JPSCが対象を女性に限定していることにある。労働市場の参入退出が発生しやすいこと、配偶状

態の変化等により、夫あるいは同居する親の収入をどのように評価するかなど、個人のライフヒストリーを通じ、安定した収入の系列を構成することが困難である<sup>2)</sup>。また、次節で説明する時間に依存する変数を組み込んだ生存解析モデルでは、共変量の観測から微小期間でのイベント発生の有無を想定している。調査脱落時の収入は観測できないため、最後に観測された収入の時間ターム（月あるいは年）から脱落までのスパンは1年以上になる。なぜ収入が変動した直後の調査ではなく、1年以上経過した調査での脱落につながるのかの説明が難しい<sup>3)</sup>。

生活基礎時間は労働時間だけでなく家事・育児時間を除いたものであるため、対象の労働力状態のみに大きく依存しない。ただ、生活基礎時間についても、対象が脱落した調査回の値は観測することができないため、1期前の値と脱落の有無をリンクさせることになる。その際の対象の行動は以下のような仮定になる。回答終了後に今後の調査継続の意思決定を行っている。あるいは、前回の繁忙の状況と回答の手間を想起し、当期調査の参加の諾否を考えている。参加の諾否が当該期の繁忙に依存しない点において強い仮定ではあるものの、全く不合理な解釈を要するわけでもない。なお、生活基礎時間は第1回調査から第21回まで継続して、同一の質問形式で調査している項目である。本稿で生活基礎時間に着目する動機は、このようなJPSC特有のデータの制限による面も大きい。

## 4. モデル

分析では時間依存変量の個人の縦断面の動きを記述する式と、脱落発生までのプロセスを記述する式をリンクさせる。一般に、Joint Modelsと呼ばれる生存時間解析の拡張モデルである（Rizopoulos 2012）。

・縦断面プロセス

$$y_i(t) = m_i(t) + \varepsilon_i(t),$$

$$m_i(t) = x_i(t)' \boldsymbol{\beta} + z_i' b_i,$$

$$b_i \sim N(0, D), \varepsilon_i \sim N(0, \sigma^2)$$

図表-3 推計結果

		モデル 1		モデル 2		モデル 3	
		推計値	p 値	推計値	p 値	推計値	p 値
縦断面プロセス	切片	10.01 (.055)	<0.0001	10.01 (.055)	<0.0001	10.01 (.055)	<0.0001
	経過年	0.02 (.013)	0.1751	0.02 (.013)	0.1966	0.02 (.013)	0.2006
	配偶状態 (無配偶 = 1)	0.38 (.059)	<0.0001	0.38 (.059)	<0.0001	0.38 (.059)	<0.0001
脱落発生プロセス	生活基礎時間 (平日：時間) コーホート B	0.08 (.003)	<0.0001	0.06 (.007)	<0.0001	0.07 (.03)	0.0293
	コーホート C			0.34 (.11)	0.0017	0.30 (.11)	0.0062
	コーホート D			0.44 (.095)	<0.0001	0.45 (.095)	<0.0001
	調査参加時年齢			0.27 (.104)	0.0086	0.26 (.104)	0.0114
						-0.05 (.012)	<0.0001
	AIC 尤度比	69284.7		69267.34 23.36	<0.0001	69265.7 27.06	<0.0001
対象数 (unique)		2,794		2,794		2,794	

注：( ) 内は標準誤差

・脱落発生プロセス

$$h_i(t) = h_0(t) \exp(\gamma'w_i + \alpha m_i(t))$$

時間依存の共変量について、モデル上は実際の観測値に測定誤差が含まれると考える。観測された値  $y_i(t)$  ではなく、観測されない真の値  $m_i(t)$  を想定し、この  $m_i(t)$  を用いて横断面と脱落発生の両プロセスをリンクさせる。横断面プロセスでは線形混合モデルを採用し、共変量ベクトル  $x_i(t)$  に対し固定効果を、時点と対象固有の要因に対しランダム効果  $b_i$  を想定した。脱落発生プロセスにおいては、ベースラインのリスク関数  $h_0(t)$  について特定の分布を仮定しない Cox 比例ハザードモデルを考える。  $w_i$  はベースラインを規定する共変量ベクトルである。実際の推計に際して、用いた変量は下記のとおりである。  $y_i(t)$  は生活基礎時間（平日、単位：時間）を用いる。  $x_i(t)$  には経過時間と配偶状態を用いた。脱落発生プロセスの共変量の  $w_i$  には、コーホートを示すダミー変量、各対象の初回調査時の年齢を採用した。

このような2段階の推定を行う理由は、同一個

人のデータを集めたパネルデータでは、変量に関し個人内での系列相関が発生しうることを考慮するためである。生存時間解析において時間依存の共変量は、外生的な (exogenous) ものと内生的な (endogenous) ものに分けられる。Kalbfleisch and Prentice (2002) では、外生的な変量を、ある時点の変数の実現値が、その後の観測値のとりうる確率に影響を与えないものとしている。例えば、気温の周期的な季節変動などがあげられる。一方で、内生的変量はバイオマーカーなど、ある時点の実現値が、その後の観測値の確率に影響を与えるものとしている。生活基礎時間については、どのように判断できるだろうか。景気の変動などで、その年の労働時間が伸縮し、生活基礎時間が変動するケースが大半を占める場合は、外生的ともいえる。しかし、子どもが生まれると一般に家事・育児時間は増大し、結果として生活基礎時間は減る。その後、子どもの手が離れるまでは、生活基礎時間は減少したままと思われる。このように生活基礎時間の減少の多くが、子どもの誕生などに付随して発生しているならば、

内生的な変量といえる。どちらの要素が強いかは極めて実証的な話であるが、今回は変量に内生性が存在する可能性を考慮して、モデルに縦断的側面を付与したJoint Models を用いた推計を行うことにする。

## 5. 結果

推計結果を図表-3に示した。図表-3に示した3つのモデルの違いは、脱落発生プロセスの記述に採用した共変量の違いである。モデル1は、生活基礎時間だけを採用したものである。モデル2は、モデル1にコーホート属性の違いを共変量として追加したもの。モデル3はモデル2に対象者の調査参加時点での年齢を追加したものである。定式化したモデルのあてはまりの相互評価、ならびにコーホートと調査参加時年齢の係数に説明力があるかを検定する。モデル評価をAICの基準で測れば、モデル1よりもモデル2、モデル3のAICが順次小さくなっており、予測やあてはまりの面からはモデル3を選択することが好ましい結果となっている。図表-3の尤度比は、モデル1を基準にモデル2と3の値を計算したものである。コーホート属性、調査参加時年齢の係数を0とする帰無仮説を、尤度比をもとに検定すると、モデル2、モデル3とも1%水準で帰無仮説を棄却する結果となっている。これらの結果を勘案すると、コーホート属性、調査参加時年齢を加えた定式化は妥当といえる。以下では、両者を含めたモデル3の推計結果をもとに述べていく。

まず生活基礎時間が脱落に与える影響についてみる。生活基礎時間の係数は0.07であり、5%水準で有意な結果となっている。基礎時間1時間の増加は、 $\exp(0.07)=1.07$ 倍程度の脱落リスクを高めることになる。生活基礎時間の長さを回答者の時間的余裕とみなせば、得られた係数は時間的余裕がある／余裕ができるほど脱落リスクが高まることを示しており、当初の予想に反する結果である。結果の精査は今後の課題となるが、生活基礎時間の観測時点と脱落の1年のずれが影響している可能性はある。比較的時間に余裕の

あった時期の値のみがカウントされ、脱落した時の繁忙状態がカウントされていない可能性はある。また、体調の悪化が、仕事、家事・育児時間の減少につながり、生活基礎時間を増加させるケースも考えられる。健康状態とどの程度リンクしているかは別途、精査する必要がある。いずれにせよ、前期の生活基礎時間を繁忙の指標として用い、脱落への影響を解釈することは難しい結果となった。

コーホートの推計値は、最も早生のコーホートAを基準にしたものである。後生のコーホートB、C、Dともに係数が正になっており、コーホートAよりも脱落リスクが高いことを示している。コーホートBでは $\exp(0.30)=1.3$ 倍程度、コーホートCでは $\exp(0.45)=1.6$ 倍程度、コーホートDでは $\exp(0.26)=1.3$ 倍程度、それぞれコーホートAよりも脱落リスクが高い。この結果は、蓄積したデータから異世代間で同一年齢の対象を抽出した場合に、コーホート間での構成比が当初のものとは異なることを意味している。

最後に調査開始時年齢の係数をみると、負の値をとっている。年齢が1歳上がるごとに5%ほど脱落リスクが減少する結果である。逆に考えると、同一コーホート内では年齢の若い対象者ほど、わずかながら脱落しやすい傾向にある。摩耗したサンプルを扱う場合には、前述のコーホート間の相違だけではなく、コーホート内の年齢構成にも留意する必要がある。このような結果となった一つの理由として、年齢が低い層ほど、結婚、出産等を調査参加後に経験する確率が高いことが考えられる。調査の拒否理由が判明しているものうち、結婚、出産が占める割合は一定数に上る(村上2003)。

ただ一方で、坂口(2013)では、対象者の脱落によってJPSCの配偶構成が想定母集団と乖離していかないことから、結婚という理由のみならず、無配偶という状態そのものが有配偶に比して脱落に結びつきやすいことを示唆している。前述の生活基礎時間についても、その縦断面での動きには配偶状態が影響を及ぼし、無配偶ほど生活基礎時間が長い傾向にある。有配偶では結婚やその後の

出産、子育てに付随する家事・育児時間の増加が、生活基礎時間を減少させているとも考えられる。今回の結果だけからは憶測の域を出ないが、無配偶女性を若い時期にいかにかに調査に引き止められるかが、サンプル保持の観点からは鍵となるかもしれない。

## 6. まとめと今後の課題

本稿では、対象の調査参加時年齢と属するコーホート、および生活基礎時間の変動が脱落に与える影響を生存時間解析の枠組みで推計した。調査期間の短いコーホートDに合わせ、すべてのコーホートとも初回参加から6回分の動向を分析の対象にしている。

調査参加時年齢については、若い方の脱落リスクが高い結果が得られた。属するコーホート間の相違は、最も早生のコーホートAを基準にすると、続くコーホートB、C、Dでは脱落のリスクが高くなっている。これら2つの結果は、脱落により摩擦した残存サンプルにおいて、コーホート内、コーホート間の構成が初回抽出時よりずれていることを意味している。全てのコーホートを用いて、平均値等の集計値を算出する際には、構成のずれを補正する必要がある。ただ、これら構成のずれが存在することが、必ずしもJPSCを用いた回帰分析の信頼性を損ねるわけではない。構成のずれが推計に与える影響は、個々の分析が焦点をおく事象や、モデルの定式化に依存している。確実に言えることは、年齢や出生年に強く依存する事象を扱う場合は、定式化に細心の注意が必要である。例えば、加齢にともなう意識変化に焦点をあてる場合、構成のずれを意識しないモデルの定式化を行うと、出生年による意識の違いが推計値にバイアスを与えることになる。

生活基礎時間の推計値は予想に反し、時間の長さが脱落リスクを高める方向に働いていた。今後、原因について精査する一方で、脱落の状況をうまく説明できる生活基礎時間以外の共変量を見つける必要もある。しかしながら、JPSCの調査期間の20余年を通して、継続的に捕捉している項目は

限られている。また、女性のライフステージを通じて安定している系列の事象、および調査20余年の環境、時代変化を通じて含意する内容が異なる事象<sup>4)</sup>を、調査項目そのものから探しだすことは困難である。既存の複数の変数群から潜在変数を構築することが妥当な選択肢となるであろう。付随して、参加拒否者に対しても脱落時点の状況を可能な範囲で聞き出すこと。また回答者に対しても調査票の記入時間や負担感など、調査のための調査といったメタ的な情報も積極的に収集するなど、調査実施面での工夫も必要である。

最後に、脱落と時間の流れに関して、調査への参加回数と脱落の間に関係がみられることに言及しておく。図表-2で示されているように、個々の対象にとって2回目の調査、すなわち最初の追跡調査で拒否される確率が高く、その後、指数関数的に減退していく。この動きをどのようにモデルに組み込み、初回の脱落にどのような要因が強く影響を与えているかを把握すればよいか。モデルの予測能力を向上させる点からも有用であるが、初回脱落をいかに抑えるかというサンプル保持の点からも必要とされる情報である。JPSCの調査設計の性質上、調査回数は対象者の年齢の増加と同期してしまう。1年の変化が意味するところを明らかにすべく、個々の要素を識別する仮説とその数理モデルを蓄積することが求められる。

### 注

- 1) 図表-2で示されているように、コーホート間で初回収率も大きく異なる。この値も出生年やサンプル抽出時の調査環境の影響を強く受けているものと思われる。見方を変えれば、調査からの脱落にコーホート間で大きな違いがあるのは、出生年や調査参加時点の社会環境の影響だけでなく、抽出サンプルのセレクションが影響している可能性も考えられる。
- 2) 収入関連の質問については無回答 (Item Non Response) も多く、欠損しているデータへの対処法をさらに考える必要もある。
- 3) パネル調査からの脱落に関して、生存時間解析の枠組みを適用した研究は、管見の限り見いだせなかった。対象の調査からの脱落は、個々の分析において関心のあるイベントの発生を観測不能にする調査打ち切りの要因であり、コントロールすべきものとして捉えられている。ただ、共変量の観測から脱落の観測まで一定の間があくことが、脱落の分析では最大の障壁とも思える。

- 4) 女性の学歴は、JPSCの調査期間を通じて、その含意が大きく変化したものの一つである。女性の大学進学率の上昇にともない、50代女性と20代女性とでは大卒者の比率は大きく異なっている。

#### 文献

- 坂口尚文, 2013, 「『消費生活に関するパネル調査』における残存サンプルの配偶率について」『季刊家計経済研究』100: 52-58.
- 坂本和靖, 2006, 「サンプル脱落に関する分析——『消費生活に関するパネル調査』を用いた脱落の規定要因と推計バイアスの検証」『日本労働研究雑誌』551: 55-70.
- 村上あかね, 2003, 「なぜ脱落したのか——『消費生活に関するパネル調査』における脱落分析」財団法人家計経済研究所編『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』国立印刷局, 115-122.

Kalbfleisch, John D. and Ross L. Prentice, 2002, *The Statistical Analysis of Failure Time Data*, 2nd ed., Hoboken: Wiley, John & Sons.

Rizopoulos, Dimitris, 2012, *Joint Models for Longitudinal and Time-to-Event Data*, Boca Raton: CRC Press.

さかぐち・なおふみ 公益財団法人 家計経済研究所 次席研究員。主な論文に「パネルデータからみた女性の仕事・結婚・出産」(『季刊家計経済研究』92, 2011)。労働経済学専攻。(sakaguchi@kakeiken.or.jp)