

Discussion Paper Series

University of Tokyo
Institute of Social Science
Panel Survey

東京大学社会科学研究所 パネル調査プロジェクト
ディスカッションペーパーシリーズ

子どもに関する項目のデータハンドリングに関する
提案：生年・年齢・同別居を中心に

A Proposal for Data Handling of Items on Respondents' Children in JLPS:
Focusing on Birth Year, Age, and Residential Status

柳下実 (東京都立大学大学院)

石田賢示 (東京大学社会科学研究所)

Minoru YAGISHITA, Kenji ISHIDA

July 2021

No.144

子どもに関する項目のデータハンドリングに関する提案： 生年・年齢・同別居を中心に

柳下実（東京都立大学大学院）
石田賢示（東京大学社会科学研究所）

要約

本稿は、働き方とライフスタイルの変化に関する全国調査（以下、JLPS）の子どもにかんする項目のうち、とくに年齢・同別居項目を利用する際の指針を示す。子どもの年齢にかんする項目は、対象が同居子か、すべての子どもかなど、wave間で差があることや、JLPSは調査年によって調査月が異なることで、利用する際のハードルが高い可能性がある。本稿では、子どもにかんする項目を概観し、論理的には調査時の子どもの年齢がどのように増加するのかを示す。そのうえで、分析者が可能なデータハンドリングとして、子どもの生月・生年・同別居・年齢の無回答や誤答を修正する指針と、実際にハンドリングした例を示す。最後に、子どもについての設問の改善可能性について議論する。

謝辞

本研究は、日本学術振興会（JSPS）科学研究費補助金・特別推進研究（25000001, 18H05204）、基盤研究（S）（18103003, 22223005）の助成を受けたものである。東京大学社会科学研究所（東大社研）パネル調査の実施にあたっては、社会科学研究所研究資金、株式会社アウトソーシングからの奨学寄付金を受けた。パネル調査データの使用にあたっては東大社研パネル運営委員会の許可を受けた。本研究の遂行にあたって日本経済研究センター 2020年度研究奨励金の支援を受けた。

1. 問題の所在

本稿は、働き方とライフスタイルの変化に関する全国調査（Japanese Life Course Panel Surveys, 以下 JLPS と略す）の同居者項目のうち、とくに子どもにかんする情報を利用する際の指針を示す。子どもの成長が親に与える影響は興味深い問いであり、横断データを用いて知見が蓄積されてきた（たとえば、稲葉（2011））。パネルデータを用いた研究としては、家計研パネルを用い、長子の入園・入学に着目した研究がある（田中 2010）。JLPS においても、対象者の同居子や子どもにかんする情報が蓄積されてきた。JLPS を用いた、子どもの成長というライフコース上の変化に着目した研究が可能であると考えられる。

しかし、子どもの情報、とくに同居の末子年齢や長子年齢¹の影響を分析するうえでは、検討する必要がある問題が何点かあると考えられる。たとえば、子どもにかんする項目は同居子について尋ねた wave や、すべての子どもについて尋ねた wave など、wave 間で項目に差があることや、調査年によって調査月が異なることで利用する際のハードルが高い可能性がある。

具体的にみるため、同居の末子年齢の移動表を確認してみよう。年齢を1歳刻みで示すと非常に大きな表になるため、ここでは同居末子年齢を、同居子なし、0～2歳、3～5歳、6～8歳、9～11歳、12～14歳、15～17歳、18歳以上と単純化して示した。表側が $t-1$ 時点、表頭が t 時点のカテゴリを示す。緑で示したセルは、移動の中でも生じやすいと考えられる移動を含む。

たとえば、 $t-1$ 時点で同居子なしだった人が、 t 時点で同居子なしであったり、0～2歳の子どもがいる（子どもが生まれた）という移動は生じやすいと考えられる。同様に、 $t-1$ 時点で末子年齢が何歳であろうと、0～2歳への移動というのは、新たに子どもが生まれたと考えられる。同様に $t-1$ 時点で末子年齢が0～2歳だった人が、 t 時点で0～2歳であったり、3～5歳であるのは、理解しやすい。

一方で、なぜそのような移動が生じたのか、一見して理解しにくい移動もある。たとえば、 $t-1$ 時点で末子年齢が0～2歳であったのに、 t 時点では同居子なしに移動しているケース（子どもがいなくなった？）や、 $t-1$ 時点では同居子なしだったのに、 t 時点では3～5歳の子どもの出現しているケース（報告忘れ？再婚？）である。こうした移動は、離婚や再婚、対象者の単身赴任などで、実際に移動が生じたケースと、無回答や誤記によってデータのうえでのみ移動が生じたケースに分けられる。

¹ 本稿で年齢とは age at last birth を指す。

表1. 同居末子年齢の遷移表

	同居子なし	0～2歳	3～5歳	6～8歳	9～11歳	12～14歳	15～17歳	18歳以上	計
同居子なし	17,655	761	35	42	48	26	37	44	18,648
%	94.68	4.08	0.19	0.23	0.26	0.14	0.2	0.24	100
0～2歳	68	4,134	1,555	10	8	9	1	1	5,786
%	1.18	71.45	26.88	0.17	0.14	0.16	0.02	0.02	100
3～5歳	60	323	2,694	1,320	8	5	3	1	4,414
%	1.36	7.32	61.03	29.9	0.18	0.11	0.07	0.02	100
6～8歳	59	82	18	2,361	1,137	8	3	8	3,676
%	1.61	2.23	0.49	64.23	30.93	0.22	0.08	0.22	100
9～11歳	45	27	2	9	1,997	945	8	9	3,042
%	1.48	0.89	0.07	0.3	65.65	31.07	0.26	0.3	100
12～14歳	45	10	3	4	19	1,536	669	14	2,300
%	1.96	0.43	0.13	0.17	0.83	66.78	29.09	0.61	100
15～17歳	51	0	2	2	1	13	982	355	1,406
%	3.63	0	0.14	0.14	0.07	0.92	69.84	25.25	100
18歳以上	79	1	1	3	2	10	7	879	982
%	8.04	0.1	0.1	0.31	0.2	1.02	0.71	89.51	100
計	18,062	5,338	4,310	3,751	3,220	2,552	1,710	1,311	40,254
%	44.87	13.26	10.71	9.32	8	6.34	4.25	3.26	100

JLPS の同居末子年齢の遷移表では複数の要因により、複雑な移動が生じる。第一の要因は、調査年によって調査タイミングや調査形式が異なることである。第二は、年齢や生年の誤答・無回答である。たとえば、3歳なのに2歳と回答することで、生じるはずの移動が生じなかったり、2011年生まれなのに2013年生まれと回答することで、2歳分移動したりというケースである。第三は、同居についての誤答・無回答である。同居しているにもかかわらず別居と回答することで、前の wave では子どもがいたのに次の wave ではいないという移動が生じる。同居子の情報にかんするこれらのエラーは、本来はデータクリーニングの段階で解決するのが望ましい。しかし、エラーが生じた回答者についてどの情報が正しいのかを容易に判別できない場合も多い。したがって、本稿での議論とデータハンドリングが、同居子の情報を用いる分析において網羅的に適用可能であるかについては留保がつく。しかしながら、同時に本稿の議論は JLPS あるいは類似の方法で子どもの情報を測定する調査のデータクリーニングに対しても、実践的な示唆をもたらすものと思われる。

本稿では、同居子や子どもにかんする項目を概観し、論理的には調査時の子どもの年齢がどのように増加するのかを示す。そのうえで、データ利用者が可能なデータハンドリングとして、子どもの生年・生月・年齢および同別居の無回答や誤答を修正する指針を示す。最後に、子どもについての設問がどのように改善可能か議論する。

2. 調査形式と調査のタイミング

JLPS で同居の長子年齢や末子年齢を作成する際に利用する設問の、wave 1 から 13 までの構成は以下ようになる。もっとも大きな wave 間の差は、すべての子どもについて尋ねた wave 1 や wave 7 以降と、同居子のみについて尋ねた wave 2 から wave 6 というように、設問の形式が異なる点である。すべての子どもについて尋ねた wave では年齢が問われていないため、年齢を計算する必要がある。そのため wave 1 や wave 7 以降で年齢を計算する際は、一般に調査終了時の年齢が計算されることが多いと考えられるのに対し、wave 2 から 6 までの年齢は調査回答時のものである。また wave によって調査月が異なる。

wave 1 や wave 7 以降で子どもの年齢を計算する際には、調査年から生年を引くよりも、調査年の一年前から生年を引く方が正しい年齢が計算できる可能性が高い。JLPS は調査が一年の前半におこなわれることが多く、調査終了時点までに誕生日が来ていない子どもが多いと考えられる。調査年から生年を引いて計算した年齢は、その年の 12 月 31 日にすべての子どもが到達する年齢のため、調査時点の年齢よりは一年多くなってしまう人が多い。より正確な計算は、wave 10 から尋ねられた生月を、それ以前の同一の生年の子どもの生月に代入することで可能になる。また、調査年に生まれた子どもについては、調査年から生年を引いて計算する必要がある。年齢の計算については、後ほど詳しく取り扱う。

表 2. 子どもの年齢・性別・同別居にかんする設問と調査タイミング (wave 1～13)

	設問構成	生年	生月	年齢	性別	同別居	調査年	調査月
wave 1	すべての子ども	○			○	○	2007 年	1～4 月
wave 2	同居子			○	○		2008 年	1～4 月
wave 3	同居子			○	○		2009 年	1～3 月
wave 4	同居子			○	○		2010 年	1～3 月
wave 5	同居子			○	○		2011 年	1～3 月
wave 6	同居子			○	○		2012 年	1～3 月
wave 7	すべての子ども	○			○	○	2013 年	1～3 月
wave 8	すべての子ども	○			○	○	2014 年	1～3 月
wave 9	すべての子ども	○			○	○	2015 年	1～3 月
wave 10	すべての子ども	○	○		○	○	2016 年	4～6 月
wave 11	すべての子ども	○	○		○	○	2017 年	3～5 月
wave 12	すべての子ども	○	○		○	○	2018 年	1～3 月
wave 13	すべての子ども	○	○		○	○	2019 年	1～3 月

移動表における一見して理解しにくい移動には、JLPS の調査タイミングも影響する。以下の表は 2007 年の各月に生まれた子どもが各 wave の調査タイミングで何歳かを示した。これをみると 0 歳, 1 歳, 2 歳と 1 年ずつ歳を重ねる子どももいれば, 4 月生まれのように, wave 2 から 3 では, 1 歳から 1 歳へと歳を重ねず, wave 9 から 10 では 7 歳から 9 歳と, 2 年分歳を重ねる子どももいる。調査タイミングが wave によって異なることも, 一見して分かりにくい移動の発生に寄与する。

表 3. JLPS の各 wave 調査タイミングで 2007 年の各月に生まれた子どもは何歳か

	生まれ月							
	1	2	3	4	5	6	7~9	10~12
wave 1 (2007 年 4 月末)	0 歳				NA			
wave 2 (2008 年 4 月末)	1 歳				0 歳			
wave 3 (2009 年 3 月末)	2 歳			1 歳				
wave 4 (2010 年 3 月末)	3 歳			2 歳				
wave 5 (2011 年 3 月末)	4 歳			3 歳				
wave 6 (2012 年 3 月末)	5 歳			4 歳				
wave 7 (2013 年 3 月末)	6 歳			5 歳				
wave 8 (2014 年 3 月末)	7 歳			6 歳				
wave 9 (2015 年 3 月末)	8 歳			7 歳				
wave 10 (2016 年 6 月末)	9 歳						8 歳	
wave 11 (2017 年 5 月末)	10 歳					9 歳		
wave 12 (2018 年 3 月末)	11 歳			10 歳				
wave 13 (2019 年 3 月末)	12 歳			11 歳				

注：調査タイミングは各調査年の調査最終月とした。NA: not applicable.

上記の例をもとにした架空データから年齢の遷移表を作成し, 下に示した。2007 年の各月に生まれた 12 人の子どもの年齢が, wave 1 から 13 まで実際どのように変化するかを示した。表側に t 時点の年齢を, 表頭に $t+1$ 時点の年齢を示した。架空データの再現コードは末尾に示した。

本架空データは測定誤差がないため, 対角線上の移動, つまり NA から 0 歳, 0 歳から 1 歳といった, 1 年ずつ年を重ねる移動が多い。しかし, 1 歳から 1 歳への移動や, 7 歳から 9 歳への移動, 9 歳から 9 歳や 10 歳から 10 歳への移動がみられ, wave が一つ進むといても年齢が進まなかったり, 2 年分進んだりといったケースもあることがわかる。JLPS では

データがどれほど整っていたとしても、調査最終月時点の子どもの年齢の遷移表では、対角線上のセルに分類されない移動が生じる。

表 4. 年齢の遷移表 (2007 年生まれ 1 月~12 月に 1 人ずつ)

t	$t+1$ の年齢												
	0	1	2	3	4	5	6	7	8	9	10	11	12
NA	8												
0		12											
1		1	12										
2				12									
3					12								
4						12							
5							12						
6								12					
7									9	3			
8										9			
9										1	12		
10											2	12	
11													3

注. 架空データ. 回答は調査タイミングの最終月と仮定した.

上に示した架空データとは異なり、実際のデータでは同じ調査回答者の回答時期が wave によって異なることで、子どもの年齢の増加がずれることもある。子どもの誕生日を 2007 年 2 月 15 日と仮定して、wave 2 と wave 3 の回答時期について、誕生日前と誕生日後とを組み合わせた 4 パターンを考える。wave 2・3 とともに誕生日前に回答したケースや、誕生日後に回答したケースでは 0 歳から 1 歳、また 1 歳から 2 歳といった 1 歳分の移動が観測される。一方で、wave 2 は誕生日後・wave 3 は誕生日前であると、1 歳から 1 歳であり、0 年分の移動、wave 2 は誕生日前・wave 3 は誕生日後であると、0 歳から 2 歳といった 2 年分の移動が観測される。

表 5. 子どもの誕生日を 2007/2/15 と仮定，w2 調査 2008 年 1~4 月，w3 調査 2009 年 1~3 月

	w3 誕生日前 (1 歳)	w3 誕生日以後 (2 歳)
w2 誕生日前 (0 歳)	0→1 (+1)	0→2 (+2)
w2 誕生日以後 (1 歳)	1→1 (+0)	1→2 (+1)

年齢に移動が生じないことや 2 年分の移動は，ほぼ同じ年度に調査がされている wave 11 から 12 を除き，分析者が回答時期を指定できる wave 7 以降は生じないが，回答者が回答の際に子どもの年齢を計算して示す wave 2 から 6 では生じる．そのため，上記の例のように，回答時期が wave によって異なるだけで，回答時期に誕生日（とくに 1 から 3 月生まれ）が含まれる子どもの年齢は 0 歳増加から 2 歳増加までずれが生じる．ここまで，調査タイミングや回答タイミングに起因する年齢のズレを述べたが，以下ではこれら以外に起因するズレ，すなわち生年・生月・年齢・同別居について，無回答・誤答と判断できるケースについて取り扱い例を示す．

3. データの準備

無回答・誤答の例を示す前に，データの準備をすることで，ハンドリングが容易になる．wave 2 から 6 については，設問が子どものみではなく，同居者にかんする項目で子どもについても回答する形式になっている．そのため，同居者の項目から同居子を抜き出したほうが，その後のハンドリングが簡単になる．同居子を抜き出す方法はいくつかあると思われる．一例として，同居者が 8 人分変数として保存されている wave を念頭に置くと，8 人分の変数それぞれについて，子どもであったら 1，子どもでなかったら 0 と判定する変数を作成する．判定した全員分の変数を組み合わせることで，どのようなパターンで情報が入力されているのかが明らかになる．たとえば，同居者が 3 人いて，1 人目が配偶者，2 人目・3 人目が子どもの場合は，8 人分組み合わせた変数には”01100000”という値が入る．このパターンに沿って，データハンドリング用に新しく作る同居子 1 番目の変数には，もともとの同居者 2 人目の値を代入し，同居子 2 番目の変数には，もともとの同居者 3 人目の値を代入する．こうした作業をすべてのパターンについておこなうことで，同居者から同居子を抜き出すことができる．

この作業をした後の注意点は，同じ位置に必ずしも同じ子どもについての情報が入っているとは限らない点である．すべての子どもについて回答するよう求めている wave 1 や 7 以降であっても，書き忘れや，離家した子どもを書かないことなどに起因する子どもの位置のずれがある．また，同居子について回答を求めた wave 2 から 6 については，離家によると思われる年長の子どもの非同居が生じると，今までは年長の子どもを第一子の位置に書

いていたが、非同居が生じた wave では年少の子どもを第一子の位置に書くため、位置がずれる。また、ある wave では年齢が高い順で報告していたのが、別の wave では年齢が低い順に報告していたり、ある wave では男女混合で報告していたのが、別の wave では男女別に報告したりすることで、同じ子どもであっても位置がずれることが見受けられた。こうした子どもの位置をすべてそろえるのは大きな労力がかかるため、今回おこなったハンドリングでは、すべてそろえることはしなかったが、ハンドリングを容易にするために一部についてはそろえた。また、ハンドリングの参考になる変数としては、婚姻状態・同居者数・子ども数・子どもとの同別居（息子・娘）である。子どもについての情報は、第一子（年齢・性別・同別居・生年・生月）、第二子……と並べ替えるとハンドリングが容易になると思われる。

4. 生月の無回答・誤記の確認と修正

次に、生月について取り扱う。生月は wave10-13 で回答があり、ハンドリングの対象となる wave が少ないため、最初に取り扱うのがよいと思われる。生月についても無回答や、ある wave では7月生まれと答えていたり、8月生まれと答えていたりなど、wave 間で相違がある場合がある。たとえば、きょうだいで同じ月を書いてしまうという回答者の誤りに起因するとおもわれる相違や、回答者の数字の記入の仕方によっては、調査票点検、データ入力時に「1」を「7」と読み違えてしまったことに起因するとおもわれる相違がみられた²。wave 10 と 11, 12, 13 や、wave 11 と 12 など各 wave を比較し、無回答やズレを抜き出し、一貫した回答に修正することが考えられる。

次に、wave 1 や wave7-9 の生月に wave 10 以降の回答を代入する。この作業をすることで、生月の回答を求めている wave 1 や 7-9 であっても、より正確な年齢の計算ができ、wave 2~6 の年齢の齟齬を判定する際に判断が容易になる。ただし、同年生まれで生まれ月が異なる子どもが別々に存在する場合に問題が生じる。たとえば第一子は 2004 年 1 月生まれ、第二子は 2004 年 12 月生まれといった場合や、再婚後の子どもと再婚前の子どもの生年が同じという場合である。ただ、こうしたケースは非常に少ないと思われる。

また、生月を代入した結果、ある子どもが調査時点には存在しないはずだったことが明らかになる場合もあり、そうした場合は該当する wave の値を欠損にする必要がある。たとえば、2013 年 7 月生まれなのにもかかわらず、wave 7（調査時期 2013 年 1~3 月）で生年を「2013 年」と回答していた場合が考えられる。生月を代入する前は、調査年生まれで齟齬

² こうしたエラーは調査票点検段階で検出できることが望ましいが、点検時に過去のすべての wave の調査票と照合することは現実的ではない。

はないが、生月を代入した結果、wave 7 の調査時点である 2013 年 1～3 月には生まれていなかったことが明らかになる。

5. 生年の誤答・無回答

次に、生年の誤答・無回答へ対応する。まず、無回答への対応だが、特に同居と回答しているのにもかかわらず、生年・生月無回答へ対応する必要がある。これは、前年と回答内容に変更がないので、回答者としては質問への回答を飛ばしたくなり、無回答が生じると考えられる。これは、同居人数・子ども数の変動や、前の wave と後の wave の回答状況を参考に、代入できる部分については前後の情報で代入することができる。この作業では、すべての子どもについて、各 wave で第 n 子について同居と回答するが、年齢や生年を無回答のケースをワイド形式のデータで判定し、当てはまるケースの全 wave の回答をロング形式で確認し、代入できる部分について代入する。ここでの「ワイド形式」とは、一人ひとりの回答者がケースの単位となっており、各 wave の情報が列方向に蓄積されているデータの形式である。一方、「ロング形式」では各 wave の回答がケースの単位をなす。同一回答者にかんする各 wave の回答がネストしたデータの形式となっている（より詳しい説明は三輪・林編（2014）を参照）。

次に、生年の誤答への回答に対応する。まず、wave 1 や 7 以降で、同じ子どもについて答えていると思われるにもかかわらず、生年がずれるという傾向が見られた。これは、年号と西暦に起因するものが多いと思われる。昭和は西暦と 5 年ずれる（昭和 60 年が 1985 年）ので、記入ミスはあまりないと考えられるが、西暦と平成は 2 年ずれるので（平成 5 年が 1993 年）、混同する例がみられた。生年が 2 年早くなったり、遅くなったりという記入の変動は回答者が記入する際に、西暦と年号で混乱して回答にズレが生じたと推察される。この部分のハンドリングについては、ワイド形式のデータで第 n 子について wave 1 と 7～13、wave 7 と 8～13 などを対照し（以下の表参照）、同じ位置で回答されているにもかかわらず、生年が異なる対象者の ID を確認する。次に、ロング形式のデータで確認した対象者の全 wave の値をみながら、回答位置がずれたために生年のずれが生じたのか、それとも回答に問題があるのかを判定する。修正できる回答について修正する。

表 6. 生年のずれを確認する際の対照例（すべての子どもについて対照する）

	w7	w8	w9	w10	w11	w12	w13
w1	w1 & w7	w1 & w8	w1 & w9	w1 & w10	w1 & w11	w1 & w12	w1 & w13
w7		w7 & w8	w7 & w9	w7 & w10	w7 & w11	w7 & w12	w7 & w13
w8			w8 & w9	w8 & w10	w8 & w11	w8 & w12	w8 & w13
w9				w9 & w10	w9 & w11	w9 & w12	w9 & w13
w10					w10 & w11	w10 & w12	w10 & w13
w11						w11 & w12	w11 & w13
w12							w12 & w13

6. 同別居の誤答・無回答の可能性

wave 1 や 7 以降など、すべての子どもの同別居について回答する場合は、同別居についての回答が正しいとは限らない。wave 2 から 6 についても同様の懸念はあるが、確認することが難しい。wave 1 と 7 以降では、別の設問である、誰と同居しているのかについての設問で、娘や息子と同居していると回答しているのに、同居子の設問では無回答であったり、別居と答えていたりするケースがある。こうした回答はどちらかが誤りだと考えられる。同居者数を答えている場合は、それと前後の wave の同別居の傾向を見て、無回答が同居か別居か、また別居という回答が誤っているかどうかを判断する。「別居」と答えていて、誤りではない回答の例としては、単身赴任が考えられる。単身赴任と思われるケースは、ある wave で婚姻状況は変化しないが、同居者数が 1 人で一人暮らしになり、同居子数も 0 となるものの、数年立つと元の状況に戻るといったパターンがみられる。

7. 年齢の誤答・無回答

wave 1, 7 以降の生年と同別居のハンドリングが終わったら、wave 1~13 の年齢のハンドリングに移る。まず、wave 7 以降の同居長子年齢・同居末子年齢を確定させるために、wave 間（wave 12 と 13 など）で比較し、wave 12 の同居長子・末子年齢が wave 13 の同居長子・末子年齢から 1 を引いたものと一致しないケースについて確認し、誤りがないか確認する。これを wave 11 と 12, 10 と 11……と繰り返す³。wave 7 以降のハンドリングが終わり次第、wave 2 から 6 の年齢のハンドリングに移る。

wave 2 から 6 でもおこなう作業は同様であり、同居末子年齢や同居長子年齢を計算し、

³ wave 11 と 12 はほぼ同じ年度で調査がされているので、注意が必要である。

ワイド形式のデータで wave 1 と 2 を比較し、wave 2 の末子年齢から 1 を引いたものが、wave 1 の末子年齢と一致しない ID について、ロング形式のデータで全 wave の回答をみながら修正する。修正する順番は、wave 1 と 2 を比較し修正、wave 6 と 7 を比較し修正、その後、2 と 3、3 と 4 と順に修正するのがよいと思われる。wave 1 と 3 の比較など、隣り合わない wave でも比較したほうがよいが、作業が多くなる。そうした比較をするかは分析者の判断による。

また wave 1 と 2 や wave 6 と 7 を比較する際には、wave 10 以降に回答がないか、あっても無回答で生月を答えておらず、さらにその子どもが早生まれの場合に問題が生じる。第一に、そうした場合に wave 1 と 2、wave 6 と 7 で同じ年齢が計算される可能性がある。例えば、2007 年 3 月生まれだが、生月を答えていないケースを考える。wave 1 では 2007 年生まれなので、2007 から 2007 を引いて、0 歳と年齢は計算される。wave 2 回答時に、たとえば 2008 年 2 月に回答したとすると、該当の子どもの年齢は 0 歳であるので、0 歳と回答する。そのため、wave 1 で 0 歳、wave 2 で 0 歳という状況が発生する。第二に、そうした場合に wave 1 と 2、wave 6 と 7 で 2 歳異なる年齢が計算される可能性がある。2006 年 1 月生まれのケースでは、2006 年生まれという情報しかない場合は、wave 1 では 2006 から 2006 を引いて、本当は 1 歳なのに、年齢が 0 歳と計算されるが、wave 2 回答時に、2008 年 2 月に回答したとすると、親による回答は 2 歳となる。そのため、早生まれで生月を回答していない場合、wave 1 から 2 および wave 6 から 7 において、同じ年齢が続くか、2 歳飛んで回答されることがある。そうしたケースについては、早生まれを判定する変数を wave 1 および wave 7~13 まで作成し、早生まれの場合は調査年と同じ年数を生年から引くことで、wave 2 から 6 と整合的な年齢を計算することができる。また、wave 2 から 6 の年齢は調査時点の年齢なので、wave 1 のみ生年を回答しており、wave 2 以降の年齢の回答と齟齬がある場合は、wave 1 の生年を誤って回答したと判断するのがよい可能性が高い。

加えて、早生まれの場合は wave 2~6 の回答時期がおおむね 1 月から 3 月であるため、1 歳も進まなかったり、2 歳進む年があったりという状況が発生する。調査時点の年齢の影響を見たい場合には問題がないが、3 月末時点での満年齢（学年）の影響を考慮したいときには問題となる。そのため、wave 2 から 6 の同居子の年齢については、調査時点の年齢に修正するのか、つまり論理的にあり得ない回答のみを修正するのか、それとも 3 月末時点の満年齢に修正するのか、どちらかを選択する必要が出てくる。最も理想的な方法は、wave 2~6 についても、生年や生月を割り当てることだが、必要とされる労力が非常に多くなると推測されるため、今回はそうした方法を取っていない。生年や生月を割り当てることで、分析する際に年齢をより柔軟に計算することができる。

ここまでハンドリングをしたうえで、生年にもとづく年齢の計算は以下の 4 つのパターンに分かれる。①全員が早生まれでないとは仮定し、調査前年から生年を引く、②次に、生月

を回答しており、早生まれの人については調査年から生年を引く、③ハンドリングで早生まれと判定された人については、調査年から生年を引く、④調査年に生まれた人については、調査年から生年を引く。この計算で、おおむね一貫した年齢が計算できると思われる。

また、同じ子どもについての情報が同じ位置にある、つまり同じ変数に同じ子どもについての情報が入っている場合は、wave 1 の年齢に 1 を足したり、wave 7 の年齢から 1 を引いたりすることで wave 2 から 6 の年齢が容易に計算できる。しかし、上述の子どもの離家などの理由でずれが生じている場合で、一括して計算したい場合は、子どもについての情報の位置をそろえる必要がある。確認せずに計算すると、正しい値を間違った値で上書きする可能性がある。

8. ハンドリング後の遷移表とハンドリングが推定に与える影響

実際にハンドリングした一例を以下に示した。上記に示したのは指針であり、同じ指針に従ってもハンドリング方法はさまざまあると思われる。そのため、一例としてみていただきたい。

表7. 同居末子年齢の遷移表（データハンドリング後）

	同居子なし	0～2歳	3～5歳	6～8歳	9～11歳	12～14歳	15～17歳	18歳以上	計
同居子なし	17,606	754	14	17	17	9	18	35	18,470
%	95.32	4.08	0.08	0.09	0.09	0.05	0.1	0.19	100
0～2歳	32	3,941	1,578	1	1	0	0	0	5,553
%	0.58	70.97	28.42	0.02	0.02	0	0	0	100
3～5歳	32	352	2,807	1,317	1	0	0	0	4,509
%	0.71	7.81	62.25	29.21	0.02	0	0	0	100
6～8歳	21	88	0	2,507	1,159	0	0	1	3,776
%	0.56	2.33	0	66.39	30.69	0	0	0.03	100
9～11歳	21	25	0	0	2,138	958	0	3	3,145
%	0.67	0.79	0	0	67.98	30.46	0	0.1	100
12～14歳	23	5	0	1	0	1,650	679	2	2,360
%	0.97	0.21	0	0.04	0	69.92	28.77	0.08	100
15～17歳	34	2	0	0	0	1	1,042	364	1,443
%	2.36	0.14	0	0	0	0.07	72.21	25.23	100
18歳以上	79	1	0	0	0	3	1	914	998
%	7.92	0.1	0	0	0	0.3	0.1	91.58	100
計	17,848	5,168	4,399	3,843	3,316	2,621	1,740	1,319	40,254
%	44.34	12.84	10.93	9.55	8.24	6.51	4.32	3.28	100

まず、緑色のセルの部分については、同居子なし→同居子なしや0～2歳→0～2歳のように該当ケースが減ったセルもみられたが、それ以外の大部分のセルは該当ケースが増えた。

ハンドリング前は他のセルに分類されていた移動が、ハンドリングによって分類しなおされたためと考えられる。次に表側の子どもがいたにもかかわらず、次の wave で同居子なしと答えた移動については、18 歳以上については変動がないが、それ以外については半減した。表頭の同居子なしと答えていて、次の wave で 3 歳以上の子どもがいると答えた移動についても半減した。ハンドリング後は緑色のセルを除く上三角・下三角の部分に分類される移動はほとんどなくなった。ハンドリングで誤答や無回答による移動が、より生じやすいと考えられる緑色の部分に再分類されたことが分かる。

次に、ハンドリングが推定に与える影響をみる例として、心理的な尺度として抑うつ尺度を、行動面の尺度として家事頻度を取り上げる。もし無回答や誤答がランダムなのだとすれば、推定結果への影響はおもに標準誤差に現れる一方、ランダムではない無回答・誤答ならば、回帰係数への影響が生じると推測される。独立変数として同居末子年齢を投入し、統制変数として wave ダミーのみを加えたシンプルなモデルである。推定は固定効果モデルでおこなった。多くの統制変数を投入したモデルとは、ハンドリングが係数に与える影響が異なるかもしれない点に留意が必要である。

表8. 抑うつ尺度と家事頻度にかんする固定効果モデルの結果と比較

	抑うつ尺度			家事頻度		
	元データ	ハンドリング後	係数の差	元データ	ハンドリング後	係数の差
末子年齢 (基準: 同居子なし)						
0 ~ 2歳	-0.376 *** (.072)	-0.402 *** (.074)	6.5%	3.804 *** (.206)	3.949 *** (.218)	3.7%
3 ~ 5歳	-0.129 † (.078)	-0.159 † (.083)	18.9%	2.976 *** (.207)	3.154 *** (.223)	5.6%
6 ~ 8歳	.016 (.086)	-.056 (.092)	127.9%	2.386 *** (.213)	2.562 *** (.231)	6.9%
9 ~ 11歳	.136 (.093)	.107 (.101)	-27.1%	1.785 *** (.225)	1.925 *** (.245)	7.3%
12 ~ 14歳	.103 (.102)	.083 (.110)	-24.7%	1.254 *** (.244)	1.312 *** (.265)	4.4%
15 ~ 17歳	.215 † (.119)	.154 (.129)	-39.6%	.785 ** (.269)	.801 ** (.288)	2.0%
18歳以上	.201 (.145)	.189 (.156)	-6.3%	-.442 (.323)	-.573 † (.334)	22.9%
観測数	49455	49455		39412	39412	
個人	8071	8071		8085	8085	
sigma_u	3.083	3.082		7.792	7.778	
sigma_e	2.399	2.399		3.742	3.736	

注. 括弧内はロバスト標準誤差である。Wave ダミーを投入した。† $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

まず、抑うつ尺度の結果を見てみると、有意水準はおおむね一致している。15～17歳の係数については、元データでは10%水準で有意だが、ハンドリング後では有意でない。係数の差は、ハンドリング後の係数を基準とし、ハンドリング後のデータから得られた係数から元データから得られた係数を引き、それをハンドリング後のデータから得られた係数で割った。抑うつ0～2歳を例にすると、 $\{-.402 - (-.376)\} / (-.402) \approx 0.065$ となる。抑うつでは、有意でない変数の係数を除外して、0～2歳の0.1%水準で有意な係数でも、6.5%の係数の差がある。元データを用いた推定の方が全体的に正の方向に係数の値が高い。家事頻度では、18歳以上のみ有意水準が異なる。係数にかんしては元データのほうが全体的に0に近づく傾向が見られる。家事頻度では両データで有意な係数で係数の差は2.0%から7.3%であった。標準誤差にかんしては、抑うつ・家事頻度のどちらでもハンドリング後よりも元データのほうが小さい⁴。

抑うつ尺度・家事頻度についての単純な固定効果モデルでは、元データでもハンドリング後のデータでも、ほぼ異ならない結果が得られた。ただし、今回の推定で用いたモデルは統制変数が非常に少ない。多数の統制変数を入れたモデルでは、異なる推定結果が得られる可能性がある。ハンドリング後のデータを用いた推定を真の値と仮定すると、元データを用いた推定では有意な係数でもバイアスが生じており、分析の目的によるがある程度はハンドリングをした方が望ましいと考えられる。

9. 結論——子どもについての設問の改善可能性——

本稿では、同居子項目、とくに同居子の年齢に着目した分析をする際に、留意すべき点について論じた。同居末子年齢の遷移表からは、1年ごとの調査において生じにくいと考えられる移動が生じているように見えるが、調査時点や回答タイミングが異なることで、そうした移動が生じることを指摘した。そのうえで、ハンドリングする際のデータの加工について述べ、生月・生年、同別居、年齢のハンドリングについて指針を示した。次に、実際に例として指針をもとにハンドリングしたデータと元データで、遷移表や固定効果モデルの結果を比較し、ハンドリングが与える影響の例を示した。同居の長子年齢や末子年齢に着目した分析はJLPSの長所を生かす分析となると思われるが、同居子の年齢は上述のように扱いづらい側面がある。本稿の指針によって扱いやすくなることを期待する。

そのうえで、ハンドリングをするなかで明らかになった、子どもについての設問の改善可能性について、3点述べる。1点目として、子どもの生年を回答する際に平成や令和といっ

⁴ 係数の差の検定では(Clogg et al. 1995)、元データとハンドリング後で末子年齢において有意な係数の差はなかった。

た元号に基づいて回答することで生じる、生年の wave 間の変動を減らすことが挙げられる。子どもの年齢については正しく回答できる親が多いと思われるが、生年についてははっきりと記憶しているか分からない人がいるという印象を受けた。調査回答時に「子どもは今 n 歳だから、平成 m 年生まれ」などと計算して、回答している可能性がある。平成や令和は西暦とのズレが少なく、その際に計算ミスなどで生年の変動が生じている可能性がある。JLPS の第一子の生年について確認すると、調査票上で「西暦・昭和・平成」と横に並んでも、縦に並んでもいても、「平成」に基づいた回答が多い。そのため、調査票に「平成 20 年は 2008 年」など、調査年で一番回答が多そうな年齢層の和暦と西暦を示すことは考えられるが、そうした介入によって回答ミスが減るのか、増えるのかは明らかでないため、導入する前に検証する必要がある。

また、子どもの生年の誤答の可能性はパネル調査だけでなく、横断調査においても問題となる。パネル調査であれ、横断調査であれ、対象者の親や子どもの生年については、対象者本人の生年よりも測定の際の誤差が大きくなることが考えられる。調査対象者の親や子どもの生年をどのようにすれば正確に回答してもらえるのかは、今後の課題だろう。

2 点目として、回答者がすべての子どもを回答しているわけではない点にも改善の余地があると考えられる。「すでに独立している方も含めてください」というリード文があるものの、離家と推測されるが、回答されたり、回答されなかったりと、回答に変動が生じる子どもがみられた。また、年少の子どもでも途中から報告されなくなる子どもも存在した。報告されなくなる子どもについては、生きており、調査時点で離家や入院などで別居のため、回答されていない場合と、すでに亡くなっており回答されない場合が考えられる。JLPS では親や配偶者の死亡については、経験したかどうかを尋ねているが、子どもの死亡については尋ねていない。発生件数は低いかもしれないが、対象者に大きな影響を与えるイベントではあり、項目として検討する余地はあるかもしれない。

3 点目として、JLPS では離別への移行もある程度生じており（不破・柳下 2017）、離別後の再婚によるステップファミリーが一定数おり、実子かどうかという項目を導入することで、現在同居していなくても親子関係がある、または同居していても実子でない子どもについての回答が促される可能性がある。アメリカでは、2010 年の国勢調査で、世帯主の子どものうち、実子が 93.0%、養子が 2.3%、ステップチルドレンが 4.7%である（Kreider & Lofquist 2014）。たとえば American Community Survey の子どもについての質問項目では、子どもを実子の息子・娘（Biological son or daughter）、養子の息子・娘（Adopted son or daughter）、継子の息子・娘（Stepson or stepdaughter）、義理の息子・娘（Son-in-law or daughter-in-law）と分類している（Kreider & Lofquist 2014）。今回のハンドリングでは再婚の結果、実子と思われる子どもについて回答しなくなり、再婚後のステップチルドレンとみられる子どもを新たに回答していると推測されるケースがみられた。日本の量的調査はこうしたステップファ

ミリーやステップチルドレンを捉えるのが難しい質問項目を採用してきた（野沢 2008）。JLPS に質問項目として「実子か、そうでないか」という項目を加えることで、再婚後の回答者などが回答しやすくなる可能性もあるが、回答者が質問を理解できるのかという問題もあり、小規模な実験や調査などでこうした項目の導入可能性を探る必要があると推察される。

参考文献

- Clogg, C., E. Petkova, and A. Haritou, 1995, “Statistical methods for comparing regression coefficients between models,” *American Journal of Sociology*, 100(5): 1261–293.
- 不破麻紀子・柳下実, 2017, 「離死別者の親同居」『東京大学社会科学研究所 パネル調査プロジェクト ディスカッションペーパーシリーズ』 No. 103.
- 稲葉昭英, 2011, 「NFRJ98/03/08 から見た日本の家族の現状と変化」『家族社会学研究』 23(1): 43–52.
- Kreider, Rose M. and Daphne A. Lofquist, 2014, *Adopted Children and Stepchildren: 2010*, Report Number P20-572.
- 三輪哲・林雄亮編, 2014, 『SPSS による応用多変量解析』 オーム社.
- 野沢慎司, 2008, 「ステップファミリー研究の動向——アメリカからの視点」『家族社会学研究』 20(2): 71–6.
- 田中慶子, 2010, 「子どもの入学による家庭生活の変化——教育費負担・主観的 well-being・生活時間」『家計経済研究』 88: 41–8.

移動表（仮想例）再現用 Stata コード

```
set obs 12
gene id = _n
gene ymonth = _n
gene age1 = 0 if ymonth < 5
replace age1 = - 88 if ymonth > 4
gene age2 = 1 if ymonth < 5
replace age2 = 0 if ymonth > 4
gene age3 = 2 if ymonth < 4
replace age3 = 1 if ymonth > 3
gene age4 = 3 if ymonth < 4
replace age4 = 2 if ymonth > 3
gene age5 = 4 if ymonth < 4
replace age5 = 3 if ymonth > 3
gene age6 = 5 if ymonth < 4
replace age6 = 4 if ymonth > 3
gene age7 = 6 if ymonth < 4
replace age7 = 5 if ymonth > 3
gene age8 = 7 if ymonth < 4
replace age8 = 6 if ymonth > 3
gene age9 = 8 if ymonth < 4
replace age9 = 7 if ymonth > 3
gene age10 = 9 if ymonth < 7
replace age10 = 8 if ymonth > 6
gene age11 = 10 if ymonth < 6
replace age11 = 9 if ymonth > 5
gene age12 = 11 if ymonth < 4
replace age12 = 10 if ymonth > 3
gene age13 = 12 if ymonth < 4
replace age13 = 11 if ymonth > 3
reshape long age, i(id) j(wave)
xtset id wave
xttrans age, fre
```

東京大学社会科学研究所パネル調査プロジェクトについて

労働市場の構造変動、急激な少子高齢化、グローバル化の進展などにもない、日本社会における就業、結婚、家族、教育、意識、ライフスタイルのあり方は大きく変化を遂げようとしている。これからの日本社会がどのような方向に進むのかを考える上で、現在生じている変化がどのような原因によるものなのか、あるいはどこが変化してどこが変化していないのかを明確にすることはきわめて重要である。

本プロジェクトは、こうした問題をパネル調査の手法を用いることによって、実証的に解明することを研究課題とするものである。このため社会科学研究所では、若年パネル調査、壮年パネル調査、高卒パネル調査、中学生親子パネル調査の4つのパネル調査を実施している。

本プロジェクトの推進にあたり、以下の資金提供を受けた。記して感謝したい。

文部科学省・独立行政法人日本学術振興会科学研究費補助金

基盤研究 S：2006 年度～2009 年度、2010 年度～2014 年度 基盤研究 C：2013 年度～2016 年度 特別推進研究：2015 年度～2017 年度 若手研究 A：2015 年度～2018 年度
基盤研究 B：2016 年度～2020 年度 特別推進研究：2018 年度～2024 年度

厚生労働科学研究費補助金

政策科学推進研究：2004 年度～2006 年度

奨学寄付金

株式会社アウトソーシング（代表取締役社長・土井春彦、本社・静岡市）：2006 年度～2008 年度

東京大学社会科学研究所パネル調査プロジェクト ディスカッションペーパーシリーズについて

東京大学社会科学研究所パネル調査プロジェクトディスカッションペーパーシリーズは、東京大学社会科学研究所におけるパネル調査プロジェクト関連の研究成果を、速報性を重視し暫定的にまとめたものである。



東京大学社会科学研究所 パネル調査プロジェクト
<https://csrda.iss.u-tokyo.ac.jp/panel/>