

# 日本におけるパネル調査が抱える課題の包括的検討

田辺 俊介

(早稲田大学文学学術院 准教授)

## 1. はじめに

近年、社会科学の研究において、パネルデータは因果関係の推論などのためにも重要なデータと見なされるようになってきている。2000年代に入ると日本においても数多くのパネル調査が実施されるようになり、データの蓄積も増えてきている。パネルデータの持つ豊富な情報量を生かすため、パネルデータの分析に特化した複雑な分析手法の開発も進んでいる。

ただし、そのような複雑な分析を支えるデータの「精度」に関する議論、なかでも特にデータ収集や管理に関する議論は、日本においては個別調査に関するものが多く（例えば坂口 2011）、パネル調査一般について論じたものとしては保田（2012）による希少な論考以外では、管見の限りではあまりなされていない。さらに、データ収集に関するインフラやデータ公開までも含めて総合的に論じられることは少ない。それに対して近年、郵送調査などの自記式調査一般の技法については発展著しく、Dillman（2007）が主張する「Tailored Design Method」では、調査を対象者との「社会的交換」（Social Exchange）と見なした上で、回収率や回答の精度を高めるための総合的な指針が提唱されている。

そこで本稿では、パネルデータ収集や管理に関わる課題についてできるだけ幅広く論じることで、その完成までの道のりは長いとしても、パネル調査版のある種の Tailored Design Method 作成の第一歩となることを目指す。パネル調査は、横断

調査よりも同一対象者との長期にわたる社会的交換を続ける必要がある。そのような対象者との関係構築という側面に着目しながら、具体的には予算や運営体制のようなインフラ面からはじめ、実査やサンプル維持などデータ収集過程に関連する課題を論ずる。さらに対象者の情報管理、データクリーニング、データ形式や公開方法という広くデータ管理に関わる諸点について、対象者も含む社会とパネル調査の関係の問題として考察する。その中で同時に、それら諸課題への個別対策についても試論を行う。

## 2. パネル調査をめぐる課題

横断調査でも、調査を支える組織や予算のような調査インフラは不可欠であり、また実査の方法や回収率など考慮すべき諸事象は数多い。しかしパネル調査では、その「継続性」と関連し、横断調査とは異なる考慮すべき諸点がさらに多く存在する。

### (1) 調査インフラについて

#### (a) 予算の確保

横断調査であればどんな大きな調査でも、その調査年度にまたがる数年分の予算を獲得することで、調査の全工程を完結させることが可能である。しかし、同一対象への継続的な追跡調査を行うパネル調査を実施するためには、その調査が継続する期間にわたる長期的な予算の確保が必要となる。そして、日本における多くのパネル調査で問

題になっているのは、そのような継続的な予算の確保である。

諸外国の著名なパネル調査の多くは、政府が直接管理・運営する研究所などが主体となって実施しているため、安定的な資金拠出が確定している調査も多い。それに対して日本のパネル調査の多くは、大学などの研究者グループが獲得した時限的な予算によって運営されている。日本のパネル調査の嚆矢とされる家計経済研究所の「消費生活に関するパネル調査」も、公益財団法人格を持つとはいえ、国や地方政府が管轄するわけではない、一研究所によって運営されている。そのような事情により、多くのパネル調査が研究上の内在的な理由ではなく、予算という外在的事情によって継続年数に制限を受けることも少なくない。そのため、今後日本においてパネル調査をより発展させていくためには、現在進行中の調査も含めて、データ収集体制を維持するための安定的で継続的な予算確保が、解決がもっとも望まれる課題の一つである。

この予算に継続性と安定性がない場合、調査の継続期間という対象者との信頼関係構築において非常に重要な点を、対象者に伝えることができない。実際、著者も関わったパネル調査でも予算の関係上継続期間を明確にできず、回答者から「あと何年続くのですか」との問い合わせや、「終わりの見えない調査に疲れたので、今年限りにしてください」との声が寄せられていた。そのように継続期間の問題に直結する安定的かつ継続的予算の確保は、パネル調査にとって第一の課題である。

## (b) 運営体制・組織

前記のような予算の確保がなされたとしても、次に問題となるのはそれを制度的に受け続け、実査を担う研究体制をいかに構築するかである。この点も日本では、パネル調査を行っているグループは少なくないが、その実査に特化した調査部門などがある組織は少なく、その基盤が不安定であることは否めない。パネル調査は継続性が重要であることから、効率性と正確性を考えるならば、そのスタッフの継続性も必要だが、その確保や維

持が困難な状況なのである。一方、官公庁が直接行っているパネル調査、例えば厚生労働省が行っている「21世紀出生児縦断調査」などでも、調査開始当初は官庁内にパネル調査の管理や分析手法の知識・技術・経験の蓄積が十分ではなく、大学や研究機関に所属する専門家の協力である程度改善されたという（北村・金子 2013）。ただし、後に述べるデータ管理、特にクリーニング作業の体制などについては、官庁内での専門職員の不在が問題になっているとも予想される。

欧米の先行する代表的なパネル調査では、実査専門の研究機関などが存在しており、体制が整っている事例が多い。例えばアメリカを代表するパネル調査の1つ National Longitudinal Surveys（以下、「NLS」と略記）は、米国労働統計局が主導する調査であるが、その運営実施とデータのユーザー・サポートは、NLSの運営を主要目的の1つとして設立されたオハイオ州立大学内の機関 Center for Human Resource Research に委託されている（坂口・ホリオカ 2009）。あるいは、1984年に開始したドイツの代表的な家計パネル調査である German Socio-Economic Panel Study（以下「G-SOEP」と略記）については、DIW Berlin という国や地方政府から資金援助を受ける独立・非営利組織が実査を担っている（野村総合研究所 2012）。以上の事例に見られるように、一定規模の長期的パネル調査を維持・管理するためには、官公庁と大学等の研究機関が連携した、継続性のある体制作りが求められていると言える。

このような調査運営組織の継続性と安定性は、調査対象者との関係構築においても重要である。日本では多くのパネル調査を大学に所属する研究者が行っていることから、「大学名」を前面に出すことで、対象者からの一定の信頼感が担保されている。しかし調査実施主体の組織が不安定で、例えば名称変更などが行われれば、調査対象者の不信感を醸成し、結果的に継続調査への非協力につながる危険性は高い。

さらに実査を調査会社に委託している場合、委託調査会社の選択や関係構築にもいくつかの注意が必要である。まず実査を担う調査会社こそ、対

象者と直に接触する組織であり、対象者との関係構築において決定的な役割を果たすことになる。そのため、信頼できる調査会社を選定することはもちろん、調査継続期間中は調査会社の担当者との密なコミュニケーションが必要となることも多い。例えば、調査員に対する苦情などが対象者から寄せられた場合、調査会社とともにできる限りの個別対応をすることは、対象者との信頼感を維持・発展させるために必須である。

委託調査会社については現実問題として、日本において全国規模の調査をパネル調査として引き受け、実施可能な調査会社は数社しかない。そのため、選択肢の数はそもそも限られている。しかしそのことは逆に考えれば、その数社が多くのパネル調査を引き受ける中で、パネル調査の知識やノウハウを蓄積していくこととなる。そのことによって、調査会社に委託して行うパネル調査一般について、その精度などの底上げ効果も期待したい。

## (2) 実査に関する課題

### (a) Computer Assisted Interviewの導入

保田(2012)が指摘するように、Computer Assisted Interview(以下「CAI」と略記)と総称される、データ収集時にコンピューターを用いる手法の導入において、日本におけるパネルデータ収集方法の技術面の現状は、先進的な諸国に比べてはなはだ遅れた状態である。欧米の著名なパネル調査の多く(例えば前述のNLS等)では、CAIは1990年代より活用されている。それに対して日本ではごく一部の例外を除き<sup>1)</sup>、著名なパネル調査のほとんどが紙に印刷された調査票を基に調査が行われている。

この点はパネル調査に限った問題ではなく、日本の社会調査ではCAIの導入が進んでおらず、ほとんどの社会調査においてデータ収集方法は旧来からの紙と鉛筆(Paper and Pencil Interview、以下「PAPI」と略記)の時代が続いている。しかし、パネル調査ではCAIの導入がなされていないことが、後に述べるクリーニングの問題などとも関連し、パネル調査の発展に対する大きな阻害要因ともなっている。

パネル調査にCAIを導入することの利点は前述の保田(2012)に詳しいので、ここでは要点のみを手短にまとめておこう。回答不要な箇所を飛ばすことができるため、PAPIでは陥りがちな調査票の膨大化が防げる。前回の回答を事前に入力しておくことで対象者の負担低減や、回答エラーの回避につながる。データがその場で電子化されていることから、回答間の論理的矛盾もその場で指摘可能である。さらに用紙上に表現される範囲の情報よりも、ビジュアルとして理解しやすく、回答しやすい調査票が作成できることから、対象者が正確な回答をしやすい効果も期待できる。

以上のように利点の多いCAIについては、保田の(CAIの)「活用を1つの理想像(規準)として共有して議論を行うことが、日本のパネルデータの収集・管理を改善し、より望ましい調査環境を実現する議論を進めるためのステップになる」(保田2012: 96)という主張に、著者も全面的に賛同する。そのため、日本でも継続中の多くのパネル調査についても、CAIを利用した実査形式への移行が望まれるところである。

確かに継続中のパネル調査への導入は、調査モード変更の影響なども想定され、簡単な話ではない。しかし海外では先行事例は少なくない。例えば1968年に開始したNLSは、1990年代半ばにPAPIから、訪問調査員がコンピューターに回答を入力するComputer Assisted Personal Interview(以下「CAPI」と略記)に移行している(坂口・ホリオカ2009)。同じくアメリカの著名なパネル調査であるPanel Study of Income Dynamics(以下「PSID」と略記)でも1993年にComputer Assisted Telephone Interview(CATI)を導入している(村上・ホリオカ2008)。また前述したドイツの世帯の消費や健康に関するG-SOEPも、1998年からCAPIを導入している(野村総合研究所2012)。

このように海外の著名なパネル調査のほとんどでCAIが用いられている状況を考えれば、CAI導入の遅れはそのまま日本のパネル調査の後進性を示すとも言える。そのため、少なくともこれから始めるパネル調査であれば、CAI導入を検討す

ることは必須であろう。

CAIの導入にあたっての注意点もいくつか議論されているが<sup>2)</sup>、最も大きい阻害要因はCAIの導入に伴うさまざまなコスト問題であろう。例えば、調査員を使う調査では、調査員の人数分必要な端末の値段のような金銭的成本に加え、「それまでの紙の調査票に慣れた調査員の再教育」(野村総合研究所 2012: 125)が必要となる。

しかしこの点については、端末価格の下落傾向や、直感的な操作が可能なタブレット端末の普及などを考えれば、CAI導入の環境は急速に整いつつあると言えるだろう。さらに調査体制のところでも述べたように、パネル調査を行える調査会社が寡占状態であることは、逆にその数社でのCAIの活用が始まれば、多くのパネル調査でも利用可能な環境が整う可能性もある、とも考えられよう。

### (b) 調査員の身分・教育

訪問面接調査や留置調査の場合、調査員は対象者との社会的交換の最前線に立つ存在であり、調査員への信頼感などはそのまま調査の精度と直結する重要な要素である。

その調査員の身分について、官庁統計の調査員であれば非常勤とはいえ公務員扱いであり、調査対象者から一定の信頼を得やすいと思われる。それに対してパネル調査の委託先の調査会社の調査員は、基本的に一私企業の非正規職員である。そのため、対象者が安心できるような調査員に対する一定の身分保証なども、社会調査一般に関わるインフラ整備として必要だと思われる。

また調査員の教育も、対象者との良好な関係構築のためには不可欠である。もちろん日本のパネル調査でもある程度のインストラクションは行われているだろう。しかし、例えば、「欧州における健康、加齢および退職に関する調査」(SHARE)では、対象者情報の管理やコンタクト方法まで、2日間で合計750分にもわたる総合的な調査員への教育を行っている(野村総合研究所 2012: 121)ことを考えれば、調査員、ひいては調査自体の質の向上のためにも、その教育をより徹底する必要があると思われる。

### (3) サンプル維持に関する課題

#### (a) 脱落防止

パネル調査の特有の問題点として、対象者の脱落(attrition)問題は、横断調査にはない、パネル調査特有の大きな問題である。この点は、統計的な補正方法なども議論されているが、ここでは調査方法的側面の改善案を論じる。

脱落防止の施策として、ほとんどの調査が何らかのインセンティブを提供している。多くの調査では金券等の金銭的な誘因を用いているが、例えば同一コーホートを追跡する調査(「出生児縦断調査」等)では、そのライフステージに応じた物品(乳児も使用可能なオーガニックコットン等)を提供している。

またいくつかの調査では、金銭的インセンティブだけでなく、「調査への共感を促すようなノベルティグッズ(調査名称入りのボールペン、ノート、キーホルダーなど)の開発が行われている」(野村総合研究所 2012: 121)。それらは、ある種の帰属意識構築を通じた調査へのコミットメントの強化である。また、ある海外のパネル調査では、その対象者たちのみが参加できるネット上のコミュニティを作り、そこでのやりとりを通して対象者の心理的コミットメントを高める、という試みを行っているという。そのようにインセンティブを単純な経済的交換だけではなく、調査対象者との社会的交換の過程で必要な存在との視点から、品物や提供のタイミング<sup>3)</sup>が判断されるべきである。

また直接のインセンティブではないが、調査結果の情報発信強化も広い意味での調査対象者の回答継続の誘因になりうる。人は、「意義ある」と思う活動への参加から心理的報酬を得ることができる。そのため、自らが参加を続けるパネル調査の結果がマスメディアに採り上げられたり、政策選考の資料として用いられられたりしたなどの情報を知れば、そのことも調査継続の1つの誘因になるだろう。そのように考えると、調査成果の情報発信も、遠回りであっても対象者との社会的交換という意味で重要な要素である。そしてその要素を良化するためにも、後節で述べるデータ管理の側面も重要になってくる。

**(b) 補充サンプル**

脱落は完全に防止できるものではないため、調査継続期間が長くなっていくほど、サンプルの代表性の問題が大きくなる。そこで脱落分を補うサンプル追加を行うこともある。その追加の方法も、脱落が多い属性（例えば若年層）だけを補充するのか、もともとの母集団を反映した標本を再度抽出するのか、状況や調査関心に合わせた方法を採用する必要がある。また追加されたサンプルは当然のことながら、追加された時点からの情報しかないため、使用可能な分析に限られる点にも注意が必要である。

あるいは調査継続期間が長い場合は、当初の調査対象母集団には含まれなかった新規コーホートの追加が必要になることもある。この点についても、調査の関心ごとに具体的な追加方法は異なるであろうが、調査関心に合わせた適切な補充が必要なのは言うまでもない。

**3. パネルデータの管理をめぐる課題**

パネル調査のデータは、一時点のクロスセクショナルなデータと異なり、調査自身が継続し、さらに継続 Wave の調査が行われるたびに、そのデータ自体も増えていく。そのため、パネル調査におけるデータ管理には独自の注意点が存在する。

**(1) パネル対象の管理****(a) 対象者の個人情報保護**

対象者の住所と氏名は個人情報保護法の対象となる「個人情報」であり、どんな社会調査であっても対象者名簿の管理に細心の注意を払う必要があることは言うまでもない。それでもクロスセクショナルな調査であれば、調査終了後即廃棄するなど、流出の可能性を最小限に減らすことができる。

しかしパネル調査の場合、継続調査のためには、それら情報を保持し続けなければならない。そのため、そのような対象者氏名や住所等コンタクトに必要な個人情報の管理と保護を徹底するためにも、前述の調査インフラ面の安定性は非常に重要である。個人情報を管理する組織の安定性に加え、

担当者の不要な変更などはできるだけ避け、セキュアで一貫した管理を行うことが必要である。

**(b) コンタクトの維持**

対象者のコンタクト情報は、常に最新の情報へと更新する必要がある。例えば1年おきの調査において、1年に一度だけ調査時点のみの接触では、少なからぬ対象者の住所変更を把握できず、その結果としての標本脱落の危険性が高まる。そのため、多くのパネル調査では、毎年の調査速報や年賀状の送付など、調査と調査の間の時期に対象者にコンタクトをとり、住所変更等を把握する施策を行っている。ただし、それら方策は予算との兼ね合いもあり、やみくもに回数を増やすことができない場合も多い。そのため調査ごとに、対象者の特質なども考慮しつつ、適切な方法を用いる必要がある。

例えば、調査設問の一つとして「引越し予定」を尋ね、引越し予定がある対象者に限って、引越し先の住所その他連絡先を教えてもらうための特別なコンタクトをとるなどの方法もあろう。その際、英国の代表的なパネル調査 The British Household Panel Survey (BHPS) が行っているように、引越し先住所の連絡にギフト券を送るなども良い方策と思われる。

また対象者の住所等の名簿面でのコンタクト確保だけでなく、対象者との人間関係的な側面の維持も重要である。その点、調査員を用いる場合は、同じ調査員を毎年派遣することで、調査へのコンタクトを、調査員に対する信頼感や安心感を基に維持することも、地味ながら有効な方策と言えるであろう。

**(2) データクリーニングの問題****(a) Wave間の情報の整合性**

一般的な社会調査においても、調査後のデータのクリーニングについては多大な労力を要する。しかしパネルデータに関しては、同一対象を継続的に追跡することから生じる独自の問題として、Wave間の情報の整合性が問題となる。例えば著者も関わった調査でもデータ納品段階では、毎年

数名の「性転換者」と「生年月日変更者」が存在した。それらは「配偶者の代理回答」（回答の性別が変わる）や「家族内の別人の代理回答」（生年月日が変わる）が原因と推察されるが、このような同一対象の回答であるかの同定問題が、パネル調査に独特なクリーニング必須事項である。さらにこの点は、どんなクリーニング作業にも優先して行い、必要ならば再調査を行う必要のある問題である。

同一対象の回答であっても Wave 間をまたがる情報の矛盾は、パネル調査では度々発生している。例えば、記憶違いによって父母の年齢が「若返り」ことや、既婚者が（離別や死別ではなく）「未婚者」になることなどである。このようにパネル調査では、追加 Wave が実施されるたびに、過去のデータも含めたクリーニングが必要になることも多い。そのため、複数年度をまたがるクリーニング作業を矛盾なく実施するためには、リスト化された詳細な作業記録とログの作成は必須である。

### (b) クリーニングの方法

前述のようなさまざまな矛盾のクリーニング作業は、担当設問を決めて分担して行う方法が採用されることが多い。その利点としては、担当者の専門性が生かせることや、作業工程を分割するため作業の一部に遅れがあっても他の部分のデータは利用できることなどである。しかし欠点も多く、分担箇所をまたぐ矛盾の処理について処理同士の矛盾が生じたり、処理の順番によって処理をした以外の箇所に新たな矛盾を発生させたりする。それを防ぐためには、作業手順や作業状況に関する綿密な全体的統括が不可避となり、その管理コストも少なくない。

そのため保田（2012）は、Fellegi-Hout の原則（Fellegi and Hout 1976）をクリーニングの一般哲学として再考すべきと主張する。特に上記の設問分担方式に対して、原則3の「データの修正はそれぞれのルール毎に考慮するのではなく、全ルールの適用結果を見てケース単位で修正を考慮しなければならない」（保田 2012: 92）に則り、ケース単位の処理を推薦する。「全ての論理エラー」

を検出した上で、ケースを単位として担当し、そのケースに生じている論理エラー全体を一括して処理していく、という方法である。確かにこの方法であれば、前期設問分担の方法に比して、処理間の矛盾や処理の順番の問題も生じにくい。ただし、全ての論理エラーを検出するプログラム作成のコストや、ケース単位で分担可能なだけの分担者の確保など、理想的方法ではあるが、その課題も少なくない。

この点については、同じく保田（2012: 92-93）で述べられているように、論理エラーを検出するプログラムを組み込んだCAI導入によって、エラーの発生を未然に防ぎ、かつケースごとに調査現場で対象者とともに修正することが、事後的な負担軽減や正確な修正という意味でも、最善の方策である。

## (3) データファイルの管理形式

### (a) 変数名

変数名は、内容を示す名前とするか、調査票の間番号とするかの2通りの方法が考えられる。パネル調査では意味ある変数名とした場合、それに加えて Wave 番号などを付与しておけば、多くの統計ソフトウェアでそれらを「パネル化された変数」として認識させることが容易であり、パネルデータとして活用しやすくなる。例えば、年齢に関する変数は「age」と命名し、実際の変数名としては第1波であれば「age01」、第2波であれば「age02」とするなど、「有意義な変数名 + Wave 番号」という管理方式である。

しかし実際に分析に用いる場合には、調査票との対応関係がないため、整理されたデータインデックスがない場合、使用時に不便が生じる。あるいは内容を示す名称といっても多くの統計ソフトウェアでは変数名の文字数や文字種が限られるため、短縮形の英語名をつける必要があるが、その時点で命名自体に少なからぬ労力が必要となる。

一方、調査票の間番号を用いる場合は、初期のデータ処理や管理が容易である。例えば婚姻状態に関する変数が第1波の間25ならば「01q25」、第2波では間32なら「02q32」などとする命名法で

ある。そのような変数名であれば、命名に悩むことは少なく、また調査票との対応関係も単純である。しかしそれらを「パネルデータ」として用いる場合には、毎年の調査票の構成が全く同じでない限り、対応する問番号が違い、変数名が異なることから、まずは変数名をパネル化された変数として変更した上で処理する必要が出てくる。

### (b) 変数リスト

パネル調査は、詳細な情報を得る分、1つのWaveでも変数の数が多くなりやすい傾向がある。そのもととも多い変数が、Waveを重ねることで累積的に増えていく。仮に全く同じ設問のみを尋ね続けるならば、Waveが異なる同内容の変数の増加であり、特に問題はない。しかしほとんどのパネル調査では、年度ごとに調査票の練り直しが行われ、新規項目や逆に削除項目、あるいは一部変更される項目など、Waveごとに質問項目が異なっている。そのため、Waveごとに含まれる変数と含まれない変数、あるいはWaveをまたぐ変数の比較可能性を確認する必要が出てくる。

それらを確認・管理するためには、適切な変数リストの作成が必要となる。そのリストについてはExcelなどの表形式でも対応は可能であるが、可能であれば変数リストと調査票がリンクするような、ハイパーリンク形式（Webページの構成のようなもの）で作成すれば、拡張性が良いだろう。加えて、パネルデータの基礎的な理解の助けとなる項目ごとの基礎集計、特にWaveごとの集計を軌跡として表示するような機能等は、後に述べるユーザビリティと関連して望まれるものである。

### (c) データ形式の問題

データをどのようなデータ形式で管理していくかも、パネル調査では考慮が必要な問題である。この点は後に述べる提供データ形式とも関連する。

まずWaveごとか、複数Waveの累積か、という問題がある。単年度ごとであれば、ファイルサイズも比較的小さく、また1Waveごとの修正などを反映したデータ更新などにも対応しやすい。しかし利用に際しては、データ合併の手間がかかる。

一方、累積データの利点としては、Wave間をまたぐパネル分析にそのまま利用できることや、先述のWave間をまたぐ情報の整合性のクリーニングなどにも利用しやすい。しかし、複数Waveに同時に修正がされた場合のデータバージョンの管理が難しく、また年数を重ねた調査ではデータが膨大で利用したい変数を見つけにくくなる。そのように一長一短であることから、用途に合わせた管理が必要となる。

次にデータ形式として、ワイド形式か、ロング形式か、という点も一考する必要がある。1つのケースの回答が横へ横へと入力されていき横に幅が伸びていく形式はワイド形式と呼ばれる。それに対してパネルデータの分析手法の多くは、別時点の回答を別のケースとして縦に並べていくロング形式を用いる。この点については、多くのパネルデータの分析手法がロング形式のデータを前提としているため、パネルデータ特有の成果を得るためには、ロング形式のデータを利用するのは確かに近道である。

しかし、ロング形式のデータにもいくつかの問題点が存在する。まず、パネルデータの分析手法によっては（例えば成長曲線モデルのような構造方程式モデリングを用いる分析など）ワイド形式のデータの方が望ましい。またロング形式はデータが肥大化しやすい傾向がある。例えば5,000人を対象とした調査の20年分の調査データであれば、（脱落がない理想的な状態を仮定した場合であるが）ロング形式では $5,000 \times 20 = 10$ 万行分のデータが縦に並ぶこととなる。つまり、ロングデータはそれだけデータサイズが肥大化しやすい傾向がある。過剰なデータ量は分析者の認識にも負担をかけることにもなる。また先述のクリーニングの段階では、Waveごとの回答を別ケースとして扱うロング形式よりも、同ケースとして扱うワイド形式の方が作業しやすい。

そのため、累積させたワイド形式のデータからパネルデータとして分析に用いる変数だけを取り出し、その上でロング形式に変換するという方法が、利便性や効率性の面で最適ではないか、と思われる。

#### (4) データリリースや公開に関する問題

##### (a) 秘匿性の問題

ここまで述べてきたように、パネル調査はそのデータを獲得し、科学的分析に堪える水準まで整えることに多大な資金と労力を必要とする。だからこそ、そのデータが広く研究者コミュニティに開放され、多くの研究者が利用することでそのデータから得られる知見を生産的に増産する必要がある。そのため、ほとんどのパネル調査は、一定期間の後に自らの組織を通じて、あるいはデータアーカイブ等の専門機関を通してデータを公開している。

その公開に際して考慮すべき問題となるのは、データの秘匿性である。秘匿性とは英語の confidentiality の訳語であり、統計的調査の調査対象者の個人が特定されないという「秘密保持契約」といっても良いだろう。調査対象者は、個人を特定されないことを大前提として調査に協力している。そのため、万が一にもパネル調査データに秘匿性の問題があって個人が特定された場合、それは広くパネル調査全体の対象者と調査者の間の信頼関係に甚大な悪影響を与える可能性がある。

その秘密保持については横断調査データでも考慮すべき点は少なくないが、パネル調査の場合、継続した期間の回答が蓄積していく分だけ、個人特定の危険性が高まっていく。1時点の情報ではなく、変化を伴う情報は、その分個人を特定する危険性が高まる。特に Wave 間をまたぐ情報集積によって起こりうる個人特定の危険性に配慮しつつ、秘匿化処理を行う必要がある点を強調しておこう。

##### (b) 公開方法の問題

貴重なデータであるからこそ、パネル調査はデータを公開し、多くの研究者が二次分析できるようにデータを提供することが、半ば義務となっている。ただし、その公開方法についても、少なからぬオプションが存在する。さらにパネル調査は追加 Wave がくるたびに、データの修正が行われることがある。その場合、その修正データの配布方法も問題となる。

まずデータ提供の仕組みとして、データアーカイブ（以下「DA」と略記）に寄託することで配布するか、もしくは独自にデータ提供するか、という選択が必要となる。DAに寄託した場合、種々の事務的作業はDA側に依存できる。しかし独自公開に比べると、データの変更などがあった場合の対応は遅くなる。

また独自配布を行う場合は、データの提供方式についてもCDなどの媒体を郵送で送付するのか、それともインターネットを通じた配布を行うのか、など選択肢は少なくない。さらに提供するデータ形式についても、最も汎用性が高いのはCSVその他テキスト形式であるが、その場合は変数名その他のメタ情報は別添となる。あるいは特定の統計ソフトウェアに対応した形式での提供となれば、そのソフトウェア利用者以外には使いにくいデータともなる。よってより広い利用を促すためには、複数の形式でのデータ提供を準備することが望ましい。

##### (c) データのユーザビリティの問題

年々巨大化していくデータをどのような形で提供するのか。データをたとえ提供しても、それが実際に容易に使える形でなければ、分析成果にも結びつきにくいだろう。この点は、脱落防止の節でも述べたように、ただ利用者に利するというのではなく、データの適切な利用による良質な成果は調査の高評価へとつながり、それがひいては対象者が調査に参加することの「動機付け」にもなっていく。そのことでさらに良質なデータが生まれることで、質の高い成果が生まれる。そのような好循環のためにも、ユーザビリティは決して疎かにできない問題なのである。

この点については、これまで述べてきた変数リストやデータの形式、あるいは公開方法など総合的に勘案しつつ、利用者にとって利用しやすいデータの提供を目指すべきである。

## 4. おわりに

本稿では、パネルデータの収集や管理の課題に

ついで Tailored Design を目指すため、一定程度網羅的に諸問題を列挙した。それら全体に通底する問題は、継続的な調査を支えるためのインフラの不在ではないか、と思われる。そのため、個々のパネル調査は個別諸問題に適切な対処を行いつつ、同時に望ましい調査インフラの構築を進めていくことが必要不可欠である。

本稿が、今後パネル調査を始めようとする人たちの参考となるだけでなく、現在パネル調査に携わっている人々にとっても、自分たちの営みの課題を検討する一助となれば幸いである。さらにはパネルデータのユーザーにとっても、使用するデータの背景にある諸問題を理解することが、よりよいデータ利用につながることを希望する。

#### 注

- 1) CAIの日本での実用例としては、早稲田大学21世紀COEの一環として行われた「日本人の社会意識に関する世論調査」や、独立行政法人経済産業研究所と一橋大学、東京大学が共同で行っている「くらしと健康の調査」などがある。ただし、前者については政治学的関心に基づく調査であるため、選挙前と選挙後のような短い2時点のパネル調査であり、後者も今のところ2007年・2009年・2011年の3Waveで、比較的实施回数は少ないままである。
- 2) CAIの導入に関し、パネル調査独特の注意点も存在する。CAIでは多くの場合、評価項目の順序による順序効果バイアスを除去するため、選択肢や質問項目を自動的にランダムに入れ替える「ランダムマイゼーション」が用いられる。しかし仮にパネル調査で調査ごとにランダムマイゼーションを行い、同じ対象者でもWaveごとに選択肢や質問項目の配置が異なってしまうと、評価基準点が最初の項目に依存する「アンカリング効果」が変わり、パネルデータとして利用できなくなる危険性が指摘されている(西澤2012)。
- 3) パネル調査ではないが、日本における代表的な社会調査の1つ日本版総合的社会調査(JGSS)では、半額を前渡し、半額分を後渡しという方法を用い、回収率が向上したという(保田ほか2008)。これは前渡しによって「受け取ったものへは何かを返さないといけない」と感じる人々の互酬性の規範を利用した方策である。確かに回収率向上の効果は高いが、パネル調査で活用する場合、継続的に用いることが対象者との信頼関係に与える影響を慎重に見極める必要があるであろう。

#### 文献

- 北村行伸・金子隆一, 2013, 「縦断調査の厚生労働政策への応用に向けて」『厚生指標』60(2): 38-49.
- 坂口尚文・チャールズ ユウジ ホリオカ, 2009, 「アメリカのパネル調査(2)——National Longitudinal Surveys (NLS)」『季刊家計経済研究』81: 52-64.
- 坂口尚文, 2011, 「消費生活に関するパネル調査」の現状と課題」『中央調査報』647: 5707-5711. (<http://www.crs.or.jp/backno/No647/6471.htm>)
- 西澤由隆, 2012, 「Computer Assisted Personal Interviewにおけるランダムマイゼーションの是非」『政策科学』19(3): 461-481.
- 野村総合研究所, 2012, 『日本におけるパネルデータの整備に関する調査報告書』.
- 村上あかね・チャールズ ユウジ ホリオカ, 2008, 「アメリカのパネル調査——Panel Study of Income Dynamics」『季刊家計経済研究』79: 55-61.
- 保田時男・宍戸邦章・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(2): 129-136.
- 保田時男, 2012, 「パネルデータの収集と管理をめぐる方法論的な課題」『理論と方法』27(1): 85-98.
- Dillman, Don A., 2007, *Mail and Internet Surveys: The Tailored Design Method - 2007 Update with New Internet, Visual, and Mixed-Mode Guide*, Hoboken: Wiley.
- Fellegi, I. P. and D. Hout, 1976, "A Systematic Approach to Automatic Edit and Imputation," *Journal of the American Statistical Association*, 71: 17-35.

たなべ・しゅんすけ 早稲田大学文学学術院 准教授。主な著書に『外国人へのまなざしと政治意識——社会調査で読み解く日本のナショナリズム』(編著, 勁草書房, 2011)。経験社会学、社会意識論専攻。