

## NFRJ-08Panel における調査票の設計 ——研究課題とクリーニングを視野に——

保 田 時 男

### I. 本レポートの狙い

NFRJ-08Panel (全国家族調査パネルスタディ; National Family Research of Japan, 2008-2012 Panel Study)<sup>(1)</sup> は、学会員の有志によって進められているパネル調査プロジェクトである。この調査は NFRJ08 (第3回全国家族調査) の回答者のうち継続調査への協力を応諾してくれた者を対象としており (最終的に 1879 名)、2009 年 1 月の NFRJ08 を第 1 波として、1 年ごとに 2013 年 1 月の第 5 波までの実施を予定している。

本レポートでは、NFRJ-08Panel の (とくに第 2 波の) 調査票をどのような意図と手続きで設計したかを順に説明する。調査票の設計においては、クリーニングの方法も一つの大きなポイントとなったので、第 5 節ではクリーニングの手続きについても簡単に説明する<sup>(2)</sup>。このような裏方の手続きは公になりにくいですが、实际的に調査研究を行う際には重要なプロセスである。学会で共同利用されるデータの収集過程を周知するとともに、類似の調査プロジェクトの参考資料となることを企図している。

### II. 設計に際しての主な問題点

NFRJ-08Panel で第 2 波の調査票を設計する際には、大きく 3 つの外的な問題が存在した。第 1 の問題は、第 1 波の調査票がパネル調査を意図

やすだ ときお: 大阪商業大学総合経営学部

してデザインされていないということである。第 1 波となる NFRJ08 は、繰り返し横断調査 (repeated cross-sectional survey) であり、後にパネル調査が続くようにはデザインされていない。しかも、その調査デザインはかなり複雑で、複数のダイアド関係 (配偶者、父母、義父母、3 人までの子ども、3 人までのきょうだい) を網羅的、並列的に尋ねるというものである。この第 1 波を前提にしながらパネル調査を設計することには、独特の困難が感じられた。

第 2 の問題は、第 2~4 波が郵送調査という点である。当初、第 1 波と第 5 波のみが本格的な調査として計画されており (これらは留置調査)、第 2~4 波は住所確認を兼ねたごく簡易な調査が予定されていた。しかしながら、パネルデータは 3 回以上の回答が積み重なることでその価値が格段に高まることなどを考慮して、第 2~4 波も 10 ページ程度の本格的な郵送調査を実施することになった。郵送調査では回答者と直接的な接触を取ることができないため、より慎重に調査票をデザインしなければ回答の不備が続出する。実行委員会では、類似のパネル調査をレビューする研究会を開くなどの検討を行ったが、郵送によるパネル調査は参考にできる先例がなく<sup>(3)</sup>、また、本格的なプリテストを行う時間的、予算的な余裕もなかったため、手探りで郵送調査を行わざるをえなかった。

第 3 の問題は、データ・クリーニングをいかに行うかという問題である。一般に、複雑な内容の

調査であるほどクリーニングの作業量は増大する。パネル調査はいうまでもなく非常に複雑な調査である。過去の回答との整合を確かめなければならず、またデータが一度に揃うのではなく1年おきに段階的に収集される。パネルデータのクリーニングにどれほどのコストがかかるかという正確な情報は確認できていないが、政府が行う一般的な調査において、クリーニングのコストは調査コスト全体の20%以上を占めるといわれている (Groves et al. 2009)。パネルデータのクリーニングにはさらに大きなコストがかかることが予想される。にもかかわらず、NFRJ-08Panel ではクリーニングのための予算を確保することはできなかった。これは、申請した研究費が減額されたことと、当初明確には予定していなかった第2～4波の実施を優先させたことからやむをえないものの、大きな現実的な問題と考えられた。つまり、調査票のデザインの段階からクリーニングのコストを減らすことを強く意識する必要があった。

### III. 研究課題の絞り込み

以上のような困難があったものの、第1波の調査がすでに行われていることには、よい面もあった。第1波 (NFRJ08) のデータから、パネル調査への協力応諾者についてその詳細な特徴を検討することができたからである。そこで、第2波の調査票デザインにあたっては、あらかじめ研究会のメンバーから寄せられた研究課題の原案を考慮することに加えて、第1波のデータを参考にすることで効果的な研究課題を絞り込むことから出発することにした。NFRJ 自体は研究課題を絞り込まない汎用的な家族調査であるが、NFRJ-08Panel の第2～4波は郵送調査なので、分量的に全てをカバーすることはできない。効果的なポイントに研究課題を絞り込む必要があった。

パネル調査への協力応諾者の属性分布を確認すると、50～60代で男女ともにボリュームが大きいことがわかった。すでに親が死亡している者が多く、大多数が有配偶者で、成人している子ども

がいる割合が比較的高い。これらのことから、NFRJ-08Panel では、回答者の親ダイアドを追及するよりは、子ダイアドに焦点を当てるのが適切と判断できた。しかも、それは子育て期の子どもとの関わりというよりは、子離れや世代間援助といった成人子を視野に入れた関わりである。具体的には、子世代に起こるイベント (離家、就職、結婚、離婚、出産等) が親世代つまり回答者の生活や意識 (働き方、家事、夫婦関係、メンタルヘルス等) とどう関係するのかを明らかにすることを主な研究課題として調査票をデザインすることになった。

このデザインは、子育て期や高齢期を対象にした一般的なパネル家族調査を想像した場合には期待外れのものと考えられるかもしれない。しかし、むしろNFRJ-08Panel ならではのデータを収集できるため、両者の研究を繋ぐ貴重な資料となると考えられる。また、第5波では第1波と同様の詳しい留置調査を予定しているので、ある程度広い研究課題にも応えることができるはずである。なお、子ダイアドの他にNFRJの特徴の一つであるきょうだいダイアドに焦点をあてることも検討されたが、短期間での変化に乏しいことが予想されたので、取り上げなかった。また、新たに孫ダイアドを調査することも検討したが、自立的な年齢に達している孫が少数であり、調査デザインも複雑になるために断念した。

### IV. 第2波調査票の作成

具体的に第2波の調査票を作成するにあたって最初に考えたことは、調査票の種類を複数用意する必要があるかという点である。NFRJ08 ではライフステージに応じた調査内容を盛り込むために若年票・壮年票・高齢票の3種類が用意されたが (保田 2007)、NFRJ-08Panel では内容が多岐にわたるわけではないので、そのような必要がないと判断した。しかし別の問題として、基本的な有配偶者を想定したデザインとなるため、多くの質問がスキップされる未婚者や離死別者にとっ

ては自分と関係の薄い調査と感じられるのではないか、ということが懸念された。また、枝分かれの結果として見た目の調査分量が多く感じられることも懸念された。そこで、「有配偶票」と「無配偶票」を用意し、枝分かれによる質問のスキップを減らすことにした。第1波のデータから有配偶・無配偶の区別をすることが可能なため、このような分割ができる。無配偶者が第2波までの間に結婚した場合は、第2波回収後に即座に「新婚票」を送付し、結婚や配偶者についての情報を追加で収集することにした。データの取り回しはやや煩雑になるが、郵送調査の回収率を上げるためには回答者と関連の深い調査票を印象づけることが重要である。その意味もあり、時事的に人々の関心を引きやすい質問を導入質問として毎回挿入することになった。第2波の調査票では、婚活と夫婦別姓に関する意識を尋ねている。

第2波調査票の具体的な調査項目は表1のとおりである。大部分は第1波 (NFRJ08) の調査項目から選択されたものであるが、取捨選択の際に特に考慮したのは以下の4点である。第1に、焦点をあてる子ダイアドについては、第1波の調査項目の大部分をそのまま継続調査することにし

た。ただし、仕事の有無や援助行動の有無など、幼い子どもについて尋ねることが奇異な項目が多いため、全項目を18歳以上の子どもについて尋ねるものと全ての子どもについて尋ねるものに分け、再構成した。また、子どもの配偶状態を明確に捉える質問を追加した。全10ページの調査票のうち2ページを費やして、3人の子どもそれぞれの変化をかなり克明にとらえている。

第2に、基本的な生活評価項目は毎回必ず尋ねることにした。どのような変化を扱うにしても、回答者がその変化を最終的にどの程度高く (あるいは低く) 評価しているかは、重要な分析側面となる。そのため、生活満足度と配偶関係満足度をすべて第2波の調査票に含めた。スペースを取るうつ尺度をそのまま継続調査しているのも同じ理由である。

第3に、影響の大きい統制変数と考えられる項目を毎回必ず尋ねることにした。具体的には、就業状態 (本人・配偶者・子)、収入 (本人・配偶者・世帯)、健康状態 (本人・配偶者・親) などである。子ダイアドに焦点をあてるため、親ダイアドについてはほとんど何も尋ねていないが、その生死・健康状態だけはとらえることにした。

表1 NFRJ-08Panel 第2波の調査項目リスト (有配偶者票)

問1. 性別	問9 付問4. 配偶者の変化した現在の仕事	問20. 子の年齢
問2. 生年月・年齢	問10. 配偶者の健康状態	問20. 子が18歳未満か以上か
問3. 婚活についての意識	問11. 配偶者との会話	問20. 子の仕事の有無 (以下18歳以上の子のみ)
問4. 夫婦別姓についての意識	問12. 配偶者からの情緒的援助	問20. 子の婚姻状態
問5. 去年の出来事	問13. 本人・配偶者の家事	問20. 子の子 (孫) の有無
問6. 今年の出来事	問14. 配偶者満足度	問20. 子との会話
問7. 就業状況	問15. うつ尺度	問20. 子からの金銭/情緒/実践援助
問7 付問1. 労働日数/月	問16. 不安感	問20. 子への金銭/情緒/実践援助
問7 付問2. 労働時間/日	問17. 本人健康状態	問21. 父/母/義父/義母の健康状態
問7 付問3. ワークライフバランス	問18. 生活全体満足度	問22. 同居人数
問7 付問4. 去年と比べた仕事の変化	問19. 健在の子ども数	問23. 同居者
問7 付問5. 変化した現在の仕事	問20. 子の性別 (以下1～3人目までそれぞれ)	問24. 本人/配偶者/世帯年収
問8. 婚姻状態	問20. 子の居住場所	問25. 回答の日付
問9. 配偶者の就業状況	問20. 子と一緒に遊ぶ	
問9 付問1. 配偶者の労働日数/月	問20. 子に知識や技能を教える	
問9 付問2. 配偶者の労働時間/日	問20. 子と一緒に夕食	
問9 付問3. 配偶者の去年と比べた仕事の変化		

第4に、できる限り同じ情報を2度尋ねないことに細心の注意を払った。精確なデータを手に入れるという視点からいえば、同様の事柄についてやや異なる角度から2度、3度と質問を重ねることは望ましいことである。しかし、郵送調査ではできる限り分量を削減しなければならない。また、同じ情報を繰り返し尋ねることはデータ内部に明確な矛盾を生じさせる原因となるので、クリーニングのコストを爆発的に増大させる。これらのことから、NFRJ-08Panelでは結果的に多少データの精度を落とすことになったとしても、同じ情報を2度尋ねないことを優先させている。なお、性別と年齢は第1波の情報で十分なので繰り返し尋ねる必要のない項目であるが、最低限の整合を確かめるために、本人と子どもの性別・年齢については毎回尋ねることにした。

ところで、以上のように第1波から取捨選択した調査項目を第2波で尋ねることは、1年間の変化を2時点の「状態(status)」の違いからとらえようとするものである。一方、変化をとらえるには別のやり方として「出来事(event)」を直接尋ねる方法がある(Bakeman and Gottman 1997)。変化を状態の違いから把握する方法で十分かという点について、実行委員会で議論を重ねた結果、やはり重要な点については出来事の有無を直接尋ねておくべきという結論に至った。たとえば、1年の間に離婚と再婚が続けて発生した場合、状態の違いからは変化がとらえられないからである。第2波では、12個の出来事(転居、持ち家の入手、本人・配偶者の定年、子・孫の誕生、配偶者との離別・死別、本人・配偶者の親の死亡、子の死亡、その他の親類の死亡)について複数回答でその有無を尋ねている。また、本人と配偶者の仕事の変化の有無についてはとくに調査項目を立てている。この点については、同じ変化を状態と出来事の2側面からとらえているので「同じ情報を2度尋ねない」という方針から外れるが、パネル調査において変化を精確にとらえることの重要性を優先させた。

## V. クリーニング法の模索

クリーニング作業のコストを抑えることを第2波調査票の作成の一つのポイントとしたので、最後にNFRJ-08Panelのクリーニング方法について触れておく。クリーニングの問題はNFRJ-08Panelに限らない一般的な問題であり、とくに1990年代から応用統計学者によってクリーニング法の開発が急速に進んでいる(Herzog et al. 2007)。彼らが開発しているのは、主にセンサスのような超巨大データを自動的に高速にクリーニングするアルゴリズムである。社会学者が扱うような比較的小規模(数千ケースまで)で、かつ調査内容が複雑なデータに対して完全に同じ方法を適用することは難しいが、Fellegi and Hout (1976)の原則などベースとなっているクリーニングの哲学は有効であり、大きく三つの示唆を得ることができる。(1)クリーニングは、異常を検出するeditingの段階と、異常値を別の値に修正するimputationの段階に分かれ、これらは完全に独立した作業である。(2)editingを行うためには、異常な値の範囲や組み合わせ(editルール)を明確に論理式で表さなければならない。(3)データの修正はそれぞれのルールごと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。

NFRJ-08Panelのクリーニングにおいては、他の調査でのクリーニング方法や第2波の回収原票の検討会を行ったうえで、Fellegi and Houtの原則にアレンジを加えた方法(保田 2010)を用いることにした。大まかな手続きは以下のとおりである。まず、editルールをより緩やかに考えて、完全な異常だけでなく異常の「可能性が高い」ケースを検出するものに拡張する。その上で可能な限りのeditルールを並べ上げ、それぞれに違反するケースにフラグを立てるようにプログラミングする。具体的なeditルールは、範囲外の値の検出や、枝分かれの間違いの検出から始めて、第1波と第2波の矛盾の検出(第1波で死亡して

いた父が第2波で生存しているなど)や異常の可能性が高いパターンの検出(回答者と子どもの年齢が近すぎるなど)まで多岐にわたる。この作業は実行委員のクリーニング責任者が行い、1,763個のeditルールを用意した。

次に、それぞれのケースについて、どのeditルールに違反しているかを参照しながら修正の必要性および修正方法を考える。修正方法をあらかじめ厳密に統制することはしない。大まかな方針だけを定めておき、個別のケースがもつ情報を生かした最適な修正を考える。この段階の作業は、20数名の研究会全体で分担された。作業はケースごとなので、1~50ケース目までというようにケース単位での分担となる。分担者は、図1のようなツールで、検出された異常の箇条書きとデータを照らし合わせ、そのケースに必要なと思われる修正プログラム(SPSSのシンタックス)を作成した。

この修正作業は、短期集中的に1カ所に集まっ

て情報を共有しながら行うことが望ましいが、NFRJ-08Panelではそれはかなわない。そのため、それぞれの作業を小グループのリーダー(実行委員)が統括し、さらにクリーニング責任者が再修正を加えた上で、実行委員会全体が必要に応じて原票を参照しながら最終的な全体調整を行った。

今回のクリーニングでは、第2波が回収できなかった場合も含めて1879ケースをクリーニング対象とした。その結果、研究会メンバーの分担作業によって681ケースに何らかの修正が加えられ、436ケースに実行委員による再修正が加えられた。このように大きな数を示すと当初データの品質に疑問をもたれるかもしれないが、中には無回答の変数の値を周辺情報から補填したような場合や、形式的なコードの置き換えも含み、大部分のケースでは小さな修正に留まっている。郵送調査で心配された不整合な回答は少なかったという印象である。第5波まで同様の回答状態であ

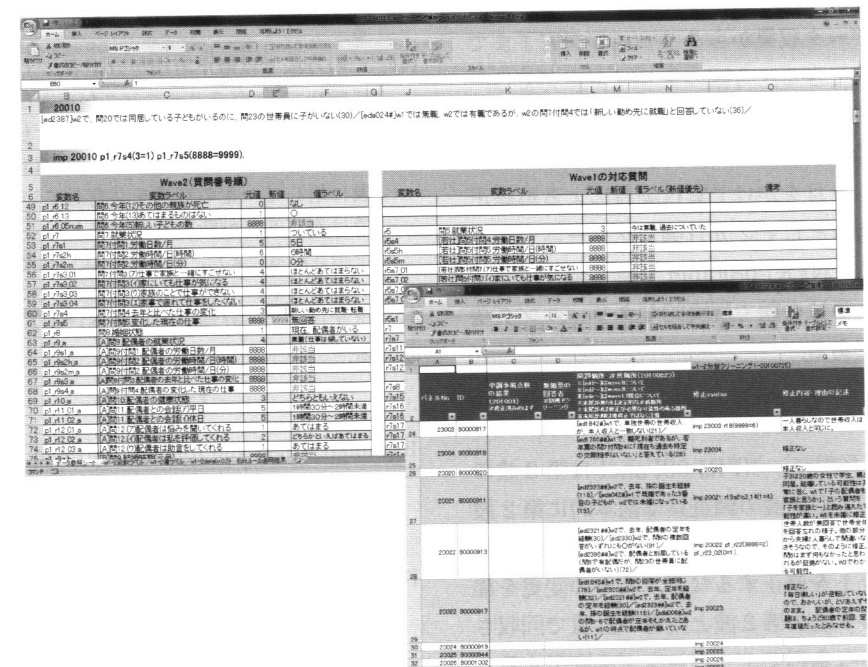


図1 NFRJ-08Panelのクリーニング(イメージ)

れば、かなり品質の高いデータが期待できる。

このクリーニング方法は、異常なケースについて一つずつ周辺情報を見ながら適切なデータに修正していくという手続きが、歪んだ粘土細工を一つずつ周辺の形に合わせて手直しする感覚に似ているので、「粘土細工アプローチ (clay modeling approach)」と呼ぶことにする<sup>(4)</sup>。粘土細工アプローチでは異常の検出とケースごとの修正を何度も繰り返す。そのため、新たな異常パターンを追加で検出することが容易であり、逆に誤って修正しすぎたものを元に戻すこともできる。修正方針自体も具体的なケースの積み重ねから徐々に固まってくる。コネコネとより適切なデータに徐々に近づくイメージである。

また、このアプローチは現実的に作業を進行させやすいという明らかなメリットがある。ケースを単位として修正作業を進めるので、分担作業が容易で、達成割合も把握できる。修正後のデータに異常が残っていないことを、同じ edit ルールの再適用で間違いなく確認することも利点である。また、edit ルールはデータ収集前から作成可能なので、データ収集からクリーニング完了までの期間を短縮できる。今回のクリーニングでは、調査会社から最終データが納品された後、約1カ月間、研究会メンバーによる分担修正作業を行い、さらに1カ月をかけて実行委員による再修正を行った。全体として2カ月強でクリーニングを完了している。NFRJ-08Panel で見る限り、粘土細工アプローチは有効で現実的なクリーニング法といえそうである。

## VI. おわりに

2011年1月現在、NFRJ-08Panel は第3波の調査中である。第3波の調査票は基本的に第2波と同様であるが、回答間違いが目立った質問についてレイアウトを修正するなど、若干の変更を加えている。第3波が第2波と同等の品質で回収され、より小さな負担でクリーニングを完了することができれば (ノウハウの蓄積ができたのでその

ように予定している)、NFRJ-08Panel の調査票設計は一応の成功を収めたと評価してよいだろう。データ収集で力を使い果たすことなく、貴重なデータを活用した分析成果が生みだされることが期待できる。

### 【注】

- (1) NFRJ-08 パネルについては、西野(2009)およびウェブページ ([http://www.wdc-jp.com/jsfs/committee/contents/NFRJ08Panel\\_profile.htm](http://www.wdc-jp.com/jsfs/committee/contents/NFRJ08Panel_profile.htm)) で詳細が示されている。なお、第3回全国調査の正式略称は NFRJ08 であるのに対して、本パネル調査の略称は NFRJ-08 Panel とハイフンが入る。「NFRJ」における「2008年抽出サンプルを第1波とするパネル調査」を示している。
- (2) 第2波の調査票デザインは、筆者と水落正明氏 (三重大大学) を中心に実行委員会で協議して決定した。また、第2波のデータ・クリーニングは、筆者と田中慶子氏 (家計経済研究所) が取りまとめにあたり、実行委員会および研究会メンバーの全体で作業にあたった。
- (3) 東京大学社会科学研究所による「働き方とライフスタイルの変化に関する全国調査 (JLPS)」の調査票は郵送配布であるが、回収方法は原則的に訪問である。
- (4) この方法では修正方法が分担者によってまちまちになり、均一なクリーニングができないと感じられるかもしれない。しかし、少なくとも今回の作業では分担者による判断の違いは意外なほど少なく、対象に対して一定の知識を共有している研究者の間では本質的な判断の齟齬が生じることは少ないと予想される。また、望ましくない齟齬はクリーニングを進めていく中で自然と収束していく。粘土細工アプローチの詳細は別の機会にまとめる予定である。

### 【文献】

- Bakeman, Roger and John M. Gottman, 1997, *Observing Interaction: An Introduction to Sequential Analysis*, 2nd Edition, Cambridge: Cambridge University Press.
- Fellegi, I. P. and D. Holt, 1976, "A Systematic Approach to Automatic Edit and Imputation," *Journal of the American Statistical Association*, 71: 17-35.
- Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Cou-

- per, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau, 2009, *Survey Methodology*, 2nd Edition, Hoboken: John Wiley & Sons.
- Herzog, Thomas N., Fritz J. Scheuren and William E. Winkler, 2007, *Data Quality and Record Linkage Techniques*, New York: Springer.
- 西野理子, 2009, 「〈NFRJ の確立〉にむけて 2——家族

パネル調査という新たな試み」『家族社会学研究』21 (1): 114-117.

保田時男, 2007, 「NFRJ08 における複数調査票の作り方」『家族社会学研究』19(2): 106-112.

———, 2010, 「調査データのクリーニング方法に関する提言: Fellegi-Hout の原則に立ち返る」第4回数理社会学会大会報告資料.