

パネル調査のデータマネジメント方策及び分析に関する総合的システムの開発研究  
第2回企画委員会議事録

1 日 時：平成17年7月27日（水）午前11時～午後1時

2 場 所：国立社会保障・人口問題研究所 第4会議室

3 出席者（13名）：

委員（五十音順）

阿藤 誠 早稲田大学人間科学学術院 特任教授

津谷 典子 慶応大学経済学部 教授

中田 正 日興フィナンシャル・インテリジェンス株式会社 副理事長（年金研究所担当）

参与

村山 令二 厚生労働省大臣官房統計情報部 人口動態・保健統計課長

西山 裕 同 社会統計課長

衛藤 久司 同 社会統計課国民生活基礎調査室長

原山 喜久男 同 社会統計課縦断調査室長

調査担当

北島 和久 厚生労働省大臣官房統計情報部 社会統計課縦断調査室 室長補佐

後藤 敬一郎 同 室長補佐

山下 りつ子 同 縦断調査専門官

議長（司会）

金子 隆一 国立社会保障・人口問題研究所 人口動向研究部長（開発研究主任研究者）

オブザーバー

首藤 健治 厚生労働省大臣官房統計情報部 人口動態・保健統計課  
疾病傷害死因分類調査室長

石井 太 厚生労働省大臣官房統計情報部 企画課審査解析室 室長補佐

4 配布資料：

(1) 企画委員会議事（付 出席者（予定））

(2) 縦断調査関連資料

21世紀出生児縦断調査 …… 第3回調査「結果の概況」、調査票（第4、5回）

21世紀成年者縦断調査 …… 第2回調査「結果の概況」、調査票（第3回）

## 5 議題及び議事：

### 【以下概略】

はじめに（金子部長）

まずこの企画委員会の性格について。本委員会は本科研プロジェクトの組織として位置づけられているが、本プロジェクトは縦断調査の実施主体の統計情報部とも密接な連携を図っており、また実施については恒常的な専門的諮問組織も設置されていないことから、本委員会がその機能を果たすことが期待されている。したがって、本日も研究プロジェクト実施に対しての助言とともに、調査を含めた全体的な視野から意見、助言を賜ればと思う。なお、実施体制については最近大きな変革があった。西山社会統計課長にご説明をお願いしたい。

### （1）縦断調査の新しい実施体制について（西山課長）

人事異動による本企画委員会のメンバーの変更は資料のとおり。また、縦断調査の実施体制を整備して社会統計課に7月1日付けで縦断調査室を創設し、人口動態・保健統計課で実施していた「出生児縦断調査」、国民生活基礎調査室で実施していた「成年者縦断調査」と「中高年者縦断調査」の3係を原山室長はじめ13名の体制で担当する。

### （2）経過報告

#### ① 出生児縦断調査（北島補佐）

第3回出生児縦断調査の概況は昨年12月に公表。子どもの就寝時間と朝食の欠食状況、テレビの視聴時間等を中心にまとめている。平成16年度の第4回調査については、調査が終了し、現在データの整備中で本年12月の公表を予定している。調査のポイントは、子どもの性格、しつけの状況、しかり方である。第5回調査は、本年調査となる。子どもの1日の生活パターンをポイントとして、1月出生児の調査は本年の8月に実施、7月出生児の調査は来年2月に実施する。

#### ② 成年者縦断調査（後藤補佐）

第2回成年者縦断調査を本年3月に公表。概況は2回目であり縦断らしい集計公表ができた。第1回で調査した結婚意欲や出産に関して、2回目の調査で結婚意欲などの意識がどう実現していくかという意識をポイントとした。第3回調査については、ほぼ第1回調査と同様の内容で調査している。第4回調査は第1回で調査した結婚意欲などの3年経過後の状況等を本年11月に調査する。また、中高年縦断調査も11月に第1回調査を実施する。

### （3）パネル調査分析システム開発プロジェクトについて（金子部長）

昨年まとめた研究成果を報告する。報告書では、事例研究レビュー、データマネジメント研究、縦断調査の不詳・脱落等の課題に対する研究の3つに分けて報告している。

## [事例研究]

我が国では事例の少ないパネル調査であるが、諸外国では調査の主流となっているパネル調査の事例レビューを行った。個別データを集約したミシガン大学の ICPSR (Inter-university Consortium for Political and Social Research)のデータアーカイブをインターネットで蒐集し、研究及び縦断調査の目的に沿ったテーマ別に「健康・医療・社会生活・家計・仕事」、「政治・選挙」、「犯罪学」の3つに分類し、主な調査を収集した。分類に関しては本年度も引き続き整理を続けて行っていく。

次に、方法論について多くの文献があるので、縦断調査の参考となる入門的なもの、初歩的なものを中心に基本的なものを取り上げ文献の内容のレビューを行った。報告ではごく一部をレビューした。また、巻末の資料編にパネル調査分析手法に関する文献集を作成した。今後も充実させたものを作成していき、調査に関わる人々の間で共有できるように整えたい。

パネル調査における統計分析モデルでは、実際にデータマネジメントを行っていく上で、どのような分析を行うかを主眼として、縦断調査に使用できる基礎的な統計分析モデルを簡単にレビューした。たとえば、まず縦断調査のデザインに沿った分析方法を探るために、基本的な調査デザインの体系を分類した。調査の体系でみると、出生児縦断調査は出生コホート調査、成年者縦断調査は年齢コホート調査となる。昨年度は少子化の原因究明あるいは健全育成に関わる分析に主力となる因果関係を分析する統計モデルについて、最も基礎的なタイプを取り上げてみた。また、欠損・脱落に関する統計的な対処としてどういったものがあるかレビューした。引き続きレビューを行ってゆき、本年度はとりわけ本格的、応用的な分析方法をレビューしていく。

## [データマネジメント]

縦断調査の実施に関わる実際的な課題について担当者にヒアリングを行い、データマネジメントの観点から問題点の所在と解消ができるか否かをまとめている。

出生児調査に関しては、年2回調査を実施しているが、スケジュールの中で限られたマンパワーでいかに効率的に行っていくかが大きな課題である。統計的分析では、単に横断的な形の発表だけでなく、時系列的な変化に関する高度な統計的分析が必要となってくる。実施・公表という切れ間のない作業サイクルの中に高度な統計的分析をどう組み込んでいくか検討しなくてはならないことが浮きぼりとなってきた。

縦断調査データの管理的システムの検討を行った。実際のデータマネジメントをいかに行っていくか、基本的な考え方を検討した上で、具体的にシステムを作成し対処できるか否かを検討した。データ分析という観点から見ると、データセットが増え密接に関連したデータが増えてくるが、統計分析をする場合はこれらを一つのデータとして扱わなくてはならない。単に直列につなげばよいわけではなく、特別な工夫をしてデータセット間の関係を図らなければならない。これが特に重要となる調査は、成年者調査であり、男性票、女

性票、これを合わせた夫婦票、配偶者票と複数のデータセットが1回の調査で発生する。調査の回数を重ねれば普通の扱いでは対処が難しい。一般的な縦断調査のデータマネジメントの必要性として、統一性、汎用性、柔軟性などいくつかの課題を特定したが、実際にシステムを作成していく段階で、1つのマスターデータを分析すればよいというわけではなく、分析の目的あるいは母集団に合わせて新たにデータセットの作成が必要になる。マスターデータとしては、オリジナルの毎回のデータセットをきちんとクリーニングしたものを管理し、集計分析する際に、目的ごとに自動的にそれらから必要な個票データを抜き出して、集計分析対象データセットを構成することによって分析につながっていくという形をとることが必要であるという結論に至った。

それを行うためにはどのようにするかはP80に概念図を示してある。図で示す左側に毎回のラウンドごとのデータセットとコード表があり、コード表が中心的な意味合いを持ち、コード表にデータに関するすべての情報を持たせることで、そのコード表を操作することにより、自動的に、必要なとき必要なデータセットができるようになる。それを行うためにはそれぞれのプログラムが必要で、図でいうと、上の「変数記述プログラム」と下の「変数記述プログラム」でコード表を情報として、そこから自動的にプログラム「データセット加工プログラム」を生成していく。そこから下の「データセット生成システム」に直接かつ瞬時につながり、集計分析対象データセットが出来上がる。実際にツールを作成し実験的に作業を行ったところ比較的うまくいった。詳細な説明は省くが、イメージとしてP82にコード表を示してある。データに関する情報すべてを表の形でExcelを使い記述する。この表からツールを使いデータ加工のためのプログラムを生成する。この様子をP84、P85に示している。P85はSASプログラムの例であるが、コード表から直接集計プログラムを自動的に生成する。P86はSPSSプログラム、P87はADAMプログラム、P88はJPERLプログラムと目的に応じた統計分析システム、集計システムのそれぞれのプログラムにはき出すシステムを開発した。

いろいろなプログラムのパッケージを生成することが可能であることがわかり、P89の例は単純集計であるが、コード表とデータセットさえあればワンタッチででき、クロス集計であれば変数を選べばクロス集計が可能となる。このシステムはデータに関しての課題を探る観点もあり、実用的かどうかはまだ不十分な点があるが、本年度の研究の一つの主眼として、実務の方と連携し、縦断調査室で実際に使えるものにしていこうと考えている。

#### [縦断調査の不詳・脱落に関する課題]

P99からはデータの不詳・脱落に関する課題であるが、ここでは不詳に関する基礎的な検討である。どの調査においても不詳は悩みの種で、単に不詳を除外してよいものか、何らかの意味を持つのかいつも悩ましいものである。縦断調査では一定のサンプルについて詳しく調べているので不詳の性質を知っておく必要がある。逆にいえば不詳の発生の仕方

を分析する手段でもある。今年度の分析としては基礎的なものとして発生パターンを分析していく。

パネル調査では、不詳が発生しても変わらない性質の属性によってフォローアップできる一面、同じ属性を調べていく上で矛盾が生じることが課題である。

P107 は脱落に関する分析で、縦断調査をする上での脱落は中心的な問題である。昨年度は出生児縦断調査の2回分のデータを得ている。その分析結果は1回目の回収率は87.7%で回収率としてはかなり高いもので、2回目の脱落率は6.6%で回収率としては悪くはないが低いともいえず、官庁の統計は比較的回収率が高い点からみるといかに脱落を減らすかが難しい課題ではないかと思われる。脱落がランダムに生じているわけではなく、父母の年齢が若い場合、父母の年齢差がある場合など属性による違いが出ている。そのほか脱落の要因として挙げられるのは婚前妊娠、片親、多胎の場合、非嫡出子の場合脱落が目立っている。逆に低い場合は、第1子の場合、祖父母と同居している場合である。また、注意していかなければならないものとして、母親が有業の場合、低所得の場合に若干脱落が多い傾向が見られ、これらはこの調査の中心課題に関わるものであるので見守っていかなければならない部分である。

P127 の居住移動については、1回目と2回目の居住地(都道府県)で異なる割合が8.3%でかなり移動が多い。移動と脱落については一定の傾向があり、移動の傾向を調べることも重要である。移動の多い傾向は、大都市居住、核家族、高学歴、父親が高収入、労働時間・通勤時間が長い、母親が専業主婦である等である。調査の目的に関わって結婚、出産に関連した移動が多いことがこの調査や他の調査からも多く見られることから、移動に伴った脱落を何とか防がないとこの調査の目的に沿わなくなることがここでの分析の結論である。

P139 は出生動態や意識の地域間格差の分析への課題である。本年度は、実際の統計分析を行いたいので、分析への準備として地域に着目して基礎的な分析を行った。

以上が昨年度の研究成果である。P151は全体の研究課程の見取り図だが、16年度及び17年度に大きく分けて示している。昨年度は主要な文献レビュー、本年は集計分析事例の検討に沿って応用していく。最終的な成果として文献資料データベース、統計分析システム、分析に関する方法論的な分析提言を挙げている。2年度のプロジェクトであるので、報告の形態をどうするかだが、現在、成果報告、ワークショップ、既存の研究会等への発表を検討しているが、具体的な形がまだ定まっていないので、どのような形が望ましいかもについてもご意見をいただきたい。

## 6 意見、質疑等：

【以下概略】

(津谷委員)

NLSY (Nation Longitudinal Survey of Youth) という調査があるが、文献に載せてい

ないが70年代から個人を追跡している調査であるので、参考になるのではないかと。

(津谷委員)

マスターデータセットとは基本的にはExcelのコード表でリレーショナルデータベースを作成され、分析の目的別にフラットファイルをライトアウトするということか。最初のデータベースを作成するときは、マンパワー等相当な時間が必要で、これをうまく作られると後々に楽になり、成年者調査だと、男性、女性が夫婦になると夫婦票、子どもが増えてくると子どもからみた父親の情報、母親の情報といった場合にデータを逆に見ることになり、そうするとユニットを換えて見なければならぬこと、最終的な効率、コストを考えるとこのシステムは重要なことと思う。CSV ファイルをそのまま Excel ファイルで、コード表も Excel ファイルできているとそのまま使用が可能であるので楽かと思う。多量のデータをカスタムメイドでフラットファイルができるし、ディスクスペースもありロジカルレコードリングスも関係なくよい方法と思う。

(金子部長)

なぜこういうシステムを考えたかという点、基本的にこれまで統計情報部の横断調査でのノウハウがあり、現在、コード表はExcelを使用しており、かつ今後の集計に関してもいままでの継続性を考慮すると、Excel データ から SAS プログラム等で集計表ができるイメージの方がよいと思われた。

(津谷委員)

父母の学歴についての調査は？

：2回目で調査している。

(津谷委員)

P 118, P 119 に示す変数の要素が70ほどあり、20個で5%のレベルで優位性の判断ミスがあると70個では4～5個となるので、基本的、客観的な情報に絞られてみたらどうか。人口学的な基本的な属性、居住地の属性、金銭的な問題、母親の就業等に絞り込むと良いのではないかとと思われる。結果の解釈やこれからどのへんが落ちやすいか調査員などに簡潔なメッセージを伝えられたらフォローアップにもなり、これからの調査のゆがみも矯正されると思われる。

(阿藤委員)

海外でのパネル調査の事例で、オーストラリアで赤ちゃんの追跡調査をしていると聞いたが、調べて事例に加えたらどうか。

パネル調査のサーベイでは子どもの成長・発達、結婚・出産予定など政策評価への実例を是非示して欲しい。統計分析法のレビューでは、パネル調査では因果分析ができることが最大のメリットであることから最終報告では因果分析の実例を示し、メディア、財政・政策当局に向けてアピールしたらどうか。

統計処理の課題が一番のポイントとなるが、統計情報部で毎年される集計・公表部分とパネル調査の分析結果とどうかみ合わせるかが難しい問題であり、行政チームと研究チー

ムが常に存在しなければならない。一時的に2年間研究分析を行った、報告したということではなく、5年、10年、あるいは5年毎といった節目というような長期的な高度分析をするため長期的な計画かつ長期的な組織を考える必要がある。

データの管理システムのメインポイントは、簡易に動かせるシステムであるかどうかである。今回できたサブシステムをさらに発展させ行政が採用できるのかどうか判断が必要。行政側の仕組みと協働で作業をお願いしたい。

脱落に関する問題では、バイアスのかかった特定の層が答えないと齟齬が起きるということだが、研究者が使う分にはマイナスとならないが、行政で発表するにはマイナスとなることを意味し、1回目調査より2回目、3回目の方が、誤差が大きくなるということで、行政側が対応策をどういう方向で、どこまですべきかが行政の課題であり、研究者もそうなった場合により方法を行政側にサゼッションできる統計的な処理の問題に有意義なアドバイスをしてもらいたい。

個人情報法の問題では、居住者をいつまで追跡調査するかの問題に対してどう回避するか等、今後の問題として想定される。

(中田委員)

今後、分析を行政だけでやるのは大変である。研究者との連携が必要だろう。分析を研究者にやってもらうには、役所側が何を知りたいのかを研究者に伝え、研究内容を整理すべき。また、研究会のメンバーをたとえばこのチームだけでやるのか、オープンに声かけするのか、など考慮すべき点もある。

出生児調査で脱落の検討をされているが、同じようなことを成年者調査での脱落をみると、成年者調査は、サンプルが抽出であり推計の仕方によっては、分子の方だけでなく分母の方も確率変数となるので分子、分母も被推計で普通より分散が大きくなる。その分を考えて行う必要がある。こうした事情もあるので、役所のデータに関しては、パッケージは有効に機能しないことがあるので、注意が必要となる。

(津谷委員)

膨大なデータベースの構築は時間がかかることなので、有効活用するためにも研究を長期的なものとするべきとの阿籐先生の考えに賛成である。

2番目は、役所が報告書に出す集計とその後の2次集計やその他も含めパネル調査は因果関係を解き明かす最大の長所を活かすためには、結果を単純集計やクロス集計をどのように合致させて一般の門外漢にもわかるような分析が必要であり、解析にはトレーニングを積んでいくことが必要である。

3番目は、対象が確率サンプルですから、脱落によりエリートサンプルとなつては調査の目的からすると困る。サンプルをジェネラライズ(一般化)にする秘訣は、サンプルを常にクロスチェックを何回もベースに作ってかけていくことで、これをひとつのプロセスに組み入れることが必要な措置である。脱落が進む過程で、大規模調査(例えば国民生活基礎調査)とのクロスセクションをやったらどうか。

4番目は、個人保護法の問題や調査の協力を得るには、縦断調査の結果「・・・のようなことがわかった」というようなことをPRすることが重要で、プレゼンテーションのポイントの1つは、因果関係で意識が行動をどう限定するかである。アメリカの調査のNSFH(National Survey of Families and Households)という調査がある。これは非常に大きな調査で5年毎に行っている調査で、この中で結婚時の意識から離婚に至るといった行動を起こす「行動パターン」を追跡している。日本でもこのような「意識から行動を生む」調査を、少子化の動向を政策的な観点からみる上でも「意識の行動への影響」を今後も続けて行う必要があると思う。

(西山課長)

出生児調査は今年5回目を迎え、調査回数を重ねてくると各方面から成果はないのか?といったことを聞かれる。成年者調査では調査票の種類も多いことなどデータ整理の作業量も膨大となり、データを活かすためにも縦断調査本来の結果研究を含めて今後を見据え計画的に行う必要がある。縦断調査はいろいろな方面からも期待が大きいだけに成果についても得られるものはどのようなものを求められている。

(阿藤委員)

学者は、パネルデータを如何にアクセスできるか関心を持っている。

(津谷委員)

未婚者割合の増加が進んでいるが、人口学者からみると、「結婚したくても結婚できない」のか「結婚したくないから結婚しない」は大きな課題であり、パネル調査でたどっていけば何らかの因果関係が見つかるのではないかと。

(金子部長)

社人研で行っている「出生動向調査」で結婚意欲を年齢別にとっているが、何回か調査結果が出れば疑似コーホートでみられるが、若いときは結婚意欲が高いが、年齢が高くなると結婚しない割合が増え年齢とともに結婚意欲が下がってくることが捉えられている。

(津谷委員)

結婚の状況が低い出生率に影響を与えているので、大きな問題である。

(村山課長)

離婚数など人口動態調査では数字はわかるが、意識に関しては人口動態調査では把握できない。

(津谷委員)

パネル調査で、統計担当者の国際会議などは行っていないのでしょうか?

: 具体的には聞いたことはない。

(阿藤委員)

パネル調査は、外国では国で行っているのか?

(金子部長)

ほとんどは研究機関、主に大学で行っており、国の研究費で行っていることが多い。国



の研究費で行うということは、一般に還元するという義務付きであり、データも公開している。日本の縦断調査とはかなりシステムが違う。

(津谷委員)

大学の先生でも大きな額の研究費がとれるとそのうち一部は大学が間接経費としてとり、その分の教育義務を免除して研究に時間を振り向けることができるシステムがある。また、良質のデータをとることが業績になる。そうした仕組みでよい調査が行われる。日本のように、誰かが無理を重ねるようなやりかたでは、継続は難しい。

(金子部長)

官庁の目的外申請などの仕組みで、統計がより多く活用されるほど成果として評価されないことには、追加的業務が発生するだけという矛盾があると思われる。

(阿藤委員)

統計法の見直しの動きもある。経済財政諮問委員会の組織で、経済社会統計整備推進委員会(吉川委員会)で検討が始まっている。

：国・地方で、必要な分野を重点的に統計制度、要員配置等を含めて既存の統計を抜本的に見直し整備するというもの。

(津谷委員)

目的外申請については、国立大学の独法化などで国の機関として仕分けが難しくなることもあり、たとえば厚生科学研究費の中で申請者の資格、目的を審査して、省庁で行われている調査データを使えるようにしてはどうか。

(金子部長)

良質のデータを扱う上で研究者の側にもコンペティションがあつてしかるべき。その上で利用の便宜を図るのが望ましいと思われる。

本日はプロジェクト研究の内容から調査実施にいたるまで有益なご意見をいただき、感謝したい。本プロジェクトの残りの研究期間は短いですが、これらの意見に沿って全力で役に立つ成果を挙げるようにしたい。

以上

パネル調査のデータマネジメント方策及び分析に関する総合的システムの開発研究  
第3回企画委員会議事録

1 日 時：平成18年3月3日（水）午前11時～午後1時

2 場 所：国立社会保障・人口問題研究所 第4会議室

3 出席者（13名）：

委員（五十音順）

阿藤 誠 早稲田大学人間科学学術院 特任教授

津谷 典子 慶応大学経済学部 教授

中田 正 日興フィナンシャル・インテリジェンス株式会社 副理事長（年金研究所担当）

参与

村山 令二 厚生労働省大臣官房統計情報部 人口動態・保健統計課長

西山 裕 同 社会統計課長

原山 喜久男 同 社会統計課縦断調査室長

調査担当

北島 和久 厚生労働省大臣官房統計情報部 社会統計課縦断調査室 室長補佐

山下 りつ子 同 縦断調査専門官

研究班

金子 隆一 国立社会保障・人口問題研究所 人口動向研究部長

釜野 さおり 同 人口動向研究部第2室長

三田 房美 同 企画部主任研究官

オブザーバー

仲津留 隆 厚生労働省大臣官房統計情報部 企画課審査解析室 室長補佐

4 配布資料：

- ・ 企画委員会議事次第（本文書、付参加メンバー）
- ・ 平成17年度報告書案
- ・ 平成16～17年度総合報告書構成案
- ・ 新規プロジェクト（平成18年度厚生労働科学研究費補助金）企画案
- ・ 新規プロジェクト（平成18年度厚生労働科学研究費補助金）研究計画書（抜粋）

5 議題および議事：

## 【以下概略】

### (0) 挨拶 (金子)

忙しい中の出席に対するお礼が述べられ、会議の進め方についての説明がなされた。会の進め方は、まず、経過報告として、報告書案について簡単に説明し、その後今後のこの研究の発展のさせかた、進め方、事業終了後のあり方について企画を述べる。その後、皆様からのご意見いただきたい。

### (1) 経過報告

パネル調査分析システム開発プロジェクト報告書案について (金子)

#### 1) 平成17年度 報告書 (案) の説明がなされた。

事前に送付したものに沿って、今年度の成果を簡単に説明する。送付したものは単にまとめたもので、充分精査を行っていないので、恐縮であるが、本会議で意見を賜った後に精査してまとめた。

- ・ 目次：全体のテーマを3つに分けた。事例研究とレビュー、データマネジメント、今年度はあらためて分析事例を入れ協力者それぞれにやってもらってまとめた。
- ・ 総括については省略。
- ・ 1：諸外国に置けるパネル調査のレビュー (パート2)：一年目のものを継続して、釜野さんに作成をいただいた。基本的には国際的なアーカイブ、最大と思われる ICPSRなどを参照しながら、主要なもの、縦断調査に関係あるものを集めた。今回は、調査リストを用意し、アルファベット順になっている。主に家族関係や厚生労働関係を入れたが、他にも政治関係、パネル調査では先駆的な分野である更正プログラムや犯罪に関するものも入れている。ペーパーとして参照できるが、実際にはコンピューター上のリンクを利用した形になっている。各調査のオリジナルのサイトへのリンクも貼ってある。従ってコンピューター上の利用が最も効率的である。いずれ、ネット上等に公開することも考えている。内容については基本的に、タイトル、目的、調査項目、方法、結果 (回収など) について要約したものを載せている。そこからより詳しい情報についてはオリジナルサイトにいくという形で利用できる。特に NSFH、ウイスコンシンで行なわれたものには詳しい紹介がある。こういった形で今後も増え続けるものを入れて行って維持することが重要だと考える。
- ・ 2：パネル調査の分析の検討：メモ的なものを入れている。初年度のレビューの続きだが、応用事例を含めたらという提案があったので、試みとして行ったが、実際にやってみると、困難があることがわかった。
- ・ 3：イベントヒストリー分析：応用例への発展というステップ。鎌田さんをお願いした。基本的には山一先生のもをレビューしている。
- ・ 4：マイクロシミュレーションシステム (MS) の開発：諸外国では縦断調査とシミュレ

ーションがついになっている。方向性、分析の方向性、発展の可能性、有効性があるのでは、MSの研究へのステップとなる部分について開発を行った。基本的に、パネル調査というのは、データ構造と個人を追跡するという面で、MSの有用性が高い。脱落など統計的な困難性のため、通常の手法をそのまま当てはめていいのかということについて、MSを用いた分析法の検討がよく行なわれている。今回は、21世紀縦断調査に即したMSを行なうための基本設計とそれを支援するためのシステムを開発した。これによって、具体的には第一回の成年者調査をもとにして、2回目を再現する、そして実施と比べるなどができる。その枠組みを作った。

- ・ 5：21世紀パネル調査 PDB の開発：特に21世紀調査。複数のデータセットをタイプアップして使う必要がある。特に成年者については男性票、女性票、配偶者票、それらのペアリングを行なって夫婦を集計単位として分析する必要。ここでは、それらに対して支援を行なうシステムを作った。基本的な機能は、複数のデータセット／調査票を、一つのものであるかのように使うことができる機能である。複数の調査回／調査票に対応したデータセットを個人をキーとしながら各調査項目を厳選しデータを抽出し、滞りなく各種の集計ソフトに受け渡すという機能をもっている。毎回蓄積していくデータセットを納めておき、そこから分析に必要なデータだけを抜き出し、次の分析システムに引き渡すシステムとなっている。
- ・ 6：21世紀出生児調査における脱落についての分析：西野さんに詳細に分析を行ってもらい、分析をおこなってもらい、問題点の指摘、提言、諸外国の例をまとめてもらった。その結果によると、脱落がおこりやすい属性というのがはっきりとわかる。出生児調査、どちらかが外国人である親をもつ、ひとり親の世帯、育児へのかかわりが弱いケース、収入が少ないなど、比較的政策支援のターゲットになり得るところで脱落が起きやすい。移動とともに脱落が起きることが生じているが、これをなんとかして防ぎたい。諸外国の経験のインセンティブを高めるなどの例もあげてもらっている。それについては、調査対象者に研究、調査の意義をよく理解してもらうのが一番の基本だろうと考えている。そのためにも有効な分析が問われると考える。
- ・ 7：分析結果に影響を及ぼしかねない不詳回答の発生状況について：小山さんに分析をお願いした。たとえば、出生児調査では1月と7月では不詳の出方が違うという結果も出ている。実施のしかたに依存する可能性も指摘されている。今後の調査に参考になると思う。

次からは分析事例である。本研究では、高度な分析に必要なデータマネージメントの基礎を作り上げる努力をしてきたが、実際に分析に入らないと、どのような橋渡しができるのか、また分析の中でも基礎的なデータのクリーニング、方式などの課題が新たに生ずるので、事例として、分析に踏みこむ研究を行なった。

- ・ 8：ライフイベント因果分析：岩澤さんに分析いただいた成年者については1、2回と連続した際に生じるものをみる事ができた。調査分析の中心となるだろうライフイベント、結婚、出産について、分析していく上で、課題となるデータ上の問題を精査している。たとえば出生について、調査の間に起きた出生を捉えることであるが、要因の効果を分析する場合、本来妊娠との関係を捉えなければならない。しかし縦断調査の場合これを出生というイベントで捉えると、その要因の働き方の時期がずれてしまう問題がある。ここでの提案として、妊娠中かどうかを各調査で把握できたら、分析に有効な情報になる。実際に分析してみると、一年間におきるイベントの数は少ないこともあり、なかなか難しいこと。本格的な要因分析などは、ある程度の年数をかけないと難しいと考える。
- ・ 9：結婚意欲や出生意欲：福田さんの分析である。実際のイベントの分析は難しいが、全員を対象として分析できる意欲、結婚したいか、追加の子どもが欲しいかなどについては、分析できる。横断調査でも可能だが、縦断調査を使うと項目を多くとれる。また属性の変化が意欲に与える影響を分析できる。イベントの数が増えた場合、意欲と行動の関係を特定するという縦断調査の強い利点であるので、その基礎ともなるものである。分析結果を見ると、同棲は意欲高いが、ひとり暮らしの女性は意欲低い。パラサイト仮説によると親との同居が結婚を遅らせることになるが、女性の結婚意欲については、逆の傾向がみられた。所得については、のではというのがあるが、他の研究結果でもそうであるが、親と同居している女性は結婚意欲が高いという傾向がみられた。については、女性の場合は、非常に低いと結婚意欲が低く、400万円をピークとして、それ以上あがると再び下がるU字型の関係を示す。出生に関しては就業との関係、親との同居の関係について一定の法則性が見いだされた。夫の家事参加は、直接の影響ではないが妻の意識に影響をあたえて出生意欲に影響しているなど、やや複雑な関係がみられる。
- ・ 10：子育て意識、地域性に関して：相馬さんに分析いただいた。次世代育成支援法によって自治体、育児支援、行動計画の策定が行なわれているが、多様な施策が行なわれることが想定される。縦断調査によって、それと個票レベルの出生や意識などとの間連を調べるのは、政策評価という目的上たいへん有効だと考える。その観点から地域というものを分析軸としてみていくのが有効だろうと考えてすすめている。詳しい結果については、最終報告で報告する
- ・ 11：21世紀縦断調査の保育担当者の育児分析：元森さんに分析いただいた。出生児調査の育児に関して、回答者は必ずしも母親ではないので、それは選別しなければならない。その上で若干分析に踏み込み、子育ての負担感の規程要因の分析を行った。

以上。

## 2) 平成16～17年度 総合報告書構成 (金子)

資料の最後に、総合報告書の目次、構成案が載せてある。総合報告書とは、2年次の研究成果全体をまとめたものである。目次は初年度と今説明した本年度の研究を総合したものの。いくつかの項目の中には、ひとつにまとめたものもある。資料編では、調査実施や研究分析や進め方に関する提言を含めたこの委員会の議事録のまとめ、収集した文献資料の複製一式、システムの記述などを、今後の研究のためのセットとして載せる。

### 3) 今後の研究の進め方について (金子)

次の後継のプロジェクトをどうするかについて。本2年間の事業で、システムの基礎を作れたと思うが、やり残した部分、実際の分析につなげていく肝心な部分ではできていない。継続して研究を行うべく、新規事業として厚生労働科研に申請中。研究計画書、企画書は資料にあるとおり。資料の「研究の要約」を読み上げた。基本的には、分析法を検討し分析システムを構築する。データマネージメントから、分析の方に軸足を移して研究を進める。初年度は手法の研究、2年度はシステムの検証と確立、分析の検討を予定している。

縦断調査は有効だが弱みもある。回を重ねるごとに母集団(全国)からの代表性が落ちて行く。横断調査はそれが確保されている。両方のよいところを組み合わせ、補強してより有効な分析結果につながらないかと考えているが、本研究所では出生動向基本調査というテーマの似た横断調査を行なっている。主な役割は、モニタリング、社会の時系列的な変化を追っている。同じテーマに関して行なわれている2つの長所のある調査を連携させ、信頼性の高い結果に結びつけたい。その方法論についてはあまり事例がないので、どういことができるのかを検討し、実施してみたい。

組織については、これから固めて行く。企画委員会は同じ形だと考えているが後ほど意見をいただきたい。

ここまでは発展的な話だが、一つ申し上げておきたいのは、予算面で、厚生労働科研の全体枠が極めて厳しい。特にこの統計情報の部分についてはかなり厳しい削減がされると聞いている。今回の事業のレベルからするとかなり厳しくなると思われる。発展的な内容を企画しているが、この制約が組織的なことを含め、どのように影響してくるのかは、まだわからない。

## (2) 縦断調査、ならびにプロジェクトに関する意見交換 (全員)

阿藤：全体的なことについて、研究インフラの研究に対する評価が日本ではあまりなされていないが、極めて重要である。アメリカでなぜあのように計量分析や社会調査が盛んになったかという、アカデミズムの中心に図書館というものがあってそれらを非常に重視する。そしてコンピューターが発達すると、これらを含めた研究の共通インフラに多くの才能とお金と時間をかけている。それが汎用性の高いものになり、有効利用されている。最近ではデータアーカイブなど、立派なものを作ってい

る。同時に、官庁統計の公開や活用については1、2歩も、10歩も日本の先をいっている。そういう所が、社会全体、アカデミズム全体の研究を促している。日本はそういう部分が評価されないことが、大変残念である。このプロジェクトでは、金子部長が忙しい時間を割いて、研究の中心部分、データマネジメントシステムの開発などに大変な時間を割いてもらって、これからの日本の中心的なパネル調査を利用した研究のために有効になると思われ、感謝している。質問としては、事例研究は金子さんが中心になって開発されたそのデータマネジメントシステムに基づいて分析を進められたのか、それとは別に、個々に進められたのか。また、システムは統計情報部の方に行っていて、それで統計情報部の方が、つまりアカデミシャンでない人が使えるものになっているのかがわからないので教えて欲しい。

個々のものでは、パネル調査のメリットは、時間軸で前に起こったことが後ろに起こったことを規定するという単純な事実からくる。そうすると、意識が実際の結婚行動にどうむすびついたか、あるいは政策がどう効果があるのか、意識が変わったかどうか行動が変わったかなどを分析できること。その辺の応用の可能性について今回は調査の回数が少ないのでそこまではできていないという感じはするが、いろいろな可能性に種まきができたというところだろう。具体的な分析内容、方向性を作られたというのは大変よかった。

マイクロシミュレーション (MS) はまさにパネル調査と裏表の関係にある。それについて将来的にもやって欲しい。このまま終わってしまったらもったいない。アメリカなどでは時々このデータはどこからきているのかと思う分析発表があるがそれはおそらくMSからきている。日本では行なわれていないことで、こういう先例をつくったのはよい。レビューは、これも大変な努力で、もちろんホームページを探せば出てくるのだろうが、これを日本語に訳したのは、今後日本で関わる若い人も増えるだろうから有益であろう。日本の調査については入っていないのか。前にあったのかもしれないが、それもあつたら良かったと思うのであとで教えて欲しい。次年度以降について、予算削減で厳しいということだが、統計情報部にとっては、3つの調査、日本の官庁統計のパネルの3大部門を一手に引き受けているのであるから、それがうまくいかないというのは、大きな影響を与えるので、焦点をしばって、この調査および分析する人を育てるという観点から、ぜひ強力なサポートと後押しをしていただきたい。同時に統計情報部がこれをやっていることも社会的に意義を格段に高める。これまでも努力しているが、単発的に分析結果を発表してその時々で調査発表をしていけばいいかということ、だんだん時間がたつと、それだったら横断調査をやればいいのか、と成りかねない。そういう意味では縦断調査というのは、こんなに有効だということを示せる研究インフラを作っておく必要がある。せつかくこの2年間やったので、ぜひ次につなげるように御願いたい。

金子：ありがとうございました。本研究の意義を評価したご意見いただき、感謝したい。質問の事例研究で本システムを実際に利用しているのかということだが、初年度に開発したコード表を用いた分析支援システム、すなわち分析データの基本部分、一番面倒な変数記述をつくる部分については利用していただいているが、今年開発した部分は、とりあえずできたところなので、まだ実用のテストはしていない。今は実験的なデータで使えるのだというところまで作った段階。これからそれを実際に使える形で情報部にお持ちして使っていただく、と考えている。

釜野：3点目の、パネル調査レビューに日本のものは含めていないのかについては、「諸外国」というタイトルのためというわけではないが、入れていない。

金子：今後増えて行くと思うので、リストに含めるようにしていきたい。

中田：報告書（案）は全般的に良くまとめられていると思う。もう一つ何かアピールするプラスアルファがあるともっと良いと思う。アピールと言った時に、パネル調査の何をアピールするのかということと誰に対してアピールするのかということがあるが、何をアピールするかについては、具体的な分析結果、こういうことがわかったということになる。誰にというのは、例えば、研究者にアピールするとすれば、情報を公開して皆さん使えますとアピールすれば関心をもってもらえると思う。先日、家計研でパネル調査関係のシンポにいったが、いろいろな大学の先生が報告されていた。最近では、パネル調査関連で、文部科学省の予算がついているようで、大学の方で、盛んに行われているようだった。大学のものと規模を比べると、こちらがずっと大きいけど、逆に大学の方では調査項目が多く、ある調査では「第1回は70ページだったが、50ページに減らした」と言うことだった。大学では数は少ないが細かい情報がとれるということがある。各大学で今は調査を始めたところだろうが、そのうち結果がいろいろ出てくるだろうから、このパネル調査の結果も早めに答えを出しているいろいろアピールすることが重要だと思う。

2点目は、パネル調査の利用に対し、具体的な提言があるといい。例えば、厚生労働省での利用であることを考えてみると、マイクロシミュレーション（MS）モデルによって、政策の評価をするというのが一番ふさわしいように感じる。それから研究者に対しての情報提供ができれば良いと思う。いろいろな分析のやり方は、研究者にまかせ、統計情報部ではむしろモデルを作って分析するというのの方がいいと思う。MSモデルなどは、時間もお金もかかるので研究者レベルではしづらい。MSモデルはいろいろな形で使えそうなのでよいと思う。官庁の調査の場合は特に行政データの活用、行政データとのリンケージが目玉になると思うので、そういうのも検討してもらえたらと思う。私は、年金の調査研究をやっているが、年金だと、



働きはじめてから40年くらいのデータが必要となる。これでは、成年者調査では40年後にならないと使えないので、すぐの活用は難しい。しかし、中高年調査をやるときに一緒に年金関係の役所の既存データとリンケージをすれば年金の調査研究に利用できることになる。リンケージがうまくできれば有効性が非常に高くなる。

また、国際比較も念頭におく必要がある。国際比較で日本だけもれているのは寂しい。年金と早期退職の関係に関して、10数カ国でパネルデータを使って分析するプロジェクトをやっているが日本はデータがないのでパネルでない通常の分析となっている。年金の関係では、アメリカの研究者などが日本のパネル・データはないかと聞いてくることもある。データの利用についても、日本では使えないが、外国では自由に使えるというのは好ましくない。

次に報告書の内容についての具体的なコメントを挙げたい。

- ・「まえがき」のところで、「調査主体」と言及しているが、これには専門家は入っていないのか。役所が調査をやって、専門家は単に使いなさいというのではなく、調査自体に専門家も意見をフィードバックできるのがよいと思う。
- ・「総括」のところで、2ページ目に、「2回の出生児調査」と出てくるが、出生児調査が2回で十分のようにもとられかねないので、これは書き方を変えたらどうか。
- ・「諸外国におけるパネル調査のレビューⅡ」では、アウトプットがデータベースみたいになっているということだが、読んだときにそういう風になっているのか単にリストされているのかわからないので、明確に書かれた方がよい。
- ・「パネル調査の統計分析モデルとその応用」には説明がないので、あったほうがよい。説明がなく、内容がよくわからなかった。
- ・パネル調査の統計分析モデルと応用例：イベントヒストリー分析」は、山口先生の論文をレビューしたということだが、専門的な部分だけでなく、よくわからない点がある。縦断調査の関係でどう使えるかというのがあったらいい。
- ・「縦断調査マイクロシミュレーション分析基礎システムの開発」では、MSの説明の中に何か所かに「自動的に生成される」とあるが、「自動的に」というのはどうか。通常は何か指示をするのだから、表現を工夫したらどうか。
- ・「不詳回答の発生状況とその処理に関する検討」では、最後に、「パネル調査について考察を加える」と書いてあるが、ページが抜けているのかもしれないが、途中で終わっている感じである。
- ・最後に、分析事例として具体例をいれてもらったのは大変よかった。

金子：詳しい指摘に感謝いたします。貴重なコメントを踏まえ、最終報告に向け、精査して行きたい。

津谷：研究所で最も忙しい金子部長が、時間を割いて、ここまで引っ張ってやって来たことは素晴らしい。報告書を読み、勉強になった。以下、改善のための提案を中心にコメントしたい。レビューで挙げている調査はすべて ICPSR に入っているのか？それ以外にもあるのか。あそこのアーカイブを中心にやっているのか。

釜野：ICPSR を中心として行っている。

津谷：それなら、Director を良く知っているのだから、何かあったらどのように管理しているのかをたずねることができる。見ているといろいろなデータの出し方がある。日本においても政府統計の公開、一次データの公開について今話し合われていて、わたしもその会議にいたが、総務省の審議会吉川委員会や、今学術会議もその方向性である。21世紀COEで樋口さんのやっている慶応のパネル調査には私も入っていて、公開を前提としている。

政府統計も今後は公開という方向に進んで行くと思うので覚悟が必要。非常に状況は変わってきている。家計研、慶応のパネル、厚労省のものは、ある意味違う。特に家計研は公開が原則となっていて。しかしこれは統計法の範疇のど真ん中。当然あれは使われる。公開のためであるので、私も大学院生のためにたくさん申請している。しかしデータの規模が小さい。1500確保するようにしてやっているのだから代表性の面ではよくない。その点、このデータは素晴らしいものである。将来的に一次的データの公開ということになるだろうと考えて、大きな流れの中で考える必要がある。しかし公開となると日本ではその土壌がない。1次データを出すとなるとそれなりのルール、エシックスが必要となる。出してしまっただけではお手上げというわけにはいかない。そこでこの最初のレビュー、もう少しやるとすると、データが公開されているかどうかについてのコラムをつくって、入れてほしい。Panel Study of Income Dynamics (PSID)などは、院生などをふくめダウンロードできる。しかし、このPSIDをやめるとの話もあるらしい。今日もNIHの人と話したが、そうになったらどうするのかと思う。それとICPSRにあるものでも、データの情報元に許可を取らないとだめなものもある。いろいろなレベルがあるので、もう1コラム入れて、一次データがあるかどうかの情報をいれたらいい。ところで、ここにウェブのアドレスがあるか、調査レビューのファイル上から、データを収集した人・機関のサイトに行けるようにリンクを張るかどうかは検討中ということか。

金子：それは入っている。クリックすればリンクするようになっている。

津谷：大変いいことではないかなと思う。

津谷：イベントヒストリーについては、先ほどわかりづらいということも出ましたが、これはわかってしまえば簡単なのですが、そこが問題だ。いかにこのモデルがパネルに適しているのか、今まではパネルではないものに当てはめてきたが本来パネルのためにつくられたものを impose してきた。しかしこの調査は本当のパネルなので、10年は続くわけで、そのことをはじめに言うておけばよい。いきなりモデルの説明

をするとなんだかよくわからないとなってしまう。なぜそうなのかということきちんと書けばそんなに大きな労力なしでよくなると思う。イベントヒストリーモデルは、日本ではあまりないが、ある意味使い古されていて、これ自身は別に新しいことではなく、**public dissemination** の意味を持つ。山口先生には去年セミナーをやってもらい、大変好評だった。またいらっしゃるかもしれない。イベントヒストリーモデルは、コックスが74年に出しているのですから、もう30年。**fixed effect model** について、今後回数が重なってきてからでいいので、やれば有用性があると思う。

- ・マイクロシミュレーション (MS) については、本当に価値のあるものである。欠損値、欠測値について、ない場合のシミュレーションの理論的な値と実際の値とを比べる。それでうまくいくとなったら脱落したもの、ちょうど外国にいていかなかったとかが、これからできてきた場合にも役に立つし、たとえば項目自体がよくなって、うまくいかなかった場合これで代替するなど、いうこともできる。ただ比べるだけでなく、欠損、欠測の今後の補正も含め、日本でできる人といったら金子さんくらいしかみあたらないと思うので、がんばっていただきたい。
- ・PDB 2 1 は、おそらくこれは報告書で最も重要で貢献度が高い部分だと思う。前も言ったが、結局 CSV のファイルで、SQL を使って **relational** データベースを構築したようだが、それが一番いいと思う。いろいろなテーブルを作るものである。**Child, household** 属性, **women** テーブルなど。それを **query** で結んできて、分析の対象ごとにカスタムメイドでフラットファイルをつくっていくもの。SQL を用いれば自由自在にできる、ある程度手動でやればできるが、カスタムメイドでフラットファイルを作る事自体が、すでに研究の一部である。どういう変数が欲しいかをわかっていないとできない。それをいろいろな要求にこたえ、厚労省や社人研、あるいは金子さんのグループがやってあげることにはできない。ということは、テーブルのまま置いておき、使い方の解説をしておくということになる。そうすると、いくらもっと有用性を高めて行くといわれても、これはしかたがない。誰もができるものではない。ただ、元々のフラットファイル、毎年のものであるわけで、それを、そのまま一次データ出すかどうかは別としてそれで処理できることはして、そしてこのリレーショナルデータに関しては、それをできる人が、使うとなれば例えばきちんとしたものを作って、グループのメンバーになってもらってやってもらう。そうでないとコントロールできなくなるし、そんな下働きをやる人はいない。これの素晴らしいことは、リレーショナルファイルは、そのままエクセルのファイルを **comma separated value file** として読み込めるので、もともとのデータの構築が非常に楽なので素晴らしい。

次に、脱落の分析について：結果の要約を読むとなるほどと思うが、もっと一般の人にもわかりやすくしたらどうか。例えば調査員に、こういう人が抜け落ちやすいので、こういう人に注意してもう一回訪問するようにとか、そういう風に

使えるとよい。それには、基本的属性のテーブルが必要。年齢、学歴、住居、子どもの数、収入などで、どういう人が抜落ちやすいのかがわかるものを作っておく。具体的にこの年齢でこういう状況の人が落ちやすいとわかることが大事。育児不安があるかどうかは、相手を見ただけではわからない。年齢や学歴、居住地などは、以前のデータからもわかる。そういう人に目星をつけて注意するという意味で必要。これを見てみると、おそらく、夫が失業者のところが多いと想定される。育児不安もそうだし、いろいろな変数があるが、それと correlate している、内生性が強い。ついでに、ケースの数、log-likelihood、likelihood-ratio  $\chi^2$  など、その他の基本的な情報を入れてほしい。

ライフイベントの分析の部分は勉強になった。この調査では当然ライフイベントが中心になってくる。4次調査くらいになればもう少ししっかりとしたものができる。どういう風にデータを構築すべきか。それには2つある。一つは person-year、同じ女性に一年ごとに1ライン。もう一つはイベントを単位にしたもの。そして第1子から第2子へ。後者だと何年何月という情報がないと困る。最初に両方やってみてあまり変わらなければ person-year でやってしまえば簡単だ。しかしそうでなく結果が変わってくるとなると、考えて入れるべきだと思う。

意欲の分析は、きれいになっていてよかった。ただ結果表の統計的有意性のマーク、\*の付け方を他と合わせる必要がある。育児その他、今度の第2次のプロジェクトでは就労がはいつていたが、今回はない。厚生労働省なので、労働もあってよかったかとは思う。

最後、代表性の問題。永遠のテーマである。出生動向は5年に一度だが、こちらは毎年やっている。ひとつのやり方は、一時的であるが、ウェイトをかける。脱落はランダムではない。基になるベースの人口を用いる。ただこの調査の開始は、国調と同じなので残念だ。2年ずれていると、国調2年度というベースになるウェイトを計算できる。どこからか、何らかの形でウェイトを指定してかけてみる。多変量解析の場合は基本的にはいいが、全体の descriptive の時にはウェイトを計算することを勧める。Statistics Canada のウェブサイトにも、代表性がなくなっていく調査についてどうするのかについての研究があった記憶がある。中央統計局で多くの研究をやっている。たぶん脱落が大きな問題になったのだろう。もう少し詳しく調べてみたらいい。これは政府が行っていることなので、日本と同じような立場であり、よい参考になるのではないかと思う。

その他、調査の呼び方について、第2回、3回という言い方をすると、クロスセクションと間違えるので、第2次、3次としではどうか。

金子：多岐に渡る有益なコメントに感謝します。宿題もでたので、いろいろと調べて対処していきたい。