

201002004A

厚生労働科学研究費補助金  
統計情報総合研究事業

# パネル調査(縦断調査)に関する統合的分析 システムの応用研究

(課題番号H22-統計-一般-002)

平成 22 年度 総括研究報告書

研究代表者 金子 隆一

平成 23(2011)年3月



# 目 次

## I. 総括研究報告

研究代表者（金子隆一）	3
研究分担者（北村行伸）	13
研究分担者（釜野さおり）	15

## II. 個別研究報告

### 縦断調査支援システムの開発研究

1 諸外国のパネル調査の概要の閲覧システムの更新 （釜野さおり・鎌田健司）	25
2 データ管理・統計分析システム（PDA21）の開発 （金子隆一）	29

### 21世紀縦断調査データ特性把握に向けた研究

3 21世紀成年者縦断調査（第1回～第6回）における女性票および 男性票の脱落者・継続回答者の特性に関する分析 （釜野さおり）	37
4 成年者縦断調査の標本特性：繰り返し横断調査との比較 （岩澤美帆・鎌田健司・福田節也・三田房美）	61

### 21世紀縦断調査の各種テーマに関するシード研究

5 子どもの感染症の罹患に関する分析 （井出博生）	85
6 現代日本の若年出産：21世紀出生児縦断調査第1～7回から （相馬直子）	95

7	子育て期の保育支援資源の都市規模・地域移動経験による違い (西野淑美)	123
8	子ども観と育児方針—主として「第7回出生児縦断調査」の分析より— (元森絵里子)	133
9	子育て費用と習い事の実態：21世紀出生児縦断調査第1～7回から (相馬直子)	167
10	育児不安・負担感に関する分析手法の開発に向けて： 21世紀出生児縦断調査第1～7回より (相馬直子)	185

## 縦断調査分析方法論の整備 — 『パネルデータ分析ガイド』 —

	『パネルデータ分析ガイド』への序 (パネルデータの利点と課題) (金子隆一)	229
第1章	固定効果・ランダム効果モデル (石井 太)	243
第2章	ダイナミックパネル分析 (北村行伸)	269
第3章	同時方程式パネルデータ分析 (北村行伸)	289
第4章	生存時間分析 (石井 太)	309
第5章	離散時間ハザードモデル (福田節也)	323
第6章	SURF モデル (福田節也)	337
第7章	共分散構造分析 (鎌田健司)	355
第8章	傾向スコア・脱落サンプルバイアスの検定法 (鎌田健司)	379

### Ⅲ. 資料編

- 1 パネル調査（縦断調査）に関する統合的分析システムの応用研究  
[研究過程の流れ図] ..... 399
- 2 国民のライフコース事象、21世紀縦断調査、高度統計分析研究、  
ならびに関連施策に関する相関図[事業概要図] ..... 400
- 3 「平成22年度中間・事後評価委員会」報告資料 ..... 401

※ 本報告書は本プロジェクト研究に基づき、研究代表者の監修により編集されたものであり、  
記述に関する一切の責任は研究代表者にある。

## 研究組織

### ○ 研究代表者

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

### ○ 研究協力者

(企画委員会委員 (50音順)) (敬称略、所属等は平成23年3月時点による)

阿 藤 誠 早稲田大学人間科学学術院 特任教授

津 谷 典 子 慶應義塾大学経済学部 教授

中 田 正 (株) リソースネット 顧問

(データ管理・分析法研究担当)

北 村 行 伸 一橋大学経済研究所 教授

釜 野 さおり 国立社会保障・人口問題研究所 人口動向研究部第二室長

岩 澤 美 帆 国立社会保障・人口問題研究所 人口動向研究部第三室長

阿 部 彩 国立社会保障・人口問題研究所 社会保障応用分析研究部長

石 井 太 国立社会保障・人口問題研究所 国際関係部第三室長

三 田 房 美 国立社会保障・人口問題研究所 企画部主任研究官

守 泉 理 恵 国立社会保障・人口問題研究所 人口動向研究部主任研究官

鎌 田 健 司 国立社会保障・人口問題研究所 人口動向研究部研究員

井 出 博 生 東京大学医学部附属病院 助教

藤 原 武 男 (独) 国立成育医療センター研究所 成育社会医学研究部長

福 田 節 也 マックスプランク人口研究所 研究員

西 野 淑 美 東洋大学社会学部社会学科 専任講師

相 馬 直 子 横浜国立大学大学院 准教授

元 森 絵里子 明治学院大学社会学部 専任講師

### ○ ご協力いただいた方々

(敬称略、所属等は平成23年3月時点による)

廣 松 毅 情報セキュリティ大学院大学 教授、東京大学 名誉教授

(厚生労働省大臣官房統計情報部)

小 野 暁 史 人口動態・保健統計課長

青 木 重 仁 社会統計課長

上 田 響 国民生活基礎調査室長

市 川 隆 之 社会統計課 縦断調査室長

(同部21世紀縦断調査担当)

吉 川 英 樹 社会統計課 縦断調査室 室長補佐

松 岡 芳 弘 社会統計課 縦断調査室 室長補佐

巽 恵美子 社会統計課 縦断調査室 縦断調査専門官

ならびに、社会統計課縦断調査室の皆さん

# I. 總括研究報告（要旨）

研究代表者 金子 隆一  
（国立社会保障・人口問題研究所）



厚生労働科学研究費補助金（政策科学総合研究事業(統計情報総合研究事業)）  
総括研究報告書

パネル調査(縦断調査)に関する統合的分析システムの応用研究  
(平成 22 年)

研究代表者 金子隆一 国立社会保障・人口問題研究所

研究概要

本研究事業の目的は、厚生労働省が実施しているパネル調査（21 世紀出生児縦断調査、成年者縦断調査、中高年者縦断調査、（以下、21 世紀縦断調査））に対し、高度統計分析に基づく政策的応用ならびに学術的研究を行うためのインフラストラクチャーを構築することである。

パネル調査は、施策効果測定や行政ニーズ把握等に有効であるが、横断調査と異なる独自のデータ管理法と分析法を要する。しかし、21 世紀縦断調査は政府統計初のパネル調査であり、行政目的に向けたパネル調査のデータ管理、集計・分析システム等の調査インフラは皆無であった。これに対し本プロジェクトは先行事業において、調査データの管理から統計分析手法の適用までを統合的に実施するシステム開発に取り組み、縦断調査データを用いた分析研究のインフラ構築に努めてきた。本年度の研究事業においては、それらシステムの実用化に向けた統合・総括を図り、調査実施者、一般研究者を使用者と想定した実装を目指した。これらによって、調査実施者に対しては年々蓄積される調査結果の速やかで有効な結果公表と行政的応用に資するとともに、研究者に対しては分析手法ガイドの作成や分析対象の体系化などを行いライフコース研究の学術的基礎の確立に資するものである。このような研究インフラのシステム開発に際しては、当該調査の調査項目全般に渡る基本的な分析工程の確認が必要となるため、実際に多くの事例的な分析を行っているが、データの二次利用に関する法制的な制約からこれらは個別の学術研究結果としては発表しないこととしている。ただし、これらの多くは個別テーマ分析の具体的な課題や注意点に関する情報を提供し、また今後本格的な研究へと発展しうるものであり、さらには調査項目全体の分析の体系化に資するものであるため、本事業ではシード研究と位置づけて報告書に含めている。

本事業では、具体的に 5 つの分野に分けて作業が進められた。すなわち

- (1) 海外パネル調査情報閲覧・検索システムの開発、
- (2) データ管理・統計分析システム(PDA21)の開発、
- (3) データ特性把握－脱落・移動・回答不詳の分析、
- (4) 分析法ガイドブック作成－縦断調査分析の手引き、
- (5) 各種テーマに関するシード研究－ライフコース的体系化、

である。(1)においては海外パネル調査情報を網羅し、キーワード検索が可能な日本語の情報ベースを提供しており、今回内容の追加更新に加えてシステムの改善を行った。

(2)においては累積される縦断調査データを管理し、複数の調査回や調査票にまたがった集計や統計分析を効率的に行えるよう支援するシステムの開発が進められた。SAS, SPSS, Perl, ADAM など様々な統計ソフト等による分析の基礎となるプログラムを自動生成するなどの機能があり、今年度においては全変数をグループ化してそれぞれコード番号を付して管理する方式を導入するなどして分析をより効率的に行えるようにし、またインターフェースをわかりや

すくするなど実用面における改良を行った。

(3)については、脱落を始めとする 21 世紀縦断調査のデータ特性のモニタリングを継続したほか、同時期に定期的実施された横断調査である出生動向基本調査の結果と比較することにより、脱落による偏りの程度を定量的に評価するなど、国際的に見ても例のないデータ特性把握の試みを行った。

(4)では、今後の 21 世紀縦断調査の分析研究に携わる分析者、研究者のために、縦断調査特有の分析方法論の基礎について解説し、21 世紀縦断調査の例を用いて解説を行った。全 7 章により構成され約 160 ページの解説書としてまとめた者である。各章の内容は以下の通りである。

- 第 1 章 固定効果・ランダム効果モデル
- 第 2 章 ダイナミックパネル分析
- 第 3 章 同時方程式パネルデータ分析
- 第 4 章 生存時間分析
- 第 5 章 離散時間ハザードモデル
- 第 6 章 SURF モデル
- 第 7 章 共分散構造分析
- 第 8 章 傾向スコア・脱落サンプルバイアスの検定法

本ガイドブックは、取り組みの難しいパネルデータの分析法を平易に解説し 21 世紀縦断調査データを対象とした今後の分析研究の推進に資するものと期待される。

(5)のシード研究において、次世代健全育成の領域では、子どもの感染症について、ワクチン全員接種、任意接種による罹患率の違いが明らかにされた。すなわち、全員接種の風疹、麻疹の罹患率に対して任意接種の水痘、おたふくでは圧倒的に高い罹患率を示し、また保育所での集団感染の可能性や若い父母経由の感染の可能性が示唆され、感染症法、予防接種法の制度設計に資する分析結果が得られた。また、同様に次世代健全育成の観点からは、親が子どもに望む性格が、その後の育児・教育方針、行動に強く反映されていること、同時に子ども自身の生活・学習状況にも影響していることが見いだされ、次世代健全育成に関する施策や教育施策の立案に資する結果が得られている。また、少子化メカニズム解明の領域では、母の就労における祖父母と保育サービスの補完関係、さらにはその追加出生との関係についての複雑な因子の絡み合いを、縦断調査データを用いることでよく整理できることを示した。これらは公的育児支援制度設計に資する結果である。このように政策的応用に直結する分析の基礎的な研究が実施されており、重要な提言となっている。

21 世紀縦断調査は、国の重要課題を中心テーマとしながら、対象年齢層の異なる三つの調査の組み合わせによって国民生活をライフコースの視点から体系的に捉える点で諸外国にも例を見ない独自性を持つ。とりわけ青年層の就業、結婚等を取り巻く状況の急速な変化や「団塊」世代の高齢層への移行など、今後わが国の社会経済に大きな影響を与えるだろう事象を対象としており、その実態や意識を時系列的にモニタリングする本調査の行政的価値は計り知れない。これまで本事業ではその分析を真に科学的なものとするための研究インフラの整備に勢力を傾けてきたが、現在それは一定の成果としてまとめ、今後の応用に附することが可能な段階に至って来ている。今後は開発されたインフラの活用しつつ、主にシード研究を元に発展研究の推進を行う。とりわけ政策応用研究の展開については、本格的な研究成果の提供を行うことが出来るものと期待される。



## A. 研究目的

本研究事業の目的は、厚生労働省が実施しているパネル調査（21世紀出生児縦断調査、成年者縦断調査、中高年者縦断調査、（以下、21世紀縦断調査））に対し、高度統計分析に基づく政策的応用ならびに学術的研究を行うためのインフラストラクチャーを構築することである。

パネル調査は、施策効果測定や行政ニーズ把握等に有効であるが、横断調査と異なる独自のデータ管理法と分析法を要する。しかし、21世紀縦断調査は政府統計初のパネル調査であり、行政目的に向けたパネル調査のデータ管理、集計・分析システム等の調査インフラは皆無であった。これに対し本プロジェクトは先行事業において、調査データの管理から統計分析手法の適用までを統合的に実施するシステム開発に取り組み、縦断調査データを用いた分析研究のインフラ構築に努めてきた。本年度の研究事業においては、それらシステムの実用化に向けた統合・総括を図り、調査実施者、一般研究者を使用者と想定した実装を目指した。これらによって、調査実施者に対しては年々蓄積される調査結果の速やかで有効な結果公表と行政的応用に資するとともに、研究者に対しては分析手法ガイドの作成や分析対象の体系化などを行いライフコース研究の学術的基礎の確立に資するものである。

とりわけ 21 世紀縦断調査は、国の重要課題を中心テーマとしながら、対象年齢層の異なる三つの調査の組み合わせによって国民生活をライフコースの視点から体系的に捉える点で諸外国にも例を見ない独自性を備えている。わが国は世界的にみても最低位の出生率と最高位の平均寿命を合わせ持ち、その結果として世界一の高齢化率を擁する。こうした各国の将来を先取りした形の日本社会やその中での国民のライフコ

ースのあり方については国際的にも注目を集めているところである。また、青年層の就業、結婚等を取り巻く状況の急速な変化や極端な「団塊」の世代の存在などの特異性は、独自の課題をもたらしている。こうした状況の深層にある国民生活の実態や意識を時系列的のモニタリングする形式の本調査の行政的、歴史的価値は計り知れない。ただし、テーマ、項目、データ構造等において独自の点を多く持った当該パネルデータの管理・分析は、既存既成の手法だけでは不十分であって、その真価を引き出すためには独自の仕組みが求められる。本研究では統合化した分析システムを構築し、高度統計分析を適用する仕組みを実用化することで、迅速で信頼性の高い結果提供に資することを目指すものである。

## B. 研究方法

本事業では、具体的に5つの分野に分けて作業が進められた。すなわち

- (1) 海外パネル調査情報閲覧・検索システムの開発
  - (2) データ管理・統計分析システム(PDA21)の開発
  - (3) データ特性把握－脱落・移動・回答不詳の分析
  - (4) 分析法ガイドブック作成－縦断調査分析の手引き
  - (5) 各種テーマに関するシード研究－ライフコース的体系化
- である。

(1)においては海外パネル調査情報を網羅し、キーワード検索が可能な日本語の情報ベースを提供しており、今回内容の追加更新に加えてシステムの改善を行った。

(2)においては累積される縦断調査データを管理し、複数の調査回や調査票にまたがった集計や統計分析を効率的に行えるよう支援するシステムの開発が進められた。

SAS, SPSS, Perl, ADAM など様々な統計ソフト等による分析の基礎となるプログラムを自動生成するなどの機能があり、今年度においては全変数をグループ化してそれぞれコード番号を付して管理する方式を導入するなどして分析をより効率的に行えるようにし、またインターフェースをわかりやすくするなど実用面における改良を行った。

(3)については、脱落を始めとする 21 世紀縦断調査のデータ特性のモニタリングを継続したほか、同時期に定期的に実施された横断調査である出生動向基本調査の結果と比較することにより、脱落による偏りの程度を定量的に評価するなど、国際的に見ても例のない試みを行った。

(4)では、今後の 21 世紀縦断調査の分析研究に携わる分析者、研究者のために、縦断調査特有の分析方法論の基礎について解説し、21 世紀縦断調査の例を用いて解説を行った。全 7 章により構成され約 160 ページの解説書としてまとめた者である。各章の内容は以下の通りである。

- 第 1 章 固定効果・ランダム効果モデル
- 第 2 章 ダイナミックパネル分析
- 第 3 章 同時方程式パネルデータ分析
- 第 4 章 生存時間分析
- 第 5 章 離散時間ハザードモデル
- 第 6 章 SURF モデル
- 第 7 章 共分散構造分析
- 第 8 章 傾向スコア・脱落サンプルバイアスの検定法

本ガイドブックは、取り組みの難しいパネルデータの分析法を平易に解説し 21 世紀縦断調査データを対象とした今後の分析研究の推進に資するものと期待される。

(5)の事例研究(シード研究)において、次世代健全育成の領域では、子どもの感染症について、ワクチン全員接種、任意接種による罹患率の違いが明らかにされた。すな

わち、全員接種の風疹、麻疹の罹患率に対して任意接種の水痘、おたふくでは圧倒的に高い罹患率を示し、また保育所での集団感染の可能性や若い父母経由の感染の可能性が示唆され、感染症法、予防接種法の制度設計に資する分析結果が得られた。また、同様に次世代健全育成の観点からは、親が子どもに望む性格が、その後の育児・教育方針、行動に強く反映されていること、同時に子ども自身の生活・学習状況にも影響していることが見いだされ、次世代健全育成に関する施策や教育施策の立案に資する結果が得られている。また、少子化メカニズム解明の領域では、母の就労における祖父母と保育サービスの補完関係、さらにはその追加出生との関係についての複雑な因子の絡み合いを、縦断調査データを用いることでよく整理できることを示した。これらは公的育児支援制度設計に資する結果である。このように政策的応用に直結する分析の基礎的な研究が実施されており、重要な提言となっている。

## C. 研究成果、ならびに D. 考察

### (1) 海外パネル調査情報閲覧・検索システムの開発

本事業では、諸外国のパネル調査に関する情報を日本語によって検索・閲覧するシステムを開発している。これはわが国においてパネル調査の企画あるいは結果の分析を行う際、諸外国における先行事例や比較対象を探索したり、分析法を参照したりする際に用いることを目的としている。

先行事業においては、ミシガン大学の大規模なデータバンクと、小規模であるものの社会科学系の縦断調査等のデータをまとめて所有している、ハーバード大学のデータバンク(Murray Center)に所蔵されている各種パネル調査の概要をベースに、日本語によってウェブサイトを通じて検索、

閲覧できる形式のシステムを開発した。これらの全情報はキーワードによって検索することが可能であり、該当する調査のリストやここの情報を得ることができる。本システムで各調査について抽出できる情報は、調査対象国・地域／調査対象年齢／調査継続期間・調査頻度、調査の目的・サマリー・調査主体、具体的な調査内容・調査項目、母集団・標本抽出法・サンプル数・調査方法・その他関連情報、回収数、脱落等に関する情報、ならびにそれら情報の出典とウェブ上のリンク先である。

本年度においては、直近におけるパネル調査（28 調査）に関するデータを追加し、すでに記載済みの調査（8 調査）について情報更新を行った。

## (2) データ管理・統計分析システム (PDA21)の開発

現在開発中の縦断データ管理・分析支援システム (PDA21) は、複数回、あるいは複数調査票に対応するオリジナルのデータセット群から、分析に必要な変数群を指定することによってそれら当該データが抽出し、必要な変数群のみから群から成る分析用の一時ファイルを自動的に生成し、分析用のソフトウェア（通常市販のパッケージ）やマイクロシミュレーション用のプログラム（現在は C++のみ想定）へと受け渡される。

通常市販パッケージ等で分析プログラムを作成する際には、データセットからプログラム上の変数に対応するデータを読み込み、カテゴリーを表す数値を分析に対して適切な値に変換するなどの作業をプログラムとして記述する必要がある。この工程は大きく分けると(1)分析に必要な変数をデータセットから読み込む、(2)読み込んだ変数を分析用に変換する、(3)分析用変数を用いて集計・分析を行う、という3段階で構成されるが、これを手動で行う場合は、(1)

～(3)に関してすべてプログラムを記述しなければならず、用いる変数の数が多数となると多大な時間と労力を要することが普通である。とりわけ縦断調査の場合、多数の調査票に対応するデータセットからそれぞれ変数を読み込んで(1)、比較可能となるよう変換作業を行う(2)ため、その分、多量の作業が発生し、またその過程でミスも起こり易い。そこで(1)～(3)の工程を可能な限り自動化し、集計・分析までが連続的で円滑な流れとするように支援するシステムが望まれる。

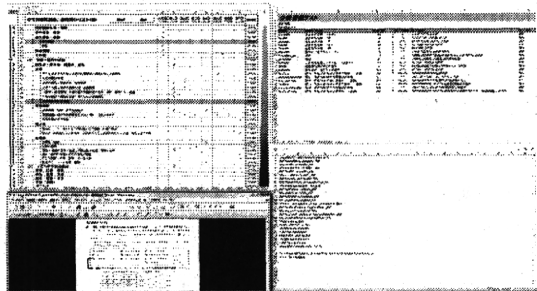
本システムは、縦断調査データのリレーショナルデータベースとしての管理を含め、こうした集計・分析までの工程の支援することを目的としたシステムである。これまで先行事業において、基礎部分から応用部分へと段階的に開発を進めてきたが、本年度においては本格的な実用化を視野に入れ、新しいインターフェースの導入が行われた。具体的には、これまで使用する変数について個別に指定する方法をとっていたのに対し、ユーザーの利便性を一歩進めて、調査項目単位で指定する方式を追加した。図 A にそのインターフェースの例(出生児調査)を示す。たとえば出生児調査において、父母の年齢は多くの分析で用いられる変数であるが、個別の調査における父の年齢、母の年齢として指定する代わりに、「父母に関する項目」を指定することによって、父母の生年月日、年齢、国籍、学歴等の複数の変数が複数の調査票にわたって選択される。これは「家族・保育の状況」というさらに上位にある分類項目を指定することによって、兄弟姉妹や祖父母の情報等をも含む多数の変数がパッケージとして選択されることになる。これら分類の上位から下位にわたる項目の指定は、ワンクリックで行われる。この作業工程は調査原票を参照にしたリ、自動指定された変数が逐次表示され確



認しているなど、支援的な環境下で行われる(図B)。これらの指定の後、次段階の実効を指示すると、システムは自動的に指定された項目に含まれる調査回、変数のデータをリレーショナルデータベースから読み込んで、分析用のデータセットを生成し、またこれに対する集計・分析操作を行う統計ソフトウェアの基本プログラムを同時に生成する。ユーザーはこの基本プログラムに対して、集計形式を変更したり、分析内容を追加するなど自由に編集を行うことによって、どのような目的の集計・分析でも行うことが出来る。

図A

図B

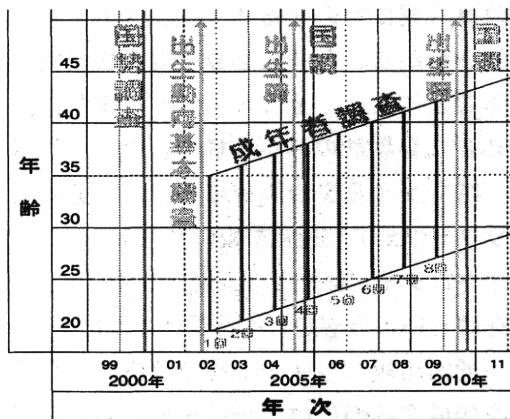


### (3) データ特性把握—脱落・移動・回答不詳の分析

調査データの分析を行い科学的な知見を得ようとする際、調査方法や調査テーマ、設問の設計の仕方などに関わる様々な制約によって、得られたデータには独自の特性が生ずる。これらの特性を正しく考慮しないと、分析結果の正しい解釈は得られない。

たとえば、パネル調査データに共通の特性として、調査回による標本脱落の進行が有る。脱落がランダムではなく、分析テーマに連関する形で生ずる場合には、標本集団は次第に母集団の代表が損なわれて行くことになり、これを考慮しないと分析結果の解釈を誤ることになりかねない。こうした点に鑑みて、本事業では21世紀縦断調査における脱落を始めとするデータ特性のモニタリングを継続して行って来た。本年度もこれを継続して実施すると共に、成年者縦断調査に関して、同時期に実施された全国標本の横断調査である出生動向基本調査における共通の調査項目に関する結果を比較することにより、脱落による偏りの程度を定量的に評価する試みを行った。これは国際的に見ても例の見られない試みである。図Cには成年者縦断調査と出生動向基本調査の実施時期ならびに調査対象年齢をレキシス図上に対比して示している。

図C



成年者縦断調査は2002年11月以降毎年実施されているのに対し、出生動向基本調査は2002年、2005年、2010年の6月に実施されており、本年度の研究においては2005年調査を中心に比較を行い、多項ロジットモデルなどを用いて成年者調査データの特徴を分析した。対象とした項目は(1)出生コーホート、(2)配偶関係、(3)学歴、(4)就業形態、(5)従業者数、(6)職種、(7)親との同別居

である。比較の結果によれば、成年者調査は男女とも1974・76年生まれを中心に中間の年齢層で過少、20代前半コーホートで有配偶者割合が少ない傾向がある。また男女とも在学中が多く、男性で大学・大学院卒が少なく、女性では専門学校が多い。就業形態では派遣・契約社員、女性ではアルバイト・パートなども含めた非正規雇用者が少ない。従業者数では100-999人の企業勤務で少なく、官公庁勤務が多い。また、有配偶、独身とも親と同居していない標本が少ない傾向にあり、とりわけ非同居の独身者が過少であった。

#### (4) 分析法ガイドブック作成—縦断調査分析の手引き

パネル調査(縦断調査)では、対象に生ずるライフコース事象の発生メカニズムや因果関係に対する強力な推論が行える。しかし、その有効性を十分に引き出すためには横断調査とは異なる統計手法が必要となる。本年の事業においては、そうしたパネルデータ特有の分析手法を概説し、入門者から本格的な分析研究を目指す者までを対象とする実践的なガイドブック(『パネルデータ分析ガイド』)の編纂を行った。

わが国では従前においてパネルデータの蓄積が遅れていたが、最近に至って多くのパネル調査が創設され、それらデータに対する分析への関心は高まっている。とりわけ21世紀縦断調査は国が行う初の公的パネル調査であり、国民生活の多様な側面を大規模な標本と複数の調査でカバーする画期的なものである。『パネルデータ分析ガイド』はこの21世紀縦断調査を中心的な題材として、パネルデータ分析に必要な統計手法について基礎から実践までのガイドを試みるものである。

その内容は、

- 第1章 固定効果・ランダム効果モデル
- 第2章 ダイナミックパネル分析
- 第3章 同時方程式パネルデータ分析
- 第4章 生存時間分析
- 第5章 離散時間ハザードモデル
- 第6章 SURF モデル
- 第7章 共分散構造分析
- 第8章 傾向スコア・脱落サンプルバイアスの検定法

の全8章により構成され、約160ページの解説書となっている。また各章は取り組みの難しいパネルデータの分析法を平易に解説しており、21世紀縦断調査データを対象とした今後の分析研究の推進に資するものと期待される。それらの概要については、北村による分担研究報告書を参照のこと。

#### (5) 事例研究とその体系化

本年度の事例研究については、21世紀縦断調査データをライフコースの観点から様々な調査項目の体系化を念頭に、出生児縦断調査データを中心に分析を行った。出産、育児、子どもの育ちに注目し、医療面からは子どもの感染症、規範面からは若年出産および子ども親と育児方針、心理面からは育児不安・負担感、経済面および社会資源面からは子育て費用と保育資源についての分析を行った。具体的には以下のテーマに関する事例研究を実施した。

- (1) 子どもの感染症の罹患に関する分析
- (2) 現代日本の若年出産
- (3) 子育て期の保育支援資源の都市規模・地域移動経験による違い
- (4) 育児不安・負担感分析手法の開発
- (5) 子育て費用と習い事の実態
- (6) 子ども親と育児方針

これらの事例研究の概要については、釜野による分担研究報告書を参照のこと。

## E. 結論

本事業では、三つの 21 世紀縦断調査を中心とするわが国で実施されるパネル調査に対して、分析研究のためのインフラストラクチャーを構築、提供することが主眼である。しかし、そのためには非常に多方面からのアプローチが必要となる。

まず、そのような調査を企画・実施するにせよ、データの分析を行うにせよ、手本となる先行研究の情報が体系的に得られなければならない。これに対しては、パネル調査に関して豊富で長い経験を有する欧米における研究の展開を一望することが出来れば有効であろう。そうしたことから本事業では、まず (1) 海外パネル調査情報閲覧・検索システムの開発を行った。すでに海外に存在する複数の情報ベースをもとに、わが国における利便性を高めるため、日本語による検索・閲覧可能な情報ベースの構築を行った。基本的なシステムはすでに完成しており、今後は情報を更新しつつ公開する予定であり、広くわが国におけるパネル調査の普及と理解に資するものと期待される。

次にパネル調査データの横断調査に無い特徴として、複数回または複数の調査票のデータを同時に分析することが必須だということが挙げられる。これはデータ管理や分析のためのデータ抽出に関して、横断調査よりも複雑で多量な操作を要することを意味する。したがって、これら多数回あるいは大量のデータを効率的に管理、操作するための支援システムが必要となる。それは単なるデータベースではなく、データ管理から高度な統計分析に至るまでを連続的で円滑な流れとするような仕組みが望まれる。これに対して本事業では、(2) データ管理・統計分析システム (PDA21) の開発を行った。これは年々蓄積される調査票情報 (調査データ) をリレーショナルデータベース

によって管理を行い、追加的に更新してゆくことを可能としながらも、それらの全累積データを有機的に融合し、あたかも一つのデータであるかのように扱うことを可能にするシステムである。具体的には、調査項目 (あるいは調査テーマ) を指定することによって、関連する調査回の関連する変数群が自動的に抽出され、分析のための一時的なデータセットが生成されるというものである。この際、データセットの生成と同時に同データセットの変数を市販の (あるいは既存の) 統計ソフトウェアに読み込むプログラムを同時に生成することによって、後は集計プログラムを修正したり、市販統計ソフトウェアによる高度な統計分析を指定するだけで、その分析結果を得ることが出来る。現状では用意されている市販統計ソフトウェアとしては、SAS、SPSS があり、クロス集計ソフトウェアとしては ADAM、また汎用ソフトウェアとしては Perl などがある。また、C++ によるマイクロシミュレーションプログラムに対しても出力が可能である。また、このシステムに登録が必要な各回、各変数の情報はすべて 21 世紀調査の担当部局である統計情報部縦断調査室が結果公表のために作成する情報セットであるため、追加的な労力投資は最小限に抑えられることも本システムの特徴のひとつである。基本システムはすでに完成しており、本年度のインターフェース部分の改良によりさらに利便性が増したことから、今後は実用によって 21 世紀縦断調査データの分析の中心的ツールとして効率化に役立つとともに、細部についての改良を加えることによって有用性を増し、引いては一般のパネル調査においても有効に活用されることが期待される。

さて、ひとたび調査データ準備が整い、集計や統計分析のための基本プログラムが自動生成された後は、実際にプログラムを



実行し、集計結果や分析結果を得ることが出来る。しかしながら、現実の調査データには調査方法や調査テーマ、設問の設計の仕方などに関わる様々な制約によって、理論的統計モデルが想定するような状態には無いことが多く、手にした集計・分析結果をそのまま解釈できることは少ない。具体例としては、パネル調査データに共通の特性として、観察対象の不均衡な脱落による標本の母集団代表性の喪失や回答不詳による情報の歪みの存在が挙げられる。これら攪乱要因が存在する下では、分析結果を単純に解釈することは誤った結論を導くおそれがある。場合によっては、脱落などを前提とした特殊な分析手法が必要となる。したがって、こうした歪みが存在するか否かといったデータ特性に関する情報は、科学的な分析には欠くことができない。こうした点を鑑みて、本事業では 21 世紀縦断調査における脱落・移動・回答不詳をはじめとする (3) データ特性把握のモニタリングを継続して行って来た。本年度もこれを継続して実施しており、とくに調査が捉える就業、結婚、転居などのライフコース事象との関係について分析を行った。また、本年度のこの分野における特筆すべき試みとして、成年者縦断調査に関して、同時期に実施された全国標本の横断調査である出生動向基本調査における共通の調査項目の回答結果を比較して、その脱落・回答不詳等による偏りの定量的評価を行った。こうした試みは、調査項目について比較可能な縦断調査と横断調査が同時に実施されていることが前提となり、そうしたケースはきわめて少ないことから実際に行われることは非常に珍しく、この種の調査では国際的にも例を見ない。その結果をみると、成年者縦断調査では第 4 回までに脱落した標本は、70 年代半ば生まれ、20 代前半で結婚した男女、大卒男子、非正規雇用者、従業者数

中程度の企業勤務者、親と同居していない者が相対的に多く含まれていることがわかった。また学歴に関しては在学生在が脱落しやすいことがわかっているが、そうした標本はその後大卒者となっているはずであり、分析時点での学歴構成は本来と異なっているという認識が必要である。逆に脱落があっても構成に影響の少ない項目も多い。同調査では第 4 回までに標本数は 6 割程度に減少しているが、職業などでは脱落の影響がほとんど検出されない。これらの知見は、様々なライフコース事象の発生メカニズムと要因の統計的分析ならびに分析結果の解釈に際して、十分理解しておく必要があり、場合によっては分析手法の選択を替える必要があるだろう。

それではパネル調査データに対して、どのような分析手法を採用すべきだろうか？縦断調査には横断調査とは異なる分析手法が必要となることはわかるが、これまでパネルデータに乏しかったわが国では、その分析方法について、必ずしも知識が普及しているとは言えない。本事業では、こうした点を補う必要性の認識に立って、本年度において (4) 分析法ガイドブック作成－縦断調査分析の手引きの編纂を行った。これは今後 21 世紀縦断調査をはじめとするパネルデータの分析研究に携わる分析者、研究者のために、縦断調査特有の分析方法論の基礎について解説し、21 世紀縦断調査の例を用いて解説を行ったものである。それは全 7 章にわたり、主立った統計手法を取り上げ、21 世紀縦断調査データを中心的な題材として取り上げている。本書は 21 世紀縦断調査等のデータ分析の実践的な手引きとして広く活用されることが望まれ、ひいては研究者の層を大いに広げることが期待される。

以上の事業によって 21 世紀縦断調査をはじめとするパネル調査データ分析を支援

する多面的なインフラが一通り構築されたことになる。しかし、それらを実践的なものとして用いるためには、実際の応用に付すことが必要である。現実の分析過程においては、当初簡単には想定できないような課題が多く存在する者である。そこで本事業では、21世紀縦断調査の基本的なテーマについて実際に基礎的な分析を行い、そうした課題について検討することとした。それらは本事業における分野としては、(5)各種テーマに関するシード研究－ライフコースの体系化であり、個々の取り組みは事例研究と呼んでいる。本年度は6個の研究が行われた。それぞれの研究概要については、後述の釜野による分担研究報告書を参照されたい。しかし、それら事例研究の役割は、単に分析上の課題を発見することに止まらない。それらは各種テーマの本格的分析研究に取り組む際の基礎的知見を提供し、また有効な方向性を示すいわばシード研究としての役割を担っている。とりわけそれらのテーマは、当該調査が捉える国民のライフコースを体系的に捉えるように配置されており、今後行政において重要になると考えられるライフコース縦断的な施策の構想・立案ないし評価について、実証データによる科学研究との連携の在り方の実例を示すことが意図されている(巻末資料「国民のライフコース事象、21世紀縦断調査、高度統計分析研究、ならびに関連施策に関する相関図」を参照)。ただし、それらは本事業におけるデータ二次利用上の法制的制約により、学術研究成果として公表することは行わず、これらを発展させた個々の本格的分析研究は本事業のスコープを超えたものとして位置づけられる。それらは今後、本事業の後継事業等において、テーマごとに本格的な研究へと発展することが期待される。

以上を総括すると、分担研究報告を含め

た本事業内のすべての研究分野の課題について、本年度において一定の到達点に至ったといえる。今後は本事業成果の活用により、年々蓄積される21世紀縦断調査に対し、パネルデータ特有のデータ管理ならびに統計分析が効率的になされ、速やかな結果公表、高度な分析研究が展開されることが期待される。また、縦断・横断両調査の比較・連携など新しいアプローチの開発により、学術的観点からも、方法論、分析結果の双方において国際的に価値の高い貢献が期待される。

#### F. 研究発表

本事業における研究インフラのシステム開発に際しては、当該調査の調査項目全般に渡る基本的な分析工程の確認が必要となるため、実際に多くの事例的な分析を行っているが、データの二次利用に関する法制的な制約からこれらは個別の学術研究結果としては発表しないこととしている。ただし、これらの多くは個別テーマ分析の具体的な課題や注意点に関する情報を提供し、また今後本格的な研究へと発展しうるものであり、さらには調査項目全体の分析の体系化に資するものであるため、本事業ではシード研究と位置づけて報告書に含めている。

#### G. 知的所有件の取得状況

なし

厚生労働科学研究費補助金 政策科学総合研究事業（統計情報総合研究事業）  
分担研究報告書

パネル調査(縦断調査)に関する統合的分析システムの応用研究  
(平成 22 年度)

研究分担者 北村行伸 一橋大学経済研究所

研究要旨

本分担研究では、パネルデータの分析方法について検討し、厚生労働省統計情報部で実施されている 21 世紀縦断調査の分析を行う上で有用と考えられる、いくつかの分析手法の理論とその応用例を 160 ページにわたる「パネルデータ分析法ガイド」としてまとめた。本ガイドは、今後パネルデータ分析を手がける人々による使用も念頭において作成したものである。具体的には、第 1 章固定効果・ランダム効果モデル、第 2 章ダイナミックパネル分析、第 3 章同時方程式パネルデータ分析、第 4 章 生存時間分析、第 5 章 離散時間ハザードモデル、第 6 章 SURF モデル、第 7 章 共分散構造分析、第 8 章 傾向スコア・脱落サンプルバイアスの検定法から成る。

本研究により、21 世紀縦断調査に即した分析手法について、データ作成者およびデータ利用者に対してわかりやすくかつ専門性を失わない形での普及が可能となった。

A. 研究目的

本研究の目的は、厚生労働省による 21 世紀縦断調査の 3 調査（成年者縦断調査、出生児縦断調査、中高年者調査）の分析に適した分析手法に関して、理論面、実用面の双方を重視しながら、わかりやすく説明する方法を検討し、それを実践することである。

B. 研究方法

まず当縦断調査に適した方法を挙げ、それぞれについて、その手法に長けている専門家に協力を依頼し、適切な文献を吟味し、それらを参考にしながら、理論的な説明を作成した。次に、実際に手法を用いることを念頭に置き、R や STATA などのコードを実際に試しながら検討を行った。また、

可能な限り、21 世紀縦断調査を実際に分析した例を示した。

C. 研究成果

ガイド作成チームで検討を重ねた結果、以下のような内容を含めることにした。

第 1 章固定効果・ランダム効果モデル（1.1 通常の線形回帰モデル 1.2 回帰モデルの残差 1.3 パネルデータの表示法 1.4 固定効果モデル 1.5 ランダム効果モデル 1.6 具体的な分析例 1.7 出生児縦断調査への応用例）

第 2 章ダイナミックパネル分析（2.1 はじめに 2.2 ダイナミック・パネルデータの考え方 2.3 最尤法推定と操作変数法推定 2.4 一般化積率法推定 2.5 STATA コード）



第3章 同時方程式パネルデータ分析 (3.1 はじめに 3.2 同時方程式パネルデータ分析の考え方 3.3 単一方程式推定 3.4 内生性検定 3.5 不均一分散検定 3.6 弱相関の操作変数の問題 3.7 STATA コード)

第4章 生存時間分析 (4.1 生存時間分析の基本量 4.2 カプラン・マイヤ推定量 4.3 Cox 比例ハザードモデル 4.4 出生児縦断調査への応用例)

第5章 離散時間ハザードモデル (5.1 イベントヒストリー分析の概要 5.2 離散時間モデルの概要 5.3 人・期間別データの作成方法 5.4 離散時間ロジットモデルの分析プログラムと出力例 5.5 成年者縦断調査への適用例 (結婚))

第6章 SURF モデル (6.1 はじめに 6.2 離散時間ロジットモデルにおける競合イベントの取り扱い 6.3 SURF モデルの概要 6.4 2段階推定による SURF モデルの適用手順 6.5 2段階推定による SURF モデルの利用における留意点 6.6 2段階推定による SURF モデルのプログラムと出力例 6.7 成年者縦断調査への適用例)

第7章 共分散構造分析 (7.1 共分散構造分析 7.2 縦断調査を用いた場合の共分散構造分析モデル 7.3 共分散分析の問題点とモデルの適合度評価 7.4 R による分析例—自己回帰モデルによる共分散構造分析—)

第8章 傾向スコア・脱落サンプルバイアスの検定法 (8.1 はじめに 8.2 傾向スコアおよび IPW 推定量 8.3 脱落発生が初婚発生へ及ぼす因果効果 8.4 おわりに)

#### D. 考察

ガイドの作成を通じて、21世紀縦断調査のデータ分析の際の手法が多様化することで、その活用可能性が広がり、その結果として、得られる知見の幅も広がることが予想される。

#### E. 結論

本研究において作成したガイドを参考に、研究者に限らず、さまざまな立場の利用者によって、実際に研究が行われることで、本研究の成果が試され、またその意義が高まる。

#### F. 研究発表

本事業における研究インフラのシステム開発に際しては、当該調査の調査項目全般に渡る基本的な分析工程の確認が必要となるため、実際に多くの事例的な分析を行っているが、データの二次利用に関する法制的な制約からこれらは個別の学術研究結果としては発表しないこととしている。ただし、これらの多くは個別テーマ分析の具体的な課題や注意点に関する情報を提供し、また今後本格的な研究へと発展しうるものであり、さらには調査項目全体の分析の体系化に資するものであるため、本事業ではシード研究と位置づけて報告書に含めている。

#### G. 知的所有権の取得状況

なし

厚生労働科学研究費補助金 政策科学総合研究事業（統計情報総合研究事業）  
分担研究報告書

パネル調査(縦断調査)に関する統合的分析システムの応用研究  
(平成 22 年度)

研究分担者 釜野さおり 国立社会保障・人口問題研究所

研究要旨

本分担研究では、過去の関連事業で行ってきた 21 世紀縦断調査のデータをライフコースの観点から体系的に分析するシステムの構築を引き継ぎ、そのシステムの応用を念頭に、出生児縦断調査データを用いていくつかの側面からの分析を行った。出産、育児、子どもの育ちに注目し、医療面からは子どもの感染症、規範面からは若年出産および子ども観と育児方針、心理面からは育児不安・負担感、経済面および社会資源面からは子育て費用と保育資源についての分析を行った。

出生児縦断調査の協力者の子どもは本研究で分析に含めた第 7 回ではすでに就学年齢に達している。本研究を通して、若年出産であったか否か、出産時の父母の年齢、住居の移動の有無、子どもが小さいときの子ども観、など、それまでの行動・環境が、現在の子どもの状態を規定していることが多くのテーマで確認された。経年でデータをとり続けることの意義が、明らかになってきているということである。分析システムを応用して、過去からの関係性を見いだす分析を進めることで、実施回数が少なかったときには得られなかった貴重な知見が蓄積しつつある。今後のさらなる継続による貢献が期待される。

A. 研究目的

本研究の目的は、前年度までの関連プロジェクトにおいて開発されてきた 21 世紀縦断調査の分析システムを応用する一貫としてライフコースの観点から出産、子育て、子どもの育ちを捉える一連の研究を、21 世紀出生児調査のデータを用いて行うことである。分析した内容は以下のとおりである。

- (1)：子どもの感染症の罹患に関する分析 [井出]
- (2)：現代日本の若年出産 [相馬]
- (3)：子育て期の保育支援資源の都市規模・地域移動経験による違い [西野]
- (4)：育児不安・負担感に関する分析手法の開発へ向けて [相馬]

(5)：子育て費用と習い事の実態 [相馬]

(6)：子ども観と育児方針—主として「第 7 回出生児調査」の分析より— [元森]

それぞれの研究目的を下記に述べる。

(1)医学的な視点から、ワクチン接種（強制的に）が取り入れられていないことと、患者数の関係は未だ明らかにされていないことを踏まえ、ワクチンで予防可能な感染症の罹患について明らかにすること、届出精度（感染症法で主要な感染症毎に届出制度が定められている）について参考となる情報をみいだすこと、子どもの生活環境以外の因子が罹患リスクを高めているかを確認するため社会経済的因子等と罹患の関係について分析することを目的とする。

(2)日本では若年出産に着目した研究が近年いくつかなされているが、パネルデータを用いたその実態把握はなされているとはいいがたい。地方自治体の子育て支援の現場でも、若年出産に着目した支援の必要性が提起されており、国際的にも若年出産(teenage pregnancy)の問題は、重要な社会政策の課題となっている。そこで若年出産に着目し、社会経済状況や子育ての実態とその特徴を把握することを目的とする。

(3)第7回までのデータを用いて、都市規模・地域移動類型別に、母親が得られる保育支援や子育てへの感じ方を描く。そのことを通じて、都市規模による母親の就業率の差を、各親が得られる保育支援の資源の差に分解して読み取ることを目的とする。大都市部の子育て期の母親の有職率が相対的に低いことの一つの説明として、母親が就労するにあたり、日中子供を見てくれる支援者の確保しやすさが影響する、主な支援者は祖父母と保育所等のサービス、祖父母という資源の調達可能性は、同居・近居しているかに左右される、とすると、同居・近居の少ない大都市では、祖父母の支援は得にくいことが予想される。祖父母の支援が得られなければ保育サービスへのニーズが高まるだろうが、供給量が追い付かなければ、有職率自体が抑制されるだろう。このような予想されるメカニズムを、データによって裏付けていくことを目指す。

(4)少子化の要因として、子育ての経済的負担が挙げられ、政策手段として、児童手当制度における乳幼児加算や中間層への対象拡大が行われてきた。そこでは就学前の誰の経済的負担や負担感に焦点を当てることが問題となる。支援拡充を考えるうえでも、どういう特徴をもつ層の家庭が経済的負担感を持っているか、実際に家計が苦しいのかについての実態把握が基礎となる。ここでは子育て費用(保育料、習い事)の

実態を把握し、経済的支援の対象について考察することを目的とする。

(5)母親が育児に関して、不安・負担・ストレス等を抱えていることは広く社会問題となってきた。既存調査からも、「専業主婦の方が就業女性よりも育児不安が高い」と指摘され、とくに専業主婦の育児負担・不安感の緩和、在宅子育て支援が政策課題となってきた。これまでは、一時点の育児不安感や負担感の度合いの測定をもとに議論されてきた。パネル調査を用いると、子どもが成長するにつれての変化を読み解くことができる。母親の育児不安・負担感の変化の特徴を明らかにすること、ならびに、育児不安や負担感に関する分析手法の検討を行う。

(6)親の子ども親が親の学校や学習への関わり方および子どもの側の生活や学習の様子にどう影響しているか明らかにする。まず過去に未就学時点の第6回までを分析した内容を要約する。その上で、小学1年生時の調査となる第7回調査から、親の子ども親によって、就学後の教育方針や子ども自身の生活や学習行動「子どもについての悩み」に差異があるかを検討する。

## B. 研究方法

(1)感染症による医療機関の受診状況が尋ねられている第3回～第7回の個票から、水ぼうそう、百日ぜき、風しん、はしか、おたふく、川崎病を取り上げ、罹患状況を分析した。定期接種となっているのは、風しんとはしかであり、水ぼうそうとおたふくは任意接種となっている。全サンプル47,015人から同一の感染症について複数回の罹患を報告しているサンプルを除き、報告回数から観測された人年数(person years)を求めた。第三回から第七回調査における罹患者数を求めると共に、罹患者数が多い水ぼうそうとおたふくについて、性別、親の国籍、出生時の父母の年齢(10代、