

Ⅲ. データベース定義

1-1. データベース名

No.	データベース名	定義名	保存先	備考
1	パネル調査データベース	paneldb	(PostgreSQL インストール先) ¥8.0¥data¥base¥ (OID)	OID はシステムで自動でつけられる。

1-2. データベース属性

No.	定義名	オーナー	エンコーディング	Template	テーブル空間	備考
1	paneldb	nipssr	EUC_JP	Template1	pg_default	

1-3. グループ

No.	グループ名	グループID	役割	メンバー	備考
1	admins	100	スーパーユーザーグループ	postgres	
2	users	101	一般ユーザーグループ	nipssr	

1-4. ユーザー

No.	グループ名	ユーザーID	グループ	役割	ユーザー権限		備考
					DB作成	ユーザー作成権限	
1	postgres	1	admins	スーパーユーザー	○	○	
2	Nipssr	100	users	一般ユーザー			

1-5. E-R図

2. テーブル/ビュー一覧表

No.	テーブル/ビュー名		所有者	行長	行数	列数	概要
	名称	定義名					
1	調査票一覧テーブル	survey_list	nipssr			5	
2	調査票情報テーブル	ct_ (調査票番号)	nipssr			8	例) 調査票番号が「af01」の場合、テーブル名は「ct_af01」。
3	コードマスタテーブル	ms_ (調査票番号)	nipssr			4	例) 調査票番号が「af01」の場合、テーブル名は「ms_af01」。
4	出生児調査テーブル	bdt_ (調査票番号)	nipssr			※調査票により異なる	例) 調査票番号が「bk01」の場合、テーブル名は「bdt_bk01」。
5	成年者調査テーブル	adt_ (調査票番号)	nipssr			※調査票により異なる	例) 調査票番号が「af01」の場合、テーブル名は「adt_af01」。
6	夫婦調査テーブル						

※行長・行数は調査票により異なる。

3-1. 調査票一覧テーブル

(1) テーブルの概要

テーブル名称：調査票一覧テーブル

テーブル定義名：survey_list

調査票一覧テーブルにはデータベースに格納されている調査票の一覧情報を格納する。
調査票が追加されたら調査票番号・調査名などを追加する。

(2) データ項目定義

No.	データ項目名		データタイプ	NULL値	主キー	繰り返し回数	備考
	名称	定義名					
1	連番	list_id	integer	No			
2	調査票番号	survey_id	varchar(20)	No	○		
3	調査票名	survey_name	varchar(100)	No			
4	調査種類	survey_type	varchar(1)	No			a：成年者調査 b：出生児調査
5	調査データテーブル名	dt_table	varchar(30)				

(3) インデックス一覧

No.	インデックス名	所有者	インデックスタイプ	アクセス方法	列リスト
1	uix_ (テーブル定義名)		unique/No	btree	survey_id
2	ix_		dupls/No	btree	

(3) アクセス権一覧

No.	ユーザ名	選択	更新	挿入	削除	インデックス	変更	備考
1	users	Yes	Yes	Yes	Yes	No	No	
2	admins	Yes	Yes	Yes	Yes	Yes	Yes	

3-2. 調査票情報テーブル

(1) テーブルの概要

テーブル名称：調査票情報テーブル

テーブル定義名：ct_ (調査票番号)

例) 調査票番号が「AF01」の場合、「ct_AF01」

調査票情報テーブルには各調査票のKEY項目および調査項目の情報（変数名・項目名・幅など）を格納する。

調査票情報テーブルは調査票ごとに作成される。

(2) データ項目定義

No.	データ項目名		データタイプ	NULL値	主キー	繰返し回数	備考
	名称	定義名					
1	調査票番号	survey_id	varchar(20)	No	○		
2	変数名	item_id	varchar(20)	No	○		
3	項目名	item_name	text	No			
4	位置	it_position	integer	No			
5	幅	length	integer	No			
6	コード範囲	range	varchar(50)				
7	内容	content	text				
8	変換情報	chg_info	varchar(1)				数値として取り扱いたい項目の変換のため。 c : 文字、n : 整数

(3) インデックス一覧

No.	インデックス名	所有者	インデックスタイプ	アクセス方法	列リスト
1	uix_ (テーブル定義名)		unique/No	btree	survey_id, item_id
2	ix_		dupls/No	btree	

(3) アクセス権一覧

No.	ユーザ名	選択	更新	挿入	削除	インデックス	変更	備考
1	users	Yes	Yes	Yes	Yes	No	No	
2	admins	Yes	Yes	Yes	Yes	Yes	Yes	

3-3. コードマスタテーブル

(1) テーブルの概要

テーブル名称：コードマスタテーブル

テーブル定義名：ms_ (調査票番号)

例) 調査票番号が「AF01」の場合、「ms_AF01」

コードマスタテーブルには各調査票の各調査項目のコード情報（変数名・コード・カテゴリなど）を格納する。

コードマスタテーブルは調査票ごとに作成される。

(2) データ項目定義

No.	データ項目名		データタイプ	NULL 値	主キー	繰返し回数	備考
	名称	定義名					
1	調査票番号	survey_id	varchar(20)	No	○		
2	変数名	item_id	varchar(20)	No	○		
3	コード	item_code	varchar(20)	No	○		
4	カテゴリ	category	text	No			

(3) インデックス一覧

No.	インデックス名	所有者	インデックスタイプ	アクセス方法	列リスト
1	uix_ (テーブル定義名)		unique/No	btree	survey_id, item_id, item_code
2	ix_		dupls/No	btree	

(3) アクセス権一覧

No.	ユーザ名	選択	更新	挿入	削除	インデックス	変更	備考
1	users	Yes	Yes	Yes	Yes	No	No	
2	admins	Yes	Yes	Yes	Yes	Yes	Yes	

3-4. 出生児調査テーブル

(1) テーブルの概要

テーブル名称：出生児調査テーブル

テーブル定義名：bdt_ (調査票番号)

例) 調査票番号が「bk01」の場合、「bdt_bk01」

出生児調査テーブルには出生児各調査票の調査データを格納する。

テーブル定義は調査票ごとに異なり、各調査票の変数記述表に基づき定義される。

(2) テーブル作成後のデータ登録時の注意事項

複数調査間でのデータ抽出をできるようにするため、各出生児調査のレコード数を合わせる必要がある。

新しい出生児調査テーブル作成した後、以下の手順でデータ登録を行なう。

① キー項目の全件登録

新しいテーブルに1回目の出生児調査データのキー項目を全件登録する。

例) 登録SQL文 1回目の出生児調査

② 登録対象レコードの削除

新しく登録する出生児調査データで①で登録したキー項目の一致するレコードを削除する。

パネル調査データ作成ツール（出生児版）で作成したDELETE文（SQLファイル）を実行する。 例) 第2回出生児縦断調査の場合 del_bdt_bk02.sql

③ 調査データの取り込み

パネル調査データ作成ツール（出生児版）で作成した調査票データ CSV ファイルを取り込む。

※詳細については操作マニュアル資料を参照。

(3) データ項目定義

No.	データ項目名		データタイプ	NULL 値	主キー	繰返し 回数	備考
	名称	定義名					
1	調査年	(変数記述表の変数名)	varchar(2)	No			キー項目
2	調査番号	(変数記述表の変数名)	varchar(2)	No			キー項目
3	認識番号	(変数記述表の変数名)	varchar(6)	No			キー項目
4	調査項目	(変数記述表の変数名)	(変数記述表の 各変数の幅)				調査票により項目数は異なる。

(4) インデックス一覧

No.	インデックス名	所有者	インデックスタイプ	アクセス 方法	列リスト
1	uix_ (テーブル定義名)		unique/No	btree	survey_id, 調査番号, 認識番号
2	ix_		dupls/No	btree	

(5) アクセス権一覧

No.	ユーザ名	選択	更新	挿入	削除	インデックス	変更	備考
1	users	Yes	Yes	Yes	Yes	No	No	
2	admins	Yes	Yes	Yes	Yes	Yes	Yes	

3-5. 成年者調査テーブル

(1) テーブルの概要

テーブル名称：成年者調査テーブル

テーブル定義名：adt_ (調査票番号)

例) 調査票番号が「AF01」の場合、「adt_AF01」

成年者調査テーブルには成年者各調査票の調査データを格納する。

成年者調査には女性票・男性票・配偶者女性票・配偶者男性票の4種類がある。

テーブル定義は調査票ごとに異なり、各調査票の変数記述表に基づき定義される。

(2) データ項目定義

No.	データ項目名		データタイプ	NULL値	主キー	繰り返し回数	備考
	名称	定義名					
1	調査票番号	(変数記述表の変数名)	varchar(1)	No			キー項目 1:女性票 2:男性票 3:配偶者女性 4:配偶者男性
2	KEY番号(世帯情報) 地区番号	(変数記述表の変数名)	varchar(5)	No			キー項目
3	KEY番号(世帯情報) 単位区番号	(変数記述表の変数名)	varchar(2)	No			キー項目
4	KEY番号(世帯情報) 世帯番号	(変数記述表の変数名)	varchar(2)	No			キー項目
5	KEY番号(世帯情報) 該当者番号	(変数記述表の変数名)	varchar(1)	No			キー項目
6	KEY番号(世帯情報) 配偶者番号	(変数記述表の変数名)	varchar(1)	No			キー項目
7	調査項目	(変数記述表の変数名)	(変数記述表の 各変数の幅)				調査票により項目数は異なる。

(3) インデックス一覧

No.	インデックス名	所有者	インデックスタイプ	アクセス方法	列リスト
1	uix_ (テーブル定義名)		unique/No	btree	調査票番号、KEY番号(世帯情報) 地区番号、KEY番号(世帯情報) 単位区番号、KEY番号(世帯情報) 世帯番号、KEY番号(世帯情報) 該当者番号、KEY番号(世帯情報) 配偶者番号
2	ix_		dupls/No	btree	

(3) アクセス権一覧

No.	グループ名	選択	更新	挿入	削除	インデックス	変更	備考
1	users	Yes	Yes	Yes	Yes	No	No	
2	admins	Yes	Yes	Yes	Yes	Yes	Yes	

IV. ファイル定義

1-1. ファイル一覧表

No.	ファイル名		種別	格納場所	概要
	名称	定義名			
1	抽出結果 CSV ファイル	out_(日付 YYYYMMDD).csv		(データ抽出ツール 実行フォルダ)	例) 2005年12月16日の場合 「out_20051216.csv」
2	抽出条件コントロールファイル	panel_ctl.txt		(データ抽出ツール 実行フォルダ)	※常に上書きされるので注意
3	テーブル作成 SQL ファイル	crtbl_(調査票番号).sql		(データ作成ツール 実行フォルダ)	例) 調査票番号が「af01」の場合 「crtbl_af01.sql」
4	調査票情報データ	ct_(調査票番号).csv		(データ作成ツール 実行フォルダ)	例) 調査票番号が「AF01」の場合 「ct_AF01.csv」
5	コードマスタデータ	ms_(調査票番号).csv		(データ作成ツール 実行フォルダ)	例) 調査票番号が「AF01」の場合 「ms_AF01.csv」
6	出生児調査	bdt_(調査票番号).csv		(データ作成ツール 実行フォルダ)	例) 調査票番号が「bk01」の場合 「bdt_bk01.csv」
7	成年者調査	adt_(調査票番号).csv		(データ作成ツール 実行フォルダ)	例) 調査票番号が「AF01」の場合 「adt_AF01.csv」

付属資料

```

/*****/
/****| データ作成用SQLプログラム |****/
/*****/

```

/* 出生児票 */

```

INSERT INTO bdt_bk03_1 (a001, a002, a003)
SELECT a001, a002, a003
FROM bdt_bk03_0;

```

/* 女性票 */

```

INSERT INTO adt_af02 (af001, af002, af003, af004, af005, af006)
SELECT af001, af002, af003, af004, af005, af006
FROM adt_af01;

```

/* 男性票 */

```

INSERT INTO adt_am02 (am001, am002, am003, am004, am005, am006)
SELECT am001, am002, am003, am004, am005, am006
FROM adt_am01;

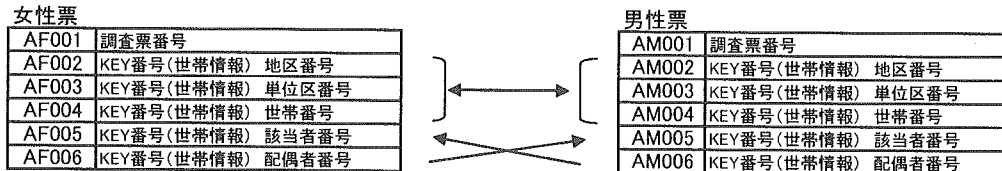
```

■ 女性票、男性票、および配偶者(男性・女性)票からの夫婦データ抽出の考え方について

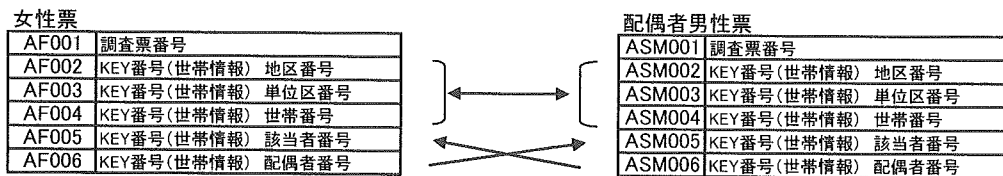
● 夫婦データについて

1. 夫婦データの組み合わせ
夫婦の組み合わせは以下の3種類しかない。

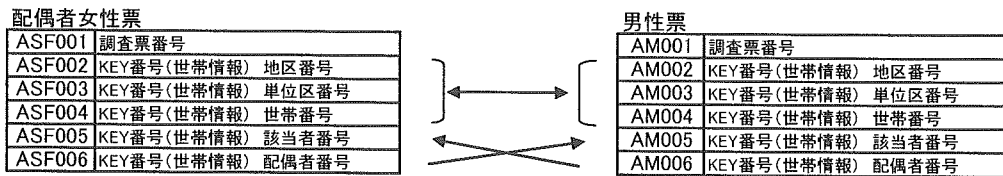
(1) 女性票と男性票



(2) 女性票と配偶者男性票



(3) 男性票と配偶者女性票



2. 夫婦データの推移

調査回が重なるにつれて調査票データは下記のように変化する。

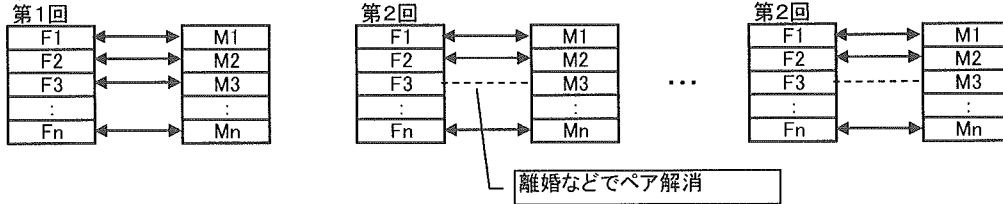
女性票・男性票

第1回	問1	問2	第2回	問1	問2	...	第n回	問1	問2
F1	1	1	F1	1	1		F1	1	1
F2	2	2	F2	2	2		F2	-	-
F3	3	3	F3	-	-		F3	-	-
⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮
F _n	n	n	F _n	n	n		F _n	n	n

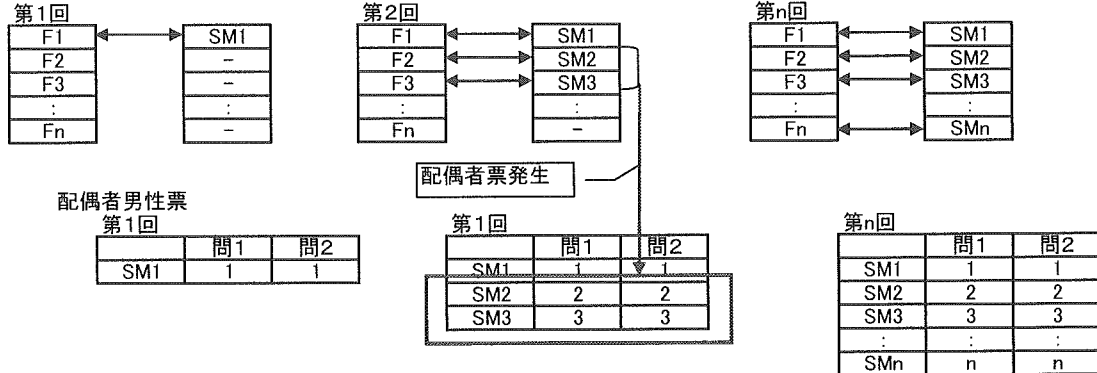
調査対象の人数は変わらない。← 今回のDB化により
回を追うごとに欠落していく。
増えることはない。

夫婦データ

(1) 女性票と男性票



(2) 女性票と配偶者男性票



3. あらかじめ用意するデータ

夫婦データの抽出を行なうにあたって、あらかじめ用意しておくデータは以下の通り。

(1) 調査票セット一覧

回	女性票	男性票	配偶者女性票	配偶者男性票
1	AF01	AM01	ASF01	ASM01
2	AF02	AM02	ASF02	ASM02
3	AF03	AM03	ASF03	ASM03
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

調査票一覧(作成済み)

No.	調査票番号	調査票名
1	AF01	第1回成年者縦断調査(女性票)
2	AM01	第1回成年者縦断調査(男性票)
3	ASF01	第1回成年者縦断調査(配偶者女性票)
⋮	⋮	⋮
⋮	⋮	⋮

(2) 女性票ベース夫婦データ管理テーブル

	世帯キー 002~005	第1回		第2回		第3回		⋮	第n回	
		配偶者 番号	調査票 番号	配偶者 番号	調査票 番号	配偶者 番号	調査票 番号		配偶者 番号	調査票 番号
FAさん	1001101011	1	AM01	1	AM02	1	AM03	⋮	1	AM0n
FBさん	1001101012	2	ASM01	2	ASM02			⋮		
FCさん	1001101013							⋮		
FDさん	1001101014	4	AM01	4	AM02	4	AM03	⋮	4	AM0n
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

調査回が増えるたびに列を増やしていく ---> テーブル再作成・データ再作成要

(3) 男性票ベース夫婦データ管理テーブル

	世帯キー 002~005	第1回		第2回		第3回		⋮	第n回	
		配偶者 番号	調査票 番号	配偶者 番号	調査票 番号	配偶者 番号	調査票 番号		配偶者 番号	調査票 番号
MAさん	1001101011	1	AF01	1	AF02	1	AF03	⋮	1	AF0n
MBさん	1001101012	2	ASF01	2	ASF02			⋮		
MCさん	1001101013							⋮		
MDさん	1001101014	4	AF01	4	AF02	4	AF03	⋮	4	AF0n
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

調査回が増えるたびに列を増やしていく ---> テーブル再作成・データ再作成要

4. データ抽出の考え方

◆前提条件

- (1) 夫婦の組み合わせは以下の3種類しかない。
 - ① 女性票対象者と男性票対象者
 - ② 女性票対象者と配偶者男性票対象者
 - ③ 男性票対象者と配偶者女性票対象者
- (2) 各回の成年者調査データがDB化されていること。
- (3) 前述の調査票セット一覧・女性票ベースの夫婦データ管理テーブルが作成されていること。

◆データ抽出手順

- (1) ベースとなる男性票あるいは女性票の抽出項目指定(変数名指定)

例) 女性票ベース

調査票番号	変数名
AF01	AF010
	AF020
AF02	AF030
	AF040

- (2) ペアとなる調査票の抽出項目指定(変数名指定)

例) 女性票ベースの場合、男性票あるいは配偶者男性票の項目指定(変数名指定)

調査票番号	変数名
AM01	AM015
	AM025
ASM02	ASM035
	ASM045

- (3) データ抽出処理実行

以下の抽出処理でデータを抽出する。

- ① (2)の指定条件および調査票セット一覧から抽出対象調査回を検索

例) AM01とASM02が指定されているので

- 調査票一覧から AM01 - 第1回の調査票セットに属する。
ASM02 - 第2回の調査票セットに属する。

回	女性票	男性票	配偶者女性票	配偶者男性票
1	AF01	AM01	ASF01	ASM01
2	AF02	AM02	ASF02	ASM02
3	AF03	AM03	ASF03	ASM03
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

水色が対象調査票番号

- ② 夫婦データ管理テーブルから抽出対象となる夫婦キーを検索

例) 夫婦管理テーブルの第1回調査票番号が"AM01"の条件で検索
夫婦管理テーブルの第2回調査票番号が"ASM02"の条件で検索

世帯キー 002~005	第1回		第2回		第3回	
	配偶者 番号	調査票 番号	配偶者 番号	調査票 番号	配偶者 番号	調査票 番号	
FAさん	1	AM01	1	AM02	1	AM03
FBさん	2	ASM01	2	ASM02		
FCさん						
FDさん	4	AM01	4	AM02	4	AM03
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮

水色のレコードが抽出される。

③②で検索された世帯キーでベースとなる女性票から指定項目を抽出する。

例) ②で検索された世帯キー“1001101011”の場合
次のようなSELECT文で検索される。

```
SELECT ct_pf.key,  
       AF01.AF010, AF01.AF020,  
       AF02.AF030, AF02.AF040  
FROM AF01, AF02  
WHERE ct_pf.key = "100110101"  ※keyは実際には地区番号～該当者番号を連結したもの  
AND AF01.key = ct_pf.key  
AND AF02.key = ct_pf.key  
;
```

④夫婦データ管理テーブルおよび世帯キーで男性票から指定項目を抽出する。

例) ②で検索された世帯キー“1001101011”および指定項目AM01・AM015の場合

	世帯キー 002～005	第1回	
		配偶者 番号	調査票 番号
FAさん	1001101011	1	AM01
FBさん	1001101012	2	ASM01
FDさん	1001101014	4	AM01

次のようなSELECT文で検索される。

```
SELECT  
       AM01.AF015  
FROM AM01  
WHERE AM01.key = "100110101"  ※keyは実際には地区番号～世帯番号を連結したもの  
AND AM01.AM005 = 1  
;
```

調査票番号が“AM01”に一致しない場合(FBさんのペアのような場合)、NULLを出力。

③、④を対象世帯分だけ繰り返し、抽出データを作成する。

■ 操作性に関する改善案

● 抽出条件指定の改善

抽出条件の指定はユーザーが調査票番号・変数名を直接入力して指定している。入力ミスなどで正しく抽出条件を指定されず、検索に失敗するようなことも起こりうる。

(改善案)

画面上で項目一覧から抽出条件の指定をできるように入力方法を簡略化する。以下に画面イメージを示す。

(サンプル画面)

The screenshot shows a Microsoft Excel window titled '抽出条件指定' (Extraction Condition Specification). The main table has the following structure:

抽出項目	抽出条件
調査票番号	調査票番号
A002	A015 <
A006	A015 >= 45
A007	

Below the table, there are two buttons: '抽出項目に追加' (Add to Extraction Item) and '抽出条件に追加' (Add to Extraction Condition). A list of variables is shown below these buttons:

- A001 (調査票)
- A002 (調査番号)
- A003 (調査番号)
- A004 (住所地(外国))
- A005 (住所地(都道府県))
- A006 (住所地(市区町村))
- A007 (住所地(市町村))
- A008 (住所地(性別))
- A009 (出生年月日(年))

Annotations and their corresponding actions:

- ① 調査票を選択する (Select the survey form)
- ② 対応する変数名(調査項目)一覧が表示される。 (A list of corresponding variable names (survey items) is displayed.)
- ③ 一覧から抽出項目あるいは抽出条件に指定する項目を選択する。 (Select an item from the list to specify as an extraction item or extraction condition.)
- ④ 選択された項目を抽出項目あるいは抽出条件に追加する。 (Add the selected item to the extraction item or extraction condition.)

II. 個別研究報告

(縦断調査データの課題と分析への課題)

8 21 世紀出生児縦断調査における脱落要因の分析

福田 節也

はじめに

21 世紀出生児縦断調査は、全国の 2001 年 1 月 10-17 日および 7 月 10-17 日に出生したすべての子を調査客体としたパネル調査である。パネル調査では、同一個人あるいは世帯を対象として複数年にわたり継続的にサーベイを行う。パネル調査における利点として、同一対象の時間的変化を特定できること、また横断調査に比べてより直接的に標本のもつ異質性を統制し、説明要因と非説明変数との因果関係を考察できることなどが挙げられる（樋口他，2004）。しかし、一方でパネル調査特有の問題として、調査拒否や転居等の理由によりサンプルが脱落してしまうパネル減少（panel attrition）がある。サンプルの脱落が非無作為に生じる場合、データの代表性が損なわれ、データから得られる知見に偏りが生じることとなる。また、サンプルの脱落は調査回数とともに累積する為、脱落が著しい場合には調査の長期的な実施が困難となる。したがってパネル調査においては、いかに脱落を減らすかが調査実施上の要諦となる。

本稿では、21 世紀出生児縦断調査における脱落の発生状況について考察し、同パネル調査における脱落要因の分析を行う。また、脱落への対処法についても述べる。本稿における考察を通じて、21 世紀出生児縦断調査における脱落の防止に資する提言を行うことが目的である。

1. 21 世紀出生児縦断調査における脱落の発生概況

本節では、21 世紀出生児縦断調査の第 1 回調査（以後、第 1 波）および第 2 回調査（以後、第 2 波）の確定データを用いて脱落の発生概況およびその留意点について報告する。

表 1-1. 21 世紀出生児縦断調査第 2 波における脱落の発生割合

年度	回収数	回収率	脱落数	脱落率
1	47015	(87.8%)	-	-
2	43926	93.4	3089	6.6

表 1-1 において明らかなように、同調査第 2 波における脱落サンプルは 3089 ケース、割合にして初年度サンプルの 6.6%となっている。一般にパネル調査では、毎年継続的に調査することについて調査対象者より了解をとる必要がある。そのため、初年度の調査では調査協力を得にくいのが、第 2 年度以降の調査では協力が得やすいことが指摘されている（坂本，2004）。厚生労働省による統計調査であることを反映し、当調査では初年度の回収率が

87.8%と高いことが指摘できる。わが国における先駆的なパネル調査として知られる『消費生活に関するパネル調査』（財団法人家計経済研究所により 1993 年より実施）では初年度（1993 年）の回収率が 41.4%，以後 1994 年から 2003 年までの各年度調査の平均回収率が 95.5%（いずれも初年度調査よりサンプルとなっているコーホート A の値）となっている（家計経済研究所，2004）。したがって，21 世紀出生児縦断調査においては初年度の回答率が高い代わりに，脱落率に関しては同様の調査と比較して同程度か若干高い傾向にあるといえる。

次に，パネル調査における脱落の留意点について述べたい。パネル調査における脱落は累積するという点に注意が必要である。表 1-2 は，21 世紀出生児縦断調査における脱落の状況ならびに 4 回調査以降の回収率を 95%と仮定した場合の推計値を表したものである。

表 1-2 をみると，各年度の回収率を 95%に仮定した場合，5 年度調査における累積脱落率の推計値は 17.8%，10 年度調査においては 36.4%にも達している。このことは，初年度調査の対象サンプルのうち 5 年度調査まで継続して回答している者は約 82%，10 年度調査では約 64%にまで減少することを意味している。さらに初年度の回収率が 87.8%であったことを考慮すると 5 年度調査では標本母集団の 72.1%（ $=87.8 \times 0.822$ ），10 年度調査では 55.8%（ $=87.8 \times 0.636$ ）しか捕捉されていないことを意味している。

パネル調査における脱落がすべての回答項目に対して無作為に生じる場合（missing completely at random）には，各変数の分布は不変であるため統計量の推定において脱落によるバイアスは生じない。ただし，脱落によりサンプル数が減少するため，推定される統計量の標準誤差が大きくなるが，このことは統計的検定に影響を及ぼさない（Allison, 2002）。しかし，ほとんどすべての場合において，脱落は調査対象者のなんらかの属性に依存して生じる。例えば，共働きや高所得などの特定の傾向をもつサンプルほど調査から脱落する確率が高い。したがって，脱落の発生したデータを用いる場合，実際には脱落と関連のある変数の統計量の推定にはバイアスが生じることとなる。

表 1-2 21 世紀出生児縦断調査における脱落の状況ならびに推計値

年度	回収数	回収率 ^{*1}	脱落数	累積脱落率 ^{*2}
1	47015	(87.8%)	-	0.0
2	43925	93.4	3090	6.6
3 ^{*3}	42803	97.4	1122	9.0
4	40662.9	95.0	2140.2	13.5
5	38629.7	95.0	2033.1	17.8
6	36698.2	95.0	1931.5	21.9
7	34863.3	95.0	1834.9	25.8
8	33120.1	95.0	1743.2	29.6
9	31464.1	95.0	1656.0	33.1
10	29890.9	95.0	1573.2	36.4

* 4 年度以降は回収率を 95%と仮定した場合の推計値

*1 当年 (t 年) の回収数 / 前年 (t-1 年) の回収数 × 100

*2 当年 (t 年) までの脱落数累計 / 47015 人 (初年度回収数) × 100

*3 回収数には第 2 回調査に無回答であったサンプルも含む

一方で、多変量解析を行って回帰係数を推定するような場合には、パネル調査における脱落は係数の推定にそれほど大きな影響を与えないとの結果もある (Lillard & Panis, 1998; Alderman, et. Al., 2001; Falaris, 2003)。しかし、21 世紀パネル調査が主として記述統計を旨とする官庁統計であることを考慮すると、脱落による標本の歪みは当調査に基づき公表する統計データの信頼性や妥当性に大きな影響を与えるものと思われる。また、脱落は調査の長期的な継続を困難にするという点からも望ましいものではない。そこで本稿では 21 世紀出生児縦断調査の第 2 回までのデータを用いて、脱落発生のメカニズムに関する考察を行う。

2. 21 世紀パネル調査における脱落要因の 2 変量解析

本節では、21 世紀出生児縦断調査の第 2 回調査における脱落が調査対象者の属性とどのように関わっているのかを明らかにする。具体的には、第 1 波における対象者の属性や回答状況と第 2 波における脱落との相関関係について考察する。

2-1. 量的変数と脱落の関係

はじめに、第 1 波の量的変数と第 2 波における脱落との 2 変量の間を考察する。表 2-1 は、第 1 波における量的変数について、第 2 波まで回答を継続したサンプルと脱落したサンプルの平均値を比較したものである。

夫妻の属性と脱落との関係を見ると、脱落サンプルは継続回答サンプルに比べて父母の年齢が若いもしくは父母の同居期間（おそらくは結婚期間）が短いといった特徴がある。また、夫婦の年齢差が大きいことも脱落サンプルの特徴といえる。さらに脱落サンプルは継続回答サンプルに比べて父母ともに年収が低いという傾向がみられる。前述の家計研パネルにおいては年収の高いサンプルほど脱落する傾向があるが（坂本, 2004）、これとは逆の結果となっている。また、出生児の父母が喫煙者である場合、脱落サンプルほど 1 日の喫煙本数が多い傾向にある。特に脱落者と継続回答者の間で母親の喫煙本数の差が大きい。

子の属性についてみると、わずかではあるが対象児の体重や母親の妊娠週数が短い場合に脱落する傾向があるようである。出生児が未熟児であるほど脱落する傾向が高いのであろうか。このことは後に質的変数と脱落の関係において考察する。また、脱落サンプルには母の出産した子の数が多いという傾向が認められる。しかし、出生児の兄弟の数は脱落に影響していない。このことは双子などの多胎児の出産が脱落を促すことを示唆している。これについても後に考察する。

さらに父母の育児意識や育児・家事頻度と脱落の関係について着目する。脱落サンプルには母親の育児・家事頻度が低い傾向が認められる。特に、育児頻度の差が大きい。一般にパネル調査からの脱落は、調査の主目的によって異なることが指摘されている。例えば、調査の目的が政治意識に関するものである場合、政治意識の高い対象者ほど継続回答する傾向があるだろう。本調査の主目的が子育てに関する意識や行動であることを考慮すると、

母親の子育てへの意識や参加度が低い場合に脱落が生じやすいのかもしれない。こうした傾向は、脱落サンプルほど子育てで意識して行っていることや子どもをもってよかったことの得点が低いことにも表れている。先の喫煙本数と脱落の関係についても、母親の喫煙本数が育児意識の低さを反映していると考えられることもできる。しかし、父親の家事・育児頻度は脱落に影響していない。母親のみが調査の回答者である割合は約9割(87.6%)であり、調査継続に関する意思決定は主に母親によって行われていることが示唆される。

表 2-1 回答継続者と脱落者の比較 (量的変数)

変数	継続回答サンプル (1)	脱落サンプル (2)	差 (2) - (1)	
子の体重 (g)	3035.92 43915	3018.99 3086	-16.93	*
妊娠週数 (週)	38.89 43895	38.83 3084	-0.06	*
この母の出産した子の数 (出生子)	1.68 43926	1.73 3089	0.05	**
妊娠満22週以後の死産子の数	0.01 43926	0.01 3089	0.00	
父の年齢 (歳)	31.66 43431	30.03 2971	-1.63	***
母の年齢 (歳)	29.50 43926	27.60 3089	-1.90	***
同居期間月数	44.94 43246	40.22 2944	-4.72	***
きょうだい等の人数 (兄弟の人数)	0.66 43926	0.69 3089	0.03	
父母の喫煙状況 (お母さん) 1日の本数	11.32 7071	12.92 1053	1.59	***
父母の喫煙状況 (お父さん) 1日の本数	18.72 26324	19.48 2125	0.76	***
母の家事・育児の点数 (育児)	16.69 42917	16.51 2969	-0.18	***
母の家事・育児の点数 (家事)	17.08 43002	16.98 2998	-0.09	***
父の家事・育児の点数 (育児)	10.51 41525	10.52 2733	0.01	
父の家事・育児の点数 (家事)	6.48 42035	6.34 2787	-0.14	
子育てで意識して行っていること(尺度)	2.78 43835	2.70 3065	-0.08	**
子どもをもってよかったこと(尺度)	3.29 43809	3.19 3060	-0.11	***
子どもをもって負担に思うこと(尺度)	1.76 43681	1.75 3040	-0.01	
母親の年収 (万円)	101.35 42352	75.12 2867	-26.23	***
父親の年収 (万円)	457.37 41214	378.35 2634	-79.02	***
その他の年収 (万円)	11.22 43406	12.11 2984	0.89	
夫婦の年齢差(絶対値) (歳)	3.10 43431	3.35 2971	0.25	***
無回答項目総数	3.53 43926	5.35 3089	1.82	***

各行の上段は実数値、下段はサンプル数を表す。

***: 0.1%水準、**: 1%水準、*: 5%水準で統計的に有意(t検定)

調査票における無回答項目の数¹⁾もまた脱落に影響を与えている。脱落サンプルほど初年度調査における無回答項目数が多い傾向にある。無回答項目の多さは、調査に対する理解度や関心の低さを反映するものと思われる。したがって、育児意識とは別に調査に対する協力度が低い場合にも脱落が生じやすいものと思われる。しかし、同時に当調査では父母のどちらかが出生児と同居していない場合や出生児が非嫡出である場合、多くの質問項目において非該当として処理されている。そのため、無回答項目の数が調査への協力度の低さを反映しているのか、それとも父母の非同居や非嫡出による影響なのかについては、2変量解析では明らかにすることができない。この問題については、後に多変量解析によって明らかにする。

2-2. 質的変数と脱落の関係

次に、第1波における質的変数と第2波における脱落との2変量の関係について考察する。表2-2は脱落率が質的変数の各カテゴリーにおいてどのように異なるのかを比較している。なかにはカテゴリーのサンプル数が非常に少ない（全サンプルの1%以下）変数もあるが、脱落に重要な影響を与えているものもあるので敢えて表に含めた。

表中のほとんどの変数において、脱落率との間に統計的に有意な関係が認められる。しかしカイ2乗検定ではサンプル数が多いほど統計的な有意性が検出されやすいという特性があるのでその解釈には注意が必要である。表の解釈においては、むしろカテゴリーの脱落率とカテゴリーのサンプル数という2つの観点から脱落への影響を判断する必要がある。主な知見について要約すると以下である。

脱落率が10%以上のカテゴリーをもち、かつそのカテゴリーに占めるサンプル数が全体の10%程度である変数として、母親の喫煙状況、父母の喫煙状況、母親の就業状況、そして婚前妊娠の有無を挙げることができる。喫煙本数は脱落サンプルほど多い傾向がみられたが、今回の分析でも母親の喫煙が脱落と関係することが明らかとなった。母親が室内で喫煙する割合は全サンプルの12.2%にも上り、脱落率は13.4%となっている。母親の喫煙は父親の喫煙と強い相関をもっており、両親ともに喫煙する場合の脱落率は12.6%となっている。次に、第1波において母親が求職中である場合にも高い脱落率がみられる。母親が「仕事を探している」と回答した割合は9.5%であったが、そのうちの10.7%が第2波において脱落している。就業による多忙によるものであろうか。そして近年増加傾向にある婚前妊娠はサンプルの13.3%を占めており、うち10.7%が第2波で脱落している。なお、本分析における婚前妊娠とは、夫婦の同居期間が妊娠週数より短い場合と定義している。婚前妊娠は早婚とも関連しているので（厚生労働省大臣官房統計情報部、2002）、夫婦の年齢が若く、同居期間が短いほど脱落しやすいという先の分析結果は、婚前妊娠と関係しているのかもしれない。これについては後に多変量解析によって確認する。

¹⁾ 無回答項目数の定義については付表1を参照のこと。

次に、カテゴリーにおける脱落率が10%以上で、かつそのカテゴリーのサンプル数が1000以上（全サンプルの約2%）である変数として、平日日中の保育者、父親との非同居、家族形態、そして父親の就業状況および労働時間が挙げられる。平日の日中に子どもを施設やベビーシッターに預けている場合に脱落率が高くなる。また、対象児が父親と同居していない場合や母子世帯において脱落率が約17%と高い傾向がある。父親が無職やパートである場合や労働時間が20時間以下である場合にも脱落率が10%を超えている。先の分析では脱落者のほうが継続回答者よりも年収が低いという傾向がみられた。民間の調査機関が行う調査とは異なり、当調査では謝金などのインセンティブが作用しないことが低所得層や母子のみの世帯における脱落を促している可能性がある。

さらに、カテゴリーに占めるサンプル数は少ないが、脱落率に非常に大きな影響を与えている変数として、嫡出子・否の別、父母の国籍、ならびに母親との非同居が挙げられる。非嫡出児の脱落率は約20%にも上る。また、父母ともに外国籍であるサンプルは217サンプルと少ないが27.7%と高い脱落率を示している。これらの世帯はもともと今回のような標本調査において捕捉することが困難な層であり、パネル調査における脱落率も大きいものと思われる。さらに54サンプルと少数ではあるが、対象児が母親と同居していないケースも約30%と高い脱落率を示している。したがって、当調査のサンプルは比較的標準的かつ安定的な世帯属性をもつサンプルに収束していくことが予測される。

その他の知見として、低体重児、早産、もしくはこの両方である未熟児といった出生児の属性が脱落率と正の関係をもつことが明らかである。このことは脱落サンプルほど出生児の体重が軽く、妊娠週数が短いという先の分析結果と一致する。なぜ未熟児に脱落が多いのかについては、脱落理由を精査しない限り断定はできないが、乳幼児の死亡率に関連しているのかもしれない。つまり、未熟児で出生したサンプルほど死亡による脱落が多い可能性がある。データでは「子どもをもって負担に思うこと」として、「子どもが病気がちである」を挙げる割合が、未熟児で7.3%、未熟児以外では3.4%と未熟児の方が2倍以上高くなっている。また、統計的な有意性は5%と低いが、多胎児であるほど脱落率が高いという傾向がみられる。これは出生児の兄弟の数は脱落に影響しないが、母親の出生児数が多いほど脱落がみられるという量的変数の分析結果を支持するものである。多胎児の回答者もまた「子どもが病気がちである」を挙げる割合が6.0%と高い傾向にある。また多胎児の場合、回答者は出生児と同じ数の調査票に記入しなければならないため負担が大きい。このことが脱落に繋がりやすいということも考えられる。