



北京大学中国社会科学调查中心
Institute of Social Science Survey, Peking University

家庭关系库的结构及运用

—中国家庭追踪调查 (CFPS)

CFPS项目办公室

戴利红



目 录



2010年家庭关系库的结构



2012年家庭成员问卷设计理念



2012年家庭关系库的构建思想



2012年家庭关系库的结构



家庭关系库的运用

2010年家庭关系库的结构

• 基因家庭=基线家庭 • 基因成员=基线成员

T1表：与家庭有血缘/婚姻/领养关系的同住成员列表及信息

个人样本编码	家庭样本编码	户内码	出生年份	性别	最高学历	其它基信息
100000101	100000	101	1978	男	本科
100000102	100000	102	1980	女	大专
...

独立的基因家庭

每个基因成员
基本信息

T3表：与家庭成员不同住直系亲属列表及信息

Fid	户内码	出生年份	性别	最高学历	其它基信息
100000	201	1952	男	初中
100000	202	1953	女	高中
...
...

T

3

T2表：家庭成员直系亲属关系表

家庭编码	个人编码		父亲编码		母亲编码		配偶编码		孩子1编码	...	孩子10编码
100000	10000	101	-8	201	-8	202	10000	102	-8
	0101						0102		
100000	10000	102	-8	203	-8	204	10000	101	-8
	0102						0101		
...

家庭关系库：T表=T1表+T2表+T3表

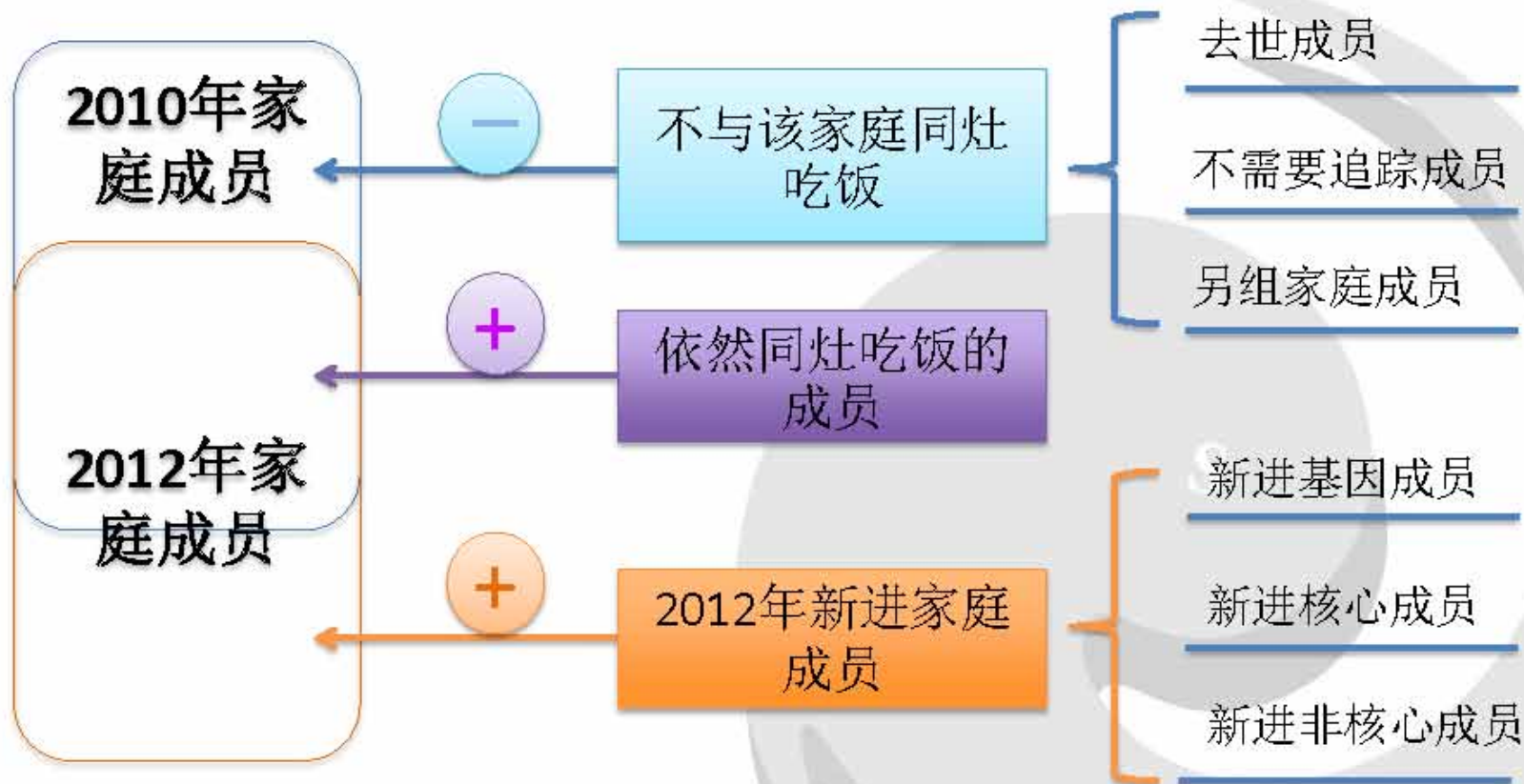
100000	$\left[\begin{array}{l} 100000101 \quad 101 \quad 1978 \quad 1 \quad 6 \quad \dots \\ 100000102 \quad 102 \quad 1980 \quad 0 \quad 5 \quad \dots \end{array} \right]$	$\left[\begin{array}{l} -8 \quad 201 \quad 1952 \quad \dots \\ -8 \quad 203 \quad 1954 \quad \dots \end{array} \right]$	$\left[\begin{array}{l} \dots \\ \dots \end{array} \right]$
100000			
.....
200000	$\left[\begin{array}{l} 200000101 \quad 101 \quad 1992 \quad 1 \quad 4 \quad \dots \\ \dots \end{array} \right]$	$\left[\begin{array}{l} 200000102 \quad 102 \quad 1970 \quad \dots \\ \dots \end{array} \right]$	$\left[\begin{array}{l} \dots \\ \dots \end{array} \right]$
.....			

↑
家庭成员的编码及信息

↑
父亲的编码及信息

↑
其他的编码及信息

2012年家庭成员问卷设计理念





2012年家庭关系库的构建思想

- 人员构成
- 个人信息
- 流动的家庭
 - 流失的家庭
 - 另组的家庭
- 流动的家庭成员
 - 外出的个人
 - 另组的个人
 - 新进基因成员
 - 新进核心成员



T1&T2表
T3表

2012年家庭关系库的结构

直接由T1_2012\T2_2012、T3表生成的变量

变量名	变量标签	变量名	变量标签
pid	个人样本编码	TB3_A12_p	个人婚姻状况
fid12	2012年家庭样本编码	TB4_A12_p	个人最高学历
fid10	2010年家庭样本编码	qa301_a12_p	个人户口类型
code_a_p	个人家庭内部编码	qa302_a12_p	个人户口所在地
cfps_interv_p	个人产生有效问卷的情况	TB6_A12_p	个人是否在家住
TB1A_A_p	个人属相	TB601_A12_p	个人离家的原因
TB1B_A_p	2012年个人年龄	outpers_where12_p	离家人（个人）居住区域
TB1Y_A_p	个人出生（年）	TB602ACODE_A12_p	离家人（个人）省国标码
TB1M_A_p	个人出生（月）	leavingtime_y_p	个人离家的年份
TB2_A_p	个人性别	leavingtime_m_p	个人离家的月份
ALIVE_A_p	个人是否健在	co_a12_p	个人是否与该家庭同灶吃饭
deathreason_p	个人去世的原因		

关键变量

cfps_interv_p

个人产生有效问卷的情况

0:没有任何有效问卷

1:2012年产生有效问卷

5:只有2010年有有效问卷

TB6_A12_p

个人是否在家住

1: 在家居住

0: 不在家住

co_a12_p

个人是否与该家庭同灶吃饭

1: 是同灶吃饭;

0: 不同灶吃饭

综合变量

1. fam_sum: 同灶吃饭成员的总人数
2. GeneType: 基因成员类型
3. CoreMember: 是否是核心成员
4. gene_n: 同灶吃饭中基因成员人数
5. core_n: 同灶吃饭中核心成员人数

- 0: 非基因成员
- 1: 在家基因成员
- 2: 新进基因成员
- 3.需要追访的外出基因成员
- 4.需要追访的另组基因成员
- 5.不需要追踪的外出基因成员
- 6.不需要追踪的另组基因成员
- 7.死亡基因成员

2012年家庭关系库的运用

• 个人层面:

- 1、家庭成员完成个人问卷的情况: cfps_interv_p
- 2、家庭成员的构成和流动人员:

Tb6_a12_p= (0-1):物理是在家居住
Tb601_a12_p=N: 离家原因
C0_a12_p= (0-1):是否同灶吃饭
Genetype= (0-7): : 基因类型

(1, -8, 1, <=2)
在家成员

(0, N, 1, 3)
同灶吃饭的离家成员

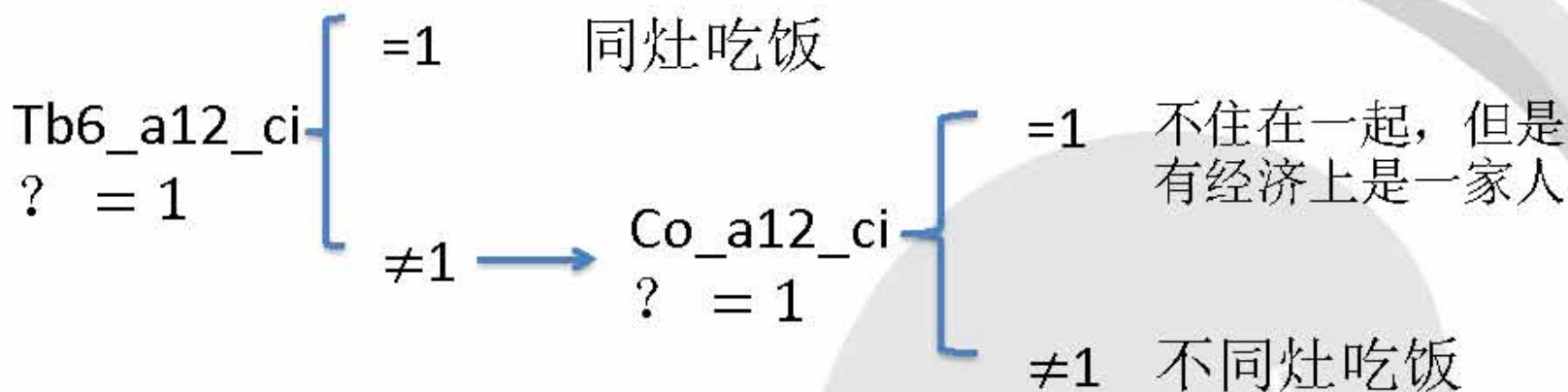
(0, N, 0, 4)
另组的基因成员

(0, N, 0, >4)
不追踪的离家基因成员

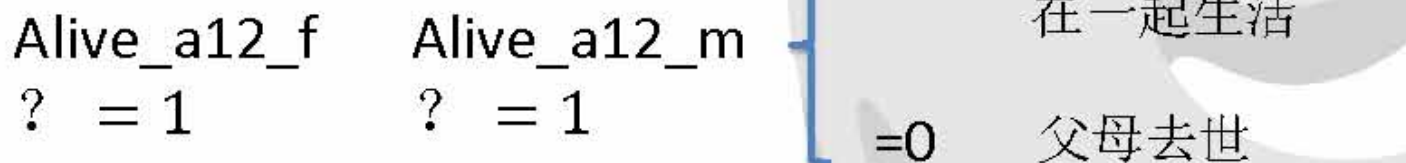


3.老人是否和孩子住在一起

A.有老人的家庭



B.没有老人的家庭



4.少儿是否和父母住在一起



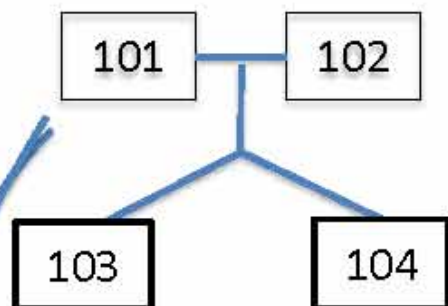
2012年家庭关系库的运用

家庭层面：

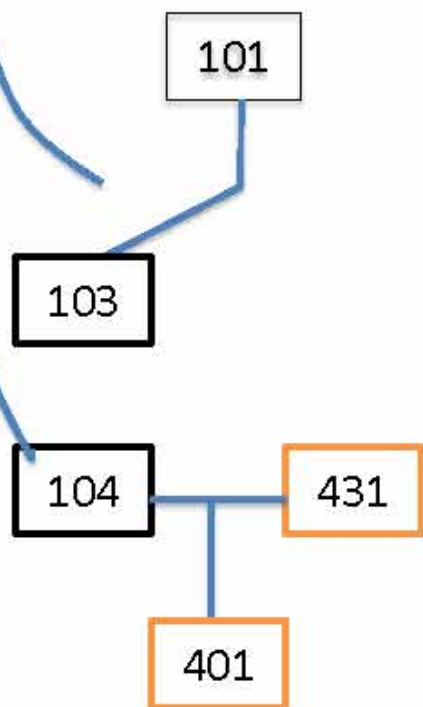
- 1、所有家庭成员和流动人员基本信息：后缀是字母“p”的变量名，例：TB1A_A_p
- 2、家庭成员关系人及其信息：例：父亲信息，后缀是“f”， TB1A_A_f；
- 3、原家庭与另组家庭的联系：fid12 \neq fid10表示分裂家庭，用fid10的数据连接全库中fid12的数据，就找到了其原家庭的信息。

家庭案例

2010



2012



Pid	Fid	fid	Code	Tb3_	Tb6_	T601	Co_a	cfps_	gene
	12	10	_a_p	a12_	a12_	_a12	12_p	inter	type
				p	p	p		v_p	
100000 101	100 000	100 000	101	5	1	-8	1	1	1
100000 102	100 000	100 000	102	-8	0	10	0	5	7
100000 103	100 000	100 000	103	1	0	2	1	0	3
100000 104	100 000	100 000	104	1	0	9	0	1	4
100000 104	200 000	100 000	104	2	1	-8	1	1	1
200000 401	200 000	-8	401	1	1	-8	1	1	2
200000 431	200 000	-8	431	2	1	-8	1	1	0



家庭关系库和其他数据库的联系





跨库使用

• 1. 与家庭经济库

家庭关系库中的家庭样本编码与家庭经济库中的样本编码进行连接，从而获得家庭成员的家庭经济情况。

• 2. 与成人库、少儿库

用成人、少儿库中的样本编码pid与家庭关系库中的pid进行连接，就找到了个人的直系亲属关系人的样本编码id_*及其基本信息，然后用id_*在与成人、少儿库中的pid进行连接，就找到了样本的全部问卷信息。

pid	fid12	ld_f	...	ld_m	...	ld_s	...	ld_c1	...
100000101	100000	-8	...	-8	...	10000 0102	...	10000 0103	...
100000102	100000	-8	...	-8	...	10000 0101	...	10000 0103	...
100000103	100000	10000 0101	...	10000 0102	...	-8	...	-8	...

pid	Fid12	Infor.	ld_f,...	ld_m,...	ld_s,...	...
100000101	100000	-8,... ..	-8,... ..	100000102	
100000102	100000	-8,... ..	-8,... ..	100000101,...	
100000103	100000	100000101,....	100000102,....	-8,... ..	