

China Family Panel Studies

CFPS

中国家庭追踪调查

技术报告系列: CFPS-38

系列编辑: 谢宇 责任编辑: 赵逸文

中国家庭追踪调查
事件历史日历记录法 (EHC)
设计方案

孙妍

2020. 12

一. EHC 通用设计

1.1 适用范围

CFPS 访问过程中需要受访者回忆两轮调查间发生的事情，为降低访问过程中因受访者遗忘事件而导致的漏报，同时校准因受访者记忆错位所导致的错报，CFPS 自 2014 年起引入事件历史日历记录法（Event History Calendar, EHC），旨在通过改变数据采集方式达到提升调查数据采集质量的目的。

EHC 工具在历轮次问卷中的运用情况如表 1 所示。2014 年 CFPS 首次引入 EHC，主要运用于 16 岁及以上的“成人自答”问卷中，用于采集成人群体迁移、婚姻及工作三个部分的信息。其中，居住地址迁移模块仅采集 CFPS 基因成员的信息，而其他两个模块则是所有回答人（基因及核心成员）均适用。基于首轮运用效果的评估，伴随着调查项目问卷整体设计方案的调整，EHC 在随后的调查轮次也有所调整。

首先，基于 CFPS2014 年数据分析发现，婚姻状态在两轮次间变化频率较低。因此，自 2016 年起，婚姻模块问题恢复传统提问方式，不再用 EHC 的形式采集数据。三个 EHC 内容模块的访问顺序也由 2014 年的迁移、婚姻、工作，调整为婚姻、迁移、工作。在 EHC 外部所采集的婚姻模块数据，作为预加载信息在 EHC 中展示，为受访者回答迁移及工作模块问题提供参考信息。

其次，2018 年项目组将 CFPS 访问过程中需要受访者自答的问卷内容作了整合，将 10-15 岁少儿自答部分的问题并入 16 岁及以上的成人自答问卷中，由此形成“个人自答问卷”。EHC 仍被运用于采集迁移及工作两个模块的信息。迁移模块沿用以往轮次的进入标准，即 16 岁及以上基因成员回答该部分内容；而工作模块则仅根据受访者回答所判定出的当前工作状态决定是否需要进入 EHC，不再设置回答人需要在 16 岁或以上的年龄限制。

再者，CFPS 自 2012 年起启用电访模式追踪离家个人，但 2012、2014 两轮次所运用的电访问卷较同期的面访问卷在内容尚有较大程度的简化。2016 年项目组在对问卷内容及提问方式调整的基础上，形成了电、面一体的共用问卷。虽然，EHC 自 2014 年引入后，始终保持着电、面两类问卷一致的设计方案；但执行上，面访访问过程中访员可以与受访者共同查看屏幕上的日历，进行数据填答；电访中，则只能依靠访员提问、受访者回答，访员根据日历所展示的信息与受访者校准各事件发生的时间点。

表 1、CFPS 历轮调查 EHC 运用情况

轮次	问卷类型	迁移	婚姻	工作
2014	成人面访/电访自答问卷	✓	✓	✓
2016	成人面访/电访自答问卷	✓	×	✓
2018	个人面访/电访自答问卷	✓	×	✓

1.2 实现方式

EHC 问卷嵌在主体问卷程序中，根据设定的位置及条件分模块激活。可参见发布版问卷获取每个模块进入顺序及条件。

值得说明的是，当访员进入顺序靠后的模块时，仍可以编辑或修改已完成模块的答案，根据受访者的回答即时修正数据。

1.3 回忆时段

为了降低受访对象的回忆负担，项目组针对受访对象项目参与经历的不同设定了不同的回忆时段起点。基本原则是，保证受访者的回忆时段覆盖时长不超过 3 年，详细回忆时段设定规则如表 2 所示。例如，当前为 2018 调查轮次，若受访对象最近一次个人问卷是在 2016 轮次完成的，则加载 2016 轮次中个人问卷的完访日期作为当前轮次回忆时段的起点；若受访对象最近一次个人问卷完访轮次早于 2016 轮次或从未接受过访问，则加载 2016 年 1 月 1 日作为回忆时段起点。EHC 回忆时段的终点是由系统在首次进入 EHC 时自动获取访问当前的日期，一旦取值当轮次后续访问过程中则不再调整。

概而言之，EHC 回忆时段是根据每名受访对象的具体情况定制，不同受访对象设定不同的回忆时段。

表 2、CFPS EHC 回忆时段设定规则

最近个人问卷完访轮次	起点	终点
上一轮次	上轮次个人问卷完访日期	首次访问 EHC 第

更早期的轮次	上一轮执行年的 1 月 1 日	一个模块的日期
从未接受访问		

1.4 时间记录方式

1.4.1 时间单位

EHC 采集模式下,所有涉及事件发生具体时间的最小单位设定为“月份”,即只记录事件发生的年月,允许模糊化回答。在 EHC 访问界面上体现为年份、季节、月份(每格代表一个月),如图一所示。每个事件发生时间信息采集完成后,会在日历表上以线条覆盖的形式呈现。

接下来我们将提问从2012年9月1日调查至今您的工作情况。

访员注意:

- 1) 请首先确认加载工作是否正确,并收集受访者当前/最近的工作。
- 2) 按工作结束时间从后往前,收集回忆时段内受访者所从事的工作。

表头 单击这里隐藏	2012年				2013年				2014年												
	秋季		冬季		春季		夏季		秋季		冬季		春季								
当前受雇状态	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月
+住址																					
Former Residence	#							#									#			#	
2014住址																					#
other residence1		#						#													
other residence2				#																	#
空白期地址1																					
+婚姻																					
SPOUSE1	#								#					#							
SPOUSE2															#	#					#
PARTNER1				#	#																
PARTNER2										#										#	
+工作																					
Former Job	#																				#
2014工作																					
T-14																					

图一、EHC 示例 1

1.4.2 时间记录方式

1.4.2.1 开始时间

(1) 开始年份

根据回忆时段确定事件起始开始年份选项具体内容，采用下拉菜单选择的方式实现。以 2014 调查轮次为例，事件开始年份下拉菜单设计如下：

- ✓ 下拉菜单选项：2012、2013、2014/2015¹、2012 以前年份、不确定年份（2012 以前）、不确定年份（2012-2014/2015）。
- ✓ 若年份选择“2012 年以前年份”展开数值型题目，值域范围 1901..2011。

(2) 开始月份

采用表格选择的方式记录事件发生的具体月份，如图二所示。

冬季	12 月份	1 月份	2 月份	退出
春季	3 月份	4 月份	5 月份	
夏季	6 月份	7 月份	8 月份	不知道
秋季	9 月份	10 月份	11 月份	

图二、EHC 月份记录示例

1.4.2.2 结束时间

(1) 事件结束年份同样采用下拉菜单的方式供受访者选择。以调查轮次为 2014 为例，事件结束年份下拉菜单选项设计如下：

- ✓ 2012、2013、2014/2015、不确定年份（2012-2014/2015）。

(2) 结束月份设计与开始月份相同，同样采用表格选择的设计方案，具体如图二所示。

1.5 回忆线索

在 EHC 中通过加载的形式导入以下几类回忆线索，辅助受访者回答：

- (1) 时间线索：日历年月信息、季节。
- (2) 重大个人事件：本轮调查前序已完访模块获取的信息，如地址迁移、婚姻状态

¹ 如果调查延续到 2015，年限增加到 2015，以下所有出现 2014 的地方采用相同方式处理。

变化。

1.6 通用加载

1. 个人最近一次完访日期。
2. 当前访问轮次中首次进入 EHC 的日期。
3. 回忆线索（参见四、回忆线索）

1.7 数据处理

1. EHC 回忆时段起点由受访者最近一次完成个人问卷的具体年月决定，用户如需要获取该信息请与项目组联系。
2. EHC 中取当前轮次中首次进入 EHC 年份及月份作为回忆时段的终点，终点年月信息分别存入变量 FIRSTENTEREH CY, FIRSTENTEREH CM。

3. 日期信息存储规则：

（1）年份信息存储规则

- ✓ 获取详细年份信息直接存储在相应年份；
- ✓ 若获取开始时间的年份信息为回忆时段前的“不确定年份”，存储为“-1900”；
- ✓ 若获取开始或结束时间的年份信息为回忆时段内“不确定年份”，存储“-2020”。

（2）月份信息存储规则

- ✓ 获取详细月份信息直接存储相应月份；
- ✓ 若获取开始或结束时间的月份信息为“春/夏/秋/冬”，对应存储为：13（春）、14（夏）、15（秋）、16（冬）；
- ✓ 若获取开始或结束时间的月份信息为“不知道”，对应存储为“-1”。

（3）结束时间选择“当前”存储规则

- ✓ 直接存储调查当前相应年份及月份；
- ✓ 增加一个以 C 结尾的 0-1 变量，用于区分结束时间在调查当前年月的状态。

数据存储示例：

例如：受访者 A、B 接受调查的时间均为 2014 年 6 月 15 日

EB202 您与“【CAPI】加载 EB200”的婚姻持续到什么时候? ____ (EEB202Y”婚姻结束日期(年)”) ____ (EEB202M”婚姻结束日期(月)”) /当前

受访者 A, 上次调查时的婚姻持续到 2014 年 6 月 10 日。在 EHC 中婚姻结束时间 EEB202Y 存储 2014, EEB202M 存储 6。

受访者 B, 婚姻持续到当前。在 EHC 中婚姻结束时间 EEB202Y 存储 2014, EEB202M 存储 6 记录时间为“当前”, 同时在 EEB202C 识别变量中存储值 1 (代表持续到当前

二. 迁移模块

迁移模块主要采集回忆时段内受访者居住地变化的经历, 每个居住地址询问搬入和搬离的时间, 是否有搬迁计划(针对现住址), 搬迁原因(前住址)。具体设计如下:

2.1 加载变量

详见问卷加载变量

2.2 提问内容

2.2.1 内容模块

- (1) 模块一、确认最近一轮访问所获得的住址

详见问卷 AR1 至 AR106

- (2) 模块二、确认当前主要居住地址

详见问卷 AR2 至 AR3

- (3) 模块三、其他居住地址

详见问卷 AR4 至 AR6

- (4) 模块四、空缺时段数据补充

详见问卷 AR7

2.2.2 根据人员类型不同有所不同，具体如下：

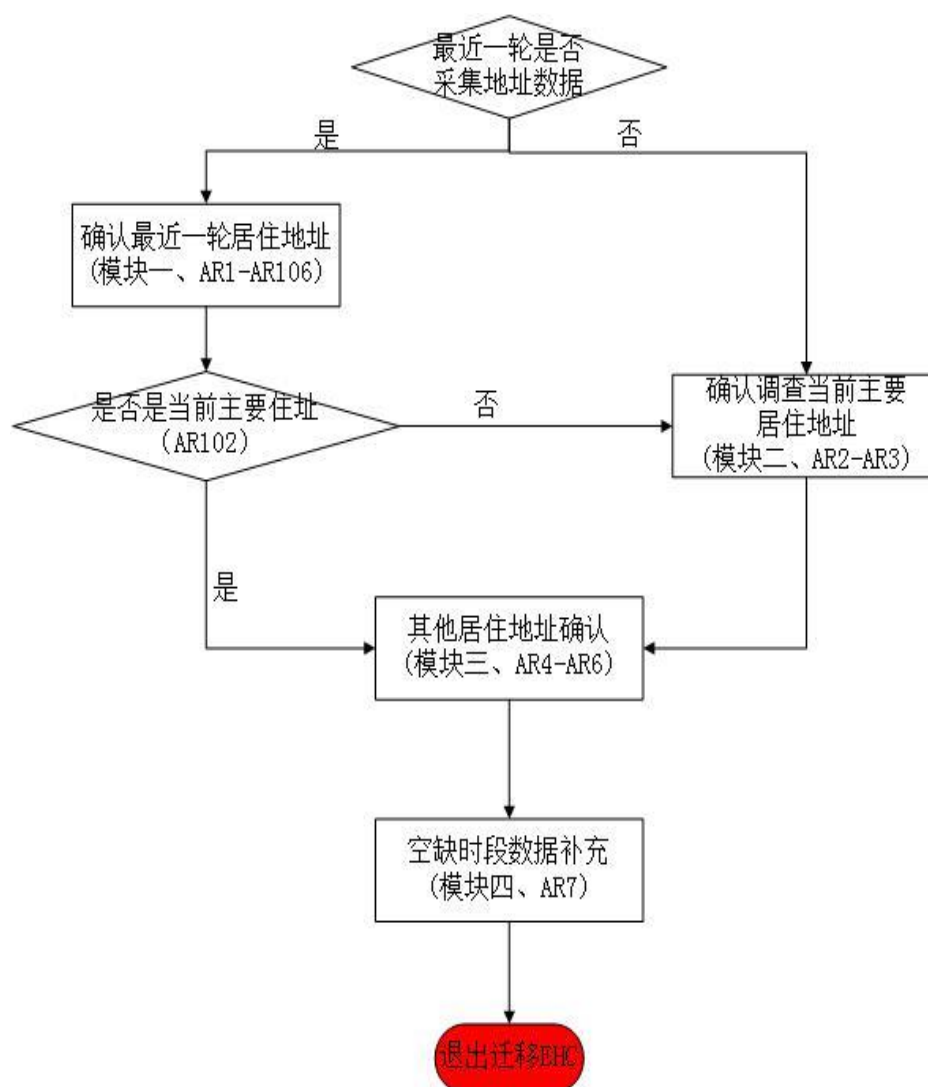
(1) 最近一轮成功完成个人自答问卷且采集到详细住址：

- ✓ 模块一、模块二、模块三、模块四。

(2) 其他成员：

- ✓ 模块二、模块三、模块四。。

2.2.3 迁移模块访问流程图



图五、迁移模块访问流程图

2.3 记录方式

整体记录方式参见图六：

	2012年-2013年					
	1月	2月	3月	11月	12月
resident 2012						
resident 2014	条件性出现					
AR4 其他地址数：8						
其他地址 1	根据 AR4 其他地址数出现，最多不超过 6 个地址					
其他地址 2						
其他地址 3						
其他地址 4						
其他地址 5						
其他地址 6						
空白期地址 1	条件性出现					
空白期地址 2						
空白期地址 3						

图六、迁移模块 EHC 记录方式

2.3 数据校验

1. 即时校验：详见问卷中题目设计。
2. 模块结束时校验：见问卷 AR7。
3. 数据有效性校验：同一个地址不同居住时段不允许交叉；允许同一时段内存在多个居住地址。

2.5 数据处理

1. 地址相关变量命名规则为：

- (1) 最近一轮居住地址及当前居住地址按照问卷所示变量名存储；
- (2) 其他地址变量名在问卷变量名基础上加“_a_N”后缀作标识；
- (3) 空白地址变量名在问卷变量名基础上加“_a_N”后缀作标识。

其他变量名参见问卷。

三. 婚姻模块

婚姻模块主要采集每类婚姻状态开始、结束时间，每段婚姻及同居经历以及配偶/同伴的信息。具体设计如下：

3.1 加载变量

详见问卷加载变量

3.2 提问内容

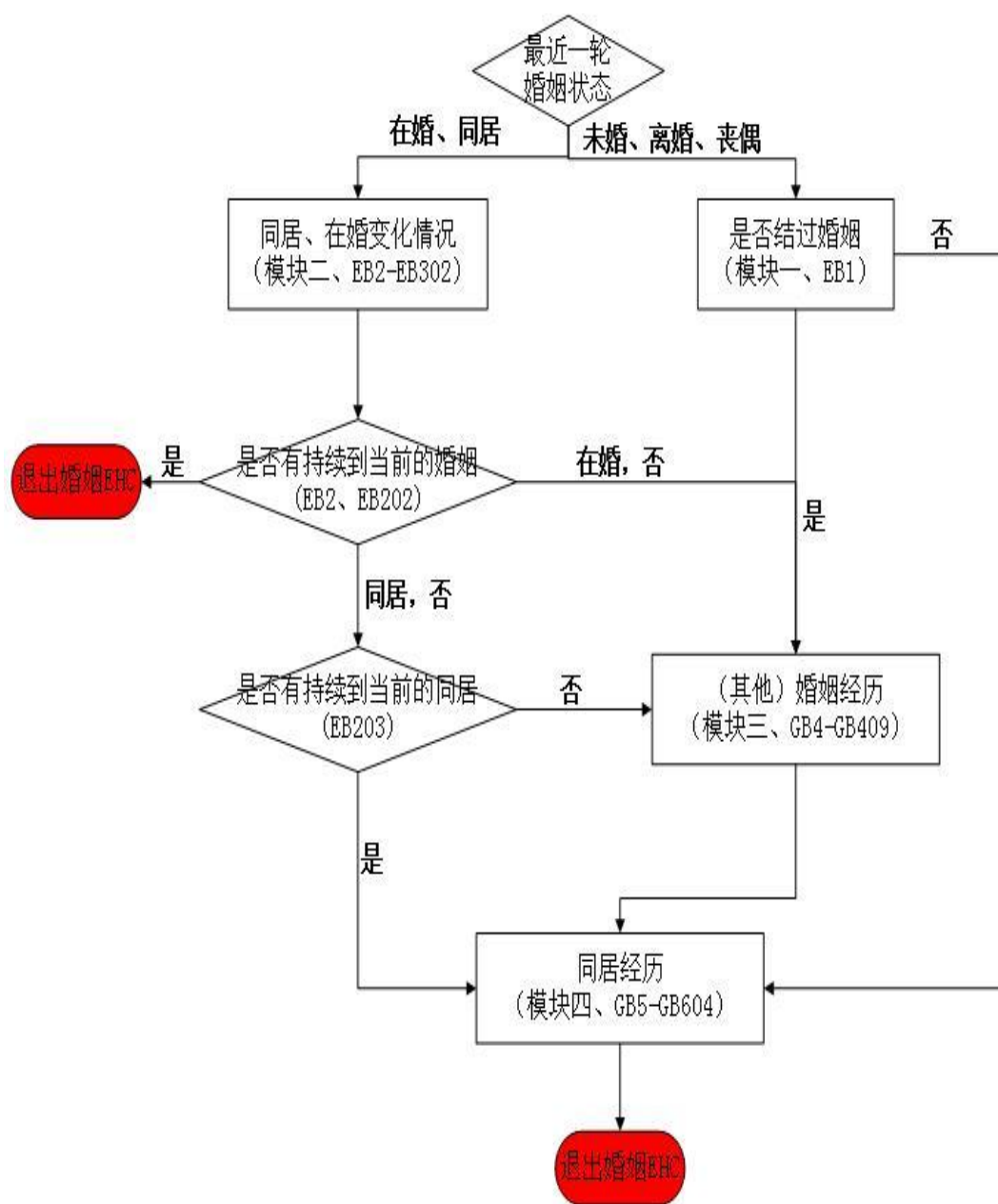
3.2.1 问题模块：

- (1) 模块一、是否结过婚
详见问卷 EB1
- (2) 模块二、当前同居或在婚的变化情况
详见问卷 GB2-GB302。
- (3) 模块三、(其他) 婚姻经历：
详见问卷 GB4-GB409
- (4) 模块四、同居经历：
详见问卷 GB5-GB605

3.2.2 根据最近一轮次婚姻状态不同提问不同内容，具体如下：

- ✓ 未婚：模块一、模块三、模块四
- ✓ 在婚：模块一、模块二、模块三、模块四
- ✓ 同居：模块一、模块二、模块三、模块四
- ✓ 离婚：模块一、模块三、模块四
- ✓ 丧偶：模块一、模块三、模块四

3.2.3 婚姻模块访问流程图



图七、婚姻模块访问流程图

3.3 记录方式

整体记录:

	2012年-2013年					
	1月	2月	3月	11月	12月
加载婚姻配偶/同居伙伴	根据确认 2012 婚姻状态, 条件性出现					
在婚模块						
配偶 1	根据经历婚姻的段数, 条件性出现					
配偶 2						
配偶 n						
同居模块						
经历同居段数 (EB501)						
同居伙伴 1	根据 EB501 数量, 条件性出现, 最多不超过 4 段同居经历					
同居伙伴 2						
同居伙伴 3						
同居伙伴 4						

图八、婚姻模块 EHC 记录方式

3.4 数据校验

1. 即时校验: 参见问卷。
2. 模块结束时校验: 详见问卷模块结束时校验
3. 数据有效性校验: 在婚时段不允许重合; 同居时段允许与其他同居或在婚时段重合。

3.5 数据处理

1. 婚姻模块相关变量命名规则为:
 - (1) 其他在婚经历在问卷变量基础上加“_a_1..N”后缀作标识;
 - (2) 其他同居经历在问卷变量基础上加“_a_1..N”后缀作标识;
 - (3) 生成以下特殊变量: cohabitN、currentmorc、currentm 具体生成规则参照问卷
 - (4) 其他变量名参见问卷。

四. 工作模块

工作模块主要采集每份工作雇主、开始时间和结束时间。具体设计如下：

4.1 加载变量

详见问卷加载变量

4.2 提问内容

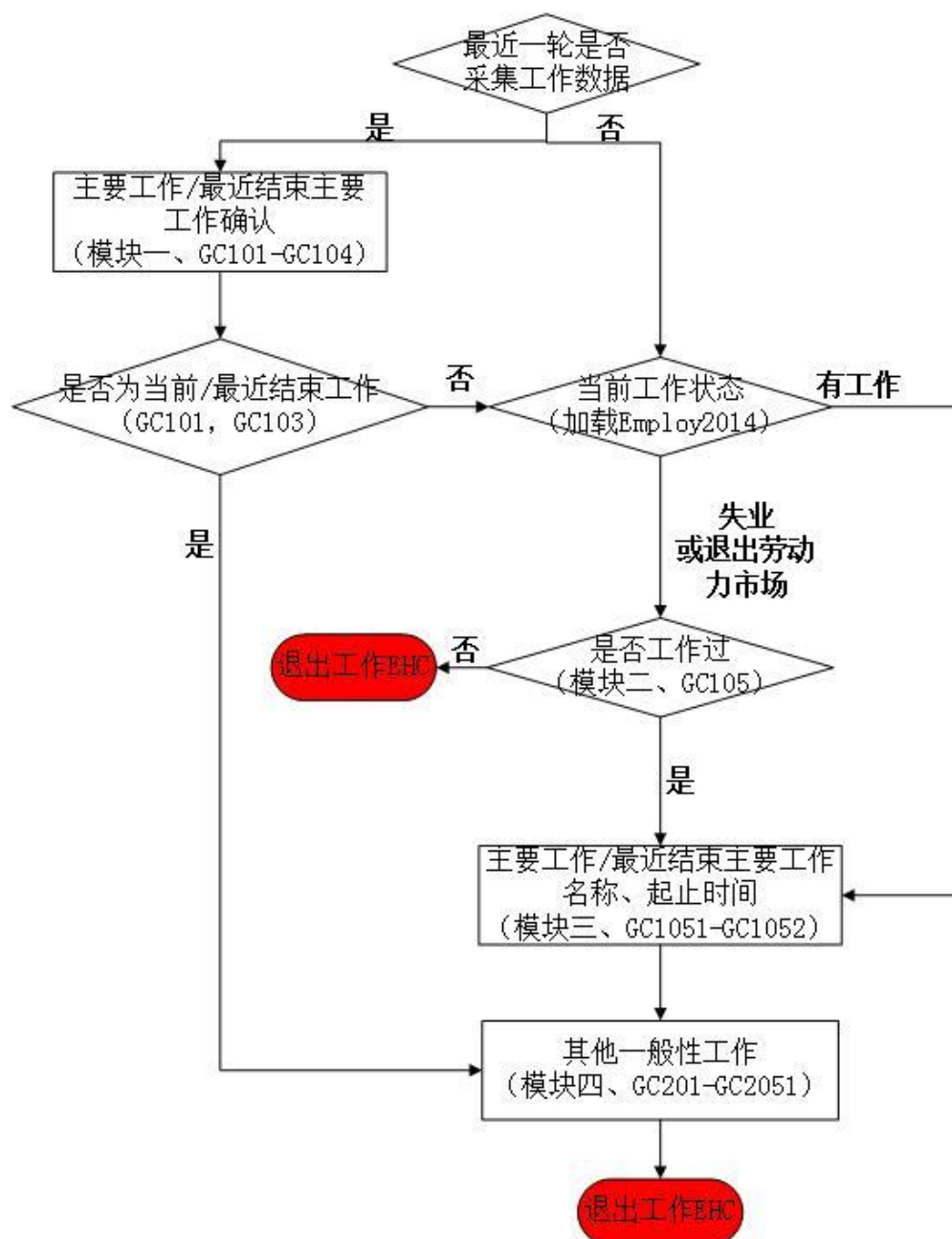
4.2.1 问题模块

- (1) 模块一、加载最近一轮调查主要工作
详见问卷 GC101-GC104。
- (2) 模块二、回忆时段内是否有过工作
详见问卷 GC105
- (3) 模块三、当前主要工作/最近结束的工作
详见问卷 GC1051-GC1053。
- (4) 模块四、其他一般工作
详见问卷 GC201-GC2021。

4.2.2 根据人员类型不同有所不同，具体如下：

- (1) 最近一轮成功完成个人问卷且最近一轮调查时有主要工作且数据有效：
 - ✓ 模块一、模块二、模块三、模块四
- (2) 其他人员：
 - ✓ 模块二、模块三、模块四。

4.2.3 工作模块访问流程图



图九、工作模块访问流程图

4.3 记录方式

整体记录方式参见图十：

	2012年-2013年					
	1月	2月	3月	11月	12月
MAINJOB_2012/update						
是否有过工作（条件性出现）						
MAINJOB_2014	条件性出现					
GC201 一般工作						
一般工作 1	根据 GC201 一般工作数出现，最多不超过 10 份工作					
一般工作 2						
一般工作 3-9						
一般工作 10						

图十、工作模块 EHC 记录方式

4.4 数据校验

1. 即时校验：详见问卷中题目设计。
2. 模块结束时校验：详见问卷模块结束时校验
3. 数据有效性校验：允许同一时段内存在多个工作。

4.5 数据处理

1. 将所有变量导入Blaise问卷，工作相关信息命名规则为：
 - (1) 主要工作相关变量参见问卷；
 - (2) 上轮调查时的主要工作，若本轮已终结，视为“一般工作”。相关信息记录在以“_b_1”为后缀的变量中。
 - (3) 本轮调查新采集的其他一般工作在问卷变量基础上加“_a_1..N”后缀作标识；
 - (4) 生成以下特殊变量：MAINJOB_2012_update, MAINJOB_2014, JOBCLASS_base, MIncomeA_n, YIncomeA_n, IncomeA具体生成规则参见问卷；
 - (5) 其他变量名参见问卷。