

不同缺失机制并存时应答偏倚校正方法 在医保学生医疗费用调查中的应用

张海霞 赵俊康 顾彩姣 崔燕 荣惠英 孟繁龙 王彤

【导读】 在研究2012年太原市城镇居民医保参保学生(幼儿园至大学)的医疗费用及其影响因素时,发现因变量数据中同时存在随机无应答偏倚(随机缺失)和选择性偏倚(非随机缺失),为此本研究提出一个多重填补与样本选择模型相结合的两阶段策略,同时校正这两种偏倚。实例中经过两阶段抽样、问卷调查,整理获得合格数据1 190例,因变量中存在2.52%非随机缺失和7.14%随机缺失。第一阶段利用完整数据对随机缺失进行多重填补,第二阶段对填补后的数据利用样本选择模型校正非随机缺失,同时建立多因素分析模型。通过1 000次两阶段校正模拟研究比较4种不同多重填补方法,得出在此缺失比例组合下预测均数匹配法与样本选择模型结合的校正效果最优。最终在实例分析中得到影响太原市居民医保参保学生年度医疗费用的因素有被调查者类型、家庭年毛收入、对医疗费用水平的承受程度、慢性病、到社区卫生服务或私人诊所就诊、到医院门诊就诊、住院、是否有应住院而未住院情况、自我医疗、可接受的自付医疗费用比例。表明应用多重填补与样本选择模型相结合的两阶段校正方法,可有效校正调查数据因变量中存在的随机无应答偏倚和选择性偏倚。

【关键词】 随机缺失; 多重填补; 样本选择模型; 两阶段校正方法

Study on correction of data bias caused by different missing mechanisms in survey of medical expenditure among students enrolling in Urban Resident Basic Medical Insurance Zhang Haixia¹, Zhao Junkang², Gu Caijiao¹, Cui Yan¹, Rong Huiying³, Meng Fanlong⁴, Wang Tong¹. 1 Department of Health Statistics, Shanxi Medical University, Taiyuan 030001, China; 2 The Health Supervision Institution of Taiyuan Municipal Health Bureau; 3 Medical Insurance Office, the Second Hospital of Shanxi Medical University; 4 Department of Urban Resident Basic Medical Insurance, Medical Insurance Management Service Center

Corresponding author: Wang Tong, Email: wtstat1@sina.com

This work was supported by a grant from the National Natural Science Foundation of China (No. 81072385).

【Introduction】 The study of the medical expenditure and its influencing factors among the students enrolling in Urban Resident Basic Medical Insurance (URBMI) in Taiyuan indicated that non response bias and selection bias coexist in dependent variable of the survey data. Unlike previous studies only focused on one missing mechanism, a two-stage method to deal with two missing mechanisms simultaneously was suggested in this study, combining multiple imputation with sample selection model. A total of 1 190 questionnaires were returned by the students (or their parents) selected in child care settings, schools and universities in Taiyuan by stratified cluster random sampling in 2012. In the returned questionnaires, 2.52% existed not missing at random (NMAR) of dependent variable and 7.14% existed missing at random (MAR) of dependent variable. First, multiple imputation was conducted for MAR by using completed data, then sample selection model was used to correct NMAR in multiple imputation, and a multi influencing factor analysis model was established. Based on 1 000 times resampling, the best scheme of filling the random missing values is the predictive mean matching (PMM) method under the missing proportion. With this optimal scheme, a two stage survey was conducted. Finally, it was found that the influencing factors on annual medical expenditure among the students enrolling in URBMI in Taiyuan included population group, annual household gross income, affordability of medical insurance expenditure, chronic disease, seeking

DOI: 10.3760/cma.j.issn.0254-6450.2015.05.024

基金项目: 国家自然科学基金(81072385)

作者单位: 030001太原, 山西医科大学公共卫生学院卫生统计教研室(张海霞、顾彩姣、崔燕、王彤); 太原市卫生局卫生监督所(赵俊康); 山西医科大学第二医院医保办公室(荣惠英); 太原市医疗保险管理服务中心居民医保科(孟繁龙)

通信作者: 王彤, Email: wtstat1@sina.com

medical care in hospital, seeking medical care in community health center or private clinic, hospitalization, hospitalization canceled due to certain reason, self medication and acceptable proportion of self-paid medical expenditure. The two-stage method combining multiple imputation with sample selection model can deal with non response bias and selection bias effectively in dependent variable of the survey data.

【Key words】 Missing at random; Multiple imputation; Heckman selection model; Two-stage method of bias correction

流行病学调查中常会出现数据缺失,目前对此有多种处理方法,但这些方法对数据的缺失机制均有严格要求。Rubin^[1]认为缺失机制主要包括完全随机缺失(missing completely at random, MCAR)、随机缺失(missing at random, MAR)和非随机缺失(not missing at random, NMAR)三类。MCAR数据缺失的概率与任何观测到或未观测到的变量值均无关,完全观测个体使用任何分析方法均有效;MAR数据缺失的概率与其他变量有关,但与该变量本身取值无关,一般是采用多重填补(multiple imputation, MI)技术^[2-3];而NMAR则对于数据缺失的概率与其他变量及该变量本身均有关,可采用样本选择模型(sample selection model)给予校正^[4-5]。本研究在实际调查中以城镇居民医疗参保学生为目标人群,由于其收入偏低属弱势人群,且较普通人群更易因贫致病和因病致贫^[6],因此在该人群的医疗费用需求调查中因变量会存在样本选择偏倚,即因经济困难等原因患病者自主选择不就医。据第四次国家卫生服务调查结果^[7],2008年我国有病但不就医的居民约占38%,应住院而未住院的病例达21%,其中城镇居民医保人群达32.5%,因“经济困难”者占70.3%。自主选择不就医者其应答表现出零消费需求的真值是未知的,并与未患病的真正零消费需求不同,应视为缺失数据^[8]。如果用直接删除或将因变量取0的方法处理此类缺失数据会忽视这种选择偏倚。同时,此类调查往往存在随机无应答偏倚^[9],而随机无应答偏倚通常假定为随机缺失。以往针对缺失数据分析的研究文献通常仅针对一种缺失机制提出方法,而上述调查的因变量中两种缺失机制并存,为此本研究在实际调查中提出一个两阶段策略,同时校正选择性偏倚和随机无应答偏倚。

基本原理

两阶段校正策略就是在辨识出因变量中随机和非随机缺失数据的前提下,在第一阶段利用完整数据对随机缺失进行多重填补,第二阶段对填补后的数据利用样本选择模型处理非随机缺失,同时建立多因素分析模型^[10]。由于实际调查中同时存在随机

缺失与非随机缺失现象,且针对随机缺失有不同的填补方法,而对不同填补方法和处理非随机缺失的样本选择模型还需组合,故本研究首先是基于调查数据中随机与非随机缺失的比例,通过模拟研究以确定第一阶段哪种随机填补方法与第二阶段的样本选择模型结合,更好分析其效果,再将该最优两阶段方法应用于实际调查数据分析。其中第一阶段利用完整数据对单纯无应答数据按照随机缺失机制分别应用预测均值匹配法(predictive mean matching, PMM)、倾向性得分法(propensity score, PS)、基于Bootstrap的EM算法(Bootstrap with expectation maximization algorithm, EMB)和马尔科夫链蒙特卡罗算法(Markov chain Monte Carlo, MCMC)进行多重填补,产生多个填补后数据集,分别与非随机缺失数据结合。第二阶段使用样本选择模型校正填补后多个数据集中的非随机缺失,最后合并多个样本选择模型拟合结果。对模拟分析后的合并结果以标准偏倚、均方误差的平方根和可信区间平均长度作为评价填补方法优劣的标准,选出最佳填补方案应用于实际调查数据分析,以正确估计潜在医疗费用需求及其影响因素。

1. MI:主要分3个步骤。第一步填补,即对每个随机缺失值均从完整数据集中通过一定方法产生 m 个填补值,并与完整数据集合并成 m 个填补后数据集。第二步分析,是对 m 个填补后数据集分别应用既定方法予以分析(本研究是将 m 个填补后数据分别与非随机缺失数据合并,采用样本选择模型校正非随机导致的偏倚)。第三步为合并,即对 m 组拟合结果进行参数合并,获得最终的统计推断^[11-12]。本研究应用PMM、PS、EMB和MCMC四种填补方法^[10-16]。

2. 样本选择模型^[5]:主要通过因变量受到某种限制但自变量可以完全观测到的样本来估计潜在总体参数,如

$$y_0 = x_1 \beta + \varepsilon \quad \text{结果等式 (1)}$$

$$d_0 = x_2 \gamma + \nu \quad \text{选择等式 (2)}$$

$$d = 1(d_0 \geq c) \quad d = 0(d_0 < c) \quad (3)$$

$$y_1 = y_0 \times d \quad (4)$$

式中 y_0 是真值, 与 x_1 建立的回归方程为结果等式; y_1 是实际观测值; d 指示 y_0 是否可被观测到, 其取值与潜变量 d_0 和 c 相关, c 为 d_0 的界值, 当 $d_0 > c$ 时, $d=1$, 此时 y_0 可被观测到, 且 $y_1=y_0$; 反之当 $d_0 \leq c$ 时, $d=0$ 时, 真值 y_0 无法观测到, $y_1=0$, 但该 0 值非真值, 应视为缺失。

样本选择模型要求 ε 和 ν 相关且 $E[\varepsilon | \nu] \neq 0$, 同时结果等式中 x_1 和 ε 相关, 因此应用最小二乘法无法得到参数 β 的一致估计。本研究应用参数两步估计法^[5]: ①选择等式(2)中 d_0 无法被观测到, 可以构建以二分类变量 d 为因变量的 Probit 模型, 然后应用最大似然估计获得参数 γ 的一致估计。②通过 γ 得到每个人的预测值 $\tau = (x_2\gamma - c)/\sigma_\nu$, 取 τ 值的密度函数与分布函数之比, 即得 $\lambda(\tau) = \frac{\varphi(\tau)}{1 - \Phi(\tau)}$ 。③在以观测值 y_1 为因变量的结果等式中加入校正项 $\sigma_\varepsilon \rho_{\varepsilon\nu} \lambda$, 应用最小二乘法得到 $\sigma_\varepsilon \rho_{\varepsilon\nu}$ 和 β 的一致估计量。

实例分析

1. 模拟研究:

(1)生成模拟数据集: 根据调查数据 ($n=1\ 190$), 因变量中随机缺失比例(7.14%)和非随机缺失比例(2.52%)生成模拟数据。为使缺失例数为整数, 故模拟时取 $n=1\ 200$, 随机缺失比例取值为 2.5%, 非随机缺失比例取值为 7.0%。模拟结果等式和选择等式中的 6 个变量为 $x_1, x_2, y_0, d_0, \varepsilon, \nu$, 其中 x_1 和 x_2 服从双变量正态分布, $\bar{x}=0, s=1, r=0$; ε 和 ν 服从双变量正态分布, $\bar{x}=0, s=1, r=0.75$ 。 y_0 由公式 $y_0 = e^{1+x_1+\varepsilon}$ 求出, d_0 由公式 $d_0 = 1 + x_2 + \nu$ 求出。对模拟数据的因变量以 2.5% 的非随机缺失比例向下截取产生缺失数据, 即以 2.5% 缺失比例给出 y_0 定义一个界值 c , 令 $d_0 > c$ 时, $d=1$ 且 $y_1=y_0$; $d_0 \leq c$ 时, $d=0$ 且 y_1 为缺失。对 y_1 进行对数转换, 使调查中因变量近似服从正态分布。第二步针对 $d=1$ 的个体以 7.0% 的比例随机产生缺失, 对应于调查中因变量的随机缺失数据^[5,10]。

(2)模拟分析: 第一阶段为随机缺失的多重填补, 即对 $d=1$ 的个体(排除非随机缺失)分别应用 PMM、MCMC、EMB 和 PS 进行多重填补, 然后将填补后的数据与 $d=0$ (仅非随机缺失)的数据合并。其中 PMM、PS、MCMC 选用 SAS 9.2 软件、EMB 法选用 R 软件进行分析^[10]。第二阶段为样本选择偏倚的校正, 即应用样本选择模型的参数两步估计, 得到校正选择性偏倚后的回归系数估计值。重复抽样 1 000 次, 最后根据 1987 年 Rubin 提出的合并参数和方差估

计值的准则^[17], 合并 1 000 个结果等式中 x_1 的回归系数(β)和标准误(s_ε)。选用 SAS 9.2 软件实现。

(3)评价标准: ①标准偏倚(standardized bias)是评价准确度的指标, 在方法评价指标中应首要观测。将 ± 0.4 作为评价准确度的上下界值, 若超出此范围, 偏倚就会对误差率、功效和可信区间覆盖率产生明显的负面影响; ②可信区间平均长度(length)是评价精确度的指标, 该值越小, 其精确度越高; ③均方误差的平方根(RMSE) = $(\hat{\beta} - \beta)^2 + [s_\varepsilon(\hat{\beta})]^2$, 该指标联合了偏倚和效率, 因此在准确度都相同的情况下, RMSE 越小, 方法精确度越高^[18]。

(4)模拟结果分析: 从表 1 可知, 当非随机缺失比例为 2.5%、随机缺失比例为 7.0% 时, 4 种填补方法的标准偏倚绝对值不等, 但均未超出所规定的界值, 其中 PMM 法的标准偏倚绝对值最小, 故其准确度最高; 4 种填补方法的 RMSE 相近, PMM 法仍最小; PMM 法可信区间平均长度最短, 故其准确度较好。综合上述结果实际调查数据应采用 PMM 法下的样本选择模型分析。

表 1 采用 4 种填补方法的样本选择模型结果等式回归系数估计值及其评价指标 (非随机缺失比例 2.5%, 随机缺失比例 7.0%)

方法	MCMC	EMB	PS	PMM
$\hat{\beta}$	1.000 1	1.000	0.993	1.000
$s_\varepsilon(\hat{\beta})$	0.031	0.031	0.031	0.031
标准偏倚	-0.047	0.006	-0.227	-0.005
RMSE	0.031	0.031	0.032	0.031
可信区间平均长度	0.119	0.118	0.120	0.117

2. 调查实例:

(1)设计: “2012 年太原市城镇居民医保参保学生医疗费用及其影响因素的调查”采用两阶段抽样设计。第一阶段为分层抽样, 将太原市参保学校分 5 层(大学、高中、初中、小学和幼儿园); 第二阶段为整群随机抽样, 以太原市参保高校整群随机抽取 1 所大学; 其余层以太原市 10 个区的学校进行两级整群随机抽样, 每层各一所。问卷设计主要包括被调查者的一般情况、家庭人员构成和健康状况、环境、2012 年医疗消费情况等。最终共收集合格问卷 1 190 份。该人群基本情况见表 2。

(2)缺失机制甄别: 因变量是个人自付的年度医疗费用, 为使其近似服从正态分布, 进行对数转换, 即 $y' = \log(y_1 + 1)$ 。 y' 共缺失 115 例, 根据问卷中 3 个问题来甄别其缺失机制: ①“年内是否患病”; ②

“年内是否就医”;③“年内医疗费用值”。随机缺失情况包括3个问题中至少有1个缺失,共85例,占总人数7.14%,占 $d=1$ (仅存在随机缺失)人群7.33%。非随机缺失为“年内患病,且年内没有就医”者,共30例,占总人数2.52%。

表2 太原市城镇居民医保学生医疗费用调查人群基本情况($n=1\ 190$)

人群类别	构成比(%)	年龄(岁)	性别比(男:女)
大学生	78.32	17~25	27.04:72.96
高中生	3.28	14~17	48.78:51.28
初中生	5.38	12~15	42.19:57.81
小学生	9.75	7~12	53.45:46.55
幼儿园	3.28	3~5	51.28:48.72
合计	100.00	3~25	31.93:61.07

(3)分析方法:

【方法一】应用PMM与样本选择模型相结合的两阶段校正方法对调查数据进行分析($n=1\ 190$)。其中①样本选择模型结果等式中的自变量包括被调查类型、性别、家庭年收入、自我感受健康状况、对医疗费用水平的承受程度、慢性病、健康体检、到社区卫生服务或私人诊所就诊、到医院门诊就诊、住院、应住院而未住院的情况、自我医疗、可接受的自付医疗费用比例;②样本选择模型选择等式中的自变量除相同于结果等式外,还增加了与承受费用相关的外生解释变量^[5],即家庭类别、家庭人口数、家庭无收入人口数、户主年龄及其受教育程度、生活满意度;③PMM填补时回归模型中自变量同选择等式中的自变量。

【方法二】将因变量中缺失数据删除,对剩余完整数据进行线性回归分析,模型中的自变量同结果等式自变量($n=1\ 075$)。

影响因素的赋值和分析结果见表3、4。

(4)结果:从表4可见两阶段校正方法和删除缺失数据后线性回归分析的结果不一致。前者筛出的变量较多,且校正项MMILLS有统计学意义,进一步提示对非随机缺失的校正有意义。两阶段校正方法在第一阶段通过PMM将随机缺失数据填补完整以校正随机无应答偏倚;第二阶段采用样本选择模型,利用完整自变量信息校正受限的因变量后,再估计潜在总体参数,从而校正选择性偏倚,相对于直接删除缺失的方法,其所利用信息更全面(两阶段校正方法 $n=1\ 190$,直接删除法 $n=1\ 075$)。

分析两阶段校正方法的结果,提示影响太原市城镇居民医保学生年度医疗费用的因素包括被调查者类型、家庭年收入、对医疗费用水平的承受程度、

表3 影响因素赋值

自变量	赋值
性别	1=男,2=女
被调查者类型	1=幼儿园儿童,2=小学,3=初中,4=高中,5=大学
自我感受健康状况	1=很健康,2=比较健康,3=一般,4=不太健康,5=不健康
对医疗费用水平的承受程度	1=承受不起,2=承受有困难,3=还可承受,4=可完全承受
慢性病	1=是,0=否
健康体检	1=是,0=否
到社区卫生服务或私人诊所就诊	1=是,0=否
到医院门诊就诊	1=是,0=否
住院	1=是,0=否
应住院而未住院的情况	1=是,0=否
自我医疗	1=是,0=否
可接受的自付医疗费用比例(%)	1= ≤ 10 , 2=10~20, 3=20~30, 4=30~40, 5=40~50, 6=50~60, 7=60~70, 8=70~80, 9=80~90, 10=90~100
家庭类别	1=城市,2=农村
户主年龄(岁)	1= ≤ 20 , 2=20~30, 3=30~40, 4=40~50, 5=50~60, 6= ≥ 60
户主受教育程度	1=文盲或少识字,2=小学,3=初中,4=中专/中技,5=高中/技校,6=大专,7=本科或以上
生活满意度	1=很满意,2=比较满意,3=一般,4=不太满意,5=不满意

慢性病、到社区卫生服务或私人诊所就诊、到医院门诊就诊、住院、应住院而未住院、自我医疗、可接受的自付医疗费用比例。

(5)结论:应用多重填补与样本选择模型相结合的两阶段校正方法,可有效校正调查数据因变量中存在的选择性偏倚和随机无应答偏倚,从而较准确估计出城镇居民医保学生的医疗费及其影响因素。

讨 论

本研究提出的两阶段校正策略可有效校正因变量中随机无应答偏倚和选择性偏倚,并通过实例分析估计出太原市城镇居民医保学生医疗费用的影响因素。其中被调查者类型和慢性病患者情况对个人医疗费用支出均有显著意义;以幼儿园儿童的医疗费用支出最高;家庭收入和对医疗费用水平的承受程度对医疗费用支出有影响(收入越高,对医疗费用水平的承受程度越高,医疗费用支出越高);社区服务站或私人诊所、医院门诊和住院等就医情况、应住院而未住院、自我医疗发生情况均直接影响医疗费用的支出。医疗卫生服务的利用情况是人群医疗费用需求的重要因素。

本研究存在不足。首先模拟研究时回归模型为

表4 医疗费用影响因素

变 量	两阶段校正后参数估计		删除缺失后参数估计	
	$\hat{\beta}$	95%CI	$\hat{\beta}$	95%CI
被调查者类型(与幼儿园儿童比较)				
小学	-0.668 ^a	-1.224 ~ -0.111	-0.643 ^a	-1.121 ~ -0.165
初中	-0.083	-0.326 ~ 0.557	-0.563	-1.083 ~ -0.043
高中	-1.363 ^a	-2.035 ~ -0.692	-1.200 ^a	-1.796 ~ -0.609
大学	-0.600 ^a	-1.061 ~ -0.059	-0.879 ^a	-1.298 ~ -0.459
性别	0.096	-0.099 ~ 0.292	0.117	-0.049 ~ 0.283
家庭年收入	0.158 ^a	0.076 ~ 0.243	-0.016	-0.074 ~ 0.042
自我感受健康状况	-0.164	-0.335 ~ 0.007	0.298 ^a	0.201 ~ 0.395
对医疗费用水平的承受程度	0.293 ^a	0.144 ~ 0.441	-0.015	-0.116 ~ 0.085
慢性病	6.121 ^a	3.921 ~ 8.320	-0.811	-1.642 ~ 0.021
健康体检	-0.084	-0.318 ~ -0.150	0.097	-0.094 ~ 0.288
到社区卫生服务或私人诊所就诊	8.532 ^a	6.563 ~ 10.501	1.718 ^a	1.554 ~ 1.882
到医院门诊就诊	9.309 ^a	7.255 ~ 11.363	2.254 ^a	2.081 ~ 2.428
住院	0.656 ^a	0.165 ~ 1.148	0.520 ^a	0.197 ~ 0.843
应住院而未住院的情况	1.944 ^a	0.976 ~ 2.912	0.765	0.108 ~ 1.422
自我医疗	9.079 ^a	7.032 ~ 11.127	1.903 ^a	1.732 ~ 2.074
可接受的自付医疗费用比例	-0.067 ^a	-0.109 ~ -0.025	-0.047 ^a	-0.083 ~ -0.011
MMILLS	-0.801 ^a	-1.059 ~ -0.544	-	-

注: ^a 差异有统计学意义($\alpha=0.05$)

单个自变量,未考虑多个自变量及其之间所存在的相关性,致使模拟研究不完全反映实际数据,应改进模拟研究。其次仅调查了参加城镇居民医保学生群体的医疗费用,应进一步调查城镇居民医保其他覆盖人群,全面准确了解该弱势人群的医疗费用需求及其影响因素。再有问卷中家庭收入的变量较敏感,自报信息其可信程度较差。

参 考 文 献

[1] Rubin DB. Inference and missing data[J]. *Biometrika*, 1976, 63(3):581-592.

[2] Little RJA, Rubin DB. *Statistical analysis with missing data*[M]. 2nd ed. New York: John Wiley & Sons, 2002.

[3] Schafer JL, Graham JW. Missing data: our view of the state of the art[J]. *Psychol Methods*, 2002, 7(2): 147-177.

[4] Xue XP, Shi DP, Wang T. The statistical model of limited dependent variable and semiparametric estimation of it[J]. *Chin J Health Stat*, 2007, 24(2):211-213. (in Chinese)
薛小平, 史东平, 王彤. 受限因变量模型及其半参数估计[J]. *中国卫生统计*, 2007, 24(2):211-213.

[5] Zhang L, Wang LN, Wang T. Likelihood estimation and two-step estimation of sample selection model[J]. *Mod Prev Med*, 2007, 34(9):1607-1609. (in Chinese)
张磊, 王琳娜, 王彤. 样本选择模型的似然估计与两步估计[J]. *现代预防医学*, 2007, 34(9):1607-1609.

[6] Li J, Sun XH. Status analysis of the urban resident basic medical insurance in Taiyuan [J]. *Chin Remedies Clin*, 2013, 13(12): 1644-1645. (in Chinese)
李静, 孙小红. 太原市城镇居民基本医疗保险现状分析[J]. *中国药物与临床*, 2013, 13(12):1644-1645.

[7] Center of Health Statistics and Information, Ministry of Health of the People's Republic of China. A analysis report of the Forth National Health Services Survey in China [DB/OL]. [2015-04-09]. <http://www.nhfpc.gov.cn/cmsresources/mohwsbwstjxxzx/cmsrsdocument/doc9911.pdf>. (in Chinese)
卫生部统计信息中心. 第四次国家卫生服务调查主要结果 [DB/OL]. [2015-04-09]. <http://www.nhfpc.gov.cn/cmsresources/mohwsbwstjxxzx/cmsrsdocument/doc9911.pdf>.

[8] Başer O, Bradley CJ, Gardiner JC, et al. Testing and correcting for non-random selection bias due to censoring: an application to medical costs [J]. *Health Serv Outcomes Res Methodol*, 2003, 4(2):93-107.

[9] Peytchev A, Baxter RK, Carley-Baxter LR. Not all survey effort is equal [J]. *Public Opinion Quarterly*, 2009, 73(4): 785-806.

[10] Zhao JK, Wang T, Rong HY, et al. A statistical simulation study of bias correction when the different missing mechanism coexist [J]. *Chin J Health Stat*, 2014, 31(4):570-574. (in Chinese)
赵俊康, 王彤, 荣慧英, 等. 不同缺失机制并存时倚倚校正的模拟研究[J]. *中国卫生统计*, 2014, 31(4):570-574.

[11] Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation [J]. *Computat Stat Data Anal*, 1996, 22(4): 425-446.

[12] Allison PD. Multiple imputation for missing data a cautionary tale [J]. *Sociol Methods Res*, 2000, 28(3):301-309.

[13] Honaker J, King G. What to do about missing values in time-series cross-section data [J]. *Amer J Political Science*, 2010, 54(2):561-581.

[14] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm [J]. *J Royal Statist Soc Ser B (Methodolog)*, 1977, 39(1):1-38.

[15] Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation [J]. *J Am Statist Assoc*, 1987, 82(398): 528-540.

[16] Schunk D. A Markov chain Monte Carlo multiple imputation procedure for dealing with item nonresponse in the German save survey [G]. MEA-Mannheim Research Institute for the Economic of Aging, University of Mannheim, 2007: 121.

[17] Donald B. *Multiple imputation for nonresponse in surveys*[M]. New York: John Wiley & Sons, 1987.

[18] Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures [J]. *Psychol Methods*, 2001, 6(4):330-351.

(收稿日期:2014-10-27)
(本文编辑:张林东)