

MSM人群HIV感染者病毒载量抽样调查 缺失数据填补方法研究

江震 斗智 宋炜路 徐杰 吴尊友

102206 北京,中国疾病预防控制中心性病艾滋病预防控制中心预防干预室(江震、斗智、宋炜路、徐杰); 102206 北京,中国疾病预防控制中心性病艾滋病预防控制中心(吴尊友)

通信作者:徐杰, Email: xujie@chinaaids.cn

DOI: 10.3760/cma.j.issn.0254-6450.2017.11.025

【摘要】 目的 探讨不同缺失数据填补法对MSM人群HIV感染者(MSM感染者)病毒载量(VL)缺失数据的填补效果。**方法** 以2013年中国16个大城市MSM感染者VL抽样检测数据为基础,采用SPSS 17.0软件,模拟完整数据集和5种不同类型的缺失数据集,采用最大期望值法(EM)、回归法、均值填补法、删除法、马尔科夫链蒙特卡罗法(MCMC)对5种VL缺失数据填补处理,从数据分布、准确度、精确度3个方面比较填补效果。**结果** VL数据呈偏态非连续分布,难以进行有效正态分布转化;不同填补方法对完全随机缺失数据填补效果均较好;对于其他类型缺失数据,回归法、MCMC较好保留完整数据主要分布特征;EM、回归法、均值填补法、删除法普遍低估数据均值,MCMC多高估数据均值。**结论** MCMC可作为首选的VL数据对数转换后缺失数据填补方法。填补数据可作为调查人群VL均值水平估算的参考依据。

【关键词】 艾滋病病毒; 病毒载量; 缺失数据; 多重填补; 马尔科夫链蒙特卡罗法

基金项目: 国家科技重大专项(2012ZX10001-007-005)

Comparison of different methods in dealing with HIV viral load data with diversified missing value mechanism on HIV positive MSM Jiang Zhen, Dou Zhi, Song Weilu, Xu Jie, Wu Zunyou
Division of Prevention and Intervention (Jiang Z, Dou Z, Song WL, Xu J), National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China (Wu ZY)
Corresponding author: Xu Jie, Email: xujie@chinaaids.cn

【Abstract】 Objective To compare results of different methods in organizing HIV viral load (VL) data with missing values mechanism. **Methods** We used software SPSS 17.0 to simulate complete and missing data with different missing value mechanism from HIV viral loading data collected from MSM in 16 cities in China in 2013. Maximum Likelihood Methods Using the Expectation and Maximization Algorithm (EM), regressive method, mean imputation, delete method, and Markov Chain Monte Carlo (MCMC) were used to supplement missing data respectively. The results of different methods were compared according to distribution characteristics, accuracy and precision. **Results** HIV VL data could not be transferred into a normal distribution. All the methods showed good results in iterating data which is Missing Completely at Random Mechanism (MCAR). For the other types of missing data, regressive and MCMC methods were used to keep the main characteristic of the original data. The means of iterating database with different methods were all close to the original one. The EM, regressive method, mean imputation, and delete method under-estimate VL while MCMC overestimates it. **Conclusion** MCMC can be used as the main imputation method for HIV virus loading missing data. The iterated data can be used as a reference for mean HIV VL estimation among the investigated population.

【Key words】 HIV; Viral load; Missing data; Multiple imputation; Markov Chain Monte Carlo

Fund program: National Major Science and Technology Project of China (2012ZX10001-007-005)

HIV社区病毒载量(community viral load, CVL)被认为是评价地区HIV传播风险的一项重要公共卫生指标。病毒载量(viral load, VL)数据缺

失是CVL收集过程中普遍存在的问题。近年来,缺失数据填补技术日趋普及,但针对不同缺失类型的数据,不同填补方法性能相差甚远^[1]。本研究结合

国家科技重大专项子项目 MSM 人群 HIV 感染者 (MSM 感染者) VL 抽样调查数据库, 探讨不同 VL 缺失数据填补方法的应用效果。

资料与方法

1. 资料来源: 数据来源于国家科技重大专项子项目“MSM 人群扩大检测扩大治疗效果评估应用研究”的 2013 年 16 个大城市 4 050 例 MSM 感染者 VL 抽样调查数据。

2. 模型构建:

(1) 完整数据库模拟: 采用马尔科夫链蒙特卡罗法 (Markov Chain Monte Carlo, MCMC) 模拟完整数据库。MCMC 是贝叶斯推断中的一种探索后验分布的方法。该方法是基于对样本分布特征的经验预测而非理论预测。其原理为: 在计算某种随机事件出现的概率, 或某个随机变量的期望值时, 通过“实验”的方法, 用该事件出现的频率估计这一随机事件的概率, 或者通过观察随机变量的某些数字特征, 将其作为问题的解。理论上, MCMC 模型需要数据符合多元正态分布假设, 随着研究的深入, 有证据认为 MCMC 可用于不符合正态分布的数据, MCMC 目前是多重填补 (multiple imputation) 主流方法。基于此^[2-3], 本研究选择 MCMC 对 HIV-VL 缺失值进行模拟, 形成完整数据库, 作为评价缺失数据填补效果评价标准。

具体方法: 在 4 050 例 VL 抽样检测数据库中, 随机抽取样本量为 100 的数据集, 根据“年龄、婚姻、性伴情况、安全性行为、是否为同伴教育员、治疗情况、最近一次 CD₄⁺T 淋巴细胞检测水平”等变量, 运用 Monte Carlo 模拟技术, 对该数据库 VL 对数值进行填补模拟 100 次, 模拟出样本量为 100 的 MSM 感染者 VL 完整数据集。

(2) 缺失数据库模拟: 完整数据产生之后, 按照不同缺失机制模拟各种类型的缺失数据库。本研究只考虑观察变量 VL 检测值的缺失, 按照缺失比例 30% 的标准 (2013 年 VL 检测值缺失率为 28.9%), 通过以下步骤^[4]产生不同机制和类型的缺失: ① 完全随机缺失 (missing completely at random) 数据库: 缺失现象完全是随机发生的, 和自身或其他变量的取值无关 (缺失和变量的取值无关), 按照机选完全随机缺失模式, 模拟 VL 检测值随机缺失率为 30% 的缺失数据库。② 随机缺失 (missing at random) 数据库模拟: 缺失值情况发生与数据集中其他无缺失变量的取值有关, 本研究中 VL 数据缺失受户籍、治疗因素

的影响, 所以按照户籍、治疗两个因素进行随机缺失模拟: 受户籍影响的随机缺失数据库: VL 缺失比例在外地户籍人群高于本地户籍人群, 将 VL 按照本地户籍 15%、外地户籍 45% 产生 VL 的缺失; 受治疗影响的随机缺失数据库: VL 缺失比例在未治疗人群高于治疗人群, 将 VL 按照治疗 15%、未治疗 45% 产生 VL 的缺失。③ 非随机缺失 (missing at not random) 数据库模拟: 数据缺失不仅和其他变量的取值有关, 也和自身的取值有关, 这种数据的缺失机制就是不能忽略的缺失。本研究按照线性、曲线两种非随机缺失机制进行模拟: 线性非随机缺失数据库: 在对应于 VL 的 4 个四分位数区间产生 VL 的缺失比例分别为 15%、25%、35%、45%。曲线非随机缺失数据库: 在对应于 VL 的 4 个四分位数区间产生 VL 的缺失比例分别为 45%、15%、15%、45%。

3. 缺失值填补方法:

(1) 最大期望值法 (maximum likelihood methods using the EM algorithm, EM): 是一种迭代算法, 每次迭代都由两步组成, E 步 (expectation step) 求出期望, M 步 (maximization step) 则将随机参数进行极大化处理。简单地说, 先给某随机变量一个初始值, 然后求出模型中各个参数估计值 (M 步), 再利用新估计出的模型对该随机变量进行估计 (E 步), 如此反复迭代, 直到模型收敛为止。本研究进行了 25 次迭代。

(2) 回归法: 以所有被选入的连续变量为自变量, 以存在缺失值的变量为应变量建立回归方程, 利用该方程对应变量相应的缺失值进行填充。

(3) 均值填补法: 根据数据库中 VL 对数值的均值来代替每个缺失值。

(4) 删除法: 删除所有 VL 缺失的样本。

(5) MCMC 多重填补: 本研究对 5 种缺失类型的数据库采取填补替代 10 次。

4. 填补效果评价参数: 利用 SPSS 17.0 软件模拟完整数据集和不同随机缺失数据集; 采用 EM 法、回归法、均值填补法、成组删除法和 Monte Carlo 法对缺失数据进行处理。从分布齐性检验、准确度、精确度三方面比较缺失数据处理效果^[5-8]。

(1) 分布齐性检验: 本研究所收集的 VL 数据对数转化后, 呈连续型非正态分布, 难以进行有效的正态分布转化。对非正态分布的数据分布的齐性检验目前还没有权威的检验方法, 目前建议的非参数检验法有 Wilcoxon 检验、Wald-Wolfowitz 游程检验 (WWR)、Kolmogorov-Smirnov 检验 (K-S)、Hollander

极端反应检验(Hollander)等。由于本研究模拟数据库总体方差、均数关系均不定,且预知差异不大,故采取 Wilcoxon 和 K-S 检验方法^[9-10],同时结合频数分布图表做辅助判断。

(2)准确度评价-标准偏倚(standardized bias): 参数估计中的原始偏倚等于重复模拟的参数平均值与总体参数值之差。偏倚对于区间估计以及假设检验的影响取决于其相对于系统变异的大小程度,故选用标准偏倚作为评价标准之一,计算公式为 $\{(B-\beta)/SE(B)\}$, B 为重复模拟填补 m 次后所有参数估计的平均值, β 为完整数据库总体参数平均值。当标准偏倚落在 ± 0.4 区间之外时,偏倚就会对功效、95%CI 覆盖率和误差率产生明显的负面影响。因此,将 ± 0.4 作为评价标准偏倚的上下界值。

(3)精确度评价-均方误差的平方根(square root of mean square error, RMSE): 均方误差(mean square error, MSE)等于参数估计值与总体参数值之间差平方的平均值。估计值的 MSE 等于偏倚的平方加上自身方差,即 $(B-\beta)^2 + [SE(B)]^2$, 该指标联合了偏倚和效率的概念。为保持与参数值相同的测量尺度,使用 MSE 的平方根,即 RMSE。在各方法的准确度都相同的情况下, RMSE 越小的方法精确度越高。

(4)95%CI 平均长度: 95%CI 平均长度表示 m 次填补数据库样本 95%CI 长度的平均值。不同填补方法比较,准确度相同,平均 95%CI 更短的方法精确度更高。

5. 统计学分析: 数据录入和整理采用 Excel 2010 软件,采用 SPSS 17.0 软件模拟完整数据集和 5 种不同类型的缺失数据,采用 EM、回归法、均值填补法、删除法、MCMC 对 5 种 VL 缺失数据填补处理和比较填补效果。

结 果

1. 分布齐性检验: VL 对数值模拟的完整数据库为偏态分布,见图 1。Wilcoxon、K-S 检验显示,仅均值填补后数据分布与完整数据库分布差异有统计学意义(表 1)。数据分布频数图显示,所有填补后的数据分布与完整数据库均发生较为明显的变化,均数填补法对分布的改变最为明显;完整数据库中有两个突出的峰值,回归法、MCMC 填补数据库大致保持了这样的分布特点(图 2)。

2. 不同填补方法处理的数据库准确度和精确度比较: 采用 5 种填补方法分别对 5 种缺失机制所构建

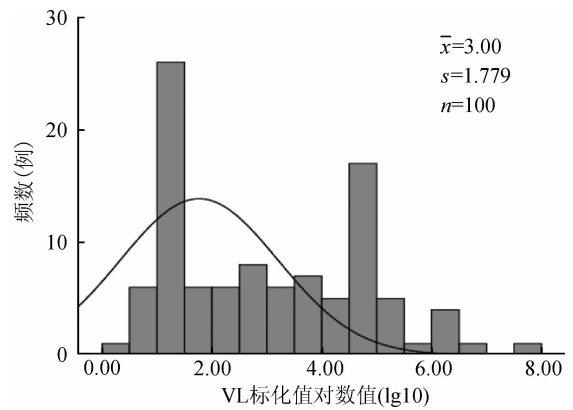


图 1 中国 16 个城市 MSM 感染者病毒载量(VL)抽样调查完整数据库 VL 标准化对数值分布

表 1 中国 16 个城市 MSM 感染者病毒载量抽样调查数据缺失处理后数据库与完整数据库的分布齐性检验($\alpha=0.05$)

不同填补方法所得数据库	与完整数据库分布齐性检验	
	Wilcoxon 检验 P 值	K-S 检验 P 值
完全随机缺失模拟数据库		
删除法	0.936	1.000
EM 模拟	0.989	0.367
回归模拟	0.871	0.906
MCMC 填补	0.941	0.994
均值填补	0.991	0.211
户籍随机缺失模拟数据库		
删除法	0.889	1.000
EM 模拟	0.800	0.367
回归模拟	0.679	0.906
MCMC 填补	0.729	0.967
均值填补	0.964	0.211
治疗随机缺失模拟数据库		
删除法	0.658	1.000
EM 模拟	0.847	0.468
回归模拟	0.856	0.967
MCMC 填补	0.406	0.813
均值填补	0.743	0.078
线性非随机缺失模拟数据库		
删除法	0.303	0.832
EM 模拟	0.335	0.078
回归模拟	0.238	0.367
MCMC 填补	0.234	0.699
均值填补	0.346	0.037
曲线非随机缺失模拟数据库		
删除法	0.303	0.832
EM 模拟	0.963	0.999
回归模拟	0.886	0.967
MCMC 填补	0.769	0.967
均值填补	0.798	0.155

注: K-S 为 Kolmogorov-Smirnov 检验

的数据库进行处理,填补后的数据库与完整数据库进行比较,其准确度、精确度以及 95%CI 长度见表 2。所有的数据填补方法所得到的数据库准确度均在 ± 0.4 以内,也即缺失数据处理结果的均值与原始数据集均值较为接近。EM、回归法、均值填补法、删

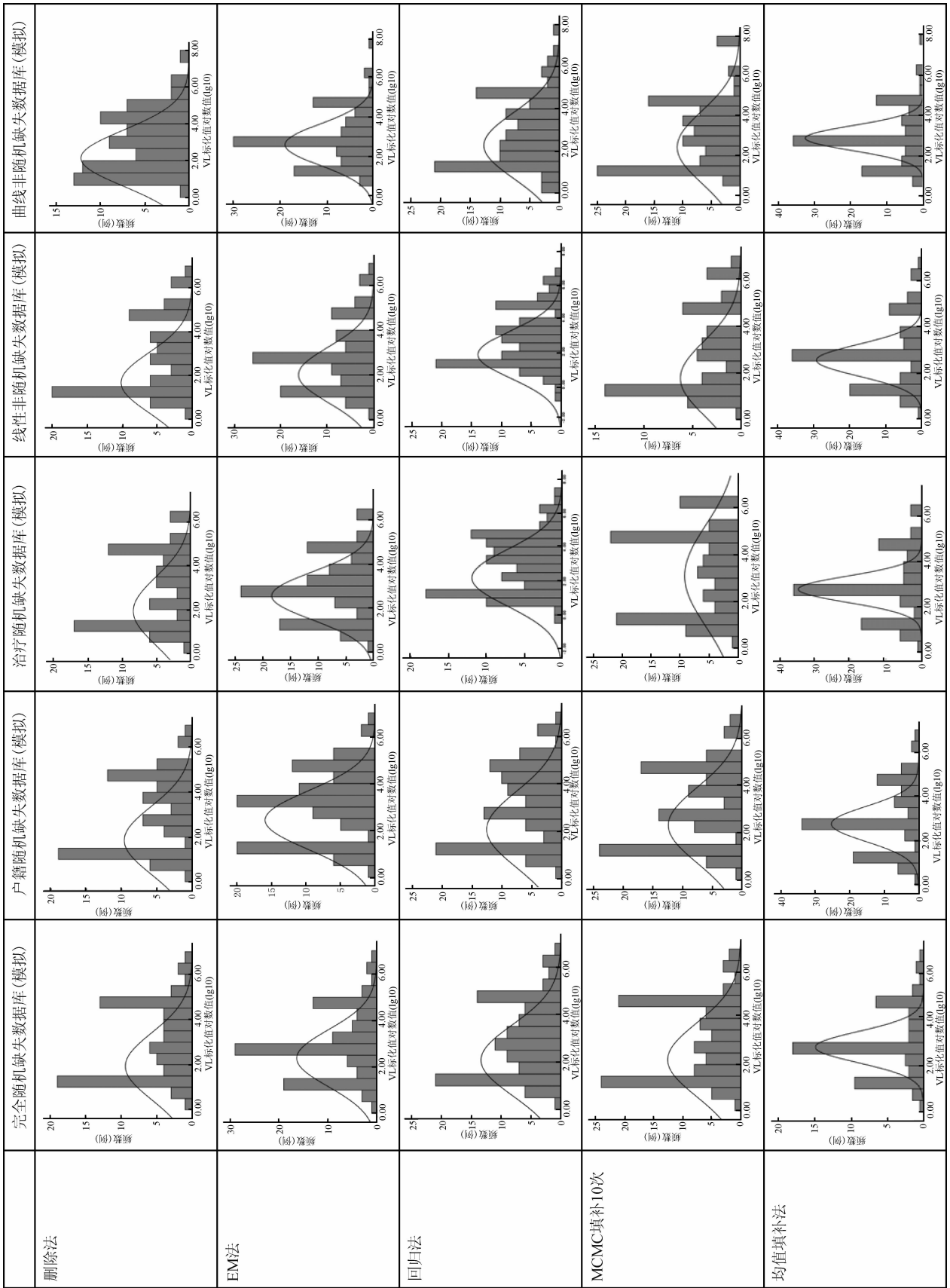


图 2 中国 16 个城市 MSM 感染者病毒载量 (VL) 抽样调查 25 个填补数据库 VL 标化对数值分布

表 2 中国 16 个大城市 MSM 感染者病毒载量 (VL) 抽样调查不同缺失数据填补处理效果比较

处理方法	完全随机缺失数据库(模拟)			户籍随机缺失数据库(模拟)			治疗随机缺失数据库(模拟)			线性非随机缺失数据库(模拟)			曲线非随机缺失数据库(模拟)							
	均值	准确度	精确度	长度	均值	准确度	精确度	长度	均值	准确度	精确度	长度	均值	准确度	精确度	长度				
EM	2.94	0.15	0.42	5.56	3.03	0.15	0.20	5.65	2.91	0.17	0.63	5.54	2.70	0.33	2.04	5.67	2.93	0.16	0.50	5.59
回归法	2.92	0.18	0.51	5.66	3.09	0.19	0.54	5.72	2.97	0.18	0.17	6.12	2.67	0.37	1.96	6.31	2.92	0.18	0.48	6.14
均值填补法	2.94	0.15	0.35	5.56	2.97	0.15	0.20	5.65	2.88	0.19	0.85	5.54	2.71	0.32	2.00	5.67	2.99	0.14	0.07	5.59
删除法	2.95	0.22	0.29	5.89	2.97	0.21	0.15	5.93	2.88	0.25	0.56	5.87	2.71	0.36	1.39	5.99	2.99	0.20	0.05	6.00
MCMC 填补 10 次	3.01	0.17	0.06	5.83	3.06	0.18	0.35	5.89	3.23	0.30	1.22	5.82	2.74	0.32	1.41	5.96	3.08	0.19	0.46	6.91

注:长度指填补后 VL 对数值的 95%CI 平均长度; EM 为最大期望值法; MCMC 为马尔科夫链蒙特卡罗法

除法普遍低估了数据均值, MCMC 多高估了数据均值; 不同填补方法处理不同缺失类型数据库的精确度差异较大, 未呈现明显规律。

讨 论

1. 缺失数据多重填补的意义及风险: 多重填补是由 Rubin 等于 1976 年提出并建立起来的一种数据扩充和统计分析方法, 其思想来源于贝叶斯估计, 认为待填补数值是随机的, 可以从相似情况或者根据后来可观测数据得到的缺损数据分布, 给每个缺失数据赋予一个模拟值。结合这种方法, 研究者可以比较容易地, 在不舍弃任何数据的情况下对缺失数据的未知性质进行推断。进行缺失数据填补, 需要警惕人为主观因素对原数据系统的影响^[2]。缺失数据多重填补方法在数理基本理论领域依然处于探索的阶段, 对其属于“艺术”还是“科学”的质疑可能还会持续很长时间^[1]。所以, 在实践领域应持谨慎态度对待缺失数据填补结果^[2]。

2. 缺失数据填补在 HIV/AIDS 领域的运用: HIV 检测治疗相关数据缺失现象极为普遍, 但缺失数据蕴含极大的利用价值。近年来, 对 CD₄⁺T 淋巴细胞计数^[12]、HIV 抗病毒治疗耐药性临床检测指标、肺结核患者 HIV 感染状态^[13]、VL 等重要目标参数的缺失填补均有较为广泛地运用^[14], 相关研究多结合 SAS、SPSS 等软件中的多重填补模块, 在展示缺失数据填补模型及预测变量的基础上, 较为系统地阐述数据所揭示的公共卫生问题, 拓展深化了人们对该领域干预实践及效果的理解。

3. VL 缺失数据填补面临挑战及应对方案: 缺失填补以及分布齐性检验, 从理论上都要求数据符合多元正态分布假设。本研究中 VL 原始及转化后数据均不符合正态分布, 这给缺失数据填补方法的选择以及填补后样本分布检验带来极大挑战。

考虑到缺失数据分布、数据缺失模式及机制共同决定了多重填补模型的选择和检验, 本研究对 VL 进行对数转换, 按照完全随机缺失、户籍所致非随机缺失、治疗所致非随机缺失、线性非随机缺失、曲线非随机缺失 5 种缺失机制, 模拟缺失数据库, 采用 EM、回归法、均值填补法、成组删除法、MCMC 多重填补对上述缺失数据库进行填补, 形成 25 个填补数据库, 通过齐性非参数检验、分布频数表、准确度、精确度等评价方法, 比较不同填补方法在不同类型数据库中的填补效果。通过全面的预实验性的缺失数据模拟, 为后续数据分析提供参考依据。

4. VL 缺失数据填补方法的选择:对于不同类型的缺失数据,分布图谱显示回归法、MCMC 较好地保留了完整数据库的主要分布特征;不同缺失数据处理结果的均值与原始数据集均值较为接近,EM、回归法、均值填补法、删除法普遍低估数据均值,MCMC 多高估数据均值;不同填补方法处理不同缺失类型数据库的精确度差异较大,未呈现明显规律。

上述研究结果说明对于 VL 缺失数据,MCMC 以及回归法可以考虑为主要的缺失数据填补方法,由于回归法要求数据符合正态分布假设,有证据认为 MCMC 可用于不符合正态分布的数据,所以,可以考虑将 MCMC 作为 VL 对数转化缺失数据填补首选^[2-3]。本研究显示,完整数据库的 VL 对数均值多介于各类型缺失样本均值与 MCMC 填补均值之间,所以,MCMC 填补数据可以作为调查人群 VL 均值水平估算的参考依据。对于具体到缺失个体值的估算,数据填补还存在较大不确定性,不建议在填补数据库中进行与个体 VL 相关的统计分析。

5. 同质性分层填补可能提高数据填补效果:研究还显示所有填补方法对于完全随机缺失型数据库,均能够较好地保持分布特征,且具有较好的准确度和精确度,其中 MCMC 表现最为突出。所以,采取同质性分层方法,将完整数据库分解成完全随机缺失子数据库,进行分层填补,有可能提高填补效果。结合本研究的前期数据分析,VL 的缺失与感染者户籍、治疗有相关性。进一步引入“治疗”“年龄”“户籍”“安全性行为”“发病时间距离基线调查时间”“最近 1 次检测 CD₄ T 淋巴细胞水平”等变量,以 $P=0.05$ 作为检验标准,数据缺失类型检验显示:4 050 例数据库 VL 缺失类型为随机缺失 ($P=0.000$);分层后,治疗样本库 VL 缺失类型为完全随机缺失 ($P=0.416$);未治疗样本库为完全随机缺失 ($P=0.177$);本地户籍样本库为随机缺失 ($P=0.023$);外地户籍样本库为随机缺失 ($P=0.023$)。说明对数据库按照治疗、未治疗进行同质分组,分别进行 MCMC 填补、合并是可行的,并且可能提高缺失数据填补效果。

志谢 本文得到国家科技重大专项 MSM 人群艾滋病干预研究课题组的 16 个现场工作组成员(卢红艳、曾吉、王娟、于茂河、徐鹏、郭伟、梅淑娟、李雪静、李一、闫红梅、刘岩琳、庄鸣华、宁镇、沈晓涛、还锡萍、闫红静、张敏、朱正平、潘晓红、王懋、罗艳、张兴亮、蒋洪林、汤恒、刘普林、李艳、徐慧芳、程伟彬、钟斐、刘少础、蓝光华、陈怡、农全兴、李恬、龚毅、何勤英、范双凤、吴国辉、欧阳琳、闵向东、章任重、梁军、常文辉、贾华、卫晓丽、吴明旭、倪明建、李凡、李瑞兰、王新迪、王云霞等)支持,以及徐晓玉、任仙龙、陈军、曹巍和 Nanci Nanyi Zhang 的大力协助

利益冲突 无

参 考 文 献

- [1] Center for Disease Control and Prevention. Guidance on community viral load; a family of measures, definitions, and method for calculation [DB/OL]. (2011-08-31) [2016-06-30]. <http://www.incarecampaign.org/files/viral-suppression/community-viral-load-guidance/>.
- [2] SAS 9.2. Summary of Issues in Multiple Imputation. SAS Help and Documentation.
- [3] 廖慧敏,林燧恒. 数据缺失机制对逐步回归变量筛选的影响 [J]. 中国卫生统计, 2011, 28(4): 400-401, 405. DOI: 10.3969/j.issn.1002-3674.2011.04.015.
Liao HM, Lin SH. The effects of different missing mechanisms on stepwise variable selection [J]. Chin J Health Statistics, 2011, 28(4): 400-401, 405. DOI: 10.3969/j.issn.1002-3674.2011.04.015.
- [4] 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. DOI: 10.1037/1082-989X.6.4.330.
- [5] Burton A, Altman DG, Royston P, et al. The design of simulation studies in medical statistics [J]. Stat Med, 2006, 25 (24): 4279-4292. DOI: 10.1002/sim.2673.
- [6] 赵俊康. 不同缺失机制并存时倚倚校正的模拟研究 [D]. 太原: 山西医科大学, 2012.
Zhao JK. A statistical simulation study of bias correction when the different missing mechanism coexist [D]. Taiyuan: Shanxi Medical University, 2012.
- [7] 花琳琳,施念,杨永利,等. 不同缺失值处理方法对随机缺失数据处理效果的比较 [J]. 郑州大学学报: 医学版, 2012, 47(3): 315-318. DOI: 10.3969/j.issn.1671-6825.2012.03.011.
Hua LL, Shi N, Yang YL, et al. Comparison of different methods in dealing with missing values of missing at random [J]. J Zhengzhou Univ: Med Sci, 2012, 47(3): 315-318. DOI: 10.3969/j.issn.1671-6825.2012.03.011.
- [8] 曾艳,庄刘,段春生. 两个非正态样本同分布检验的非参数法选择 [J]. 中国卫生统计, 2012, 29(2): 210-213. DOI: 10.3969/j.issn.1002-3674.2012.02.016.
Zeng Y, Zhuang L, Duan CS. Selection of nonparametric tests when comparing the homogeneity of two Non-normal distributions [J]. Chin Health Statistics, 2012, 29(2): 210-213. DOI: 10.3969/j.issn.1002-3674.2012.02.016.
- [9] 赵景波,李洪源,李康. 两个非正态分布资料比较方法的选择 [J]. 中国卫生统计, 2003, 20(3): 185-188. DOI: 10.3969/j.issn.1002-3674.2003.03.021.
Zhao JB, Li HY, Li K. Choice for two non-normal distribution data comparison [J]. Chin J Health Statistics, 2003, 20(3): 185-188. DOI: 10.3969/j.issn.1002-3674.2003.03.021.
- [10] 岳勇,田考聪. 数据缺失及其填补方法综述 [J]. 预防医学情报杂志, 2005, 21(6): 683-685. DOI: 10.3969/j.issn.1006-4028.2005.06.013.
Yue Y, Tian KC. Review on missing data and imputation methods [J]. J Prev Med Inform, 2005, 21(6): 683-685. DOI: 10.3969/j.issn.1006-4028.2005.06.013.
- [11] Molenberghs G, Williams PL, Lipsitz SR. Prediction of survival and opportunistic infections in HIV-infected patients: a comparison of imputation methods of incomplete CD₄ counts [J]. Stat Med, 2002, 21(10): 1387-1408. DOI: 10.1002/sim.1118.
- [12] Schumi J, DiRienzo AG, DeGruttola V. Testing for associations with missing high-dimensional categorical covariates [J]. Int J Biostat, 2008, 4(1): 18. DOI: 10.2202/1557-4679.1102.
- [13] Vinnard C, Wileyto EP, Bisson GP, et al. Winston first use of multiple imputation with the national tuberculosis surveillance system [J]. Epidemiol Res Int, 2013, 2013: 875234. DOI: 10.1155/2013/875234.
- [14] 庄严,邢艳春,马文卿. 含有缺失机制的多元纵向数据分析 [J]. 中国卫生统计, 2008, 25(5): 489-493. DOI: 10.3969/j.issn.1002-3674.2008.05.013.
Zhuang Y, Xing YC, Ma WQ. Analysis of multivariate longitudinal outcomes with nonignorable dropouts [J]. Chin J Health Statistics, 2008, 25(5): 489-493. DOI: 10.3969/j.issn.1002-3674.2008.05.013.

(收稿日期:2017-01-20)

(本文编辑:王岚)