

基于鞅残差的多重填补法介绍与 R 语言实现*

杨晓敏¹ 韩 婧¹ 田 晶² 闫晶晶¹ 韩港飞² 张岩波^{1, 3, 4△}

【提 要】 目的 简要介绍基于鞅残差的多重填补法的基本原理, 通过模拟研究证明其效能及实现方式, 为未测量混杂因素的填补提供方法学参考与思路。**方法** 基于模拟研究使用基于鞅残差的多重填补法填补未测量混杂因素, 包括非时依性变量和时依性变量, 并与传统的 Cox 模型、基于倾向性评分校准法的填补、基于对数生存时间的填补三种方法进行了比较。**结果** 模拟研究显示, 无论未测量混杂为非时依性变量或时依性变量, 基于鞅残差的多重填补法在大多数模拟情景中都可以产生更精确的估计和更小的偏倚。**结论** 在生存分析中, 相比于其他方法, 基于鞅残差的方法在计算生存时间的较复杂分布时比较灵活, 能够有效减小填补后模型分析的偏倚。

【关键词】 鞅残差 未测量混杂 非时依性变量 时依性变量

【中图分类号】 R541.6

【文献标识码】 A

DOI 10.11783/j.issn.1002-3674.2024.01.036

真实世界的纵向研究往往都是观察性研究, 暴露或干预处理都不是随机均衡的, 因此会产生很多未测量混杂, 这些都会影响对结局效应的估计^[1]。并且, 很多特征不是一成不变的, 会产生时依性混杂。对于一些慢性疾病的预后效果评估来说, 患者的一些特征和干预会随着预后情况做出相应的改变, 例如, 在药物干预和非药物干预方面, 当预后良好时, 患者可能倾向于减少药物使用的比例。如果只考虑基线协变量来推断一些干预措施对预后的影响, 可能会造成一些信息的损失^[2]。未测量混杂因素的影响可以通过一些方法比如倾向性评分校准^[3-4]、贝叶斯倾向性评分^[5]和两阶段校准^[6]来减少甚至消除。这些方法的原理是将数据库分为主数据库和一个较小的验证样本, 通过验证样本来对主数据库中的倾向性评分进行计算从而控制混杂因素^[7-8], 这些方法可以有效的控制未测量混杂因素, 但是面对纵向数据时它们的性能还未被探索。最近, Burne 等人提出一种专门针对纵向数据中存在的未测量混杂因素与时依性混杂因素的新方法, 该方法同样将数据分为主数据库和验证样本, 通过在验证样本中计算多变量 Cox 模型的鞅残差 (martingale residual, MR) 来填补主数据库中未测量的混杂因素^[9]。因此我们对该方法进行了模拟研究, 并将其与常用的其他几种方法进行了比较。

鞅残差原理与方法

1. 原理

* 基金项目: 国家自然科学基金 (81872714; 82173631); 重大疾病风险评估山西省重点实验室 (201805D111006)

1. 山西医科大学公共卫生学院流行病学与卫生统计学教研室 (030000)

2. 山西医科大学第一医院心内科

3. 重大疾病风险评估山西省重点实验室

4. 山西中医药大学

△通信作者: 张岩波, E-mail: sxmuzyb@126.com

我们假设 Cox 模型预测的结果是当前协变量和干预造成的理想结局, 而真实结果往往和预测结果存在误差, 这部分误差来自于无法观察到的混杂即为未测量混杂, 如果不考虑这些可能会错误地估计结局的风险。而鞅残差可以解释这部分误差, 它是指真实存在但却未被模型预测到的部分, 即为研究对象真实结局与模型预测结局的差值^[10]。当数据中存在缺失时, 我们将数据库分为主数据库和验证样本。验证样本不含缺失, 由 $\{t_i, \delta_i, X_i, C_i, U_i\}$ 组成, t_i 及 δ_i 表示个体的生存时间及删失时间, X_i 表示暴露, C_i 表示可测量的混杂, U_i 为在验证样本中的未测量混杂。在纵向分析中, 我们使用基于鞅残差的多重填补法来控制仅在验证样本中测量的混杂因素, 然后使用具有完整混杂信息的验证样本来估算主数据库中的未测量混杂因素^[9-11]。因此, 个体 i 在 t 时刻的鞅残差被定义为^[9-10]:

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp\{\hat{\beta}' Z_i(s)\} d\hat{\Lambda}_0(s) \quad (1)$$

$Y_i(s)$ 表示个体 i 在 s 时刻的风险, $N_i(t)$ 为个体 t 时刻的观察事件数, $Z_i(s)$ 为时间变量。

2. 方法

对于非时依性数据和内部验证样本, 首先, 我们使用全部数据中的暴露 X 和测量的混杂因素 X 拟合 Cox 比例风险模型, 获得鞅残差 M_i 。

$$\lambda(t) = \lambda_0(t) \exp\{\gamma_1 X + \gamma_2 C\} \quad (2)$$

然后将鞅残差用于未测量混杂因子的估计模型中。在验证样本中, 关于未测量混杂因素的信息是可用的, 所以我们为每个未测量混杂因素拟合单独的模型。如果未测量混杂因素是连续的, 我们使用验证样本中的暴露因素、鞅残差和所有测量的混杂因素拟合一个线性模型; 如果是分类的, 则拟合一个多因素 logistic 模型:

$$E[g(U_j)] = \alpha_0 + \alpha_1 X + \alpha_2 \hat{M} + \alpha_3 C \quad (3)$$

$$\text{logit}[P(U_j = 1)] = \alpha_0 + \alpha_1 X + \alpha_2 \hat{M} + \alpha_3 C \quad (4)$$

然后根据式(3)和(4)在主数据库中填补未测量的混杂因素。因此,只有在 t 时刻观察到的暴露和测量的混杂因素以及鞅残差的值才能获得关于未测量混杂因素的信息。

对于连续和分类的未测量混杂 U_j , 我们分别根据正态分布(5)和伯努利分布(6)进行填补,当填补结束后,我们则可以使用 Cox 模型进行最终的生存分析[式(7)]。

$$E[g(U_{ij})] = (\hat{\alpha}_0 \hat{\alpha}_1 \hat{\alpha}_2 \hat{\alpha}_3) Z'_i, V_i = \hat{\sigma}_2 z_i (Z'Z)^{-1} z'_i, z_i = (1, X_i, C_i, M_i) \quad (5)$$

$$p_i = P(U_{ij} = 1) = \text{expit} \{ \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 M_i + \hat{\alpha}_3 C_i \} \quad (6)$$

$$\lambda(t) = \lambda_0(t) \exp \{ \beta X + \gamma'_1 C + \gamma'_2 U \} \quad (7)$$

同理,当处理时依性数据时,我们首先也是利用 Cox 回归模型得到个体每次随访时间 t 的鞅残差,然后根据混杂因素的类型分别使用式(3)或(4)填补未测量混杂因素。但是在这种情况下,时依性混杂必须遵从一个假设,即该混杂只随时间变化,并不是由其他协变量的变化引起,即该混杂不作为干预和结局的中介因素。

模拟研究

本研究模拟了一个干预引起不良结局的研究来评估基于鞅残差填补方法的效果^[12-13]。我们通过改变以下参数来实现不同条件对方法性能的验证:干预的效果;结局事件的发生率;未测量混杂的大小和方向;是否违反混杂非中介的假设以及违反该假设的程度;删失情况;验证样本大小。本研究采用 R 语言进行模拟。

1. 非时依性模拟研究

(1) 理论假设

我们的模拟研究如图 1 所示,所有的变量不随时间变化且混杂因素之间是独立的,生成 1000 个模拟数据集,每个数据集样本量为 10000,并从中随机抽样出一个验证样本。每个样本的真实结局以及生存时

间通过 Cox 模型得到^[14]。通过改变参数模拟了 21 种不同的情况(表 1)来分析上述生成的数据,并且与一些常用的方法进行了比较。首先是不进行调整的 Cox 比例风险回归模型;第二种为经倾向性评分调整后的 Cox 模型^[3];第三种是使用对数生存时间和结局作为校准进行填补的方法^[15-16]。比较所有方法估计效果的偏倚,并且将不同模拟条件下基于 MR 的多重填补法与其他三种方法进行比较,得到它们之间的相对标准差(SD)和相对均方根误差(RMSE)。

(2) 模拟数据填补结果

模拟数据的填补结果如表 1 所示,基于 MR 的多重填补法(-0.002~0.021)在偏倚方面对于所有的模拟方案都强于未经调整的 Cox 比例风险回归模型(-0.186~0.368)。相比于经过倾向性评分校准后的 Cox 模型(-0.363~0.136),基于 MR 的多重填补法在大多数情况下偏倚更小,并且产生的偏倚估计都是一致无偏的,这是该方法优于倾向性评分校准的一点。虽然基于对数生存时间和结局的填补(-0.006~0.035)在偏倚方面略优于基于 MR 的多重填补,但是它要求生存时间和结局必须完整,要求较高,很大程度上受非中介假设的影响,并且一旦存在删失,偏倚会变大(方案 20, 21)。综合来看,基于 MR 的多重填补法在偏倚方面表现良好。

在所有方案中,基于鞅残差法与基于对数生存时间的填补法具有可比性(相对 SD:0.99~1.02)。对于大多数模拟条件,基于鞅残差填补的方法与基于倾向性评分校准的填补法表现相差不大,但是当改变验证样本大小并且满足非中介假设的情况下,基于倾向性评分的填补法与基于鞅残差的方法相差较大(方案 12、14, 相对 SD:1.136、1.305);除此之外,当设置条件为随机删失和 Weibull 随机删失时,两种方法也有较大差异(方案 20、21, 相对 SD:1.484、1.520)。

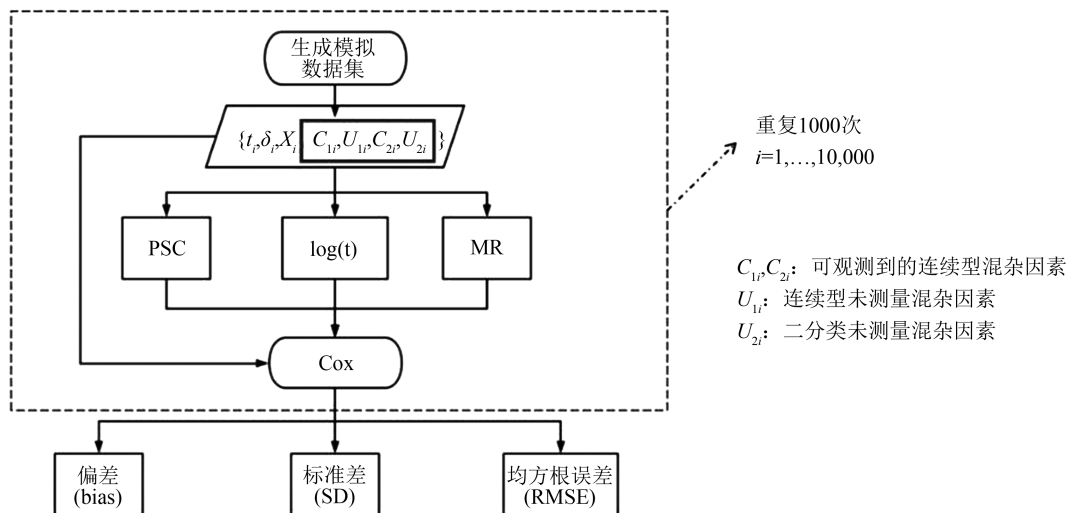


图 1 模拟研究流程图

表 1 非时依性变量模拟结果

参数变化	偏倚				相对 SD			相对 RMSE				
	Cox	倾向性评分	对数生存时间填补	基于 MR 多重填补	Cox	倾向性评分	对数生存时间填补	Cox	倾向性评分	对数生存时间填补		
1 真实 HR	0.163	-0.012	0.004	0.005	0.851	1.053	1.001	2.142	1.061	1.000		
2 1.2	0.163	-0.013	0.002	0.003	0.858	1.049	0.993	2.171	1.060	0.992		
3 1.5	0.158	-0.015	0.003	0.004	0.843	1.083	0.997	2.152	1.097	0.997		
未测量混杂强度												
4 强	0.368	-0.010	0.021	0.021	0.737	1.096	0.987	3.979	1.073	0.986		
5 弱	0.044	-0.015	-0.002	-0.002	0.988	1.020	1.003	1.143	1.039	1.003		
假设是否违背												
U1	U2	C1										
6 ↓+	↓+		-0.183	-0.305	-0.006	-0.007	0.883	0.954	1.005	1.922	2.998	1.005
7 ↓	↓		-0.046	-0.091	-0.002	-0.002	0.984	0.995	0.999	1.114	1.433	0.999
8 ↓	↓+	↓+	-0.043	-0.025	0.002	0.001	0.862	0.875	0.990	0.951	0.905	0.990
9 ↓	↓		0.029	-0.044	-0.005	-0.005	0.924	0.958	0.995	0.976	1.067	0.995
10 ↓	↓+		0.163	0.136	0.002	0.004	0.862	0.885	1.000	2.110	1.833	0.999
11 ↓	↓		0.164	0.022	0.004	0.004	0.867	1.007	1.001	2.173	1.041	1.001
VS 大小												
12 500(满足假设)			0.163	-0.033	0.004	0.004	0.715	1.136	1.009	1.831	1.185	1.009
13 500(违背假设)			-0.184	-0.363	-0.003	-0.004	0.746	1.078	1.000	1.683	3.171	1.000
14 250(满足假设)			0.164	-0.066	0.009	0.008	0.558	1.305	1.006	1.452	1.411	1.007
15 250(违背假设)			-0.186	-0.210	-0.005	-0.005	1.064	0.954	1.000	2.185	2.356	1.000
VS												
16 决定因素 C			0.160	-0.007	0.002	0.002	0.860	1.030	1.006	2.027	1.032	1.006
17 决定因素 U			0.164	-0.004	0.011	0.011	0.862	1.053	0.998	2.127	1.044	0.997
18 验证集事件率			0.163	-0.029	-0.005	0.003	0.926	1.121	1.016	2.217	1.177	1.016
删失情况												
19 管理删失			0.164	-0.011	0.003	0.004	0.845	1.046	0.993	2.108	1.053	0.993
20 随机删失			0.164	-0.007	0.035	0.021	0.717	1.484	0.933	3.591	1.327	1.126
21 Weibull 随机删失			0.164	-0.010	0.023	0.017	0.710	1.520	0.968	3.893	1.416	1.041

*: ↓, ↓: 违背假设; +: 违背假设程度强; VS: 验证样本

最后, 我们把每种方法都和基于鞅残差的填补法进行比较时, 可以看出基于鞅残差的填补法基本上在每个模拟条件下都具有最小的 RMSE, 说明该方法具有良好的稳定性与精确性。

2. 时依性数据分析

(1) 方法

本研究还模拟了时依性混杂因素和外部验证样本, 即暴露和混杂都是时依性的(它们的值可以在随访期间以 10 个相等的时间间隔更新), 并且验证样本是来自不同的源群体的外部样本, 其他设定类似于非时依性数据模拟过程。

(2) 结果

如表 2 所示, 基于鞅残差的填补法在所有情况下产生的偏倚都较小, 并且结合相对 SD 可以得到基于鞅残差填补法得到的结果都是无偏的。与上文中非时

依性混杂的情况类似, 当满足非中介假设时, 基于倾向性评分的方法产生的偏倚要小于基于鞅残差的方法, 但是一旦这个假设被打破, 基于倾向性评分的填补法性能会大打折扣, 然而基于鞅残差的方法却依旧稳定。当改变未测量混杂的强度和混杂是否为中介的假设时, 基于鞅残差的填补法的 RMSE 也是远远低于标准的 Cox 模型和基于倾向性评分的填补法。所以我们可以认为在面对时依性混杂时, 基于鞅残差的多重填补法有着较高的性能与稳定性。

讨论

本文介绍了一种新的方法来调整时间-事件分析中的未测量的混杂偏倚, 并且该方法相比于传统的 Cox 比例风险回归模型和基于倾向性评分校准后的 Cox 模型都有着较大的优势^[17]。该方法首先将从数

表 2 时依性变量的模拟结果

	真实 HR	未测量混杂	假设	偏倚			相对 SD		相对 RMSE	
				标准 Cox 模型	基于倾向性评分的填补	基于鞅残差的多重填补	标准 Cox 模型	基于倾向性评分的填补	标准 Cox 模型	基于倾向性评分的填补
1	1	中	满足	0.203	0.001	-0.007	0.862	1.048	1.165	1.048
2	1	中	违背	-0.241	0.497	-0.013	0.874	1.093	1.202	2.025
3	1	强	违背	-0.474	0.844	-0.051	0.825	1.079	1.608	2.686
4	1.5	中	满足	0.208	0.014	0.003	0.841	1.040	1.179	1.042
信息缺失										
5	1	中	满足	0.201	0.011	-0.007	0.857	1.048	1.140	1.049
6	1.5	中	满足	0.206	0.007	-0.005	0.838	1.024	1.148	1.024

据集中随机获得验证样本集,然后将其与鞅残差结合起来计算数据中的未测量混杂,从而可以更精确地得到干预或处理对结局的影响。相比于基于倾向性评分的方法和使用对数生存时间和结局事件的方法的高要求与苛刻的限制条件^[18],基于鞅残差的填补方法更加普适。即使是面对很大程度的未测量混杂或者是不满足假设条件时,该方法都能产生几乎无偏的估计。与此同时,该方法很大的一个优势是,无论未测量混杂是时依性的还是非时依性的,都可以有很好的效果。

本研究也存在一些局限性。首先,当真实的 HR 很高时,基于鞅残差的填补法也无法做到完全消除偏倚;并且本次研究只是模拟了随机删失情况,对于生存分析中可能会存在的非随机删失(比如干预或者其他原因造成的),可以在之后的研究中加以探索。

结 论

在处理未测量混杂时,不论是时依性混杂还是非时依性混杂,相比于传统的 Cox 模型、基于倾向性评分校准的填补法和基于对数生存时间的填补法,基于鞅残差的多重填补法都有着较高的精确性与稳定性,并且在未测量混杂强度增大以及不满足条件假设的情况下,基于鞅残差的方法有着更小的估计偏倚和更小的 RMSE。综上所述,基于鞅残差的多重填补法可以有效控制时间-事件分析中的未测量混杂因素,从而获得对结局更加准确的估计或预测。

参 考 文 献

- [1] 郑雪莹. 不同混杂结构条件下各倾向性评分方法的比较研究. 中国卫生统计, 2017, 34(3): 415-420.
- [2] Daniel RM, Cousens SN, De Stavola BL, et al. Methods for dealing with time-dependent confounding. *Stat Med*, 2013, 32(9): 1584-1618.
- [3] Stürmer T, Schneeweiss S, Rothman KJ, et al. Performance of propensity score calibration—a simulation study. *Am J Epidemiol*, 2007, 165(10): 1110-1118.
- [4] 王晨晨, 孙倩, 王彤. 不同混杂结构下广义倾向性评分法的模拟研究及应用. 中国卫生统计, 2022, 39(1): 7-13.
- [5] Mccandless LC, Richardson S, Best N. Adjustment for Missing Confounders Using External Validation Data and Propensity Scores. *Journal of the American Statal Association*, 2012, 107(497): 40-51.
- [6] Lin HW, Chen YH. Adjustment for missing confounders in studies based on observational databases: 2-stage calibration combining propensity scores from primary and validation data. *Am J Epidemiol*, 2014, 180(3): 308-317.
- [7] Franklin JM, Eddings W, Schneeweiss S, et al. Incorporating linked healthcare claims to improve confounding control in a study of in-hospital medication use. *Drug Saf*, 2015, 38(6): 589-600.
- [8] Wood ME, Frazier JA, Nordeng HM, et al. Prenatal triptan exposure and parent-reported early childhood neurodevelopmental outcomes: an application of propensity score calibration to adjust for unmeasured confounding by migraine severity. *Pharmacoepidemiol Drug Saf*, 2016, 25(5): 493-502.
- [9] Burne RM, Abrahamowicz M. Martingale residual-based method to control for confounders measured only in a validation sample in time-to-event analysis. *Stat Med*, 2016, 35(25): 4588-4606.
- [10] Barlow W, Prentice R. Residuals for relative risk regression. *Biometrika*, 1988, 75(1): 65-74.
- [11] Eubank RL, Speckman P. Martingale-based residuals for survival models. *Biometrika*, 1990, 77(1): 147-160.
- [12] 胡美婧, 李满娣, 林奕蝶, 等. 不同样本量条件下倾向性评分法与 Logistic 回归法估计处理效应的比较研究. 中国卫生统计, 2021, 38(6): 812-815.
- [13] 鲍晓蕾, 高辉, 胡良平. 多种填补方法在纵向缺失数据中的比较研究. 中国卫生统计, 2016, 33(1): 45-48.
- [14] Huang Y, Zhang Y, Zhang Z, et al. Generating Survival Times Using Cox Proportional Hazards Models with Cyclic and Piecewise Time-Varying Covariates. *Stat Biosci*, 2020; 1-16.
- [15] Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*, 1999, 18(6): 681-694.
- [16] White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med*, 2009, 28(15): 1982-1998.
- [17] Murad H, Dankner R, Berlin A, et al. Imputing missing time-dependent covariate values for the discrete time Cox model. *Stat Methods Med Res*, 2020, 29(8): 2074-2086.
- [18] Burne RM, Abrahamowicz M. Martingale residual-based method to control for confounders measured only in a validation sample in time-to-event analysis. *Stat Med*, 2016, 35(25): 4588-4606.

(如需要可联系作者索取 R 语言代码。)

(责任编辑:郭海强)