

病例队列研究设计中相对危险度的估计及其应用

拓嘉怡^{1,2,3} 毕京浩^{1,2,3} 李卓颖^{2,3} 沈秋明^{1,2,3} 谭玉婷^{2,3} 李泓澜^{2,3} 袁蕙芸⁴
项永兵^{1,2,3,4}

¹上海交通大学医学院公共卫生学院,上海 200025;²上海交通大学医学院附属仁济医院,癌基因及相关基因国家重点实验室,上海 200032;³上海市肿瘤研究所流行病学研究室,上海 200032;⁴上海交通大学医学院附属仁济医院,上海 200127

通信作者:项永兵,Email:ybxian@shsci.org

【摘要】 目的 系统介绍病例队列研究设计的基本原理,以及风险比(HR)的常用估计方法及其应用。方法 首先,介绍病例队列研究设计的基本原理;其次,对Prentice法、Self-Prentice法和Barlow法加权Cox比例风险回归模型进行描述和说明;最后,以上海市女性健康队列研究为例,分析全队列数据与病例队列样本中肥胖与肝癌发病的关联,并进一步比较两者在各种模型中参数估计的结果。结果 在全队列数据和病例队列样本中,发现肥胖与女性肝癌发病的关联均有统计学意义。在Cox比例风险回归模型中,全队列数据和病例队列样本的回归系数(β)随着协变量调整有所波动,但是两者的HR值较为接近;两者 β 的标准误存在差异,即病例队列样本 β 的标准误大于全队列的参数估计值,HR值的95%CI更宽。在加权Cox比例风险回归模型中,Prentice法相比Self-Prentice法和Barlow法的 β 的标准误更接近全队列的参数估计值,HR值的95%CI更靠近全队列的结果。结论 病例队列研究设计通过收集和分析子队列成员和发病者的资料,可以获得接近全队列的参数结果,同时能够节约样本量和提高研究效率。此外,在病例队列设计中可以优先选择Prentice法。

【关键词】 队列研究; 病例队列设计; 生存分析

基金项目: 国家重点研发计划(2021YFC2500404,2016YFC1302503)

Statistical methods for relative risk estimation and applications in case-cohort study

Tuo Jiayi^{1,2,3}, Bi Jinghao^{1,2,3}, Li Zhuoying^{2,3}, Shen Qiuming^{1,2,3}, Tan Yuting^{2,3}, Li Honglan^{2,3}, Yuan Huiyun⁴, Xiang Yongbing^{1,2,3,4}

¹School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China;

²State Key Laboratory of Oncogene and Related Genes, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200032, China; ³Department of Epidemiology, Shanghai Cancer Institute, Shanghai 200032, China; ⁴Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

Corresponding author: Xiang Yongbing, Email:ybxian@shsci.org

【Abstract】 **Objective** To systematically introduce the design of case-cohort study and the statistical methods of relative risk estimation and their application in the design. **Methods** First, we introduced the basic principles of case-cohort study design. Secondly, Prentice's method, Self-Prentice method and Barlow method were described in the weighted Cox proportional hazard regression models in detail, finally, the data from the Shanghai Women's Health Study were used as

DOI:10.3760/cma.j.cn112338-20210812-00638

收稿日期 2021-08-12 本文编辑 万玉立

引用格式:拓嘉怡,毕京浩,李卓颖,等.病例队列研究设计中相对危险度的估计及其应用[J].中华流行病学杂志,2022,43(3):392-396. DOI:10.3760/cma.j.cn112338-20210812-00638.

Tuo JY, Bi JH, Li ZY, et al. Statistical methods for relative risk estimation and applications in case-cohort study[J]. Chin J Epidemiol, 2022, 43(3):392-396. DOI:10.3760/cma.j.cn112338-20210812-00638.



an example to analyze the association between obesity and liver cancer incidence in the full cohort and case-cohort sample, and the results of parameters from each method were compared. **Results** Significant association was observed between obesity and risk for liver cancer incidence in women in both the full cohort and the case-cohort sample. In the Cox proportional hazard regression model, the partial regression coefficients of the full cohort and the case-cohort sample fluctuated with the adjustment of confounding factors, but the hazard ratio estimates of them were close. There was a difference in the standard error of the partial regression coefficient between the full cohort and the case-cohort sample. The standard error of the partial regression coefficient of the case-cohort sample was larger than that of the full cohort, resulting in a wider 95% confidence interval of the relative risk. In the weighted Cox proportional hazard regression model, the standard error of the partial regression coefficient of Prentice's method was closer to the parameter estimates from full cohort than Self-Prentice method and Barlow method, and the 95% confidence interval of hazard ratio was closer to that of the full cohort. **Conclusions** Case-cohort design could yield parameter results closer to the full cohort by collecting and analyzing data from sub-cohort members and patients with the disease, and reduce sample size and improve research efficiency. The results suggested that Prentice's method would be preferred in case-cohort design.

【 Key words 】 Cohort study; Case-cohort design; Survival analysis

Fund programs: National Key Research and Development Program of China (2021YFC2500404, 2016YFC1302503)

队列研究 (cohort study) 是一种常用的流行病学研究方法, 根据特定队列人群是否暴露于某种危险因素或其暴露程度分为不同的亚组, 随访各组成员结局 (如疾病) 发生的情况, 比较各组结局发生率的差异, 用来判定暴露因素与结局之间是否存在关联^[1]。由于队列研究确保了暴露出现在结局之前, 能够明确两者之间的因果关系, 因此研究证据的级别属于 II 级, 仅次于随机对照试验^[2]。

然而在实际研究工作中, 特别是在一个大样本队列中, 研究人员随访一段时间后只能得到少量结局事件, 其他多数研究对象只能获得截尾观察结果, 这时如果收集所有对象的协变量资料作统计分析, 则需要耗费大量的资源。因此在有限的预算下, 研究人员更倾向于收集部分对象的协变量来提高效率和节约成本, 此时再抽样设计便应运而生。常见的再抽样设计是巢式病例对照研究和病例队列研究设计^[3-4]。相比巢式病例对照研究, 病例队列研究设计中比较的对象是子队列非病例和子队列内及子队列外发生的全部病例, 在带有删失数据的生存资料分析中应用较为广泛。1986 年 Prentice^[5]首次提出了病例队列研究设计, 即研究人员从全队列中按照一定比例随机抽取一个子队列 (subcohort), 再将子队列之外所有发生了结局事件 (如疾病、死亡等) 的个体称为病例 (case), 子队列和子队列之外所有的病例组成一个病例队列样本 (case-cohort sample), 研究人员仅对病例队列样本的个体进行协变量或生物样品的检测。这种病例队列研究设计对于协变量收集难度较大或成本较

高的大型队列研究具有重要的理论意义和实际应用价值^[6]。但由于病例队列研究设计会系统地偏向抽取总体中的一部分关键个体, 因此如何调整有偏抽样造成的系统偏差成为此类数据分析的关键^[7]。同时, 为了调整病例队列研究设计的有偏抽样造成的系统误差, 风险比 (HR) 的估计不能采用常规的 Cox 比例风险回归模型, 需要根据病例和子队列个体对风险的不同贡献进行加权。目前病例队列研究设计应用的加权 Cox 比例风险回归模型主要有 Prentice 法^[5]、Self-Prentice 法^[8]和 Barlow 法^[9], 不同方法设定的样本风险集及其权重有所差异, 但是最终获得的统计推断结果与全队列相差不大^[10-11]。

由于病例队列研究设计能够有效降低研究成本, 提高工作效率, 在欧、美洲地区的大型队列研究中已经得到了一定范围的推广和应用^[12-13]。但目前国内的病例队列研究中应用仍然较少, 本文系统介绍其研究设计、加权 Cox 比例风险回归模型的原理和方法, 并以上海市女性健康队列为例, 简述 Prentice 法、Self-Prentice 法和 Barlow 法加权 Cox 比例风险回归模型在 SAS 软件中如何实现。

资料与方法

1. 基本原理: 处理和分析删失生存数据, Cox 比例风险回归模型是最常用的一种半参数回归模型。该模型由英国统计学家 Cox^[14]于 1972 年提出, 可同时分析多种因素对个体生存情况的影响。模型的基本形式:

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)$$

式中 $h(t, X)$ 表示具有协变量 X 的个体在时间 t 的瞬时危险率; $h_0(t)$ 表示当协变量 X 均为 0 时 $h(t, X)$ 的基准危险率; X_1, X_2, \dots, X_p 表示与生存结局可能有关的协变量, 在整个观察期间不随时间的变化而变化; $\beta_1, \beta_2, \dots, \beta_p$ 为模型的回归系数, 需要根据实际数据进行估计。目前广泛地应用偏似然函数法^[14-15]解决模型中 β 的统计推断问题:

$$L_{F(\beta)} = \prod_{i=1}^n \left[\frac{\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}{\sum_{j \in R(t)} \exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \dots + \beta_p X_{jp})} \right]^{\delta_i}$$

$$\delta_i = \begin{cases} 1, & \text{当第 } i \text{ 个个体在 } t \text{ 时刻发生结局事件} \\ 0, & \text{当第 } i \text{ 个个体在 } t \text{ 时刻删失} \end{cases}$$

式中分子反映 t 时刻观察到第 i 个个体发生结局事件的风险, 分母 $j \in R(t)$ 表示在 t 时刻所有个体的风险之和。

由于病例队列研究设计是有偏抽样设计, 并未收集全队列所有个体的有关协变量信息, 因此需要采用伪似然函数^[4-5]对病例和子队列个体的贡献进行不同的加权, 修正模型中 β 的推断问题:

$$L_{P(\beta)} = \prod_{i=1}^n \left[\frac{\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}{\sum_{j \in U(t)} w_j(t) \exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \dots + \beta_p X_{jp})} \right]^{\delta_i}$$

$$\delta_i = \begin{cases} 1, & \text{当第 } i \text{ 个个体在 } t \text{ 时刻发生结局事件} \\ 0, & \text{当第 } i \text{ 个个体在 } t \text{ 时刻删失} \end{cases}$$

式中分子反映 t 时刻观察到第 i 个个体发生结局事件的风险; 分母 $j \in U(t)$ 表示在 t 时刻经 $w_j(t)$ 加权后病例队列样本所有个体的风险之和, $w_j(t)$ 表示根据病例队列样本中个体的贡献确定的权重。

表 1 总结了伪似然函数中常用的 3 种分母加权方法^[5, 8-9, 16]。Prentice 法中子队列与子队列外 t 时刻病例发生结局事件的权重相同, 病例删失不考虑加权; Self-Prentice 法中子队列权重相同, 而子队列外所有病例均不加权; Barlow 法中子队列的非病例和病例删失时权重均为 $1/\pi$, t 时刻所有病例发生结局事件时权重为 1, 而子队列外病例删失则不考虑加权。

在加权 Cox 比例风险回归模型中, 由于有偏抽

表 1 伪似然函数中的分母加权方法

结局	Prentice 法 (1986 年)	Self-Prentice 法 (1988 年)	Barlow 法 (1994 年)
子队列外病例删失	0	0	0
子队列外病例发生结局事件	1	0	1
子队列病例删失	1	1	$1/\pi$
子队列病例发生结局事件	1	1	1
子队列非病例	1	1	$1/\pi$

注: π : 病例队列设计中子队列抽样比例

样的存在, 不同时刻的病例队列样本的权重并非相互独立(如 t 时刻病例删失和发生结局事件的权重可能不同), 因此需要使用更加复杂的方差估计方法获得模型的 β 的标准误^[17]。Barlow 和 Prentice^[18-19]提出基于分析数据中删失个体参数估计值变化的 Jackknife 方差估计作为稳健的方差估计。方差估计公式的基本形式:

$$\text{var}(\beta) = \sum_i \begin{pmatrix} \Delta\beta_i \\ p \times 1 \end{pmatrix} \begin{pmatrix} \Delta\beta_i^T \\ 1 \times p \end{pmatrix}$$

式中 $\Delta\beta_i$ 表示第 i 个个体删失后 β 的改变量, P 维向量表示协变量 P 的信息矩阵。

2. 实例数据分析: 本研究以上海市女性健康队列基线和随访数据为例, 通过分析肥胖与女性肝癌发病风险的关联为例, 比较全队列数据与病例队列样本在普通和加权 Cox 比例风险回归模型分析中结果上的差异。上海市女性健康队列是上海市肿瘤研究所于 1997-2000 年建立的自然人群队列, 队列成员包括上海市某区的 40~70 岁女性共 74 940 人。该队列的具体研究设计参见相关文献^[20-21]。基线时研究对象填写统一设计的调查问卷, 并进行了身高、体重的测量。

在 74 940 名研究对象中, 排除基线缺少 BMI 数据和罹患癌症者, 最终纳入本次分析的全队列成员 73 316 人。全队列的随访截止日期至 2017 年 12 月 31 日, 排除失访、确诊原位癌、未确诊肿瘤类型、死于肿瘤但无相关肿瘤诊断信息和日期者, 最终纳入分析共 72 867 人, 其中新发原发性肝癌患者 255 人。

为获得病例队列样本, 本研究首先根据 1998-2002 年上海市某区 40~70 岁女性的原发性肝癌发病率数据^[22], 估计全队列 73 316 人随访 20 年的肝癌发病总人数, 计算全队列 20 年的肝癌累积发病率。再根据病例队列设计的原理^[23], 从全队列中以抽样比例 $\alpha = \frac{\text{肝癌累积发病率}}{1 - \text{肝癌累积发病率}} =$

$$\frac{326.34/10\text{万}}{1 - 326.34/10\text{万}} = 0.327\%$$

随机抽取 240 人作为子队列。子队列随访期间, 排除死于肿瘤但无相关肿瘤诊断信息和日期者 3 人, 最终用于分析的子队列人群共 237 人, 其中原发性肝癌患者 1 人; 此外子队列外随访获得的原发性肝癌患者 254 人, 即子队列与子队列之外所有原发性肝癌患者组成的病例队列样本 491 人。

研究所用的 Cox 比例风险回归模型以年作为生存时间尺度, 队列进入时间为基线调查时间, 肝

癌诊断时间、死亡时间或最后一次随访时间中先发生者为退出时间。所有统计学分析采用 SAS 9.4 软件进行。全队列数据 Cox 比例风险回归模型使用 PROC PHREG 语句实现, 病例队列样本使用 Barlow 等提供的 SAS 宏^[16]建立 Prentice 法、Self-Prentice 法和 Barlow 法加权 Cox 比例风险回归模型。

结 果

肥胖与女性肝癌关联的单因素分析: 全队列数据利用普通 Cox 比例风险回归模型和病例队列样本数据采用 Prentice 法、Self-Prentice 法和 Barlow 法加权 Cox 比例风险回归模型估计 BMI 的 β 值分别为 0.082、0.084、0.086 和 0.086, 其标准误分别为 0.017、0.033、0.034 和 0.034, HR 值分别为 1.085、1.088、1.089 和 1.089。多因素模型结果显示, 全队列数据利用普通 Cox 比例风险回归模型和病例队列样本数据采用 Prentice 法、Self-Prentice 法和 Barlow 法加权 Cox 比例风险回归模型估计 BMI 的 β 值分别为 0.042、0.084、0.086 和 0.086, 其标准误分别为 0.017、0.034、0.035 和 0.035; 其 HR 值分别为 1.043、1.088、1.090 和 1.090。见表 2, 3。

表 2 全队列和病例队列样本中肥胖与女性肝癌关联的参数估计(单因素分析)

类别	β 值	标准误	P 值	HR 值(95%CI)
全队列	0.082	0.017	<0.001	1.085(1.050~1.122)
病例队列设计				
Prentice 法	0.084	0.033	0.011	1.088(1.020~1.160)
Self-Prentice 法	0.086	0.034	0.012	1.089(1.019~1.164)
Barlow 法	0.086	0.034	0.012	1.089(1.019~1.164)

注: 测量肥胖的指标为 BMI, 采用连续性变量

表 3 全队列和病例队列样本中肥胖与女性肝癌关联的参数估计(多因素分析)

类别	β 值	标准误	P 值	HR 值(95%CI)
全队列	0.042	0.017	0.017	1.043(1.008~1.079)
病例队列设计				
Prentice 法	0.084	0.034	0.012	1.088(1.019~1.162)
Self-Prentice 法	0.086	0.035	0.013	1.090(1.018~1.167)
Barlow 法	0.086	0.035	0.013	1.090(1.018~1.167)

注: 测量肥胖的指标为 BMI, 采用连续性变量; 调整因素为年龄

无论是在全队列数据还是在病例队列样本中, 肥胖与女性肝癌发病风险的关联均有统计学意义。此外, 虽然全队列数据和病例队列样本的 β 值随着协变量调整有所波动, 但是两者的 HR 值估计值差

别不大。

全队列数据与病例队列样本 β 的标准误存在差异, 病例队列样本获得的 β 的标准误均大于全队列的参数估计值, 导致 HR 值的 95%CI 更宽。此外, Prentice 法相比 Self-Prentice 法和 Barlow 法, β 的标准误更接近全队列数据的结果, HR 值的 95%CI 也更靠近全队列。

讨 论

病例队列研究设计是将队列研究和病例对照研究设计相结合, 并融合两者的优点后形成的一种研究设计方法^[23], 能够有效节约样本量和提高研究效率。本文首先介绍了病例队列研究设计的思路, 以及病例队列样本资料分析常用的 Prentice 法、Self-Prentice 法和 Barlow 法加权 Cox 比例风险回归模型的基本原理。作为实例, 本文采用 Cox 比例风险回归模型分析了全队列数据和病例队列样本中肥胖与女性肝癌发病的关联, 并比较两者在普通和加权 Cox 比例风险回归模型参数估计中的差异。结果显示, 无论是在全队列数据还是在病例队列样本中, 肥胖均和女性肝癌的发病有一定的关联^[24-25]。全队列数据和病例队列样本的 β 值随着协变量调整有所波动, 但是两者的 HR 值较为接近; 两者 β 的标准误有所差异, 病例队列样本 β 的标准误均大于全队列的结果, 导致 HR 值的 95%CI 更宽; Prentice 法相比 Self-Prentice 法和 Barlow 法更接近全队列数据的参数结果。因此在病例队列设计数据分析中可以优先选择 Prentice 法。

从本文的单因素分析与 Barlow 等^[16]的研究结果均发现, 全队列数据与病例队列样本的 β 值较为接近, 病例队列样本获得的 β 的标准误均大于全队列的参数结果。但本文在多因素模型中调整年龄变量后, 则发现全队列相比病例队列样本的 β 值有所下降, 而病例队列样本获得的 β 的标准误依然大于全队列数据的结果。因此, β 值的波动可能是由于多因素模型调整了年龄因素, 而病例队列样本由于子队列和子队列外所有病例组成, 年龄分布情况与全队列研究对象的年龄分布不同, 因此年龄调整后全队列的 β 值下降幅度大于病例队列样本, 但对 HR 值的统计推断结果影响不大。而全队列数据与病例队列样本 β 的标准误始终存在差异, 则可能是因为病例队列设计并非对整个队列的研究对象进行分析, 而是只分析子队列成员和发病者的资料,

因此样本量大小问题依然会导致病例队列样本 β 的标准误大于全队列的参数。

为了获得与全队列参数更加接近的研究结果,可以优先选择Prentice法加权Cox比例风险回归模型作为病例队列设计的分析方法。本文实例和Barlow等^[16]的研究结果均表明,Prentice法相比Self-Prentice法和Barlow法, β 的标准误更加接近全队列,因此暴露的HR值(95%CI)也会更靠近全队列,能够获得更加接近全队列的参数结果。此外,也可以通过提高病例队列设计中子队列的抽样比例获得更加接近全队列的结果。Onland-Moret等^[26]的研究发现,当选择Prentice法作为病例队列研究设计的主要分析方法时,子队列抽样比例为0.5%、1%、5%和10%时 β 的标准误分别为0.29、0.19、0.11和0.10,逐渐接近全队列的标准误(0.09)。因此在病例队列研究设计中,选择Prentice法的同时提高子队列的抽样比例,以获得更加接近全队列的参数结果,使统计推断更加准确。

综上所述,当大型队列的研究结局是恶性肿瘤等发病率较低的疾病时,病例队列研究设计可以通过收集和分析子队列成员和发病者的协变量资料,获得接近于全队列的参数结果,同时能够有效节约样本量和提高研究效率。特别是在病例队列研究设计中选择Prentice法加权Cox比例风险回归模型作为主要的HR值估计方法,并选择合适的子队列抽样比例,即可获得更加接近全队列参数的研究结果。但是由于该方法涉及较复杂的统计模型,且需要根据实际数据对程序中多个参数进行定义和修改,因此在进行大样本队列数据的分析时,应根据实际情况决定是否选用此方法并谨慎解释结果。

利益冲突 所有作者声明无利益冲突

作者贡献声明 拓嘉怡:查阅文献、编写程序、分析数据和结果解释、撰写初稿和修改文章;毕京浩、李卓颖、沈秋明、谭玉婷、李泓澜、袁蕙芸:参与方法学讨论、结果解释和文章修改;项永兵:负责课题设计和基金、整理原始数据、参与方法学讨论、结果解释、文章框架设计、内容修改和定稿

参 考 文 献

- [1] 沈洪兵,齐秀英.流行病学[M].8版.北京:人民卫生出版社,2013.
- [2] Shen HB, Qi XY. Epidemiology[M]. 8th ed. Beijing: People's Medical Publishing House, 2013.
- [3] Luchini C, Veronese N, Nottegar A, et al. Assessing the quality of studies in meta-research: review/guidelines on the most important quality assessment tools[J]. Pharm Stat, 2021, 20(1):185-195. DOI:10.1002/pst.2068.
- [4] 项永兵,高玉莹.嵌入式病例对照研究设计及其有关统计问题[J].肿瘤,1999,19(6):370-373. DOI:10.3781/j.issn.1000-7431.1999.06.017.
- [5] Xiang YB, Gao YT. Nested case-control study design and related statistical issues[J]. Tumor, 1999, 19(6):370-373. DOI:10.3781/j.issn.1000-7431.1999.06.017.
- [6] 项永兵,高玉莹.病例队列研究设计及其相对危险度估计

方法[J].肿瘤,2000,20(3):226-228. DOI:10.3781/j.issn.1000-7431.2000.03.023.

- [7] Xiang YB, Gao YT. Case-cohort study design and the estimation method of relative risk[J]. Tumor, 2000, 20(3):226-228. DOI:10.3781/j.issn.1000-7431.2000.03.023.
- [8] Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials[J]. Biometrika, 1986, 73(1):1-11. DOI:10.1093/biomet/73.1.1.
- [9] 余吉昌,曹永秀.基于病例队列数据的比例风险模型的诊断[J].数学学报:中文版,2020,63(2):137-148. DOI:10.3969/j.issn.0583-1431.2020.02.004.
- [10] Yu JC, Cao YX. Model diagnostics for the proportional hazards model with case-cohort data[J]. Acta Math Sin: Chin Ser, 2020, 63(2):137-148. DOI:10.3969/j.issn.0583-1431.2020.02.004.
- [11] 钱永春,丁洁丽. Cox模型下病例队列设计中两种伪似然推断方法及其应用[J].数理统计与管理,2020,39(5):845-856. DOI:10.13860/j.cnki.sljt.20200606-001.
- [12] Qian YC, Ding JL. Inference and application of two pseudo-likelihood-based methods for case-cohort design under Cox's model[J]. J Appl Stat Manage, 2020, 39(5):845-856. DOI:10.13860/j.cnki.sljt.20200606-001.
- [13] Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort studies[J]. Ann Stat, 1988, 16(1):64-81. DOI:10.1214/aos/1176350691.
- [14] Barlow WE. Robust variance estimation for the case-cohort design[J]. Biometrics, 1994, 50(4):1064-1072. DOI:10.2307/2533444.
- [15] Kim RS. A new comparison of nested case-control and case-cohort designs and methods[J]. Eur J Epidemiol, 2015, 30(3):197-207. DOI:10.1007/s10654-014-9974-4.
- [16] Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison[J]. Am J Epidemiol, 1990, 131(1):169-176. DOI:10.1093/oxfordjournals.aje.a115471.
- [17] Von Holle A, O'Brien KM, Sandler DP, et al. Association between serum iron biomarkers and breast cancer[J]. Cancer Epidemiol, Biomarkers Prev, 2021, 30(2):422-425. DOI:10.1158/1055-9965.epi-20-0715.
- [18] Merritt MA, Tzoulaki I, van den Brandt PA, et al. Nutrient-wide association study of 57 foods/nutrients and epithelial ovarian cancer in the European Prospective Investigation into Cancer and Nutrition study and the Netherlands Cohort Study[J]. Am J Clin Nutr, 2016, 103(1):161-167. DOI:10.3945/ajcn.115.118588.
- [19] Cox DR. Regression model and life-table[J]. J Roy Stat Soc: Ser B (Methodol), 1972, 34(2):187-202. DOI:10.1111/j.2517-6161.1972.tb00899.x.
- [20] Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study[J]. Ann Stat, 1982, 10(4):1100-1120. DOI:10.1214/aos/1176345976.
- [21] Barlow WE, Ichikawa L, Rosner D, et al. Analysis of case-cohort designs[J]. J Clin Epidemiol, 1999, 52(12):1165-1172. DOI:10.1016/s0895-4356(99)00102-x.
- [22] Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model[J]. J Am Stat Assoc, 1989, 84(408):1074-1078. DOI:10.1080/01621459.1989.10478874.
- [23] Barlow WE, Prentice RL. Residuals for relative risk regression[J]. Biometrika, 1988, 75(1):65-74. DOI:10.1093/biomet/75.1.65.
- [24] Barlow WE. Global measures of local influence for proportional hazards regression models[J]. Biometrics, 1997, 53(3):1157-1162. DOI:10.2307/2533574.
- [25] Zheng W, Chow WH, Yang G, et al. The Shanghai women's health study: rationale, study design, and baseline characteristics[J]. Am J Epidemiol, 2005, 162(11):1123-1131. DOI:10.1093/aje/kwi322.
- [26] Li ZY, Li HL, Ji XW, et al. Dose-response association between adiposity and liver cancer incidence: a prospective cohort study among non-smoking and non-alcohol-drinking Chinese women[J]. Cancer Epidemiol, Biomarkers Prev, 2021, 30(6):1200-1207. DOI:10.1158/1055-9965.epi-20-1610.
- [27] Curado MP, Edwards B, Shin HR, et al. Cancer incidence in five continents[M]. Lyon:IARC Scientific Publications,2007.
- [28] 陆伟.病例队列研究的设计及分析[J].疾病控制杂志,2001,5(2):148-150. DOI:10.3969/j.issn.1674-3679.2001.02.020.
- [29] Lu W. Design and analysis of case-cohort study[J]. Chin J Dis Control Prev, 2001, 5(2):148-150. DOI:10.3969/j.issn.1674-3679.2001.02.020.
- [30] Sohn W, Lee HW, Lee S, et al. Obesity and the risk of primary liver cancer: a systematic review and meta-analysis[J]. Clin Mol Hepatol, 2021, 27(1):157-174. DOI:10.3350/cmh.2020.0176.
- [31] Yao KF, Ma M, Ding GY, et al. Meta-analysis reveals gender difference in the association of liver cancer incidence and excess BMI[J]. Oncotarget, 2017, 8(42):72959-72971. DOI:10.18632/oncotarget.20127.
- [32] Onland-Moret NC, van der A DL, van der Schouw YT, et al. Analysis of case-cohort data: a comparison of different methods[J]. J Clin Epidemiol, 2007, 60(4):350-355. DOI:10.1016/j.jclinepi.2006.06.022.