

实现logistic与Cox回归相乘相加交互作用的临床实践宏程序

聂志强 欧艳秋 庄建 曲艳吉 麦劲壮 陈寄梅 刘小清

510080 广州,广东省心血管病研究所 广东省华南结构性心脏病重点实验室 广东省人民医院 广东省医学科学院流行病学研究室(聂志强、欧艳秋、曲艳吉、麦劲壮、刘小清),心脏外科(庄建、陈寄梅)

通信作者:刘小清, Email:drxqliu@163.com

DOI:10.3760/cma.j.issn.0254-6450.2016.05.031

【摘要】 病例对照研究常采用条件或非条件logistic分析,生存资料分析常采用Cox比例模型,但多数文献仅纳入主效应模型,然而广义线性模型不同于一般线性模型,其交互作用分为相乘交互与相加交互作用,前者只有统计学意义而后者更符合生物学意义。笔者以SAS 9.4软件编写宏,在计算logistic与Cox相乘交互项同时计算交互对比度、归因比、交互作用指数指标及利用Wald、Delta、PL(profile likelihood)3种方法的可信区间评价相加交互作用,便于临床流行病学与遗传学大数据分析相乘相加交互作用时参考。

【关键词】 logistic回归;Cox比例风险模型;交互作用

基金项目: 国家自然科学基金(U1401255);国家“十二五”科技支撑计划(2011BAI11B22, 2012BAI04B05);广东省科技计划项目(2012B032000014, 2011B031900002, 2013B030400001);广东省国际合作项目(2014A050503048);广东省医学科研基金(C2012012)

Application of SAS macro to evaluated multiplicative and additive interaction in logistic and Cox regression in clinical practices

Nie Zhiqiang, Ou Yanqiu, Zhuang Jian, Qu Yanji, Mai Jinzhuang, Chen Jimei, Liu Xiaoqing

Department of Cardiac Surgery, South China Key Laboratory of Structural Heart Disease, Guangdong Cardiovascular Disease Institute, Division of Epidemiology (Nie ZQ, Ou YQ, Qu YJ, Mai JZ, Liu XQ), Department of Pediatric Cardiology (Zhuang J, Chen JM), Guangdong People's Hospital, Guangdong Academic of Medical Science, Guangzhou 510080, China

Corresponding author: Liu Xiaoqing, Email: drxqliu@163.com

【Abstract】 Conditional logistic regression analysis and unconditional logistic regression analysis are commonly used in case control study, but Cox proportional hazard model is often used in survival data analysis. Most literature only refer to main effect model, however, generalized linear model differs from general linear model, and the interaction was composed of multiplicative interaction and additive interaction. The former is only statistical significant, but the latter has biological significance. In this paper, macros was written by using SAS 9.4 and the contrast ratio, attributable proportion due to interaction and synergy index were calculated while calculating the items of logistic and Cox regression interactions, and the confidence intervals of Wald, delta and profile likelihood were used to evaluate additive interaction for the reference in big data analysis in clinical epidemiology and in analysis of genetic multiplicative and additive interactions.

【Key words】 Logistic regression; Cox proportional hazard model; Interactions

Fund programs: National Natural Science Foundation of China (U1401255); National 12th Five-Year Support Projects of China (2011BAI11B22, 2012BAI04B05); Guangdong Provincial Science and Technology Planning Project (2012B032000014, 2011B031900002, 2013B030400001); Guangdong Provincial International Cooperative Project (2014A050503048); Guangdong Provincial Medical Scientific Research Funding (C2012012)

国外流行病学或临床研究越来越多将交互作用纳入标准化分析程序。国内常用logistic回归主效

应模型中纳入因素的乘积交互项,并根据乘积项的意义判断有无交互作用。此类分析误区是无法准确

理解统计建模交互项,错误将相乘交互等于相加交互。Rothman 等^[1]详细评价了相加交互与相乘交互的差异,认为统计建模中一般线性模型交互项反映的是因素间相加交互作用(additive interactions, INT_A),而 logistic 等广义线性模型则反映因素间统计学上相乘交互作用(multiplicative interaction, INT_M),即 logistic 乘积项仅反映统计学上的交互,而只有在生物学机制上病因因素间存在相加交互作用才可解释为相加交互作用。定量评价流行病学研究中暴露因素间及暴露因素与基因间相加交互作用需要根据 3 项指标,即交互对比度(interaction contrast ratio, ICR)又称相对超危险度比(relative excess risk due to interaction, $RERI$),归因比(attributable proportion due to interaction, AP)和交互作用指数(the synergy index, S)来计算。其中 ICR 的点估计较为简单,而 $95\%CI$ 较为复杂。当前已有文献提供计算 ICR 程序,但其算法冗繁且实际分析需要大幅修改原程序使推广受限^[2-5]。Andersson 等^[6]制作 Excel 版本工具极大推广了 $ICR 95\%CI$ 运用,邱宏等^[7]亦分别对分类交互 Delta 法进行了阐述。但 Andersson 等的方法不适合大数据多变量程序化分析,且 SPSS 进行条件 logistic 运用 Cox 分析模块无法输出协方差矩阵,Delta 法估计置信区间易造成假阴性结果。Zou^[8]亦阐述了 Wald 置信区间法需基于风险比对称的前提,故 Richardson 推荐采用轮廓似然置信区间(profile likelihood confidence intervals, PL),利用 Wald 的 ICR 点估计重构 $ICR 95\%CI$ 。logistic 回归是最常见的报道交互作用模型^[9],我们认为同时利用 Delta 法、Wald 法、PL 法可更全面评价相互作用的效果,为此本文提供 SAS 9.4 宏程序,实现非条件/条件 logistic 和 Cox 比例风险模型 $ICR 95\%CI$ 的估计,旨在为临床与遗传流行病学大数据分析因素间交互作用提供简便的方法。

基本原理

Rothman 等多数流行病学者认为,生物学交互作用的评价应该基于相加尺度的 INT_A 而非相乘尺度的 INT_M ,并以此提出构建了 ICR 、 AP 、 S 3 个指标,用于评价因素间是否有区别于相乘交互作用的相加交互作用,进而为生物学交互作用的评价提供依据。如果两因素有相加交互作用,则 $ICR 95\%CI$ 、 $AP 95\%CI$ 应不包含 0, $S 95\%CI$ 应不包含 1。假设研究多风险因素中交互作用的两暴露因素为 AB,则 OR_{00} 表示 AB 均无暴露,即 $OR_{00}=1$; OR_{10} 表示 A 暴

露、B 无暴露, OR_{01} 表示 A 无暴露、B 暴露, OR_{11} 表示 A、B 均暴露。Rothman 评价相乘交互作用 $INT_M=OR_{A \times B}=OR_{11}/(OR_{10} \times OR_{01})$,即 logistic 回归乘积项 $95\%CI$ 不包含 1,表明有相乘交互作用。

相加交互作用 3 项指标定义为

$$ICR=OR_{11}-OR_{10}-OR_{01}+1, AP=ICR/OR_{11},$$

$$S=(OR_{11}-1)/(OR_{11}+1-ICR)=(OR_{11}-1)/(OR_{10}+OR_{01})$$

如果两因素无相加交互作用,则 ICR 和 AP 的可信区间应包含 0, S 的可信区间应包含 1。则二分类 logistic 回归模型中

$$\ln[P/(1-P)]=\beta_0+\beta_1A+\beta_2B+\beta_3A \times B,$$

$$OR_{10}=\exp(\beta_1), OR_{01}=\exp(\beta_2),$$

$$OR_{11}=INT_A=\exp(\beta_1+\beta_2+\beta_3)$$

实例分析

以 2004—2014 年广东省先天性心脏病(先心)监测网数据库病例对照资料为例^[10-11],分析孕母是否被动吸烟(自报围孕期家庭、工作任一环境中接触吸烟平均时长 > 15 min/d 且持续 1 周以上)与家庭人均月收入有无相乘和相加交互作用(表 1),并比对采用 Andersson 分析结果。

表 1 广东省先心监测网病例对照研究中家庭人均月收入与被动吸烟情况

家庭人均月收入(元)	分组	被动吸烟人数	
		有	无
≤ 2 000	病例	425	1 863
	对照	210	1 773
> 2 000	病例	598	1 840
	对照	522	2 221

如表 2, 计算得到 $OR_{10}=1.383(95\%CI: 1.211 \sim 1.579)$, $OR_{01}=1.268(95\%CI: 1.160 \sim 1.387)$, $OR_{11}=2.443(95\%CI: 2.048 \sim 2.914)$, $INT_M=1.393(95\%CI: 1.116 \sim 1.739)$ 。采用 Delta 法、Wald 法和 PL 法分别估计 ICR 、 AP 、 S 及其 $95\%CI$, 其交互作用见图 1 和表 3。本例 $INT_M 95\%CI > 1$ 有相乘交互作用, 孕母被动吸烟与家庭人均月收入低有正向相乘交互作用; Delta 法、Wald 法、PL 法估计 $ICR 95\%CI > 0$ 、 $AP 95\%CI > 0$ 、 $S 95\%CI > 1$, 孕母被动吸烟与家庭人均月收入低有协同相加交互作用, 即孕母家庭人均月收入低同时存在被动吸烟者, 其胎儿患先心的风险显著增加。 ICR 和 S 的意义相同, AP 表示全部病例中可归因于两因素交互作用的病例所占比例, 本例 PL 法中 $AP=0.324$, ($95\%CI: 0.153 \sim 0.519$), 说明全部先心病例中归因于被动吸烟与家庭人均月收入低

表 2 logistic 相乘交互作用评价指标

参数	病例/对照人数	OR 值 95%CI	P 值
OR_{00}	1 840/2 221	1.000	-
OR_{10}	598/522	1.383(1.211 ~ 1.579)	<0.001
OR_{01}	1 863/1 773	1.268(1.160 ~ 1.387)	<0.001
OR_{11}	425/210	2.443(2.048 ~ 2.914)	<0.001
INT_M	4 726/4 726	1.393(1.116 ~ 1.739)	<0.001

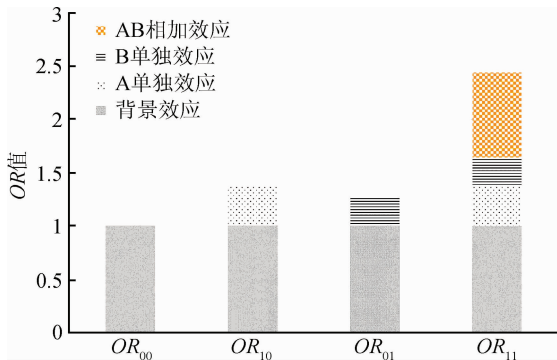


图 1 相加交互作用示意图

的相加交互作用所引起的病例占 32.4%。

讨 论

本研究采用 ICR 等指标评价围孕期被动吸烟与家庭人均月收入既存在正向相乘交互作用又存在生物学上协同相加交互作用,表明 Andersson 法、Delta 法、Wald 法、PL 法均能较好解释生物学相加作用。

统计学交互作用的线性模型一般为加法模型,乘积项表示有无相加交互作用;而广义线性属于乘法模型,乘积项表示有无相乘交互作用。生物学交互作用是多风险因素在发病的生物机制上的定性概念,两因素皆为病因前提下生物学机制的相互联系,包括协同和拮抗,生物学交互不同于统计模型中乘积项的分析。故相乘交互项 OR 值 95%CI < 1 为负相乘,OR 值 95%CI > 1 为正相乘,相加交互项 ICR 值 95%CI > 0 为协同作用, < 0 为拮抗作用。有时相乘与相加交互方向相悖,此时更宜运用生物学交互。在置信区间估计方法的选择上,Andersson 等使用 Delta 法估计,显不够稳健易造成假阴性结果,如分析广东省先心监测网数据库中孕妇被动吸烟与围孕期饮酒致新生儿先心的关系,采用 Delta 法,则 ICR = 7.635 (95% CI: -1.777 ~ 17.046), 而 PL 法为

表 3 logistic 相加交互作用评价指标及其 95%CI

参数	ICR 值(95%CI)	AP 值(95%CI)	S 值(95%CI)
Andersson 法	0.792(0.349 ~ 1.235)	0.324(0.187 ~ 0.461)	2.216(1.453 ~ 3.380)
Delta 法	0.791(0.348 ~ 1.234)	0.324(0.194 ~ 0.454)	2.215(1.452 ~ 3.379)
Wald 法	0.792(0.349 ~ 1.235)	0.324(0.187 ~ 0.461)	2.216(1.281 ~ 3.151)
PL 法	0.792(0.373 ~ 1.268)	0.324(0.153 ~ 0.519)	2.216(1.573 ~ 2.947)

ICR = 7.635 (95% CI: 1.389 ~ 32.707), 由于前者假阴性导致未正确评估出二者实际存在的协同效应。在风险比不对称时,应采用 PL 法,因 Wald 法的结果不稳定。

本研究可有效弥补其他非条件 logistic 模型的不足,稍加改写就能计算条件 logistic 和 Cox 比例风险模型交互作用参数和 P 值。Andersson 等未尝试计算条件 logistic 的 ICR 值,需计算协方差矩阵再手工输入数据至 Excel 表格,在基因研究及大型临床流行病学课题中很容易遗漏数据或难以质控。SAS 宏在大数据分析中重复操作性方面则优势明显。国外已将 logistic 模型交互作用分析纳入标准化流程中^[12],本文宏程序较高效的运算了二分类条件/非条件 logistic 和 Cox 交互效应。源程序主要展示非条件 logistic 交互,临床资料结构复杂常常需要多种统计学方法建模,例如若想进行生存资料 Cox 交互作用和条件 logistic 分析,则仅需修改过程步骤以及协方差矩阵的相关位置参数即可便捷完成计算(宏程序见附录)。

附录:SAS 9.4 宏程序

```

%macro additive2dum_LOGISTIC (dat, a, b, statues, cov1...
covN);
data P1;
set &dat; XX = &a * &b;
PROC logistic OUTEST = EST DATA = P1 COVOUT;
/*生存资料 Cox 比例风险模型则修改为 PROC PHREG;
model time * &statues (0) = &a &b XX 且 BETA3 = t(BETAP
[, 2:4]); V3 = COV [1:3, 2:4]; */
class &A (ref = first) &b (ref = first) / param = ref;
MODEL &statues (EVENT = "1") = &a &b XX &cov1...
&covN / risklimits COVB include = 3 selection = stepwise;
/*条件 logistic 则 strata duizihao; stepwise 后加 ties = discrete;
且 BETA3 = t(BETAP [, 2:4]); V3 = COV [1:3, 2:4]; */
contrast "OR10" &a 1 &b 0 XX 0 / estimate = exp;
contrast "OR01" &a 0 &b 1 XX 0 / estimate = exp;
contrast "OR11" &a 1 &b 1 XX 1 / estimate = exp;
RUN;
PROC EXPORT DATA = estOUTFILE = "D:\&b" DBMS =
excel replace; RUN;
%mend additive2dum_LOGISTIC;
%let ICR_LOGISTIC = %str(
PROC PRINT DATA = EST; RUN;
data parmscovb; set EST; drop _lnlike_;
if _type_ = 'PARMS' then output parms; elseif
_type_ = 'COV' then output covb; run;
PROC IML; USE parms; READ ALL var _all_ INTO
BETAP;
    
```



```

close parms;print betap;use covb;read all var_all_ into cov;
close covb;print cov;
BETA3=t(BETAP[,2,4]);V3=COV[2,4,2,4];PRINT
BETA3 V3;
C=BETA3[1,]+BETA3[2,]+BETA3[3,];D=EXP
(C)-1;
I=EXP(BETA3[1,])+EXP(BETA3[2,])-1;F=EXP
(BETA3[1,])+EXP(BETA3[2,])-2;H=EXP(C)/D;
ICR=EXP(C)-EXP(BETA3[1,])-EXP(BETA3[2,])+1;
AP=ICR/EXP(C);SI=D/F;
A=J(3,1,0);A[1,]=EXP(C)-EXP(BETA3[1,]);A
[2,]=EXP(C)-EXP(BETA3[2,]);
A[3,]=EXP(C);VARICR=t(A)*V3*A;Z=ICR/SQRT
(VARICR);
PVAL=2*(1-PROBNORM(ABS(Z)));L_ICR=ICR-1.96*
SQRT(VARICR);
H_ICR=ICR+1.96*SQRT(VARICR);B=J(3,1,0);
B[1,]=(EXP(BETA3[2,]-1))/exp(C);
B[2,]=(EXP(BETA3[1,]-1))/exp(C);
B[3,]=I/exp(C);VARAP=t(B)*V3*B;
L_AP=AP-1.96*SQRT(VARAP);H_AP=AP+1.96*SQRT
(VARAP);S=J(3,1,0);
S[1,]=H-EXP(BETA3[1,])/F;S[2,]=H-EXP(BETA3
[2,])/F;S[3,]=H;
VARS=t(S)*V3*S;L_LSI=log(SI)-1.96*SQRT(VARS);
H_LSI=log(SI)+1.96*SQRT(VARS);L_SI=exp(L_LSI);
H_SI=exp(H_LSI);PRINT ICR L_ICR H_ICR AP L_AP
H_AP SI L_SI H_SI Z PVAL;
run;quit;
option mcompilenote=all;
%macro addit_sign(dat,a,b,y,beta0,beta1,beta2,beta3);
PROC NLMIXED DATA=&dat DF=10000;
PARMS b0=&beta0 b1=&beta1 b2=&beta2 b3=&beta3;
odds=exp(b0+b1*&a+b2*&b+b3*&a*&b);pi=odds/
(1+odds);MODEL &y~Binary(pi);
estimate 'ICR=ICR using OR' exp(b1+b2+b3)-exp
(b1)-exp(b2)+1;
estimate 'AP using OR' (exp(b1+b2+b3)-exp(b1)-exp
(b2)+1)/(exp(b1+b2+b3));
estimate 'SI=ICR using OR' (exp(b1+b2+b3)-1)/(exp
(b1)+exp(b2)-2);run;
%mend addit_sign;option mcompilenote=all;
%macro addit_sign1(dat,a,b,y,beta0,beta1,beta2,
WaldICR);
PROC NLP DATA=&dat VARDEF=N technique=
dbldogpcov;
PARMS b0=&beta0, b1=&beta1, b2=&beta2, ICR=
&WaldICR;

```

```

b_interaction=log((ICR+exp(b1)+exp(b2)-1)/(exp(b1)
*exp(b2)));
if &y=1 then p=exp(b0+(b1*&a)+(b2*&b)+(b_
interaction*&a*&b))/(1+exp(b0+(b1*&a)+(b2*&b)+
(b_interaction*&a*&b)));
else p=1/(1+exp(b0+(b1*&a)+(b2*&b)+
(b_interaction*&a*&b)));
loglike=log(p);max loglike;PROFILE ICR/ALPHA=
0.05;RUN;
%mend;
%additive2dum_LOGISTIC(logistic,smoke,salary,y,cov1...
covN);&ICR_Logistic;
%addit_sign(logistic,smoke,salary,y,-0.1882,0.3241,
0.2377,0.3312);
%addit_sign1(logistic,smoke,salary,y,-0.1882,0.3241,
0.2377,0.7917);
利益冲突 无

```

参 考 文 献

- [1] Rothman KJ, Greenland S, Lash TL. Modern epidemiology[M]. 3rd ed. Philadelphia: Wolters Kluwer, Lippincott Williams & Wilkins, 2008.
- [2] Kalilani L, Atashili J. Measuring additive interaction using odds ratios[J]. Epidemiol Perspect Innov, 2006, 3: 5. DOI: 10.1186/1742-5573-3-5.
- [3] Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds[J]. Am J Epidemiol, 2009, 169(6): 756-760. DOI: 10.1093/aje/kwn411.
- [4] Nie L, Chu HT, Li F, et al. Relative excess risk due to interaction: resampling-based confidence Intervals[J]. Epidemiology, 2010, 21(4): 552-556. DOI: 10.1097/EDE.0b013e3181e09b0b.
- [5] Chu HT, Nie L, Cole SR. Estimating the relative excess risk due to interaction: a Bayesian approach[J]. Epidemiology, 2011, 22(2): 242-248. DOI: 10.1097/EDE.0b013e318208750e.
- [6] Andersson T, Alfredsson L, Källberg H, et al. Calculating measures of biological interaction[J]. Eur J Epidemiol, 2005, 20(7): 575-579. DOI: 10.1007/s10654-005-7835-x.
- [7] 邱宏, 余德新, 王晓蓉, 等. Logistic 回归模型中交互作用的分析及评价[J]. 中华流行病学杂志, 2008, 29(9): 934-937. DOI: 10.3760/cma.j.issn.0254-6450.2008.09.019.
- [8] Qiu H, Yu DX, Wang XR, et al. Study on the interaction under logistic regression modeling[J]. Chin J Epidemiol, 2008, 29(9): 934-937. DOI: 10.3760/cma.j.issn.0254-6450.2008.09.019.
- [9] Zou GY. On the estimation of additive interaction by use of the four-by-two table and beyond[J]. Am J Epidemiol, 2008, 168(2): 212-224. DOI: 10.1093/aje/kwn104.
- [10] Vander Weele TJ, Knol MJ. A tutorial on interaction[J]. Epidemiol Methods, 2014, 3(1): 33-72. DOI: 10.1515/em-2013-0005.
- [11] 聂志强, 欧艳秋, 陈寄梅, 等. 2004 至 2011 年广东省胎婴儿先天性心脏病危险因素分析[J]. 中华心血管病杂志, 2013, 41(8): 704-708. DOI: 10.3760/cma.j.issn.0253-3758.2013.08.016.
- [12] Nie ZQ, Ou YQ, Chen JM, et al. Risk factors of congenital heart defects in fetal and infants born from 2004 to 2011 in Guangdong [J]. Chin J Cardiol, 2013, 41(8): 704-708. DOI: 10.3760/cma.j.issn.0253-3758.2013.08.016.
- [13] Ou YQ, Mai JZ, Zhuang J, et al. Risk factors of different congenital heart defects in Guangdong, China [J]. Pediatr Res, 2016. DOI: 10.1038/pr.2015.264.
- [14] Knol MJ, Vander Weele TJ. Recommendations for presenting analyses of effect modification and interaction[J]. Int J Epidemiol, 2012, 41(2): 514-520. DOI: 10.1093/ije/dyr218.

(收稿日期: 2015-10-09)

(本文编辑: 张林东)