

doi: 10.3978/j.issn.2095-6959.2017.09.002

View this article at: <http://dx.doi.org/10.3978/j.issn.2095-6959.2017.09.002>

· AME 科研时间专栏 ·

专栏导读: AME Groups 旗下出版了 *Journal of Thoracic Disease* (《胸部疾病杂志》)、*Annals of Cardiothoracic Surgery* (《心胸外科年鉴》)、*Translational Cancer Research* (《癌症转化研究》) 和 *Annals of Translational Medicine* (《转化医学年鉴》) 等 40 余种英文医学学术期刊。2014 年, AME Groups 中文平台——“科研时间”的诞生, 为广大从事临床和基础研究的科研工作者带来了福音, 提供了更多科研交流和学习分享的机会。欢迎广大读者关注我们“AME 科研时间专栏”, 订阅我们的公众微信号(科研时间: amegroups), 给我们提出宝贵的建议和意见, 以便于将这个专栏建设得更好, 成为读者喜闻乐见的一个栏目。

使用 IRAP 2.2.0(interactive risk attributable program) 软件计算校正混杂因素的单个或多个危险因素人群归因风险

蔡东平¹, 贺永明²

(1. 苏州高新区狮山街道社区卫生服务中心, 江苏 苏州 215011;

2. 苏州大学附属第一医院心内科, 江苏 苏州 215000)

1 IRAP 软件

1.1 软件简介

人群归因风险(Population Attributable Risk, PAR)是用于统计分析人群暴露于某种/些危险因素, 发生疾病或死亡时归因于这种/些危险因素的权重指标。在国内外已发表的一些论文中, 存在多种危险因素的PAR值之和超过100%, 这与导致疾病的所有因素权重之和应为100%的逻辑不符, 提示既往在分析多因素的PAR时, 统计方法上并未考虑到多因素合并存在时多因素之间的交互作用。疾病危险因素被发现得越来越多, 在统计分析时, 疾病危险因素的模型越是完善, 其PAR值之和越接近100%。

美国国立癌症研究所(US National Cancer Institute)在上世纪90年代开发了IRAP(interactive risk attributable program)程序, 在回归分析基础上计算各因素的归因风险(attributable risk, AR)及其95%置信区间(95% confidence intervals, 95%CI), 可以矫正多因素合并存在时多因素之间的交互作用, 确保多因素的AR值之和不超过100%。

目前, 在美国国立癌症研究所网站([https://](https://dceg.cancer.gov/tools/risk-assessment/irap)

dceg.cancer.gov/tools/risk-assessment/irap)可以下载的是2002年发布的IRAP 2.2.0版, 为一个自解压文件。解压后得到四个文件(图1), 其中irap.exe是用来进行计算的程序。

名称	修改日期	类型	大小
Dos4gw.exe	1993/9/1 10:51	应用程序	239 KB
irap.exe	2002/7/17 8:11	应用程序	367 KB
irap.txt	2002/7/17 9:44	文本文档	49 KB
user-manual.wpd	2002/8/6 11:23	WPD 文件	497 KB

图 1 IRAP 软件内含文件列表

1.2 运行软件

1.2.1 直接运行

在Windows XP和Windows7(32位)操作系统下, 可以直接运行irap.exe。在IRAP界面上红色标记的字母为用于组合按键的字母。如打开“File”菜单, 使用Alt+F组合键进行操作。

1.2.2 DOS 模拟器内运行

在Windows 7(64位)及以后更新版本的操作系统下, irap.exe不能直接运行, 需要安装DOS模拟

收稿日期 (Date of reception): 2017-08-10

通信作者 (Corresponding author): 贺永明, Email: heyongming@suda.edu.cn

器。本文采用的是DOSBox，在网站www.dosbox.com下载安装程序。

DOSBox并不能直接识别irap.exe所在的文件夹，需要先将IRAP的文件夹加载到DOSBox的虚拟盘符下(图2)。为方便每次运行irap.exe，在开始→所有程序→DOSBox-0.74菜单下找到并点击“DOSBox 0.74 Options”，打开配置文件“dosbox-0.74.conf”，在该配置文件的最后一段[autoexec]下逐行添加启动irap.exe需要用到的命令(图3)。

完成以上配置后，每次点击DOSBox即可自动进入IRAP程序。不管是直接运行或在DOS模拟器内运行，IRAP的界面(图4)是一致的。

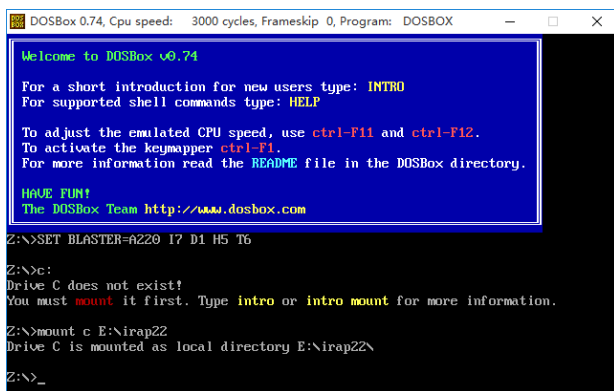


图2 DOSBox界面

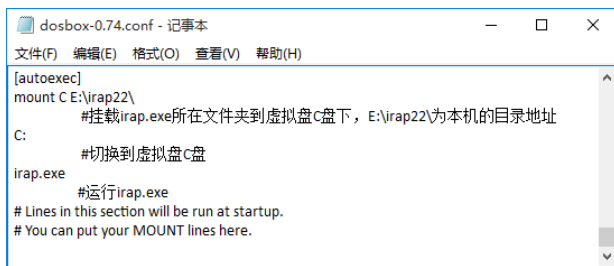


图3 DOSBox配置文件



图4 IRAP 2.2.0界面

2 实例解析

《Risk burdens of modifiable risk factors incorporating lipoprotein (a) and low serum albumin concentrations for first incident acute myocardial infarction》^[1]以多种危险因素联合矫正的logistic回归模型，通过DOSBox运行IRAP软件计算了在各年龄层男女人群，脂蛋白(a)、低血浆白蛋白水平2项新危险因素及高血压、糖尿病、缺血性脑病史、出血性脑病史、吸烟、饮酒、高密度脂蛋白胆固醇、低密度脂蛋白胆固醇、体重指数、三酰甘油、肌酐等11项传统危险因素，在首次发生的急性心肌梗死事件中各因素单独的人群归因风险(population attributable risk, PAR)及多种危险因素组合下的总人群归因风险(PARs combination for risk factors)。

2.1 文件准备

实例中使用多因素充分矫正的logistic回归分析模型涉及到17个变量(图5，表1)，其中连续变量Lp(a)，BMI，LDL-C，HDL-C，TG，ALB和Cr均被五分位分层，转变为等级变量后进入模型。

2.1.1 待分析的数据文件 (Data File)

IRAP不能直接调用Excel格式的数据，需要把Excel文件中除去变量名称这一行以外的内容以txt格式的文本文件保存，内容之间不含任何空格或制表符(图6)。从左往右一共17个数字，一行数字代表一个病例17个变量对应的具体值。IRAP正确解读Data File，需要用到变量定义文件(Library)。

2.1.2 变量定义文件 (Library)

Library的作用是让IRAP正确识别各个变量在Data File中的位置。“New library”(图7)即从头开始定义变量属性，使界面恢复到初始状态(图4)。按键Alt+V→A或直接按组合键Alt+A，进入添加变量(Add Variables)设定界面(图8)。各个变量的属性设定在IRAP主界面中“Variables”栏内显示(图9)。

Name为变量名称，IRAP不区分字母大小写，Group与group实为同一变量名；Column为该变量在Data File中起始于第几位数值；Length为Data File中数值的占用字符位数(实例中所有变量占用字符位数均是1)；Precision标记小数位数，如某变量值为1.234，其Length为4，Precision为3(实例中所有变量为整数，此处留空)；Missing Code标记缺失值，可以留空，也可以按照生成Data File时约定的缺失值代码填写；Number为第几个变量，默认跟随录入变量顺序递增。

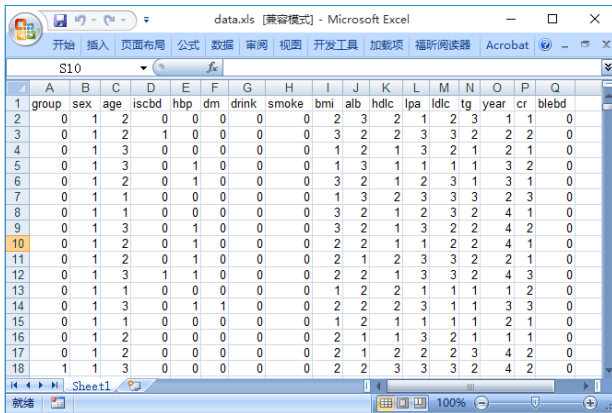


图5 Logistic回归分析使用数据

表1 变量及赋值意义

序号	变量名	意义
1	group	首发急性心梗：是1、否0
2	sex	性别：女1、男2
3	age	年龄：由低到高五分位
4	iscbd	缺血性脑病史：是1、否0
5	hbp	高血压病：是1、否0
6	dm	糖尿病：是1、否0
7	drink	饮酒嗜好：是2、已戒1、否0
8	smoke	吸烟嗜好：是2、已戒1、否0
9	bmi	体重指数：三分位
10	alb	白蛋白：由高到底五分位
11	hdlc	HDL-C：由高到低五分位
12	lpa	Lp(a)：由低到高五分位
13	ldlc	LDL-C：由低到高五分位
14	tg	TG：由低到高五分位
15	year	实验室检测年份
16	cr	Cr：由低到高五分位
17	blebd	出血性脑病史：是1、否0

将所有变量定义完成后，如图9所示，为再次运行IRAP计算时能快速调用，可将定义方案(library)保存(Alt+F→S)。默认保存位置“Directory: C:\”，此处C盘并非电脑直接管理的C盘，而是加载IRAP文件夹后对应的虚拟C盘，故变量定义文件默认保存在IRAP文件夹(图10)。保存为文件时可以不带扩展名，调用(Alt+F→O)时需输入与保存时一致的文件名(图11)。

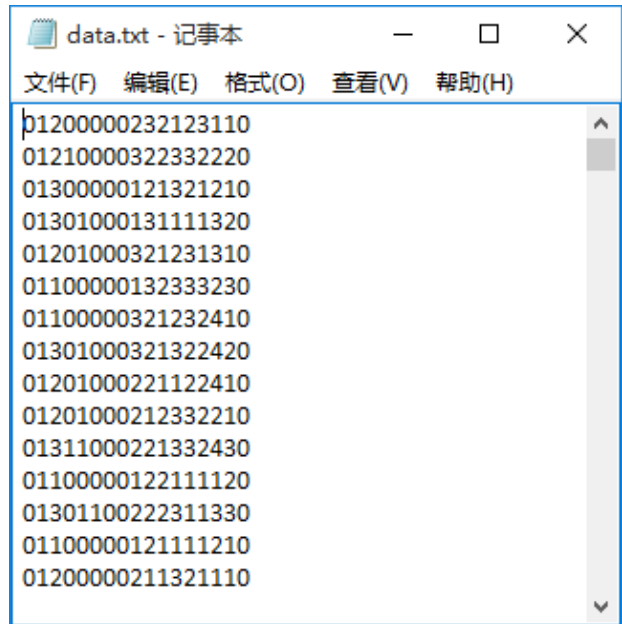


图6 Data File示意图

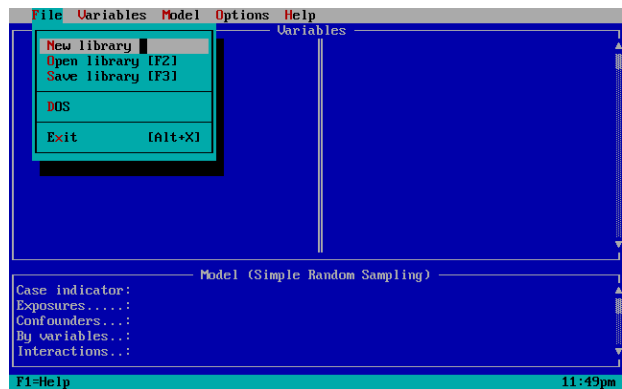


图7 操作library

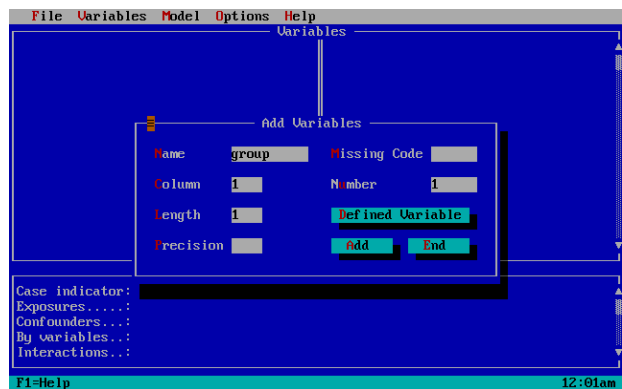


图8 添加变量设定界面

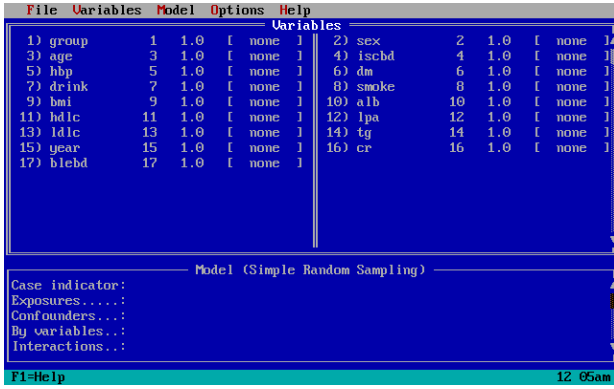


图9 变量定义一览

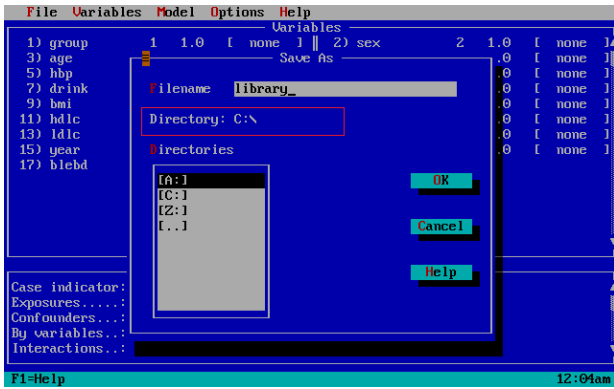


图10 Library文件保存界面

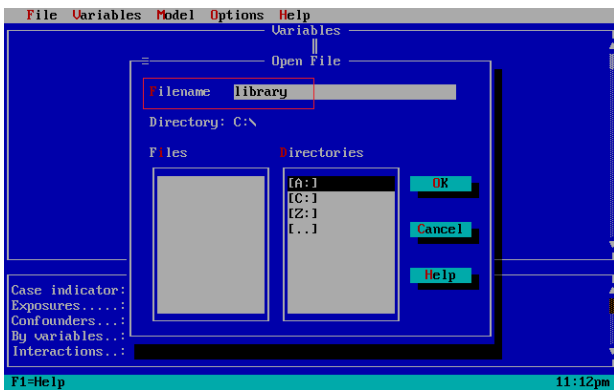


图11 Library文件引入界面

将准备好的数据文件与irap.exe、变量定义文件保存在同一个文件夹内(如图12方框内所示), 接下来进入PAR的计算过程。

名称	修改日期	类型	大小
Dos4gw.exe	1993/9/1 10:51	应用程序	239 KB
irap.exe	2002/7/17 8:11	应用程序	367 KB
irap.txt	2002/7/17 9:44	文本文件	49 KB
user-manual.wpd	2002/8/6 11:23	WPD 文件	497 KB
data.txt	2017/4/3 20:03	文本文件	142 KB
library	2017/4/3 20:03	文件	1 KB

图12 文件准备

2.2 运行

2.2.1 选择研究类型

默认类型为Simple Random Sampling, 更改后的类型名称在IRAP主界面下方的Model后的“()”内显示。实例中的研究为横断面研究, 在样本方案选项中选择Cross-sectional Analysis(Alt+M→S→r, 图13)。

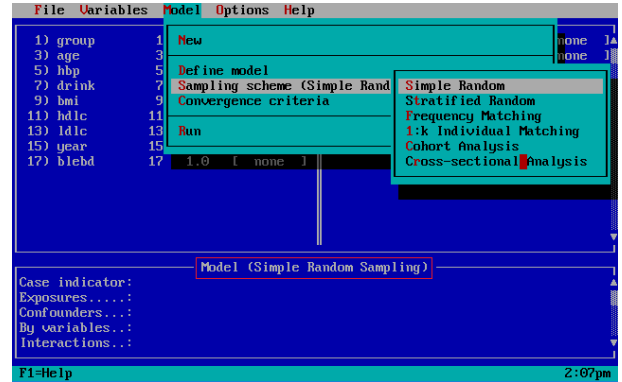


图13 研究类型选择界面

2.2.2 定义 PAR 计算模型

2.2.2.1 单因素 PAR 的计算

进入Define model界面(Alt+M→D, 图14), 对各变量进行PAR计算模型设定。“Case indicator”代表分组情况的变量, 实例中对应的是group, 序号为1; 研究的危险因素是脂蛋白(a), “Exposures”对应lpa, 序号为12。录入时不必输入完整的变量名称, 可以用变量的序号代替变量实际名称, 确认后在主界面下方显示出完整的变量名称(图15)。

2.2.2.2 合并多因素 PAR 的计算

进入Define model界面, 在“Exposures”一栏输入合并多因素的各自变量序号, “Confounders”一栏输入各混杂因素的序号(图16)。确认后得到如图17所示的PAR计算模型。

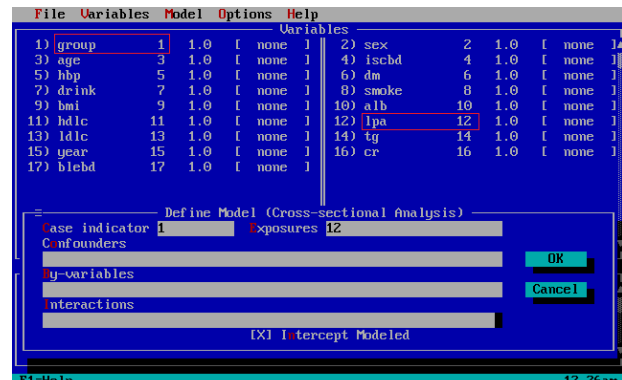


图14 定义模型输入界面

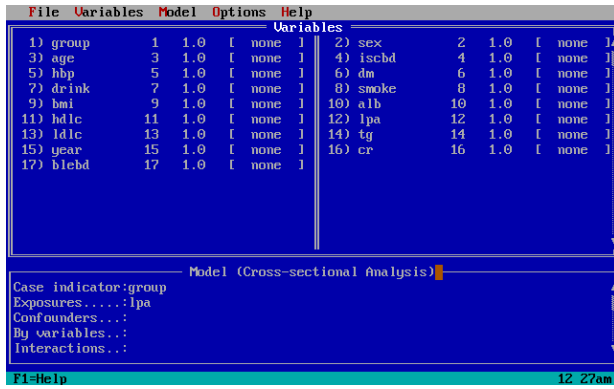


图15 单因素PAR计算模型

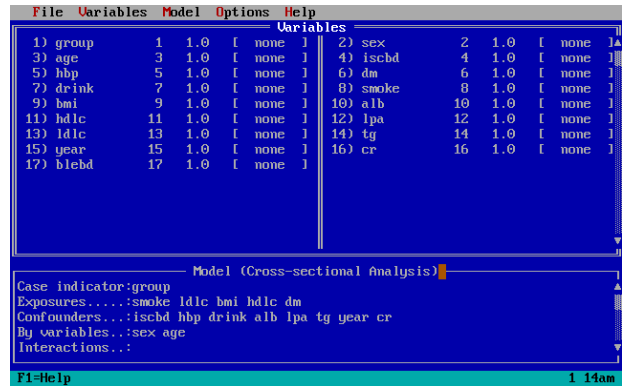


图18 亚组人群中合并多因素PAR的计算模型

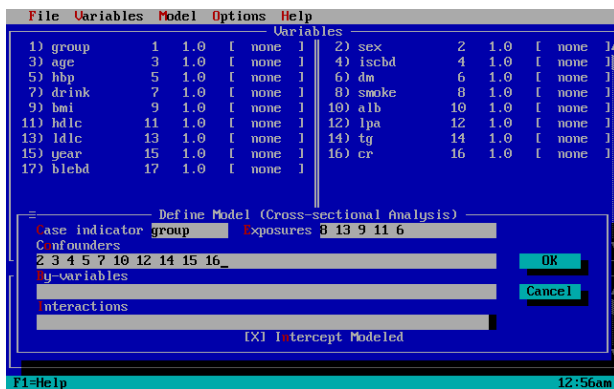


图16 合并多因素PAR计算模型的录入

2.2.3 运行计算

完成以上设定后，准备运行PAR计算模型。按Alt+M→R 或Alt+R组合键，弹出PAR计算前最后一个交互界面(图19)，在此填写Data File的完整名称和计算结果保存文件名称，选中“Partial attributable risk estimates”。建议Report file处填写扩展名“txt”，方便用文本程序打开查看。

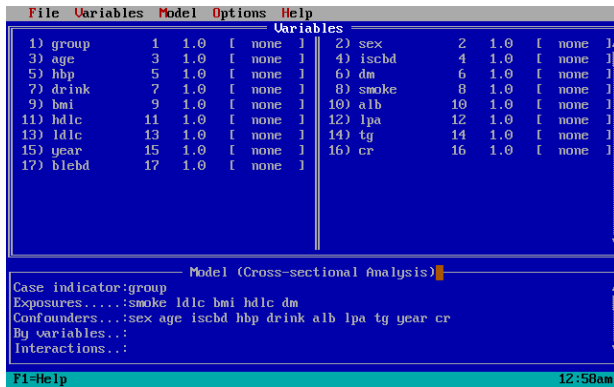


图17 合并多因素PAR计算模型

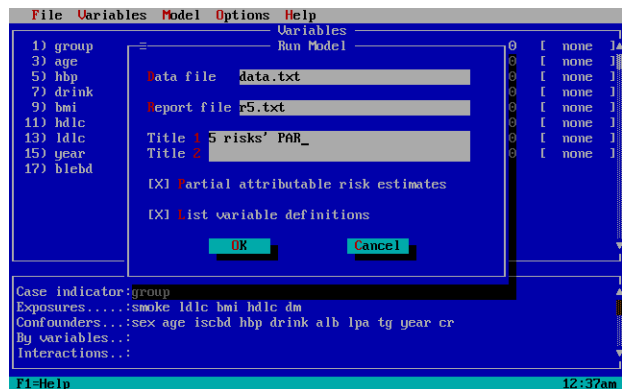


图19 Run Model界面

2.2.2.3 亚组人群中合并多因素 PAR 的计算

进入Define model界面，在“Exposures”一栏输入合并多因素的各自变量序号，“Confounders”一栏输入各混杂因素的序号，“By variables”一栏输入分亚组的变量序号。确认后得到如图18所示的PAR计算模型。

2.3 结果解析

按照以上建立的3种PAR计算模型，分别运行后在IRAP文件夹中共得到3个计算结果文件(图20)。

结果文件(Report file)第一部分内容是产生文件中列出计算结果的环境，如原始数据文件(Input data file)、研究类型(Sampling method)、PAR计算模型(含暴露因素、混杂因素、亚组分组等设置情况)和Library中的设置介绍。

名称	修改日期	类型	大小
Dos4gw.exe	1993/9/1 10:51	应用程序	239 KB
irap.exe	2002/7/17 8:11	应用程序	367 KB
irap.txt	2002/7/17 9:44	文本文档	49 KB
user-manual.wpd	2002/8/6 11:23	WPD 文件	497 KB
data.txt	2017/4/3 20:03	文本文档	142 KB
library	2017/4/3 20:03	文件	1 KB
R1LPA.TXT	2017/4/6 0:24	文本文档	4 KB
R5.TXT	2017/4/6 1:02	文本文档	82 KB
R5SA.TXT	2017/4/6 1:20	文本文档	482 KB

图20 计算结果文件

第二部分内容为IRAP计算后的结果，按照 By Variables的亚组分组给出暴露因素在每一亚组人群的总PAR(Total Attributable Risk for all combinations of exposures: …)，其中又按照暴露因素的暴露强度不同组合情况，给出每一种组合下的PAR计算结果(Partial Attributable Risk for the following exposure pattern: …)，所有组合PAR之和等于总PAR。计算结果文件的结构详见表2。

查看举例的单因素PAR的计算结果文件“R1LPA.TXT”(图21)，“Total Attributable Risk

for all combinations of exposures: AR: 0.3003, 95%CI (0.2398, 0.3609) …Partial Attributable Risk for the following exposure pattern: lpa= 2 AR : 0.0953, 95%CI(0.0611, 0.1294)…”，为实例文中表2中Lp(a)未经矫正时单一因素PAR(95%CI)的数据来源，以保留小数点后两位小数给出。

查看举例的合并多因素PAR的计算结果文件“R5.TXT”，其中“Total Attributable Risk for all combinations of exposures: AR: 0.7597, 95%CI (0.7095, 0.8099)”(图22)，是实例表3(图23)所列出Smoking, LDL-C, BMI, HDL-C&DM各合并多因素PAR(95%CI)的数据来源。

查看举例的亚组人群中合并多因素PAR的计算结果文件“R5SA.TXT”，sex=2 & age=1时“Total Attributable Risk for all combinations of exposures: AR: 0.8281, 95%CI(0.7321, 0.9241)”(表2)，是实例表3(图24)所列出Smoking, LDL-C, BMI, HDL-C&DM各合并多因素在Men_age1亚组中PAR(95%CI)的数据来源。

论文中PAR(95%CI)数据以保留小数点后两位小数给出。

表2 亚组人群中合并多因素PAR的计算结果文件解析

说明	计算结果文件内容片段
第一部分: 计算环境介绍	
页码、软件、计算时间	Page 1 IRAP v2.2.0 Thu Apr 6 01:16:23 2017
人工设定的标题	5 risks' PAR by sex & age
数据来源	Input data file : data.txt
PAR计算模型	Parameters: Sampling method : Cross-sectional Analysis Case indicator :group Exposures :smoke ldlc bmi hdlc dm Confounders :iscbd hbp drink alb lpa tg year cr By variables :sex age Interactions : '#': continuous variable
计算样本量	Records read: 7647 Records counted: 7647
提示信息	*** NOTE: Output is best viewed using 10 point Courier font. ***
变量说明(同library)	File description: 1) group 1 1.0 [none] ... 17) blebd 17 1.0 [none]
第二部分: 计算结果	
亚组结果	
亚组提示	Analysis for by-group number 1 sex = 1 age = 1

续表2

说明	计算结果文件内容片段
Logistic回归计算结果	<pre># Strata : 1 # Data cells : 804 # Covariates : 25 # Cases : 33 # Controls : 909 Logistic model CONVERGED after 27 iterations. -2 * Log-likelihood : 489.2382 Max. Absolute Score : 2.00000000e+000 ODDS RATIO 95% CI FOR O.R. VAR (VALUE) BETA S.E.(BETA) p EXP(BETA) LOWER UPPER Intercept -6.4153 0.9886 0.0000 0.0016 0.0002 0.0114 +smoke (1) -24.4062 0.0000 nan 0.0000 0.0000 0.0000 +smoke (2) 0.6820 1.1913 0.5663 1.9779 0.1915 20.4292 ... cr (3) -0.4178 0.8209 0.6101 0.6585 0.1318 3.2909 '+': Exposure parameter</pre>
AR及95%CI 合并多因素时	<pre>Total Attributable Risk for all combinations of exposures: 95% CI 95% CI (Logit transform) AR: 0.7828 Std. error: 0.1563 Robust std. error: 0.1439 (0.5007, 1.0649) (0.4068, 0.9498) # Exposed Cases (% of total) : 33 (100.00%) # Exposed Controls (% of total) : 815 (89.66%)</pre>
多因素某种暴露强度组合时	<pre>Partial Attributable Risk for the following exposure pattern: ...</pre>
多因素某种暴露强度组合时	<pre>Partial Attributable Risk for the following exposure pattern: ...</pre>
亚组结果	
亚组提示	<pre>Analysis for by-group number 2 ...</pre>
logistic回归计算结果	<pre>Logistic model CONVERGED after 6 iterations. ...</pre>
AR及95%CI 合并多因素时	<pre>Total Attributable Risk for all combinations of exposures: 95% CI 95% CI (Logit transform) AR: 0.8281 Std. error: 0.0465 Robust std. error : 0.0490 (0.7321, 0.9241) (0.7105, 0.9043) # Exposed Cases (% of total) : 347 (99.71%) # Exposed Controls (% of total) : 1149 (98.12%)</pre>
多因素某种暴露强度组合时	<pre>Partial Attributable Risk for the following exposure pattern: ...</pre>
多因素某种暴露强度组合时	<pre>Partial Attributable Risk for the following exposure pattern: ...</pre>
亚组结果	<pre>Analysis for by-group number 3</pre>
计算总耗时	Seconds elapsed: 251

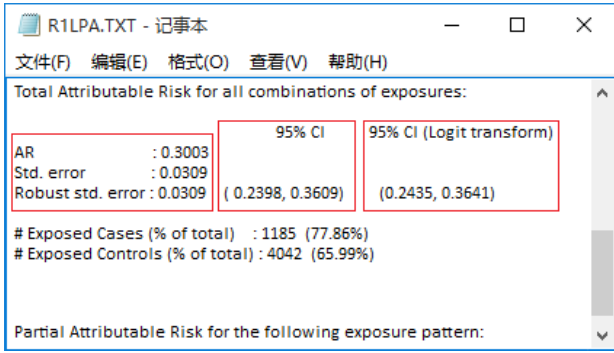


图21 单因素PAR的计算结果文件部分内容

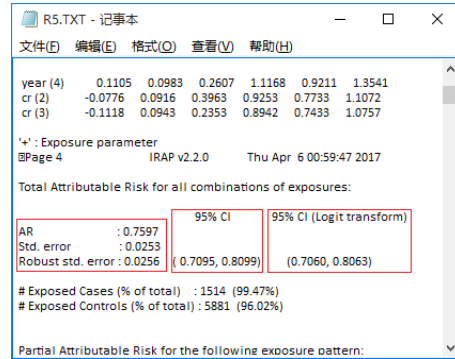


图22 多因素PAR的计算结果文件部分内容

表3 合并多因素时的PAR与多个单因素PAR之和结果比较

计算模型	Case indicator	Exposures	Confounders	Total AR
合并多因素PAR	1	6 8 9 11 13	2 3 4 5 7 10 12 14 15 16	0.7597
单因素(6)PAR	1	6	2 3 4 5 7 10 12 14 15 16 8 9 11 13	0.0805
单因素(8)PAR	1	8	2 3 4 5 7 10 12 14 15 16 6 9 11 13	0.3067
单因素(9)PAR	1	9	2 3 4 5 7 10 12 14 15 16 6 8 11 13	0.2084
单因素(11)PAR	1	11	2 3 4 5 7 10 12 14 15 16 6 8 9 13	0.1608
单因素(13)PAR	1	13	2 3 4 5 7 10 12 14 15 16 6 8 9 11	0.2466
多个单因素PAR之和				1.0030

www.nature.com/scientificreports/

Populations	PARs Combination for risk factors	PAR (95% CI)
Overall	Smoking, LDL-C, BMI, HDL-C & DM	0.76(0.71-0.81)
	LSA addition	0.89(0.86-0.92)
	Lp(a) addition	0.83(0.78-0.87)
	LSA & Lp(a) addition	0.92(0.90-0.94)
	LSA & Lp(a)	0.54(0.48-0.61)
Women	LDL-C, DM, BMI, TG & HT	0.72(0.61-0.83)
	LSA addition	0.87(0.80-0.94)
	Lp(a) addition	0.81(0.72-0.90)
	LSA & Lp(a) addition	0.91(0.86-0.96)
	LSA & Lp(a)	0.65(0.52-0.78)

图23 实例表3部分内容(1)

	LSA & Lp(a) addition	0.86(0.75-0.97)
	LSA & Lp(a)	0.63(0.40-0.87)
Men_age1	Smoking, BMI, LDL-C, DM & HDL-C	0.83(0.73-0.92)
	LSA addition	0.91(0.85-0.96)
	Lp(a) addition	0.87(0.80-0.95)
	LSA & Lp(a) addition	0.93(0.89-0.98)
	LSA & Lp(a)	0.48(0.37-0.59)
Men_age2	Smoking, DM, BMI, LDL-C & HDL-C	0.83(0.74-0.91)
	LSA addition	0.92(0.88-0.97)
	Lp(a) addition	0.87(0.80-0.94)
	LSA & Lp(a) addition	0.94(0.90-0.98)
	LSA & Lp(a)	

图24 实例表3部分内容(2)

实例文章中表3给出男女不同年龄阶段各种多因素组合时的多因素总PAR(95%CI)，其背后对应的是多次的调整PAR计算模型。建立PAR计算模型界面，在暴露因素(Exposures)一栏按统计模型输入各种因素，混杂因素(Confounders)一栏输入各矫正因素，亚组分组(By variables)一栏输入“sex age”。暴露因素(Exposures)越多，IRAP程序的计算时间花费越长，计算时间将随变量增加而延长，长时可至数小时，IRAP程序出现“假死”，需要耐心等待计算结果(图16)。

2.4 合并多因素时的 PAR 与多个单因素 PAR 之和

当研究合并多个因素的PAR时，应将多个因素同时作为暴露因素，并非分别计算单因素PAR之后求和。尽管在PAR计算模型设立时无论将因素划分为暴露因素还是混杂因素，只要纳入的因素一致，logistic回归分析的计算结果仍然一致，但对PAR的计算结果有影响。如上文“合并多因素PAR的计算”中合并多因素PAR的计算模型与该模型中涉及到的多个单因素PAR之和就有明显不同，暴露于合并5种因素时的AR值为75.97%，

而如果将这5种因素的AR值分别计算后求和则为100.30%(表3), 与所有因素AR值之和应为100%的逻辑不符。

3 总结

IRAP基于回归分析计算危险因素的归因风险度, 既可以计算单个危险因素的归因风险, 也可以计算多个因素合并存在时多因素的归因风险, 因素越全, 其结果越接近100%。尽管该程序运行

于DOS环境, 但不失为计算人群归因风险的有利工具。

参考文献

1. Yang Q, He YM, Cai DP, et al. Risk burdens of modifiable risk factors incorporating lipoprotein (a) and low serum albumin concentrations for first incident acute myocardial infarction[J]. *Sci Rep*, 2016, 6: 35463.

本文引用: 蔡东平, 贺永明. 使用IRAP 2.2.0(interactive risk attributable program)软件计算校正混杂因素的单个或多个危险因素人群归因风险[J]. *临床与病理杂志*, 2017, 37(9): 1777-1785. doi: 10.3978/j.issn.2095-6959.2017.09.002

本文首先以中文发表于【科研时间】(doi: 10.3978/kysj.2014.1.2870).