



流行病学基本方法

Ⅲ. 流行病学调查资料的调整

华北煤炭医学院 韩向午

流行病学分析时常需调整某些变量，使原因变量与结果变量的关系显示得更清楚。例如为确定吸烟对呼吸系统疾病死亡率的影响，年龄这一因素常与吸烟及死亡率纠缠在一起。因为，随年龄的增高死亡率也上升，年龄大的人吸烟者也多，不排除年龄因素就讲不清吸烟对死亡率的作用。资料调整的目的就是为了摆脱这种关系的混杂，使得能够在不受混杂变量干扰的情况下，去估价原因变量（吸烟）对结果变量（死亡率）的影响。

一、混杂作用与变量调整：某些变量的作用与其他一些变量的作用混淆在一起，影响资料的深入分析，这些变量称为混杂因素（变量），对这些变量必须进行资料调整。混杂因素经常是疾病的一个独立危险因素。混杂一词指的是某种（些）因素对所研究的原因变量（暴露因素）与结果变量（常指疾病）之间联系的干扰作用，这些因素就统称为混杂因素。混杂可使因素与疾病之间出现虚假联系（本不存在联系时）或掩盖联系的真相（本来实际上存在联系时），导致错误结论。

混杂因素必须同时与暴露因素及疾病都有联系才能成其为混杂。人的某些特征如年龄、性别、吸烟、饮酒、职业工种等，都常在流行病学资料分析中起混杂作用。因此流行病学资料中存在混杂因素总是客观存在的，不是研究人员因工作错误而带入的。现场研究中常难查清全部混杂因素，但必须把主要的足以影响研究结论的混杂因素查出来，事先设法避免其影响或事后进行资料调整以消除其影响。

1. 在设计阶段控制混杂因素：

①选择研究对象时按混杂因素给以条件限制，如只观察男性40~59岁的人；

②随机分组、均衡设计：如对年龄、性别、吸烟、身高等因素在观察组与对照组要求均衡分布；

③配对设计：按主要混杂因素进行配对，以保证

组间的可比性。

2. 在资料分析阶段消除混杂影响的方法：

①率的标准化：最常应用的是对年龄的调整，如比较不同地区的疾病率，必需消除年龄的混杂作用，以摆脱各亚组中年龄分配不均的曲解。

②分层分析：按混杂因素有无或其不同水平进行分层分析是消除混杂的简易方法。例如当年龄成为混杂因素时则按年龄分层，便可消除年龄的混杂作用。

③多元分析：当有多种混杂变量同时起作用则配对或分层分析都难以处理，如有两个主要混杂变量同时起作用可用双变量调整方法，三个或多个混杂变量可用协方差分析，多维列联表形式的资料可用Logistic回归分析，带时变协变量的资料可用COX回归模型来进行变量调整。

二、变量的直接调整与间接调整：疾病率（发病率，患病率，或死亡率）的年龄直接调整的基本概念是：假定地区甲、乙都是按标准人口的年龄分布来计算两地区的标准化率，然后比较这两个标准化率，因已考虑了年龄因素，所以甲乙两地区间的疾病率差别将不再受年龄分布不同的影响。

直接法计算地区甲乙标准化死亡率：

甲地区直接标准化死亡率：

$$\widehat{D_s R_a} = \frac{\sum_i P_{ia} n_{is}}{\sum_i n_{is}} \dots \dots \dots (1)$$

乙地区直接标准化死亡率：

$$\widehat{D_s R_b} = \frac{\sum_i P_{ib} n_{is}}{\sum_i n_{is}} \dots \dots \dots (2)$$

a: 地区甲；

b: 地区乙；

c. 选定的参考标准地区;

n_{ia} 甲地区第 i 年龄组的个体数;

n_{ib} 乙地区第 i 年龄组的个体数;

n_{is} 标准地区 i 组个体数;

X_{ia} 地区甲第 i 年龄组死亡数 (b 和 s 类推);

$P_{ia} = \frac{X_{ia}}{n_{ia}}$ 地区甲第 i 年龄组死亡率 (b 和 s 类推);

\sum_i 表示 $1, 2, \dots, i$ 个年龄组的合计。

表1是甲乙两地区死亡率直接调整率的计算。甲地区粗死亡率为 8.33% ，乙地区为 11.14% ，相对危险度 (乙/甲) 为 $11.14/8.33=1.34$ 。

1. 直接标化率的不同计算方法: 由于 $\sum_i n_{is}$ 是一个常数, (1)、(2) 两公式可改写成

表1 甲乙两地区死亡率及其直接调整率的计算

年龄组 (i)	两地区年龄别死亡率						标准人口 死亡率 P_{is}	直接标化的权重系数		甲乙两地 区的合并率 P_i	各年龄 组的估 计方差 $P_i q_i$
	甲地区		乙地区		按标准人口 $W_i = \frac{n_{is}}{\sum_i n_{is}}$	按 $P_{in}-P_{ib}$ 方差的倒数 $W_{yi} = \frac{n_{ia} n_{ib}}{n_{ia} + n_{ib}}$					
	$n_{ia}/1000$	X_{ia}	p_{ia}	$n_{ib}/1000$				X_{ib}	P_{ib}		
			(1)			(2)	(3)	(4)	(5)	(6)	(7)
<15	5524	8751	1.6	286	535	1.9	1.8	0.285	271922	1.6	0.0016
15~24	3558	4747	1.3	168	192	1.1	1.3	0.174	159996	1.3	0.0013
25~34	2677	4036	1.5	110	152	1.4	1.6	0.123	105658	1.5	0.0015
35~44	2359	6701	2.8	109	313	2.9	3.1	0.114	104186	2.8	0.0028
45~54	2330	15675	6.7	110	759	6.9	7.3	0.114	105041	6.7	0.0067
55~64	1704	26276	15.4	94	1622	17.3	16.6	0.091	89086	15.5	0.0153
65~74	1105	36259	32.8	69	2690	39.0	35.8	0.061	64945	33.2	0.0321
75+	695	63840	91.7	46	4788	104.1	96.6	0.038	43148	92.5	0.0839
合计	19953	166285	8.3	992	11051	11.1	9.4	1.000	943982	/	

$$\widehat{D_s R_a} = \sum_i P_{ia} \left(\frac{n_{is}}{\sum_i n_{is}} \right)$$

$$\widehat{D_s R_b} = \sum_i P_{ib} \left(\frac{n_{is}}{\sum_i n_{is}} \right) \dots\dots\dots (3)$$

用公式 (3) 的直接调整结果是甲地区 $\widehat{D_s R_a} = 8.8\%$ [即表1第一列 (1) 与第四列 (4) 各行乘积之和]。乙地区 $\widehat{D_s R_b} = 9.9\%$ [表1 (2) 与 (4) 各行乘积之和]。可见排除了年龄的混杂后, 相对危险度变为: $9.9/8.8 = 1.12$ 。如以甲乙两地区差别方差的倒数作权重系数, 进行调整见表2的第3、4栏, 其直接调整的结果, 甲地区 $D_s R_a = 9.8\%$ ($9255844.5/943982$) 乙地区 $D_s R_b = 11.0\%$ ($10433240.5/943982$)。但两者的比值 (相对危险度) 仍为 1.12 ($11.0/9.8$)。可见用不同的标准进行调整所得的调整率确有不同, 而它们的比值却相同。

2. 直接标化率差别的方差和显著性检验: 两个直接标化率差别方差 $\widehat{Var} (D_s R_a - D_s R_b)$ 、用下式

来计算:

$$\widehat{Var} (D_s R_a - D_s R_b) = \sum_i P_i q_i W_i^2 \left(\frac{1}{n_{ia}} + \frac{1}{n_{ib}} \right) \dots\dots\dots (4)$$

以表1 <15岁组资料代入公式 (4) 得 $0.0016 (0.285) \left(\frac{1}{272000} \right) = 0.0000000047794$, 为计算方便,

表2第五栏第一行的数字是这个值乘以100万的结果, 其余各组见表2第五栏的各行。第五栏各行合计结果

再除以100万即得 $\widehat{Var} (D_s R_a - D_s R_b)$

$$= 0.00000008196, \widehat{SE} (D_s R_a - D_s R_b)$$

$$= 0.000090532, \text{用 } Z \text{ 检验: } Z = \frac{0.0099 - 0.0088}{0.000090532}$$

$$= 12.15, P < 0.01.$$

即调整了年龄因素之后, 乙地区的死亡危险度仍高于甲地区, 差别有高度显著性。应该明确显著性是就误差而论的差别, 这样大的样本几乎总是有显著差别的, 因为误差与样本大小有关, 很大的样本即使很小的差别也会在统计学上显示出显著性。

表2

甲乙两地区死亡率的直接调整和间接调整及其误差计算

年龄组	以标准人口构成为权重系数的直接调整		以两地死亡率方差的倒数为权重系数的直接调整		两个直接标化率差异的方差	按标准人口死亡率的预期死亡数	
	甲地区 W _{xi.Pia}	乙地区 W _{xi.Pib}	甲地区 W _{yi.Pia}	乙地区 W _{yi.Pib}		甲地区 $\sum_i P_{is.nia}$	乙地区 $\sum_i P_{is.nib}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<15	0.4560	0.5415	435075.2	516651.8	0.000477794	9943.2	514.8
15~24	0.2262	0.1914	207994.8	175995.6	0.000245993	4625.4	218.4
25~34	0.1845	0.1722	158487.0	147921.2	0.000214090	4283.2	176.0
35~44	0.3192	0.3506	291720.8	302139.4	0.000349892	7312.9	337.9
45~54	0.7638	0.7866	703774.7	724782.9	0.000829269	17009.0	803.0
55~64	1.4014	1.5743	1371924.4	1541187.8	0.001423588	28286.4	1560.4
65~74	2.0088	2.3790	2130196.0	2532855.0	0.001837602	39559.0	2470.2
75+	3.4846	3.9958	3956671.6	4491706.8	0.002817479	67233.6	4443.6
合计	8.8365	9.9314	9255844.5	10433240.5	0.008195707	178252.7	10524.3

3. 间接标化率及其显著性检验：当所标化的地区不能提供人群的年龄别疾病率时，可用间接标化法。这种方法不要求地区的年龄别疾病率，但需提供①所标化地区的死亡总数，②所标化地区的人口年龄分布，③标准人口的年龄别疾病率。甲地区的间接标化死亡率：

$$ISRa = \frac{\sum_i X_{ia}}{\sum_i P_{is.nia}} \left(\frac{\sum_i P_{is.nis}}{\sum_i nis} \right) \dots\dots(5)$$

$\sum_i X_{ia}$ 即甲地区死亡总数 (166 285)， $\sum_i P_{is.nia}$ 即甲地区按标准人口死亡率应发生的预期死亡数，见表2第六栏 (178 252.7)。可见这两个数的比值即观察值与预期值的比，如再乘以100即得通常的SMR (标化死亡比)。(5)式的右半即标准人口死亡率，这样

$$ISRa = \frac{166285}{178253} (0.0094) = 0.0088, \text{同理乙地}$$

$$\sum_i P_{is.nib} = 10524.3 \text{ (表2第七栏)}.$$

$$ISRb = \frac{11051}{10524.3} (0.0094) = 0.0099, \text{甲乙两}$$

地的间接标化死亡率与以标准人口构成为权重系数的直接标化死亡率相同，一般来说所介绍的这两种方法的结果通常是差别不大的。甲乙两地区间接标化死亡率的方差和标准误：

$$\widehat{Var}(ISRa) = \frac{\sum_i X_{ia}}{(\sum_i P_{is.nis})^2} \left(\frac{\sum_i P_{is.nis}}{\sum_i nis} \right)^2 \dots(6)$$

$$\text{甲地区 } \widehat{Var}(ISRa) = \frac{166285}{(178253)^2} (0.0094)^2 = 4.62418 \times 10^{-10}$$

$$SE(ISRa) = 0.000021503907 \quad \text{同理}$$

$$\text{乙地区 } \widehat{Var}(ISRb) = \frac{11051}{(10524)^2} (0.0094)^2 = 8.816488 \times 10^{-9}$$

$$SE(ISRb) = 0.000093896154$$

也可同样进行Z检验，与直接标化率的显著性检验的结论也一致。

4. 双变量调整率及其显著性检验：当有两个重要的混杂变量会影响因素与疾病联系时，需对两个变量同时结合起来标化，而不是分开调整。因为分开调整有时会陷入令人无所适从的困境。例如研究两个煤矿尘肺患病率，粗率是乙矿显著高于甲矿，工龄调整患病率甲矿高于乙矿 (7.23%:6.23%)、工种调整患病率则相反，乙矿高于甲矿 (8.79:6.28)。当按工种工龄交叉分组进行双变量调整率分析，甲乙两矿尘肺的双变量调整患病率比较差异无显著性〔本方法详细算法参见中国卫生统计1985; 2(3):31〕。流行病学调查资料为了显示出目标变量对疾病率的影响，常遇到有两个另外的重要的较为突出的混杂变

量, 如吸烟与年龄、饮酒与年龄、吸烟与饮酒等, 有时需要结合起来同时调整, 尤其当两个混杂变量分开调整出现矛盾情况时需进行双变量调整率的分析与检验。

5. 有关调整率某些问题的进一步说明:

①疾病粗率更反映实际情况, 按混杂变量分组资料的调整只不过是计算加权平均而把“实情”概括为一个指数而已, 因此可以说调整率或直接标化率并不代表实情, 此点应该强调。

②选择不同的标准人群可以获得不同的标化结果, 因此单独一个标化率不说明问题, 其比粗率高或低都不反映事物的本质, 只有进行两个率的比较时标化率才有意义。

③计算标化率时用不同的参照标准, 如全国或全省的人口资料, 用不同的标化方法, 如以标准人口构成作权重系数的标化, 或用两组差别方差的倒数为权重系数者, 所得的直接标化率都会有所不同, 但不同方法标化后两组的比值却常相同, 如表1资料两种方法标化后比值都是1.12, 此点应很好理解。

④选择作为标准的人群应是与所分析的资料有关的, 亦即能覆盖研究地区的人群, 例如用研究单位所在的省或市的人口资料。

⑤两个地区的间接调整率的比较, 在理论上是不对的, 这与两个SMR不宜直接比较的道理相同。尽管表2资料

$$\widehat{ISR}_b / \widehat{ISR}_a = 1.125 \text{ 以及 } \widehat{SMR}_a / \widehat{SMR}_b = 1.125 \text{ (SMR}_a = \frac{166285}{178253} = 0.9329, \text{ SMR}_b = \frac{11051}{10524.3} = 1.050) \text{ 又都与 } D_s \widehat{R}_b / D_s \widehat{R}_a = 1.125 \text{ 相}$$

一致, 但不宜以间接标化率的比较代替直接标化率的比较。需知ISR和SMR都是总结性统计量, 有时难以把混杂变量不同层的差异反映在所分析的资料之中。

⑥直接标化率的显著性检验用于抽样研究, 而大规模普查或登记资料, 特别是数以百万计的人口疾病率或死亡率资料一般来说显著性检验的意义不大。样本过大会增加β类错误, 而且很大的样本即使两组只有百分之几的差别也总是在统计学上出现显著差异, 但其差别的实际意义往往很有限。

⑦一份资料中如明确查出有两个重要混杂变量, 可按这些变量的交叉分组进行双变量调整率的分析, 如有3~4个混杂变量可按多维列联表扩展为多变量调

整率的分析。当协变量超过5个以上应该用多元统计模型来处理。如是只有2个或3个重要混杂变量还是以双变量调整率分析的结果直观, 有利于结果的解释与业务内容的阐明。

三、四格表资料的混杂变量与分层分析: 研究暴露因素与疾病联系常用2×2表, 这种四格表的背后也会隐含着几个甚或许多混杂变量, 不考虑分析这些混杂变量, 因素与疾病联系的表现可能被歪曲。

1. 四格表资料的分层分析: 研究血压与冠心病的关系时, 年龄、吸烟、体型等许多因素可起混杂作用。以年龄对血压与冠心病的混杂分析为例见表3、4。

表3 年龄和收缩压与心肌梗塞的患病调查资料

		年龄(岁)		收缩压(mmHg)		合计
		≥60	<60	≥140	<140	
心肌梗塞	有	15	41	29	27	56
	无	188	1767	711	1244	1955
		$\widehat{OR} = 3.44$		$\widehat{OR} = 1.88$		

表4 冠心病患病调查资料收缩压与年龄的关系

	收缩压(mmHg)	年龄(岁)		合计	$\widehat{OR} = 3.04$
		≥60	<60		
收缩压	≥140	124	616	740	$\widehat{OR} = 3.04$
	<140	79	1192	1271	
合计		203	1808	2011	

表3、4说明年龄既与危险因素(收缩压)有联系(OR=3.44), 又与所研究的疾病有联系(OR=3.04)。可见年龄在血压与冠心病的联系研究中是一个明显的混杂因素。为了摆脱年龄的混杂作用, 常用分层方法, 按M-H分析技术(表5)。

表5不同年龄层的OR值不同, 其差异可以是因为不同年龄者血压对冠心病的发生有不同的影响, 即年龄与血压之间对疾病发生的危险性存在着交互作用, 也称为效应修正(Effect modification)。这时不同年龄层的血压对冠心病的比值比(OR)应分别表达而不宜合并表示。但其差异也有可能是抽样误差引起的, 实际上不同层的比值比没有显著差异, 这样就应将之合并表示, 为此需进行层别比值比的一致性检

表5

收缩压与心肌梗塞 (MI) 患病率的年龄别分层分析

	年龄 ≥ 60岁			年龄 < 60岁		
	MI病例	非病例	合计	MI病例	非病例	合计
收缩压 ≥ 140	9(a ₁)	115(b ₁)	124(m ₁₁)	20(a ₂)	596(b ₂)	616(m ₁₂)
< 140	6(c ₁)	73(d ₁)	79(m ₂₁)	21(c ₂)	1171(d ₂)	1192(m ₂₂)
合计	15(n ₁₁)	188(n ₂₁)	203(t ₁)	41(n ₁₂)	1767(n ₂₂)	1808(t ₂)
	$\widehat{OR}=0.95$			$\widehat{OR}=1.87$		

验, 公式为:

$$\chi^2_{\text{Homogeneity}} = \sum_{g=1}^G W_g (\ln \widehat{OR})^2 - \left(\sum_{g=1}^G W_g (\ln \widehat{OR}) \right)^2 / \sum_{g=1}^G W_g \dots\dots(7)$$

$$W_g = \left(\frac{1}{ag} + \frac{1}{bg} + \frac{1}{cg} + \frac{1}{dg} \right)^{-1}$$

a b c d 即四格表中各格子中的数值, 下标变量 g 表示不同层别. 表5中 ≥ 60岁者 g=1, < 60者 g=2. g=1 即

$$\text{年龄} \geq 60 \text{ 岁的权数 } W_1 = \left(\frac{1}{9} + \frac{1}{115} + \frac{1}{6} + \frac{1}{73} \right)^{-1} = 3.3314, \text{ 同理}$$

$W_2 = 9.9849$ $\chi^2 = 3.3314 (\ln 0.95)^2 + 9.9849 (\ln 1.87)^2 - [3.3314 (\ln 0.95) + 9.9849 (\ln 1.87)]^2 / (3.3314 + 9.9849) = 1.1457, G=2, df=2-1$ $P > 0.05$, 即分层的比值比无显著差别, 可以把不同年龄层的资料合并, 计算共同的比值比, 按 M-H 法

$$\widehat{OR}_{M=H} = \frac{(9 \times 73) / 203 + (20 \times 1171) / 1808}{(6 \times 115) / 203 + (21 \times 596) / 1808} = 1.57,$$

即消除了年龄的混杂作用后, 血压与冠心病估计的比值比为 1.57 而不是 1.88. 总的联系性的显著性检验用 $\chi^2_{M=H}$ 法, $\chi^2_{M=H} = 2.7943$. $\widehat{OR}_{M=H}$ 及 $\chi^2_{M=H}$ 两公式见本刊 1988; 9(1): 56. 比值比的 95% 可信限为 $\widehat{OR}_{M=H} (1 \pm 1.96 / \sqrt{\chi^2}) = 0.9251 \sim 2.6644$. 即排除年龄的混杂后血压与冠心病的联系还没有达到 $\alpha = 0.05$ 的显著水平, 这与 $\widehat{OR}_{M=H}$ 的 95% 可信区间包括 1 是一致的, 而与未分层资料的比值比 1.88, $\chi^2 = 4.92$ $P < 0.05$ 成鲜明对比.

2. 流行病学调查资料的类型与分层分析: 流行病学调查资料按其设计与统计分析格式的不同可分如下

类别: ①成组比较的病例-对照研究资料; ②配对或配比的病例-对照研究资料; ③现况调查的四格表资料; ④定群研究的累计发病率资料; ⑤定群研究的发病密度资料.

配对或配比资料因是按混杂因素配对的, 其未做配对的因素也不能再用配对资料进行混杂因素的分层分析, 除配对因素已消除其混杂作用外其他非配对因素的混杂或交互作用可用 Logistic 回归模型来处理. 发病密度资料因其格式与一般四格表不同, 统计处理方法也不同, 只能计算 RR 而不能计算 OR, 这类资料的分析也应分层计算相对危险度 (RR_g) 作显著性检验, 并作层别相对危险度的一致性检验, 如不同层者差别显著, 则不该求总的相对危险度, 仍应分别表示不同层的相对危险度, 以充分表达资料所蕴含的信息. 如果不同层者差别不显著 ($\chi^2_{\text{Heterogeneity}}$), 则不得不接受层别一致性的假设, 应进一步求分层资料总的相对危险度 $\widehat{RR}_{M=H}$, 进行联系的显著性检验 ($\chi^2_{M=H}$), 并求 $\widehat{RR}_{M=H}$ 95% 可信区间. 其余三类资料, 即第一及第三、四类都可同时计算 \widehat{OR} (发病率与患病率资料也可计算 RR), 也按上述程序逐步分析.

3. 不等数目多对照的配比资料分析方法: 配对是控制混杂的有效方法, 混杂变量一经配对使其在病例与对照组的分布一致则失去其混杂作用. 1:1 配对, 1:2 配比以及 1:R 多对照配比的方法教材上均有介绍, 现列出常用于队列内进行病例-对照研究的不等数目多对照的配比方法. 一般来说队列内的病例数总是有限的, 因而资料宝贵不宜轻易舍弃, 为提高统计效率常用多对照以便使样本较小的观察因素达到显著水平. 理想的是每个病例配以 4 例对照, 但在所研究的队列内有些病例配比不足 4 例, 只好配三例或两例对照, 更有甚者在队列内全部可供选配的非病例只能配

到一例合格的对照，为了尽可能地扩大样本，这样只能配到二例甚或仅一例的配比组都应全部被包括在分析之列，这时需采用不等数目多对照配比资料的设计与分析方法，举例如表6， \widehat{OR} 计算用公式(8)，

$$\widehat{OR} = \frac{\sum_{R,r=1}^R \sum (R-r+1)f_{1,r-1,R}}{\sum_{R,r=1}^R \sum r f_{0,r,R}} \quad \text{公式(8)}$$

表6 煤矿块状纤维化尘肺(PMF)与结核病联系的多对照配比病例-对照研究

病例与对照比	病例组暴露情况	每个配比组中对照者暴露(患结核病)数						合计观察组数
		0	1	2	3	4	R	
1:1	+	$f_{1,0,1}$ (1)	$f_{1,1,1}$ (0)					1
	-	$f_{0,0,1}$ (0)	$f_{0,1,1}$ (0)					
	小计	$T_1(1)$						
1:2	+	$f_{1,0,2}$ (2)	$f_{1,1,2}$ (0)	$f_{1,2,2}$ (0)				2
	-	$f_{0,0,2}$ (0)	$f_{0,1,2}$ (0)	$f_{0,2,2}$ (0)				
	小计	$T_1(2)$						
1:3	+	$f_{1,0,3}$ (4)	$f_{1,1,3}$ (4)	$f_{1,2,3}$ (2)	$f_{1,3,3}$ (1)			12
	-	$f_{0,0,3}$ (0)	$f_{0,1,3}$ (0)	$f_{0,2,3}$ (0)	$f_{0,3,3}$ (1)			
	小计	$T_1(4)$	$T_2(4)$	$T_3(3)$				
1:4	+	$f_{1,0,4}$ (5)	$f_{1,1,4}$ (8)	$f_{1,2,4}$ (7)	$f_{1,3,4}$ (2)	$f_{1,4,4}$ (0)		24
	-	$f_{0,0,4}$ (0)	$f_{0,1,4}$ (0)	$f_{0,2,4}$ (0)	$f_{0,3,4}$ (2)	$f_{0,4,4}$ (1)		
	小计	$T_1(5)$	$T_2(8)$	$T_3(9)$	$T_4(3)$			
1:R	+	$f_{1,0,R}$	$f_{1,1,R}$	$f_{1,2,R}$	$f_{1,r,R}$		
	-	$f_{0,0,R}$	$f_{0,1,R}$	$f_{0,2,R}$	$f_{0,r,R}$		
	小计	T_1	T_2	T_3	T_R		

注：括弧内数字为实际观察数

以病例与对照比1:3这项为例先说明 $f_{1,2,3}$ 与 $f_{0,3,3}$ 最前一个F标 f_1 表示一个配比组中病例为暴露者(即PMF病例是曾患结核病者)， f_0 表示病例为非暴露者(即未曾患结核病者)。最后一个F标“3”表示这两个配比组都是1:3配比者。所以 $f_{1,2,3}$ 表示此配比组为病例是暴露者，对照中有两人是暴露者，另一人是非暴露者，是1:3配比组。 $f_{0,3,3}$ 表示此配比组的病例是非暴露者，对照三人都是暴露者。表6资料代入公式(8)其分子项为 $[1 \times 1 / (1+1)] + [2 \times 2 / (1+2)] + [3 \times 4 / (3+1) + 2 \times 4 / (3+1) + 1 \times 2 / (3+1)] + [4 \times 5 / (4+1) + 3 \times 8 / (4+1) + 2 \times 7 / (4+1) + 1 \times 2 / (4+1)] = 19.33$ 、分母项为 $[0] + [0] + [3 \times 1 / 4] + [3 \times 2 / 5 + 4 \times 1 / 5] = 2.75$ 、 $\widehat{OR} = 19.33 / 2.75 = 7.0303$

$$\chi^2 = \frac{\left(\sum_{R,r=1}^R \sum (f_{1,r-1,R} - \frac{r}{R+1} T_{r,R}) \right)^2 \cdot \frac{1}{2}}{\sum_{R,r=1}^R \sum T_{r,R} \cdot r (R-r+1) / (R+1)^2}$$

df=1, (9)

表6资料代入公式(9)得 $\chi^2 = 30.9161$ 、df=1, $P < 0.01$ 。OR 95%可信区间 $\widehat{OR} (1 \pm 1.96 / \sqrt{X}) = 3.5352 \sim 13.9808$ ，结果说明接尘工人或单纯尘肺曾患过结核病者与未患结核病者相比发生块状纤维化尘肺的危险性高7倍，而且统计学上有高度显著性。

四、结束语：流行病学调查资料常需进行变量调整，用得最多的是年龄的调整，性别多用分组对比法。日常生活因素、嗜好等，则应视所调查分析的疾病及所研究的暴露因素而定。如研究职业因素与肺癌的

联系则应分组调整吸烟, 研究环境因素与肝癌则应调整饮酒量。为了较直观地对比分析2或3个协变量的影响, 可用双变量调整或用多维列联表对三个变量同时调整。协变量4~5个以上则多用Logistic回归模型来处

理。变量调整的主要目的是消除混杂影响, M-H 分层分析是消除混杂的简便方法, 除了获得 \widehat{OR} 之类信息以外, 应分层计算率直接求算RR值进行比较分析。

McAb与RPHA试剂检测血清HBsAg的比较

湖南省芷江县卫生防疫站 黄大清 戚杰

目前国内一般多用冻干HBsAg诊断血球(下称RPHA试剂)检测血清中HBsAg, 因生产厂家及标准不一而产生不同程度的假阳性和假阴性反应, 对临床诊断工作带来一定的影响。为了寻求一种优选的诊断试剂, 本文应用单克隆抗体(下称McAb血球)对338份血清标本进行了HBsAg检测, 并与RPHA试剂作了比较, 将筛出的阳性标本分别作中和试验。以探讨两种诊断试剂的特异性、灵敏性、假阳性、假阴性等问题, 现报告如下:

材料与方法

一、McAb血球: 卫生部兰州生物制品研究所出品, 批号8708, 有效期一年, 以该所血球稀释液稀释。

二、RPHA试剂: 上海生物制品研究所出品, 批号870601, 有效期一年, 以该所血球稀释液稀释。

三、HBsAg诊断血清: 卫生部北京生物制品研究所出品, 批号865, 有效期两年, 以无菌生理盐水作1:30倍比稀释备用。

四、操作方法: 按各诊断试剂说明书进行。

五、结果判定: McAb血球 $\geq 1:16$ 以上者为阳性, RPHA试剂以 $\geq 1:8$ 以上者为阳性, 阳性反应终点为卅。

六、标本来源: 健康体检的汽车司机, 共338人, 空腹抽取静脉血1~2ml分离血清供试验用。

结果

一、McAb与RPHA试剂检测结果的对比: 将血清标本分别用McAb血球及RPHA试剂作初筛试验, 阳

性标本再作中和试验, 结果表明, 338份血清标本McAb血球检出阳性45份, 阳性率为13.3%, RPHA试剂检出阳性37份, 阳性率为10.9%; McAb血球阳性检出率高于RPHA试剂, 阳性平均GMRT McAb血球为98.46, RPHA试剂为43.98, 阳性结果判定McAb血球(1:16)要高于RPHA试剂(1:8)一个稀释倍, 结果也证实了这一点, 而且McAb血球阳性标本GMRT大多集中在1:32倍之后, 统计学处理有非常显著性差异($P < 0.01$)。

二、阳性结果的比较: 两种血球共检出阳性标本49份, McAb血球与RPHA试剂检出相同的阳性标本33份, 占67.3%, McAb血球检出而RPHA试剂未检出12份, 占24.5%, RPHA试剂检出而McAb血球未检出4份, 只占8.2%, 49份阳性标本中McAb血球检出91.8%, 而RPHA试剂只检出75.5%。

讨论

通过两种诊断试剂检测结果的比较, 有明显的可比性, 具有统计学显著性意义, 证实了McAb血球的敏感性高, 特异性强。RPHA试剂阳性检出率明显少于McAb血球的阳性检出率; 对于McAb血球未检出而RPHA试剂检出HBsAg的4份标本, 可能为RPHA试剂的非特异性假阳性反应。而McAb血球检出RPHA试剂未检出HBsAg的12份标本, 说明RPHA试剂假阴性反应严重, 证明RPHA试剂特异性不强、敏感度不高; 表明这类试剂质量需急待提高。综上所述, McAb血球对应用于流行病学、三食行业人员体检及献血员筛选等是一种较理想的抗原诊断试剂。