·综述·

基于相对危险度/患病率比的模型及参数估计方法研究进展

郜艳晖 周舒冬 李丽霞 杨翌 陈跃

【关键词】 相对危险度；患病率比；常见结局

Statistical methods on the estimation of relative risk or prevalence ratio

GAO Yan-hui[1], *ZHOU Shu-dong*[1], *LI Li-xia*[1], *YANG Yi*[1], *CHEN Yue*[2]. 1 *Department of Epidemiology and Health Statistics, School of Public Health, Guangdong Pharmaceutical University, Guangdong Key Laboratory of Molecular Epidemiology, Guangzhou* 510310, China; 2 *Department of Epidemiology and Community Medicine, University of Ottawa, Canada*

Corresponding author: *GAO Yan-hui*, Email: gao_yanhui@163.com

This work was supported by a grant from the Science Foundation of Guangdong Province (No. 10151022401000018).

【Key words】 Relative risk; Prevalence ratio; Common outcome

流行病学研究中常用率比(rate ratio)或率差(rate difference)测量暴露对结局的影响。其中后者有重要的临床和公共卫生学意义，但从结局的形成机制及三级预防的角度而言则需估计率比。常用的率比指标有相对危险度(*RR*)、优势比(*OR*)和患病率比(prevalence ratio, *PR*)，这有赖于研究设计类型而选择。如横断面研究中常使用*PR*，病例对照研究中则使用*OR*，而队列研究又可分两种情况。即在封闭的队列研究中，研究个体的风险期固定，此时宜用累积发病率(cumulative incidence rate, *CIR*)；而在开放性或动态队列研究中需用发病密度比(incidence density ratio, *IDR*)，*CIR*和*IDR*常通称为*RR*。

多变量统计分析时，logistic回归主要用于病例对照研究中估计*OR*，对于开放式的队列研究，常使用Cox回归模型来估计*IDR*。实际工作中，由于logistic回归模型已被众多学者所熟识，除被广泛用于病例对照研究外，一些文献也常将其应用于封闭的队列研究或横断面研究资料，并把效应指标的*OR*习惯性地解释为*RR*或*PR*。尽管*PR*的意义与应用仍有争议[1]，本文还是将其与*RR*(主要指封闭队列研究中的*CIR*)对应以方便论述。当研究者所关心的结局为稀有事件时，利用logistic回归模型估计的*OR*值近似于*RR*/*PR*值。但当研究结局的发病率或患病率较高时，*OR*可严重高估暴露因素对结局的影响[2](图1)。如1999年*New England Journal of Medicine*报道内科医生对黑人妇女实施心导管插入术的率比白人或男性低，与白人妇女相比*OR*=0.4(95%*CI*: 0.2～0.7)，事实上应为*PR*=0.87(95%*CI*: 0.80～0.95)。该结果被媒体报道为黑人妇女接受的医疗服务比白人或男性低很多，由此引发关于种族和性别歧视问题的热烈争论[3]。此后其他刊物如*American Journal of Epidemiology*也建议队列研究或横断面研究资料中应避免使用logistic回归估计*OR*报告结果。因此在多变量状态下各种估计*RR*/*PR*的统计方法逐渐受到重视。本文主要对估计*RR*/*PR*的模型及参数估计方法进行综述。

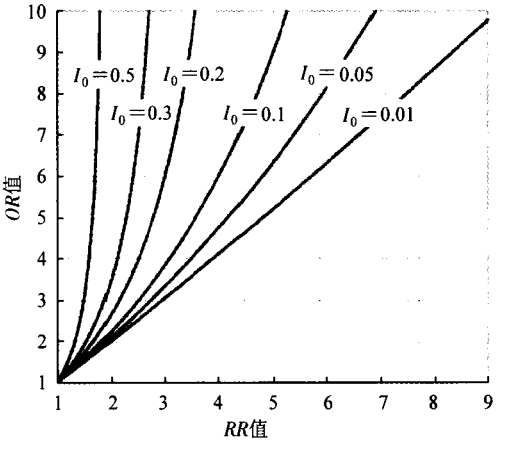

图1 各种发病率情况下*OR*和*RR*的关系[2]

1. 通过logistic回归计算*RR*/*PR*的方法：早期流行病学家估计调整的*RR*/*PR*时常使用分层分析，利用Mantel-Haenzsel方法[4]。该方法可用于协变量数不多，且均为分类变量，但如果数据中包含连续型协变量时则无法进行估计。基于*OR*和*PR*/*RR*的关系，有学者提出将调整*OR*转换为*RR*/*PR*的方法，如Thompson等[5]直接用*OR*估计*PR*，通过在Mantel-Haenzsel模型中对各层的*PR*估计值进行加权平均。而此类方法中较受关注的是Zhang和Kai[6]建议通过logistic回归得到*OR*后再转化为*RR*/*PR*，即

$$RR/PR = \frac{OR}{(1-P_0)+(P_0 \times OR)} \quad (1)$$

式中P_0为非暴露组结局事件的发(患)病率。其置信区间(*CI*)由调整*OR*值的上下置信限代入式(1)得到。但随后研

DOI: 10.3760/cma.j.issn.0254-6450.2013.09.018
基金项目：广东省自然科学基金(10151022401000018)
作者单位：510310 广州，广东药学院公共卫生学院流行病与卫生统计学系 广东省分子流行病学重点实验室(郜艳晖、周舒冬、李丽霞、杨翌)；加拿大渥太华大学社会医学和流行病学系(陈跃)
通信作者：郜艳晖，Email: gao_yanhui@163.com

究显示[7]，当研究结局出现频率并非罕见时，通过公式(1)获得的调整*RR/PR*值存在高估关联强度的偏倚，这是因为公式(1)中未考虑各种协变量模式情况下暴露因素和疾病发(患)病率间更复杂的关系，也未考虑发(患)病率和*OR*值间的协方差，因而导致*CI*估计过窄，且这些偏倚并不随样本量的增加而减少。

相比较而言，直接利用logistic回归模型预测概率的比值估计*RR/PR*，则是考虑到各种协变量模式更为简便。即

$$RR/PR = \frac{P(Y/E,x_2,\cdots,x_p)}{P(Y/\bar{E},x_2,\cdots,x_p)} = \frac{1+e^{-(\beta_0+\beta_1 E+\beta_2 x_2+\cdots+\beta_p x_p)}}{1+e^{-(\beta_0+\beta_1 E+\beta_2 x_2+\cdots+\beta_p x_p)}} \quad (2)$$

式中Y表示结局变量，E表示研究的暴露因素，$x_2,\cdots,x_p$表示各协变量[7]。但该方法最大问题是计算的*RR/PR*值可随模型中协变量取值水平的改变而改变[8]。如假设E为二分类变量(1=暴露，0=非暴露)，X为一个连续型协变量。一种方法是用协变量的均数($\bar{X}$)或指定某一参考值来计算*RR/PR*[9]，即$RR/PR = \frac{1+e^{-(\beta_0+\beta_2\bar{X})}}{1+e^{-(\beta_0+\beta_1+\beta_2\bar{X})}}$，此称为条件(conditional)方法；另一方法是假设所有个体都是暴露组或非暴露组，分别计算所有个体协变量取值时预测率的平均水平，再计算*RR/PR*，即$RR/PR = \frac{\frac{1}{n}\sum_i[1/(1+e^{-(\beta_0+\beta_1+\beta_2 X_i)})]}{\frac{1}{n}\sum_i[1/(1+e^{-(\beta_0+\beta_2 X_i)})]}$，此称为边际(marginal)方法，因此该方法不需要固定协变量的取值。再有一种与边际方法类似，但按协变量取值将人群分层，分别计算假设所有个体均为暴露或非暴露时各层患病率的加权平均，权重可来自某一标准或参考人群，再计算*RR/PR*，即$RR/PR = \frac{\sum_k w_k[1/(1+e^{-(\beta_0+\beta_1+\beta_2 X_i)})]}{\sum_k w_k[1/(1+e^{-(\beta_0+\beta_2 X_i)})]}$，称为分层(stratified)方法。该方法层内个体协变量取值相同，类似于直接标准化法。当权重取各层相对样本量时，分层方法与边际方法结果相同。各种方法得到*RR/PR*的*CI*通常可用delta方法或bootstrap方法获得[10]。

除了直接利用logistic回归结果计算*RR/PR*，还有学者提出修改数据的方法，对修改后的数据进行logistic回归计算*OR*值，相当于对原始资料进行*RR/PR*的估计[11]。该方法的思路是创建一个扩展数据，将原始数据集中结局事件发生($Y=1$)的个体复制，但将其结局修改为未发生($Y=0$)。如果原始数据N个个体中有X个事件发生，发生的概率为$p=X/N$，则在扩展数据中事件发生的概率则为$p^*=X/(X+N)=p/(1+p)$。这样可求出$p=p^*/(1-p^*)$，即原始数据集中事件发生的概率等于扩展数据集中事件发生的优势。因为扩展数据集中某些个体重复了两次，一个是结局发生，另一个是结局不发生，因此该方法的运用导致重复的个体水平上有负相关。此问题可通过广义估计方程(generalized estimate equation，GEE)或多水平模型等方法解决。

2. 直接估计*RR/PR*的回归模型：上述由logistic回归模型估计*RR/PR*的方法取决于协变量的值或其基线风险。如果假设*RR/PR*固定，即不随协变量值的变化而改变，可在广义线性回归模型框架下直接估计*PR/RR*，其中研究较多的是以下4种方法。

(1)修正的Cox比例风险模型：Cox比例风险模型最初主要用于开放性队列研究中的生存时间资料，此时个体具有不同的风险期(即随访时间不同)。在自变量X条件下，t时刻的风险函数表示为

$$h(t/X) = h_0(t)e^{(\beta_1 x_1+\cdots+\beta_p x_p)} \quad (3)$$

式中$h_0(t)$为基准风险，参数e^β称为瞬时风险比(instantaneous hazard ratio，*HR*)，如假设结局事件在同一时刻发生，*HR*则为*IDR*。对封闭的队列研究，所有个体具有固定的风险期(即有相同的随访时间)，可采用Breslow[12]处理“结”的方法将式(3)中的t设为常数，此时*HR*即为*CIR*。对横断面研究，Lee等[13,14]建议可以假设所有研究对象具有固定的风险期(即相同的随访时间)，在此条件下利用Breslow方法修正Cox比例风险模型，使*HR*与*PR*的解释类同，并得到一致的点估计。然而由于该模型假定结局为Poisson分布而非二项分布，因而导致系数的方差被高估，得到较宽的*CI*[15]，从而得出不合理的推断。需要说明的是，Cox回归中处理“结”的方式还有更精确的近似法，如Efron的精确近似或某些软件可提供确切偏似然或边际似然，这些方法尽管适用于生存时间资料，但用于*RR/PR*时将产生有偏估计。修正的Cox比例风险模型在SAS中可以用PHREG过程实现，将所有观测生存时间变量设为相等，用method=breslow指定，估计参数及其*CI*。

(2)稳健Poisson回归模型：Poisson分布常用于描述某些稀有事件的发生数，如样本量很大，率很低时的二项分布也常用Poisson分布近似。当结局事件频率较高时，为直接估计*RR/PR*，建议使用稳健Poisson模型[3]。设y_i和$X_i=(x_{i1},x_{i2},\cdots,x_{ip})^T$分别是第$i(i=1,2,\cdots,n)$个观测的二分类结局变量和$P\times1$维解释变量向量，其关系可通过Poisson回归模型表示

$$\log(\hat{p}_i) = \beta_0 + \sum_{p=1}^{p}\beta_p x_{ip} \quad (4)$$

式中$\hat{p}_i=\Pr(y_i=1/X_i)$，链接函数采用了log链接，并假设误差分布为Poisson分布。回归系数β_p表示当控制其他自变量后，第p个自变量x_p每变化一个单位时$\log(\hat{p})$的相应变化。因此，与x_p相对应的相对危险性为$RR/PR=e^{\beta_p}$。同修正的Cox回归一样，由于Poisson分布的方差无边界(方差等于均数)，当应用到二项分布资料时，易出现过度离散(over dispersion)问题。也就是Poisson模型虽可获得无偏的点估计，但通常高估参数的方差，导致较宽的*CI*。与修正的Cox比例风险模型相比，Poisson回归模型能得到截距值，估计基线状态下结局事件发生的概率，但在固定风险期的条件下两模型可获得相同的点估计和方差。处理高估方差的一个简单方法就是通过Pearson χ^2或deviance调整尺度参数，用尺度参数乘以模型估计的方差。另一方法是Barros和Hirakata[15]及Zou[16]建议采用稳健Poisson回归模型来估计*PR*，如引入Huber的稳

健"三明治"方差

$$\mathrm{Var}(\hat{\beta})=A^{-1}BA^{-1} \quad (5)$$

其中，

$$A=\sum_{i=1}^{n}X_iX_i^T\hat{p}_i,\ B=\sum_{i=1}^{n}X_i(y_i-\hat{p}_i)^2X_i^T$$

模型(4)中参数$\beta=(\beta_0,\beta_1,\cdots,\beta_p)^T$及"三明治"方差可用准似然(quasi-likelihood)估计，在SAS中可用PROC GENMOD实现，通过在repeated语句中用"subject="指定个体编号变量。

Wolkewitz等[17]比较了Poisson模型中用Pearson χ^2或deviance调整尺度参数，进而调整方差的方法以及用稳健"三明治"进行方差调整，结果显示3种调整的模型均有助于方差估计，但稳健Poisson回归模型表现最好。尽管上述Cox回归中也可以采用稳健"三明治"方差，其效果和稳健Poisson回归相同。由于该模型是Poisson近似，其局限性是采用Poisson模型估计的结局概率有可能大于1[18]，用于概率预测有时并不合理。

(3) log-binomial模型和COPY算法：除Cox回归和Poisson回归模型外，也可将结局分布指定为二项分布，利用log链接建立回归模型直接估计*RR*/*PR*，即log-binomial模型(log binomial model, LBM)[19]。该模型是由Wacholder[20]建议利用GLIM基于广义线性模型的软件来实现[21]。log-binomial模型表达式同式(4)，即

$$\log P(Y=1\mid x_1,\cdots,x_p)=\beta_0+\beta_1x_1+\cdots+\beta_px_p \quad (6)$$

模型中仍然使用log链接，但误差分布指定为二项分布。回归系数β_p表示当控制其他自变量后，第p个自变量x_p每变化一个单位时$\log(\hat{p})$的相应变化，因此与x_p相对应的*RR*/*PR*=e^{β_p}；回归系数的估计可采用最大似然估计(maximum likelihood estimation, MLE)的方法，其似然函数及对数似然函数分别为

$$L(\beta;y)=\prod_{i=1}^{n}(e^{x_i\beta})^{y_i}(1-e^{x_i\beta})^{(1-y_i)} \quad (7)$$

和

$$\log L(\beta;y)=\sum_{i=1}^{n}[y_ix_i\beta+(1-y_i)\log(1-e^{x_i\beta})] \quad (8)$$

由于结局变量$Y=1$的概率介于0和1之间，所以log-binomial模型利用MLE方法估计参数β时需加一个限制条件，即

$$x_i\beta=\beta_0+\beta_1x_{i1}+\cdots+\beta_px_{ip}\leqslant 0 \quad (9)$$

最大似然估计的解应该在此条件限制的参数中。大部分统计软件在广义线性模型的最大似然估计中均使用迭代加权最小二乘法或其他Newton法，以寻找斜率等于0的点，使对数似然最大化，得到参数最大似然估计值。因此在参数估计过程中如果初始值选择不当，或数据集中所有自变量的线性组合未满足该限制条件，最大似然估计的解可能落在有限制参数的边缘，以致不能获得似然函数导数为零的极大值点，从而导致模型无法收敛[22]。该情况在模型中存在连续型协变量时尤其常见。

由于初始值选择不当可能导致不收敛，此时试将截距指定为负值，而设置$\beta_0=-4$可能效果较好[22]。如是后者，Deddens和Petersen[23]建议可先对原始数据集调整扩充后再拟合LBM，即可得到参数最大似然估计值，这种对原始数据集调整扩充的方法称为COPY算法，可有效解决模型不收敛的问题。其步骤是先将原始数据集中复制，扩大(c−1)倍，再将原始数据集中的因变量值0-1互换生成1个新的数据集，之后把扩大(c−1)倍的数据集与互换因变量值的数据集合并，最终得到COPY数据集，用于拟合log-binomial模型。参数的标准误由COPY数据集得到的标准误乘以$\sqrt{c}$得到。利用COPY数据集拟合log-binomial模型，原始数据的信息占(c−1)/c。即c越大，最大似然估计的偏倚越小，但参数估计耗时也越长，特别是当原始数据为大样本时，使用COPY算法将面对非常大的计算量。故建议c≥100。Lumley等[24]观察到如将原始数据集和因变量0-1互换后的数据集合并，并定义原始数据集中观测的权重为(c−1)/c，而因变量0-1互换后其权重为1/c，得到加权log-binomial模型的最大似然函数，即

$$L_\omega(\beta;y)=\prod_{i=1}^{n}(e^{x_i\beta})^{\omega y_i+(1-\omega)(1-y_i)}(1-e^{x_i\beta})^{\omega(1-y_i)+(1-\omega)y_i} \quad (10)$$

利用加权log-binomial模型的最大似然估计可得到和COPY算法相同的结果。由于被分析的数据集大小只是原始数据集的2倍，因此大大减少了计算工作量。近来Savu等[25]对COPY算法的收敛和惟一的近似MLE等性质给出了严格的理论证明，即在模型X满秩的假设下，运用COPY方法，其修正似然函数在限制的参数空间中具有惟一全局极大值点，且该值点落在参数空间内部而不在边界上。只要初始值和近似的惟一MLE解相邻，可以确保算法的收敛。

多项研究[23,26,27]通过模拟或实例将稳健Poisson回归和log-binomial模型(不收敛时用COPY算法)进行比较。如Petersen和Deddens[27]显示在高频率结局和中等样本量情况下，稳健Poission回归对*PR*估计的偏倚小于log-binomial模型(不收敛时用COPY算法)，但其他情况下，log-binomial模型均比稳健Poisson回归模型具有较小的标准误和更高功效，且所估计的患病概率并不大于1。log-binomial模型的另一个优点是可以使用似然比检验，相较*Wald*检验可能更好。而稳健Poisson回归其似然函数与普通Poisson回归似然函数相同，因此假设检验只能用*Wald*检验完成，而无法使用似然比检验。但运用稳健Poisson模型几乎很少有收敛问题，又可直接对原始数据分析，省去了数据操作的过程而更容易使用，因此有学者建议[26,28]，log-binomial不收敛时可首先选择稳健Poisson回归。此外，Lumley建议模型不收敛时还可使用非线性最小二乘法来估计参数，因为MLE对存在异常值或错误指定模型时会比较敏感。Yu和Wang[29]提出采用SAS统计软件中的PROC NLP(nonliear programming)过程估计*PR*，该过程可提供多种最优化技巧，如线性搜索或岭搜索的Newton-Raphson方法、Quasi-Newton方法等来对参数β的连续型非线性函数求最大或最小值，以及计算参数的上下边界，求解过程中可以指定相等或不等的约束条件。如欲求*RR*/*PR*，可在$X\beta<0$的限制下对log-binomial模型的对数似然函数求极大值，模拟研究显示如log-binomial模型收敛，和该方法估计*RR*/*PR*结果几乎相同，但该法始终收敛且预测概

率在[0,1]之间。目前SAS/STAT中的PROC GENMOD程序提供了weight语句，可实现加权log-binomial模型，因此可利用weight语句很方便地实现COPY方法，而不用事先生成c倍于原始数据集的COPY数据集。

(4) IPTW的log-binomial模型：由于Savu等[25]从理论上已证明加权log-binomial模型最大似然估计，为此又进一步提出一种用倾向性得分(propensity score)匹配的加权log-binomial回归方法，称为IPTW(inverse-probability of treatment-weighted)的log-binomial回归。该法出发点类似前述分层方法，率的标准化法可代替回归以达到调整混杂因素计算调整或标化*RR/PR*的目的。设x_1为暴露因素，其余解释变量$x_2,\cdots,x_p$为混杂因素，调整的*RR/PR*可表示为

$$RR/PR=\frac{\sum_{x_2,\cdots,x_p}P(Y=1/x_1=1,x_2,\cdots,x_p)\times P(x_2,\cdots,x_p)}{\sum_{x_2,\cdots,x_p}P(Y=1/x_1=0,x_2,\cdots,x_p)\times P(x_2,\cdots,x_p)} \quad (11)$$

式中患病率如按logistic回归模型估计，则为前述分层方法；如满足log-binomial模型假设，则调整的$RR/PR=e^{\beta_1}$。可以证明，式(11)中调整的*RR/PR*等于一个修正的联合结局-暴露-混杂变量分布P_w下的粗RR_w/PR_w，其中P_w和原分布P有关。即

$$P_w(Y,x_1,x_2,\cdots,x_p)=w_{x_1,x_2,\cdots,x_p}\times P(Y,x_1,x_2,\cdots,x_p) \quad (12)$$

式中权重$w_{x_1,x_2,\cdots,x_p}=P(x_1)/P(x_1/x_2,\cdots,x_p)$称为IPTW，与混杂变量在各比较组中的分布有关。因此调整*RR/PR*的估计可转化为在联合分布P_w下估计粗RR_w/PR_w，并采用一般的log-binomial回归估计。即

$$\log P_w(Y=1\mid x_1)=\beta'_0+\beta'_1x_1 \quad (13)$$

该模型的估计可通过对原始资料进行加权log-binomial模型的拟合得到，其中权重为式(12)中的IPTW。实际应用中，需先根据混杂变量$(x_2,\cdots,x_p)$的分布分层，但当样本量不充足时，可能某些层中出现观测数极小或为0的情况，导致层内患病率估计的不稳定。替代的办法是①先计算倾向性得分$P(x_1=1/x_2,\cdots,x_p)$，涵义为控制混杂因素后个体为暴露组的概率；②按倾向性得分的百分位数将观测(observation)分为少数几层(如分5层，每层包含20%的观测)，定义第i层的权重为$w_i=P(x_1)/P(x_1/i)$，实质上是用倾向性得分百分位数分层取代混杂变量$(x_2,\cdots,x_p)$的联合分布；③拟合加权log-binomial模型估计调整的*RR/PR*，称为IPTW估计量，其参数方差则通过稳健"三明治"方法获得。该加权log-binomial模型参数估计也可用SAS/STAT中的PROC GENMOD程序轻易实现，同时利用weight语句定义权重变量。模拟研究显示和COPY法相比，IPTW估计量对错误指定二分类结局变量和暴露因素关系比较稳健。该法也可以一种广义的形式推广到连续性暴露或多变量暴露的情况。

上述主要方法的特点及其优劣见表1。

3. 其他模型及扩展：上述IPTW估计量利用倾向性得分分层代替原始混杂变量的联合分布，也可将暴露组和非暴露组个体按倾向性得分进行匹配，利用匹配的部分观测整理成匹配四格表后直接得出调整的*RR/PR*及其*CI*，但该方法未利用所有的观测信息而损失效能。此外，在各种回归模型中，除指定log链接、误差分布为二项分布或Poisson分布外，也有其他链接函数或误差分布形式，如complementary log-log模型[链接函数是log(-log(1-P)，误差分布为二项分布]；或log-normal模型[24](链接函数为log链接，误差分布定义为正态分布)。

近年来，*RR/PR*模型已推广应用至其他特殊类型的数据，如流行病学调查或临床试验中的非独立数据。GEE和多水平模型是处理非独立数据的有力工具，Santos等[8]将logistic模型基础上的条件、边际和分层方法扩展到随机效应的logistic模型，以估计非独立横断面研究中的*PR*(其*CI*采用delta法和bootstrap法)。Zou和Donner[30]扩展了稳健Poisson模型的GEE。Yelland等[31]做了大型模拟研究比较log-binomial模型、稳健Poisson模型、修改数据后的logistic模型和log-normal模型基础上的GEE方法，显示非独立数据时log-binomial模型GEE也存在收敛问题，基于稳健Poisson和log-normal的GEE有较小的Ⅰ型错误和较高的覆盖率，推

表1　各种估计*RR/PR*统计方法的特点及其优劣

统计方法	特点和优劣
分层的Mantel-Haenzsel方法[4]	不能调整连续型协变量，可用于协变量个数不多，且均为分类变量
利用logistic回归计算*OR*，再用公式转化为*RR/PR*[6]	参数估计有偏倚，且偏倚不随样本量增加而减少
利用logistic回归估计预测概率，再估计*RR/PR*[7]	包括条件法、边际法和分层法，估计*RR/PR*时需设定协变量的取值，某种程度上带有任意性；*CI*需用delta方法或bootstrap法得到。但推广到多水平模型时较方便
扩展数据的方法[11]	扩展后的数据存在相关，需用GEE或多水平模型估计参数
修正的Cox比例风险模型[13, 14]	可得到无偏点估计，但*CI*值较宽
稳健Poisson回归模型[16]	可得到无偏点估计，引入"三明治"方差，可改善*CI*值较宽的问题，但估计的结局概率可能大于1，不用于预测，但几乎无收敛问题
log-binomial模型[20]	正确指定误差分布和链接函数，可得到无偏估计，但模型有限制条件，如初始值选择不当，不易收敛，模型中存在连续型协变量时收敛问题更常见，但预测概率在[0,1]之间
log-binomial模型COPY方法[22]	正确指定误差分布和链接函数，可得到无偏估计，相当于一种加权log-binomial模型，可改善收敛问题。理论上已证明模型有惟一且收敛的最大似然估计
log-binomial模型，用PROC NLP估计参数[29]	利用SAS的PROC NLP过程提供的多种最优化技巧估计参数，几乎无收敛问题，且预测概率在[0,1]之间
IPTW的log-binomial模型[25]	用倾向性得分匹配的加权log-binomial方法，类似标准化法，实际应用时对需调整的协变量分层求权重，当错误指定误差分布和链接函数时该法可能较稳健

荐非独立数据时选择稳健Poisson模型的GEE方法[32]。相较而言，基于前述*RR*/*PR*模型的多水平模型虽已有应用[33]，但其理论和参数估计方法的具体表现仍需进一步研究。

参 考 文 献

[1] Reichenheim ME, Coutinho ES. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. BMC Med Res Methodol, 2010, 10: 66.

[2] Schmidt CO, Kohlmann T. When to use the odds ratio or the relative risk? Int J Public Health, 2008, 53(3): 165-167.

[3] Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. N Engl J Med, 1999, 341(4): 279-283, 286-287.

[4] Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics, 1985, 41(1): 55-68.

[5] Thompson ML, Myers JE, Kriebel D. Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? Occup Environ Med, 1998, 55(4): 272-277.

[6] Zhang J, Kai FY. What's the relative risk? JAMA, 1998, 280 (19): 1690-1691.

[7] Mcnutt LA, Wu C, Xue X, et al. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol, 2003, 157(10): 940-943.

[8] Santos CA, Fiaccone RL, Oliveira NF, et al. Estimating adjusted prevalence ratio in clustered cross-sectional epidemiological data. BMC Med Res Methodol, 2008, 8: 80.

[9] Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. J Clin Epidemiol, 2007, 60(9): 874-882.

[10] Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. J Clin Epidemiol, 2010, 63(1): 2-6.

[11] Schouten EG, Dekker JM, Kok FJ, et al. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. Stat Med, 1993, 12(18): 1733-1745.

[12] Breslow N. Covariance analysis of censored survival data. Biometrics, 1974, 30(1): 89-99.

[13] Lee J, Chia KS. Use of the prevalence ratio v the prevalence odds ratio as a measure of risk in cross sectional studies. Occup Environ Med, 1994, 51(12): 841.

[14] Lee J. Odds ratio or relative risk for cross-sectional data? Int J Epidemiol, 1994, 23(1): 201-203.

[15] Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med Res Methodol, 2003, 3: 21.

[16] Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol, 2004, 159(7): 702-706.

[17] Wolkewitz M, Bruckner T, Schumacher M. Accurate variance estimation for prevalence ratios. Methods Inf Med, 2007, 46(5): 567-571.

[18] Tian L, Liu K. Re: "Easy SAS calculations for risk or prevalence ratios and differences". Am J Epidemiol, 2006, 163(12): 1157-1158, author reply 1159-1161.

[19] Skov T, Deddens J, Petersen MR, et al. Prevalence proportion ratios: estimation and hypothesis testing. Int J Epidemiol, 1998, 27(1): 91-95.

[20] Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. Am J Epidemiol, 1986, 123(1): 174-184.

[21] Deddens JA, Petersen MR. Re: "Estimating the relative risk in cohort studies and clinical trials of common outcomes". Am J Epidemiol, 2004, 159(2): 213-214, author reply 214-215.

[22] Deddens JA, Petersen MR, Lei X. Estimation of prevalence ratios when PROC GENMOD does not converge. Proceedings of the 28th Annual SAS Users Group International Conference, 2003.

[23] Deddens JA, Petersen MR. Approaches for estimating prevalence ratios. Occup Environ Med, 2008, 65(7): 501-506.

[24] Lumley T, Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. UW Biostatistics Working Paper Series. 2006: 293. http://biostats.bepress.com/uwbiostat/paper293.

[25] Savu A, Liu Q, Yasui Y. Estimation of relative risk and prevalence ratio. Stat Med, 2010, 29(22): 2269-2281.

[26] Lee J, Tan CS, Chia KS. A practical guide for multivariate analysis of dichotomous outcomes. Ann Acad Med Singapore, 2009, 38 (8): 714-719.

[27] Petersen MR, Deddens JA. A comparison of two methods for estimating prevalence ratios. BMC Med Res Methodol, 2008, 8: 9.

[28] Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol, 2005, 162 (3): 199-200.

[29] Yu B, Wang Z. Estimating relative risks for common outcome using PROC NLP. Comput Methods Programs Biomed, 2008, 90 (2): 179-186.

[30] Zou GY, Donner A. Extension of the modified Poisson regression model to prospective studies with correlated binary data. Stat Methods Med Res, 2011. [Epub ahead of print]

[31] Yelland LN, Salter AB, Ryan P. Relative risk estimation in cluster randomized trials: a comparison of generalized estimating equation methods. Int J Biostat, 2011, 7(1): 1-26.

[32] Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. Am J Epidemiol, 2011, 174(8): 984-992.

[33] Sembajwe G, Cifuentes M, Tak SW, et al. National income, self-reported wheezing and asthma diagnosis from the World Health Survey. Eur Respir J, 2010, 35(2): 279-286.

(收稿日期：2013-03-18)

(本文编辑：张林东)