

# 日本血吸虫卵计量变异的模型建立与参数估计

余金明 袁鸿昌 陈启明 杨求吉 Sake De Vlas Bruno Gryseels

**【摘要】** 目的 建立虫卵计量变异的随机模型,并估计描述虫卵计量变异特征的参数。方法 随机模型把人群中虫卵计量总变异分成两个来源:①虫卵计量在个体间的变异;②个体内虫卵计量的变异。用具有重复虫卵计数的实际资料对模型的参数进行估计。结果 参数  $M$ 、 $r$  值在各年龄组中不同, $k$  值在各年龄组中近于相同; $r$  和  $M$  值的最佳年龄分割点分别为 8 和 12 岁;模型 E 中假定  $k$  值在各年龄组中相同, $r$  值在 2~7 岁和 8 岁以上年龄组间不同, $M$  值在 2~11 岁和 12 岁以上年龄组间不同, $AIC$  值最小。结论 年龄可能为影响  $M$ 、 $r$  值的重要因素,模型 E 为最佳模型。

**【关键词】** 寄生虫虫卵计数 血吸虫,日本 模型,统计学

**Establishment of A Stochastic Model for Variation in Fecal *Schistosoma japonicum* Egg Count and Estimation of Its Parameters** YU Jinnming\*, YUAN Hongchang, CHEN Qiming, et al. \* Department of Preventive Medicine, School of Medicine, Shanghai Tiedao University, Shanghai 200070

**【Abstract】 Objective** To establish a stochastic model suitable for interpreting the variation in fecal *Schistosoma japonicum* egg count, and to estimate its parameters and to describe it. **Methods** Total variation in fecal egg count with a stochastic model can be divided into two kinds of sources, ① inter-individual variation, and ② intra-individual variation. Parameters in the model were estimated with the data collected in actual fecal egg count. **Results** Parameters  $M$  and  $r$  differed and  $k$  kept nearly the same in various age groups. The best cut-off point of  $r$  and  $M$  for age-splitting was at eight and 12 years, respectively. Parameter  $k$  was assumed the same in different age groups in the model E, and  $r$  differed in the groups of 2~7 years and 8~ years of age, and  $M$  differed in the groups of 2~11 years and 12~ years of age, with a minimum value of Akaike information criterion ( $AIC$ ). **Conclusion** Age may be an important factor contributing to the estimations of parameters  $M$  and  $r$ , and the model E was the best one.

**【Key words】** Parasite *Schistosoma japonicum* Models, statistical

过去有关血吸虫病的数学模型研究主要局限于理论研究,较少与实际相结合,现在已经发生了转变<sup>[1-3]</sup>。改良 Kato 法诊断血吸虫感染和吡喹酮治疗血吸虫病是近些年来两个重大突破,它使病原诊断从过去的定性向定量迈出了一大步。但该方法在检测低度感染人群时,会漏诊相当部分的轻度感染者。1992 年以来,中国大部分流行区已成为中低度流行区<sup>[4]</sup>。正确评估改良 Kato 法检测的结果,对血吸虫病的防治和科研都有着极为重要的意义。作者通过建立虫卵计量变异的数学模型,使得人们能正确认识单次改良 Kato 法检测血吸虫感染敏感性及

虫卵计量变异的特征,使单次改良 Kato 法检测低估感染率的缺陷得以克服。

## 材料与方 法

### 一、流行病学资料

濒临鄱阳湖滨的日本血吸虫病流行区渚溪村为本研究的现场。对该村全部常住人口连续收集 7 次新鲜粪便标本,每份制 41.5 mg 涂片 2 张。

### 二、随机模型的基本概念

1. 虫卵计量变异的总来源:像曼氏血吸虫的人群研究一样<sup>[2]</sup>,将因素(2~4)合并在一起统称为个体内变异。因此,该模型仅将影响人群虫卵计量的来源划分为两大类:①个体之间由于成虫负荷不同造成的虫卵计量变异;②个体内变异:对于某一具有特定成虫负荷的个体逐次重复改良 Kato 法检

作者单位:200070 上海铁道大学医学院预防医学教研室(余金明);上海医科大学(袁鸿昌、杨求吉);上海师范大学奉贤校区(陈启明);Erasmus University Rotterdam(Sake De Vlas);Prince Leopold Institute of Tropical Medicine(Bruno Gryseels)

测时的虫卵计量变异。

2. 日本血吸虫的负二项分布规律<sup>[5]</sup>、参数  $h$  取值 0.17<sup>[6]</sup>、配对因素等有关概念和原理详见参考文献<sup>[7,8]</sup>。

3. 模型的数学方程:本模型按照确定每份样本中出现某一特定虫卵计数机率的基本思路设计。由 4 个参数  $M, k, h, r$  值所确定,人群中虫卵计数的分布是人群中成虫对子数的分布和某一特定数量的成虫对子数所决定的虫卵计数分布相结合的结果。引进配对的因素后,将获得成虫对子数在人群中的分布。用下式表示样本中虫卵数( $y$ )的概率:

$$P(Y = y; M, k, h, r) = \sum_{x=0}^{\infty} \{ P(Y = y; hx, r) \sum_{n=2x}^{\infty} [ P(X = X | N = n) P(N = n; M, k) ] \}$$

若每个个体重复数次虫卵计数,则该模型可以向外延伸。例如在同一时间内,重复 3 次取样,得到的虫卵计数分别为  $y_1, y_2, y_3$ , 则

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3; M, k, h, r) = \sum_{x=0}^{\infty} \{ P(Y_1 = y_1; hx, r) P(Y_2 = y_2; hx, r) P(Y_3 = y_3; hx, r) \}$$

$$P(Y_3 = y_3; hx, r) = \sum_{n=2x}^{\infty} [ P(X = x | N = n) P(N = n; M, k) ]$$

由于重复改良 Kato 法调查在非流行季节进行,且全部的现场工作集中在较短的时间内完成,故可认为宿主体内拥有的成虫对子数在此期间未发生变化。因此,以上方程同样可以用于连续数次的重复调查。以上方程中各符号的意义详见文献<sup>[2]</sup>。

4. 拟合方法和模型检验:用最大似然法对模型进行参数估计,用单纯递减法(downhill simplex method)寻找参数值<sup>[9]</sup>。为对不同模型进行比较,引进 Akaike 信息标准(akaike information criterion,简称 AIC)的统计方法<sup>[10]</sup>。AIC 值为  $-2 * \lg(L) + 2 * (\text{自由参数的个数})$ ,使 AIC 的值为最小的模型为最恰当模型。AIC 的差值( $\Delta AIC$ )大于 1.84 表示差异有显著性。

### 模型的建立和分类

表 1 是寻找最佳模型的基本思路。模型 A 对 4 个年龄组的全部非固定参数进行估计,结果  $k$  值在各年龄组中变化不明显,说明各年龄组内个体间的虫卵计量变异的程度差别不大。然而, $r$  和  $M$  值在低年龄组(2~9 岁)均明显地有别于其他年龄组,低龄儿童组  $r$  值低于其他年龄组,说明在低龄儿童中个体内虫卵计量变异的程度高于其他年龄组。正如我们所期望的那样, $M$  的估计值近似等于虫卵计数的均数除以  $h/2$ ,其微小的差别是由于配对因素的影响。

为探索年龄是否为影响参数  $M$  和  $r$  的重要因素,将全部的年龄组合并(即认为各年龄组的  $M, r, k$  值相同),结果  $\Delta AIC = -21.08$ ,已远远超出显著性水准 1.84。故认为模型 B 与资料的配合程度比 A 差,说明年龄确实影响着参数  $M, r$  的值。减少参数后,已不能较好地描述资料的性质。现在,我们仅

表 1 随机模型拟合结果( $h = 0.17$ )

模型	年龄组(岁)	人数	虫卵均数	r 值	M 值	k 值	-2 * lg(L)	AIC
A	2~	69	8.08	0.30	93.94	0.21	1 524.03	10 718.21
	10~	81	16.46	0.59	206.01	0.26	2 758.12	
	20~	101	16.54	0.66	199.46	0.24	3 309.29	
	≥40	105	13.90	0.66	167.38	0.21	3 102.78	
B	所有年龄	356	14.11	0.56	171.26	0.22	10 733.29	10 739.29
C	2~	69	8.08	0.30	93.94	0.21	1 524.03	10 708.67
	≥10	287	15.56	0.64	189.66	0.23	9 172.64	
D	2~	69	8.08	0.30	91.18	0.22	10 696.96	10 706.96
	≥10	287	15.56	0.64	190.30			
E	2~	40	6.21	0.22	91.16	0.23	10 692.79	10 702.79
	8~	58	8.44					
	≥12	258	16.60					
F	2~	40	6.21	0.22	71.81	0.23	785.21	10 704.10
	8~	58	8.44	0.47	102.12	0.26	1 559.71	
	≥12	258	16.60	0.65	203.57	0.23	8 341.18	



仅把人群分成两个年龄组, 2~9 和 10 岁以上, 并假设  $M$ 、 $k$ 、 $r$  值在各年龄组中不同(模型 C)。模型的拟合优度明显地改善, 模型 C 已优于模型 A 和 B。同模型 A 相似,  $M$  和  $r$  的值在低年龄组明显低于高年龄组,  $k$  值在两个年龄组中变化不明显。因此, 考虑将两个年龄组的参数  $k$  合并(模型 D), 使被估计的参数减少 1 个, 模型的拟合优度进一步得以改进。

在模型的拟合过程中, 我们发现参数  $M$  和  $r$  的年龄分割点的变化明显地影响着模型的拟合优度。因此, 为了寻找到最佳的年龄分割点, 分别假设  $M$  的年龄分割点为 7~18 岁,  $r$  的分割点为 5~15 岁, 再按  $M$  和  $r$  分割点的各种可能的组合重新拟合模型, 获得一系列的  $AIC$  值。 $r$  的年龄分割点为 8 岁,  $M$  的年龄分割点为 12 岁的  $AIC$  值最小。说明  $r$  和  $M$  的最佳年龄分割点为 8 岁和 12 岁。重新对以上模型进行拟合(模型 E), 即两个  $r$  值(分别为 2~7 岁和  $\geq 8$  岁组), 2 个  $M$  值(2~11 岁和  $\geq 12$  岁组), 1 个  $k$  值。与模型 D 比较  $\Delta AIC = 4.17$ , 拟合优度进一步提高。模型 E 被认为是最佳模型, 其参数  $r$  在 2~7 岁组明显低于其他年龄组,  $M$  在 2~11 岁年龄组亦低于其他组。

由于  $M$  和  $r$  的年龄分割点不统一, 事实上已将人群分成了 3 个年龄组。现假设  $M$ 、 $r$  和  $k$  在 3 个年龄组内均为自由参数(模型 F), 结果  $AIC$  值较模型 E 升高,  $\Delta AIC = -1.31$ 。虽没有显著性, 但自由参数的个数已从原来的 5 个增加到 7 个, 故仍认为模型 E 为最佳模型。

## 讨 论

应用随机模型分析日本血吸虫卵计量变异, 使我们更明确地认识到虫卵计量在个体间和个体内的变化规律, 及虫负荷在人群中的分布。虫负荷在人群中的分布比虫卵在人群中的分布更能反映人群感染程度及感染情况的变化, 为决策部门提供更准确的感染资料。

必须指出反映个体内虫卵计量变异的参数  $r$  值将受到粪检质量以及在某一特定时期内一切影响虫卵计数因素的影响。由于本研究所做的 7 次重复检查均在非流行季节进行, 同时全部的现场收集标本工作都集中在较短的时间内完成, 可以排除因体内成虫对子数的变化而产生对虫卵计数的影响。另

外, 改良 Kato 法粪检均严格按照有关的质量控制要求进行。所以, 在讨论影响  $r$  值因素时可以将这些排除在外。随机模型结果提示年龄可能是影响  $r$  值的重要因素。另外在曼氏血吸虫中, 也出现了低年龄组的  $r$  值稍小, 但没有日本血吸虫明显。说明无论是曼氏血吸虫还是日本血吸虫年龄确实不同程度影响着虫卵计数在个体内的变异。其内在机制是否与宿主易感性、获得性免疫等有关, 有待研究。

曼氏血吸虫的  $r$  值是基于每份样本的粪量 50 mg, 且重复调查次数为 5 次, 调查在 3 个月内完成。是否粪样量的多少、调查间隔、重复调查的次数亦为影响因素之一, 有待研究。最有可能的解释是决定于不同种类的血吸虫所固有的属性。中国大部分已成为日本血吸虫病轻度流行区, 在应用该模型时配对的因素将不可忽略。

模型中, 假设虫卵计量在个体间变异仅仅是由于虫负荷在人群中的分布不同, 似乎有点过于简单。因为除了虫负荷以外, 还有可能有些其他的系统误差。例如, 个体每天排粪便的总量及粪便的浓缩程度。这类问题似乎可以通过变换年龄组的  $h$  值来加以解决, 具体情况有待进一步研究。

## 参 考 文 献

- 1 Anderson RM, May RM. Infectious diseases of humans dynamics and control. Oxford: Oxford University Press, 1991. 433-656.
- 2 De Vlas SJ, Gryseels B, Van Oortmarssen GJ, et al. A model for variations in single and repeated egg counts in Schistosoma mansoni infections. Parasitology, 1992, 104: 451-460.
- 3 陈启明. 流行病学数学模型. 中国公共卫生, 1992, 8: 310-314.
- 4 郑庆斯. 血吸虫病在化疗中的社会医学问题. 全国化疗措施控制血吸虫病研讨会论文汇编. 北京: 中华人民共和国卫生部血吸虫病专家咨询委员会. 世行贷款中国血防项目疾病监测控制办公室, 1994. 16-21.
- 5 余金明, 袁鸿昌, 陈启明, 等. 负二项分布在日本血吸虫卵方面的应用. 中华预防医学杂志, 1996, 30: 360-363.
- 6 余金明, 赵英林, 张志强. 日本血吸虫在人体内成虫数与粪中虫卵计数的关系. 中国农村卫生事业管理杂志, 1997, 17(4): 53-55.
- 7 Anderson RM, May RM. Helminth infections of humans: mathematical models, population dynamics, and control. In: Baker JR, Muller R, eds. Advances in parasitology. London: Academic Press, 1985. 24: 1-101.
- 8 May RM. Togetherness among schistosomes: its effects in the dynamics of the infection. Mathematical Biosciences, 1977, 35: 301-343.
- 9 Press WH, Flannery BP, Teukolsky SA, et al. Numerical recipes, the art of scientific computing. Cambridge: Cambridge University Press, 1990. 1-20.
- 10 Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Tokyo: KTK Scientific Publishers, 1986. 56-88.

(收稿: 1998-06-29 修回: 1998-10-22)

(本文编辑: 邵隽一)