

# 遗传流行病学

## genetic epidemiology

# 背景

- 遗传流行病学是一门近20多年发展起来的新兴边缘学科。它借助于群体遗传学的基本理论，分子遗传学的实验手段，流行病学的群体资料收集、处理和病因研究以及科学研究设计方法学、生物统计学与电子计算机的数理模型和数值处理方法来研究和阐明与遗传有关疾病，特别是慢性病中的遗传因素、环境因素和这两类因素的相互作用。



# 背景

- 近10年来，随着分子遗传学技术、电子计算机软件和生物统计方法的不断发展以及这些学科的相互渗透，遗传流行病学的研究内容、对象、方法和目的不断地得到充实和扩展。遗传流行病学已不是仅仅从群体水平宏观地对遗传性疾病进行研究，而是深入到基因的分子水平，例如对多个位点或基因在染色体上的线性排列问题进行研究。

# 概念

- 遗传流行病学是一门研究与遗传有关疾病在亲属中的病因,分布,控制以及在人群中遗传规律的学科。
- 这里遗传的概念是广义的,它包括了生物学的遗传和教养的传递。
- 这里的亲属也是广义的,可以近至同卵双胞胎,也可以远至同一种族。

- a discipline closely allied to traditional epidemiology that focuses on the familial, and in particular genetic, determinants of disease and the joint effects of genes and non-genetic determinants.

● -PR ● Burton et al, Lancet 2005

# 区别

## 遗传流行病学与一般流行病学,医学遗传学,群体遗传学的区别

- 与一般流行病学相比，遗传流行病学强调遗传因素和家庭相似性
- 与医学遗传学相比，强调群体
- 与群体遗传学相比，强调研究遗传因素与疾病或健康的关系

# 研究意义

- 遗传病的危害：人群中**有25—30%受累**；在未患遗传病的人群中，每个人也携带**5—8个有害基因**。
- 遗传负荷：由于环境污染等原因，新的遗传病不断出现（近**10**年来，平均每年新发现**100—200种**），遗传负荷不断增大。

# 遗传与环境的相对意义

- 了解遗传与环境在疾病发病中相互作用与相对关系意义重大，这决定了控制对策与措施。从现代病因模式看，所有的致病因素都可归入遗传和环境两个范畴。现在一般认为，绝大多数疾病均是遗传因素与环境因素相互作用的结果。

# 疾病分类

- (1) 由遗传因素决定发病。如成骨发育不全,血友病等。
- (2) 基本上由遗传因素决定的,但需要一定的环境因素才发病。如蚕豆病、半乳糖血症等。
- (3) 多基因遗传疾病。如唇腭裂等先天畸形、糖尿病等慢性病、一些传染病等。这是本学科研究的重点。
- (4) 基本上由环境因素决定发病而无明显的遗传基础。如中毒、外伤等。



# 定位

遗传流行病学是流行病学的一新分支。

- 一方面是以遗传病作为对象，研究遗传病在群体中发生的环境原因和遗传背景，从而控制其发生的一门学科；
- 另一方面它也作为一门方法学用以研究人群中各种疾病的遗传性原因。



# 研究方法

提示疾病与遗传有关的证据:

1. 患者亲属中该病的发病率或患病率比可比的非患者的亲属中该病的发病率或患病率高;
2. 某一特定人群中的病例的亲属比该人群该病的频率或危险性高;
3. 有该病家族史的健康个体的队列,在一定时间里发生该病的比例高于没有该病家族史的可比的个体的队列的发病的比例;

# 研究方法

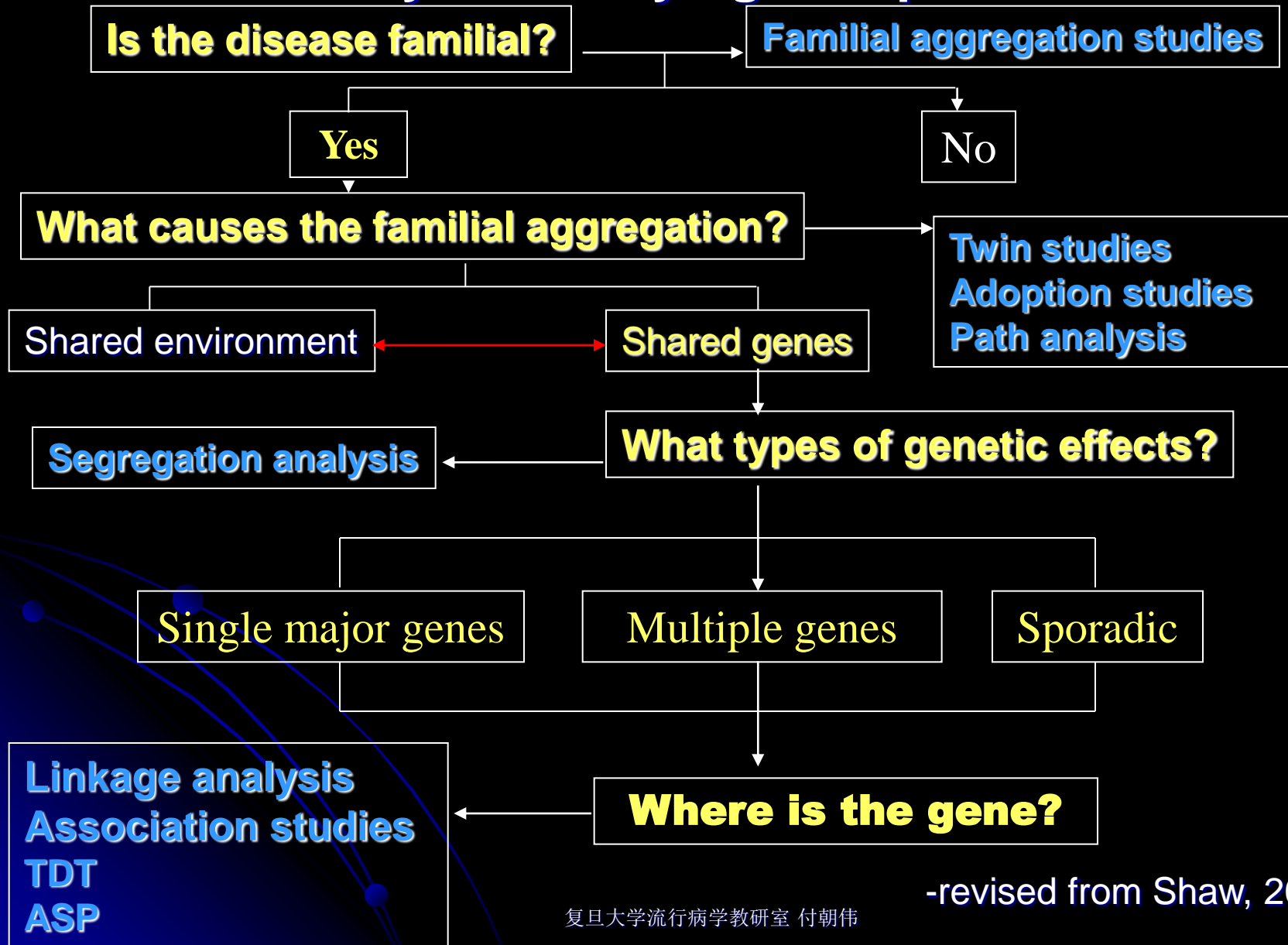
提示疾病与遗传有关的证据（续）：

4. 有血缘关系的个体之间的某性状值的相关显著地高于零；
5. 同卵双生子的同病率较异卵双生子为高；
6. 共同生活的非近亲者均不发病,而近亲中有发病者。

# 研究方法

- 流行病学的常用方法一般在遗传流行病学中均可使用，在这里就不赘述了。
- 但是遗传流行病学涉及的是一个特殊的群体即有亲缘关系的群体，因此有其独特的分析研究方法。

# Pathway for Studying Complex Disease



-revised from Shaw, 2001

# 研究方法—家族或种族聚集性分析

- ①比较患者亲属与普通人群的患病率或发病率，若前者的患病率或发病率大于后者，则提示有家族的聚集性；
- ②比较患者亲属与对照亲属的患病率或发病率，若前者大于后者也提示有家族的聚集性；
- ③患者亲属的患病率或发病率如随亲缘级数的降低而逐渐升高，则表明有家族聚集性；

# 研究方法—家族或种族聚集性分析

- ④证实有家族史患者亲属的发病风险高于从群体中随机抽取患者亲属的发病风险，也说明有家族聚集性的存在；
- ⑤对某些数量性状，如血压水平等，亲属对之间的相关大于非亲属对之间者，则提示该性状有家族聚集性。



# 研究方法—家族或种族聚集性分析

对于一种疾病在家族中的聚集现象，通常有三种解释：①患者家族中有致病基因的遗传；②由于教养传递，致病的行为、生活方式等危险因素通过学习或模仿在家庭中由上代延续到下一代；③由于共同的环境因素，即家庭中各成员暴露在同一危险因素的环境之中。

# 研究方法—家族或种族聚集性分析

- 为得到这三种解释，通常采用的研究设计或分析方法有：双生子分析法、半同胞分析法、养子分析法。



# 研究方法—家系调查和系谱分析

- 判断某种疾病或性状的遗传方式最常用的方法。
- 通过家系调查，绘制系谱图，对所研究疾病或性状有初步的认识与判断。

# 研究方法—双生子分析

- 双生子是人类性状和疾病遗传学研究的极好材料。从理论上讲，通过异卵双生子(DZ)（1/2相同遗传物质、共同生活环境）和单卵双生子(MZ)（有完全一致的遗传物质、共同生活环境）的发病一致率和数量性状相关系数的比较，可以定量估计遗传因素的作用程度。

# 研究方法—双生子分析

- 如果疾病或数量性状是完全由于遗传因素造成，则同卵双生的同病一致率或相关系数应为100%，而异卵双生子则视遗传方式不同而异，不超过50%。
- 同病率或一致率是指双生子中某疾病或性状的一致性。

$$\text{同病率} = C / (C + D)$$

# 研究方法—半同胞及养子分析

- 半同胞是指同父异母或同母异父的兄弟姐妹。半同胞分析可分析疾病或性状来自父方或母方。
- 养子分析通过与其同胞及生身父母和寄养同胞或养父母的相似性比较，可研究遗传与环境因素相对作用的大小。

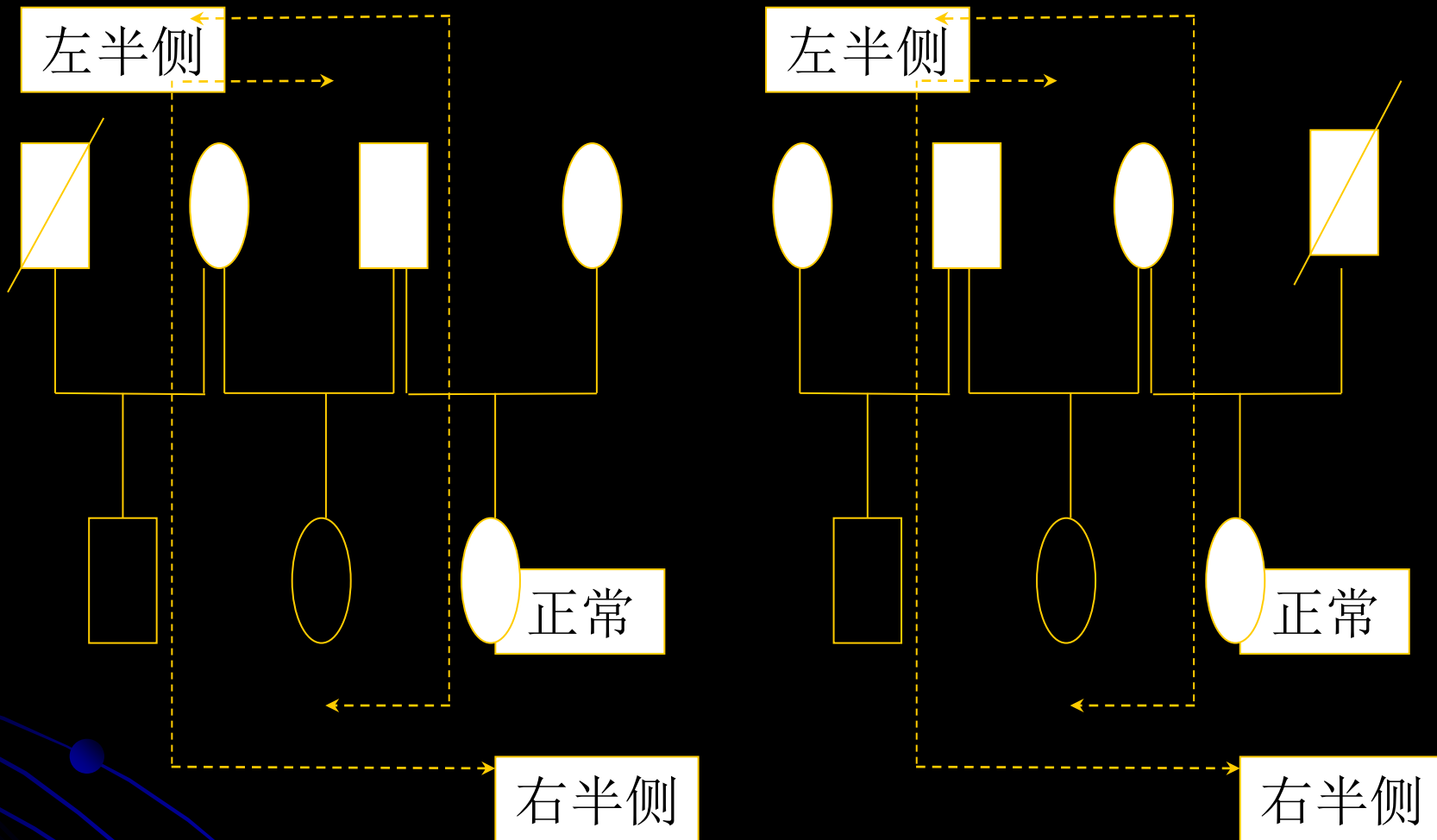


图 1 半同胞遗传分析图

1968年Denng和Heston观察了精神分裂症住院女病人产后3天送出寄养的子女47人，同时寻找条件近似，由正常的生身母亲送出的寄养子女50名作对照，观察结果如表：

	精神分裂症母亲组	正常母亲组
寄养子女数	47	50
平均年龄	38.8岁	36.3岁
精神分裂症	5	0
精神发育缺陷（智商<70）	4	0
神经症性人格障碍	13	7

1968年Kety和Rosenthal根据人口登记，检查出1924-1927年间出生，寄养在无血缘关系家庭中的寄养子女5483人，其中亲生父母一方患精神分裂症的有69人；另选86人，其亲生父母为正常者作为对照组。结果如下：

	亲生父母患精神分裂症	亲生父母正常
人数	69	86
精神分裂症	3	0
边缘性精神分裂	6	1
类分裂人格	3	4

以33名精神分裂症患者作为病例组，另选33名正常寄养子女作为对照组，分别调查了他们的亲生父母家系和寄养父母家系的精神分裂症发病情况，结果如下：

	病例组	对照组
人 数	33	33
亲生父母家系*患病率	11/173 (6.4%)	3/174 (1.7%)
寄养父母家系*患病率	1/74 (1.4%)	9/21 (2.2%)

\*：家系包括双亲、同胞及半同胞。



# 研究方法—单基因疾病的分析

- 通过对同一种病或性状进行多个系谱分析后加以综合, 检验其分离比, 是否符合由某种特定遗传方式决定的理论值, 即分离分析 (segregation analysis)。
- 分离比是指患者在同胞总数中的比值。

# 研究方法—单基因疾病的分析

- 对常染色体隐性遗传来说，当以子代来估计双亲为杂合子家系的分离比时，往往偏高，因为某些没有患病子女的家系常被遗漏，也就是说统计到的子女总数低于实际的子女总数，这种遗漏称为截尾。如果如果病例能够完全确认或选出，这种调查法一般称为完全截尾确定法。如果人群中该病的频率较低，病例收集较不完全，称为不完全确定法。这时家系被调查的概率与家系的大小及病例的多少有关，偏倚更大，需加以校正。

# 研究方法—单基因疾病的分析

- 假定分离比为 $p$ ，理论上，生有 $s$ 个子女的家系，其分布符合 $(p+q)^s$ 的二项分布。

人群中每家同胞组患者与非患者的分布(随机调查)	样本中患者与非患者的分布(完全确定法)	样本中患者与非患者的分布(不完全确定法)
<p style="text-align: center;">* 1</p> <p>每家系无病人的概率 <math>(1 - \theta)^3</math></p> <p style="text-align: center;">000 000</p>	<p style="text-align: center;">Complete ascertainment 或 Complete Truncate <math>\pi = A/R = 1</math></p>	<p style="text-align: center;">Incomplete ascertainment</p> <p style="text-align: center;"><math>1 &gt; \pi &gt; 0</math></p> <p><math>\pi</math> 为病例被确定的比例, A 为被确定的病例数, R 为同胞组内的病例数</p>
<p>每家系一人有病其余二人无病的概率 <math>3(1 - \theta)^2\theta</math></p> <p style="text-align: center;">000 000</p>	<p style="text-align: center;">000 000</p>	<p style="text-align: center;">000 000</p> <p>在不完全确定条件下不同病人非病人组合的家系出现概率取决于家内病人人数 <math>3(1 - \theta)^2\theta</math></p>
<p>每家系一人无病二人有病的概率 <math>3(1 - \theta)\theta^2</math></p> <p style="text-align: center;">000 000 000 000 000 000 000 000 000</p>	<p style="text-align: center;">000 000 000 000 000 000 000 000 000</p>	<p style="text-align: center;">000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000</p> <p>每家两名病人时其被确定的概率两倍于 1 名病人者。 <math>6(1 - \theta)\theta^2</math></p>
<p>三家系成员皆病的概率 <math>\theta^3</math></p> <p style="text-align: center;">000</p>	<p style="text-align: center;">000</p>	<p style="text-align: center;">000 000 000</p> <p>每家三名病人时其被确定的概率三倍于一名病人者 <math>3\theta^3</math></p>

[1]\* 2

患者:非患者

48:144

\*1 注: $\theta = 0.25$

\*2 注:家庭同胞组为 3 的各家庭在:[1]全人群中  
有病无病的分布比例;

[2]完全确定的样本人群中有病无病的分布比例;

[3]不完全确定或单确定样本人群中有病无病的分布比例。

[2]

患者:非患者

48:63

● 为患者

[3]

患者:非患者

72:72

○ 为非患者

# 研究方法—单基因疾病的分析

## 1、完全确认（complete ascertainment）

$\pi=A/R=1$  其中 $\pi$ 为患者被确认为先证者的概率； $A$ 为相互独立确认的患者数； $R$ 为调查得到的患者总数。

1) 先验法（a priori method）

2) 最大似然法（maximum likelihood method）

3) 单病例法（singles method）

# 研究方法—单基因疾病的分析

2、不完全确认 (incomplete ascertainment)

$\pi$  大于0小于1。

1) 同胞法 (sib method) ( $\pi$  趋进于0时)

2) 通用先证者法 (general proband method)

3) 最大似然法 (maximum likelihood method)

3、子女总数校正法 (刘祖洞氏校正法)



# 常染色体隐性遗传(例)

## -单病例法 (Li-Mantel-Gart法)

- 此法计算较简单，结果较精确，比先证者法更接近于用最大似然法所估计的值。但要确定单病例的同胞组并计算其数，故本法称单病例法，用于完全确认。

# 单病例法 (Li-Mantel-Gart法)

其计算公式如下:

$$p = \frac{R - J}{T - J}$$

式中p是欲待校正的估计分离比；R是患儿总数；T是同胞总数；J是家庭内同胞中只有一例患者的家庭数。

标准误的估计公式=

$$\sqrt{\frac{(R - J)(T - R)}{(T - J)^3}}$$



表1 Ellis-Van Creveld 综合征调查资料

同胞组大小	家庭数	同胞数	患儿数	仅发生一例患儿的家庭数
2	2	4	3	1
3	3	9	6	1
4	6	24	7	5
6	4	24	5	3
7	2	14	6	0
8	2	16	4	1
9	1	9	3	0
11	1	11	2	0
12	1	12	2	0
14	1	14	1	1
合计	23 (N)	137 (T)	39 (R)	12 (J)

# 单病例法 (Li-Mantel-Gart法)

$$p = \frac{R - J}{T - J} = \frac{39 - 12}{137 - 12} = 0.216$$

$$s.e._p = \sqrt{\frac{(R - J)(T - R)}{(T - J)^3}} = \sqrt{\frac{27 \times 98}{125^3}} = \sqrt{0.0015527} = 0.0390$$

$$p = 0.216 \pm 2s.e._p = 0.1372 \sim 0.2948$$

估计的p值包括0.25，可初步认为本病为一常染色体隐性遗传病。

# 研究方法—单基因疾病的分析

常染色体显性遗传：

1、 $\chi^2$ 检验法

2、Roberts法

3、Smith无偏分析

# 研究方法—多因子疾病的分析

- 值得注意的是，现今的多因子遗传模型与过去经典的多基因遗传模型完全不同。多因子遗传模型包括了遗传因素（主基因及其遗传规律），同时包括了广义的非随机环境因素（家庭内可测的环境因素和可观察、描述的教养性传递因素）和随机环境因素。

# 研究方法—多因子疾病的分析

- 多因子疾病同单基因病不一样，后者符合孟德尔遗传规律，而多因子多基因病却不一样，故它们又称非孟德尔遗传病或可称之为复杂疾病。复杂疾病的研究方法有遗传度估计方法、通径分析方法以及病例对照研究的关联分析或连锁不平衡分析等。

# 研究方法—多因子疾病的分析

- **易患性 (*liability*)** : 遗传因素和环境因素共同决定的个体所具有的易于或不易于罹患某病的属性变量称为**易患性**。人类易患性的变异在群体中呈正态分布。
- **易感性 (*susceptibility*)** : 单纯由遗传因素决定的患病风险称为**易感性**。

# 研究方法—多因子疾病的分析

- 阈值 (threshold): 当易患性超过一定值时即出现疾病或性状, 该值称作阈值
- 阈值模型 (threshold model): 患者亲属的患病率大于一般人群, 易患性平均值比一般人群易患性平均值更接近阈值 (见图1)

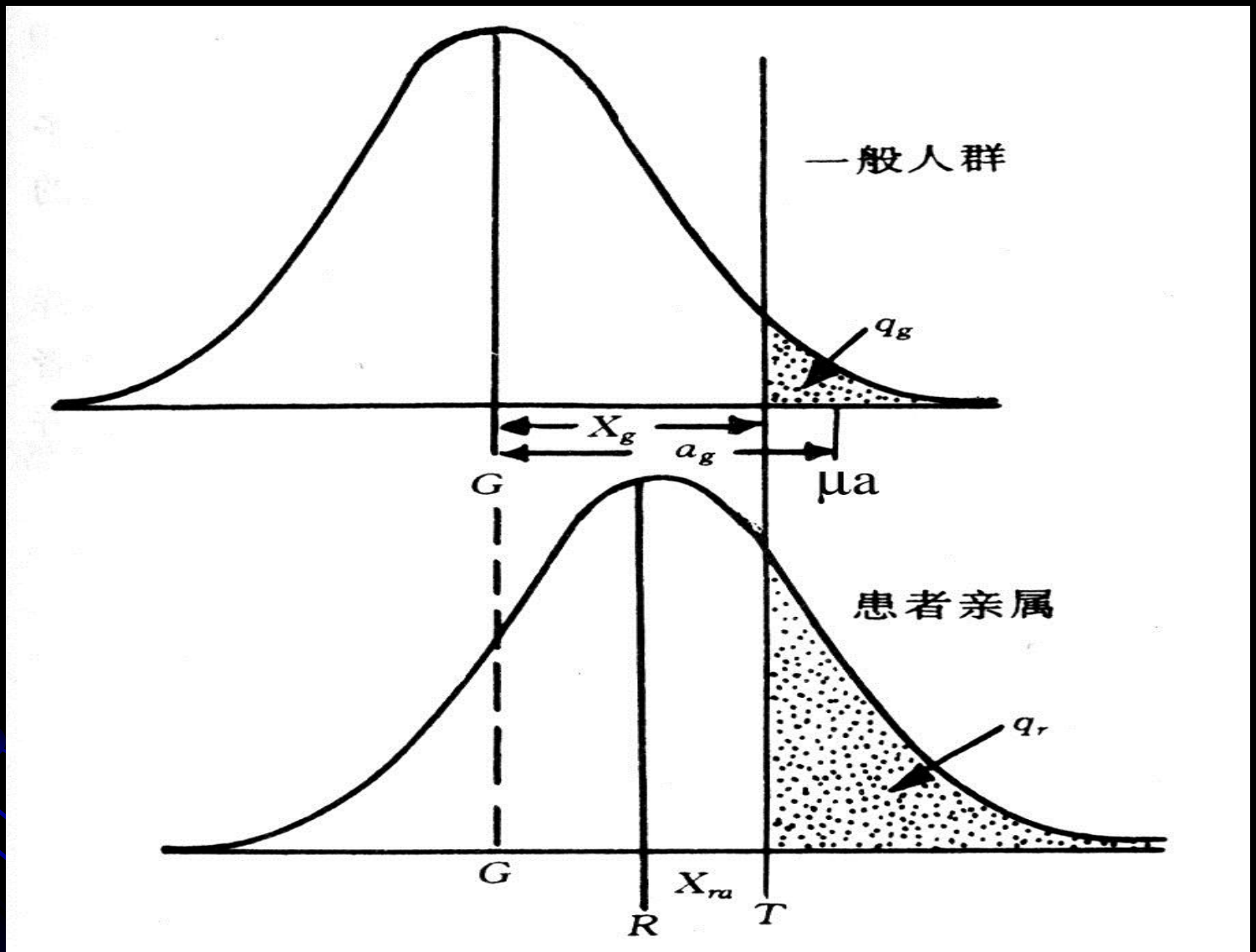


图2 一般人群和患者亲属易患性分布图



# 图注

- **G**: 一般人群易患性均值; **R**: 患者亲属易患性均值; **T**: 阈值;  $X_g$ : 一般人群易患性均值与阈值的偏差;  $X_{ra}$ : 患者亲属易患性均值与阈值的偏差;  $\mu_a$ : 患者的易患性均数;  $a_g$ : 一般人群易患性均值与患者易患性均值的偏差;  $q_g$ : 一般人群患病率;  $q_r$ : 患者亲属患病率

# 研究方法—多因子疾病的分析

- **遗传度 (*heritability*,  $h^2$ )**: 在多因子遗传病中, 遗传因素在疾病发生中作用的程度。遗传度越高, 表明遗传在某病的发病中作用越大。

# 研究方法—多因子疾病的分析

- 双生子分析估计遗传度 — 主要是用于离散性变量（质量性状）

用于前瞻性调查时：

表2 双生子分布表

	单卵双生 (MZ)	双生子 I		异卵双生 (DZ)	双生子 I		
		+	-		+	-	
双生子 II	+	$a_M$	$a_M$	双生子 II	+	$a_D$	$a_D$
	-	$c_M$	$c_M$		-	$b_D$	$b_D$

# 研究方法—多因子疾病的分析

单卵双生一致率

$$(C_{MZ}) = \frac{a_M + d_M}{a_M + b_M + c_M + d_M}$$

异卵双生一致率

$$(C_{DZ}) = \frac{a_D + d_D}{a_D + b_D + c_D + d_D}$$

# 研究方法—多因子疾病的分析

- 通过一致率的计算，可按Holzinger式子求得某病广义遗传度H值：

$$H = \frac{C_{MZ} - C_{DZ}}{100 - C_{DZ}}$$

# 研究方法—多因子疾病的分析

表3 几种疾病MZ同DZ发病一致性的比较及广义遗传度H值

疾 病 种 类	MZ一致率 (%)	DZ一致率 (%)	遗传度H值 (%)
地方性甲状腺肿	71	70	3.3
先天性幽门狭窄	67	8	66.0
十二指肠溃疡	50	4	51.7
糖 尿 病	65	18	57.3
精神分裂症	69	10	65.6

# 研究方法—多因子疾病的分析

- Falconer法—连续变量

Falconer提出一个式子，可由相关系数  $r$  求得狭义遗传度  $h^2$  值：

$$r = \frac{X_g - [X_{ra} \sqrt{1 - (X_g^2 - X_{ra}^2)(1 - X_g / a_g)}]}{a_g + X_{ra}^2 (a_g - X_g)}$$



# 研究方法—多因子疾病的分析

同卵双生 $h^2=r$

一级亲属 $h^2=2r$  (父母、子女、同胞)

二级亲属 $h^2=4r$  (祖父母、外祖父母、叔伯、姑、姨、舅、孙子女、外孙子女)

三级亲属 $h^2=8r$  (表兄弟姐妹、堂兄弟姐妹等)

- 式中 $r$ 为相关系数； $h^2$ 为遗传度；2, 4, 8为亲缘系数 $1/2$ ,  $1/4$ ,  $1/8$ 的倒数；

# 研究方法—多因子疾病的分析

- $X_g$  为一般群体易感性均值与阈值的正态偏差；
- $X_{ra}$  为亲属易感性均值与阈值的正态偏差；
- $a_g$  为一般群体易感性均值与患者易感性均值间的正态偏差。
- $X_g$ 、 $a_g$  值是查“正态分布X和a值表”而得。
- $X_{ra}$  值是根据双生子的发病一致率 ( $C_w$ ) 查“正态分布X和a值表”而得。

# 研究方法—多因子疾病的分析

举例：已知某种多基因遗传病的群体发病率（ $q\%$ ）为0.5，首先查“正态分布 $X$ 和 $a$ 值表”： $X_g=2.576$ ， $a_g=2.892$ ；其次100对MZ中有5对发病，故 $C_w$ 即MZ发病的一致率为5%；查“正态分布 $X$ 和 $a$ 值表”：

$$X_{ra}=1.645$$

# 研究方法—多因子疾病的分析

$$r = \frac{2.576 - \{1.645 \sqrt{1 - (2.576^2 - 1.645^2)[1 - (2.576/2.892)]}\}}{2.892 + 1.645^2(2.892 - 2.576)} = 0.36$$

$$r = \frac{X_g - X_{ra}}{a_g} = \frac{2.576 - 1.645}{2.892} = 0.32$$

同卵双生的 $h^2=r$ ，故该多基因遗传病的遗传度经利用双生子一致率估算得32%-36%

# 研究方法—多因子疾病的分析

- 其他如通经分析、连锁分析、复合分离分析等
- 复合分离分析（complex segregation analysis）是遗传流行病学研究的重要方法，是人类和医学遗传学上经典的分离分析方法的拓展。它以综合方式分析研究人类一些质量性状尤其是数量性状在家系中的传递方式，旨在从多基因的影响中分离出主基因（major gene）的作用。

# 再发危险率估计

## 1. 单基因遗传病

Bayes逆概率法和概率树法。

## 2. 多基因遗传病

(1) 一级亲属发病率较单基因病低, 一般在1-10%。

(2) 一级亲属发病率高与遗传度和一般人群发病率之间有一定关系。

(3) 当群体发病率为0.1-10%时, 遗传度如果是70-80%, 则患者一级亲属的发病率近于发病率的平方根。

# 表4 患者一级亲属发病率 与遗传度和群体发病率的关系

遗传度 (%)	群体发病率 (%)		
	0.1	1	10
50	1	5	20
60	2	6	24
70	3	8	28
80	4	10	30
90	6	13	33
100	8	16	36



# 再发危险率估计

- (4) 在亲属发病率中, 往往单卵双生子 > 一级亲属 > 二级亲属 > 三级亲属 > 一般人群, 但下降幅度各不相同。
- (5) 在一个家庭中罹患成员越多, 则再发危险率也越高。
- (6) 疾病越严重, 再发危险率也越高。
- (7) 再发危险率与疾病的性比率有关。

# 展 望

近年来，遗传流行病学的研究有三大趋势。趋势之一是以Bonney, Elston等人的logistic回归模型为代表的遗传分析数理统计模型的建立和不断完善。他们最大的特点是将严格的流行病学调查、设计思想和近年来的新的生物统计方法与群体遗传学的基本理论有机地结合在一起，将形式多样的调查资料用一种数理模型分析，以便获得更多的信息，并同时进行分离分析、连锁分析和非随机环境因素分析。

# 展 望

趋势之二是对主基因的易感性、基因型与环境因素的互相作用的研究，其代表人物是Khoury等人。他们将生态遗传学的理论与方法引入遗传流行病学中。由于该研究受到调查设计的严格限制，发表的研究论文并不多。

# 展 望

趋势之三是以Risch等人为代表的复杂病因疾病的连锁分析策略研究。他们在1990年发表系列文章，阐述利用亲属对间的表型和标记基因的信息，对复杂性、慢性疾病的易感性主基因进行定位。该连锁分析可以不知道易感性主基因的遗传方式，我们将其称为逆向遗传分析（inverse genetic analysis）。

# 展 望

最后，不管将来遗传学的发展如何，遗传流行病学仍不能忽视两个“E”即流行病学及环境因子。遗传流行病学的大部分还是一门观察性医学，流行病学也是，我们应重视流行病学中述及的一些原则。如应重视设计，仔细寻找对照以避免系统误差，尚可运用病例对照研究，相关性研究，全基因组相关性筛选等。还应关注生物信息学，比如数据分析、计算机软件开发等。

# 展 望

遗传流行病学也有微观的一方面，即运用实验手段为进一步进行遗传分析提供分子水平的实验数据，复杂疾病病因的探索也离不开动物实验，在动物身上寻找同人类同源的较保守的有关基因是研究复杂疾病有关基因的一个重要手段。