

·论 著·

推导IBS评分在全同胞对人群中概率分布的计算公式

赵琪¹, 李燃², 金云舟¹, 孙宏钰², 赵书民³

(1. 上海晶准生物医药有限公司, 上海 201114; 2. 中山大学中山医学院法医学系, 广东 广州 510080; 3. 江苏东南证据科学研究院有限公司, 江苏 南京 210042)

摘要: 目的 推导通过STR等位基因频率计算生物学全同胞对间状态一致性(identity by state, IBS)评分概率分布的通用计算公式。方法 依据孟德尔遗传规律和生物学全同胞(full sibling, FS)的父母为无关个体这一假设, 推导得到不同基因型组合的无关个体生育两名子代时, 子代不同基因型组合的IBS评分及所对应的概率。结果 以 f_i 表示某STR基因座第 i 个等位基因的频率($i=1, 2, \dots, m$), 则生物学全同胞对在该基因座出现2个相等等位基因的概率(p_{2FS})的计算公式为: $p_{2FS} = \frac{1}{4} \times [1 + 2 \sum_{i=1}^m f_i^2 + 2 (\sum_{i=1}^m f_i^2)^2 - \sum_{i=1}^m f_i^4]$; 在该基因座出现1个相等等位基因的概率(p_{1FS})的计算公式为: $p_{1FS} = \frac{1}{2} \times [1 + \sum_{i=1}^m f_i^2 - 2 (\sum_{i=1}^m f_i^2)^2 - 2 \sum_{i=1}^m f_i^3 + 2 \sum_{i=1}^m f_i^4]$; 在该基因座出现0个相等等位基因的概率(p_{0FS})的计算公式为: $p_{0FS} = \frac{1}{4} \times [1 - 4 \sum_{i=1}^m f_i^2 + 2 (\sum_{i=1}^m f_i^2)^2 + 4 \sum_{i=1}^m f_i^3 - 3 \sum_{i=1}^m f_i^4]$; p_{2FS} 、 p_{1FS} 、 p_{0FS} 的和为1。对于包含 n 个STR基因座的多重分型系统, 生物学全同胞间的IBS评分符合二项分布: $IBS \sim B(2n, \pi_1)$ 。其中总体率 π_1 的计算公式为: $\pi_1 = \frac{1}{n} \sum_{i=1}^n p_{2FSi} + \frac{1}{2n} \sum_{i=1}^n p_{1FSi}$ 。结论 生物学全同胞鉴定中的备择假设为两名被鉴定人为生物学全同胞, 对任意STR基因座组合、IBS评分所对应的备择假设概率均可通过本文所推导的公式直接进行计算, 计算结果是进行证据解释的基础。

关键词: 法医遗传学; 全同胞; 状态一致性; 二项分布; 概率

中图分类号: DF795.2 文献标志码: A doi: 10.12116/j.issn.1004-5619.2019.06.003

文章编号: 1004-5619(2019)06-0657-05

Equation Derivation of the Probability Distribution of IBS Score among Full Sibling Pairs

ZHAO Qi¹, LI Ran², JIN Yun-zhou¹, SUN Hong-yu², ZHAO Shu-min³

(1. Shanghai Cubicise Biotechnology Co. Ltd, Shanghai 201114, China; 2. Department of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China; 3. Southeast Academy of Forensic Evidence (Jiangsu) Co. Ltd, Nanjing 210042, China)

Abstract: **Objective** To derive the general equation of the probability distribution of identity by state (IBS) score among biological full sibling pairs by calculating STR allele frequency. **Methods** Based on the Mendelian genetics law and the hypothesis that parents of biological full siblings (FS) were unrelated individuals, the IBS score and corresponding probability of different genotype combinations in the offspring when unrelated individuals of different genotype combinations give birth to two offsprings were derived. **Results** Given f_i ($i=1, 2, \dots, m$) as the frequency of the i th allele of a STR locus, the probability of sharing 2 alleles (p_{2FS}), 1 allele (p_{1FS}) or 0 allele (p_{0FS}) with biological full sibling pairs on the locus can be respectively expressed as follows: $p_{2FS} = \frac{1}{4} \times [1 + 2 \sum_{i=1}^m f_i^2 + 2 (\sum_{i=1}^m f_i^2)^2 - \sum_{i=1}^m f_i^4]$, $p_{1FS} = \frac{1}{2} \times [1 + \sum_{i=1}^m f_i^2 - 2 (\sum_{i=1}^m f_i^2)^2 - 2 \sum_{i=1}^m f_i^3 + 2 \sum_{i=1}^m f_i^4]$ and $p_{0FS} = \frac{1}{4} \times [1 - 4 \sum_{i=1}^m f_i^2 + 2 (\sum_{i=1}^m f_i^2)^2 + 4 \sum_{i=1}^m f_i^3 - 3 \sum_{i=1}^m f_i^4]$. The sum of p_{2FS} , p_{1FS} and p_{0FS} must be 1. As for the multiple genotyping system that contained n STR loci, IBS scores between biological full sibling pairs conform to binomial distribution: $IBS \sim B(2n, \pi_1)$. The population rate π_1 , can be given by the formula: $\pi_1 = \frac{1}{n} \sum_{i=1}^n p_{2FSi} + \frac{1}{2n} \sum_{i=1}^n p_{1FSi}$. **Conclusion** The alternative hypothesis in biological full sibling testing is that two appraised individuals are biological full siblings. The probability of the corresponding alternative hypothesis of any STR locus combination or IBS score can be directly calculated by the equations presented in this study, and the calculation results are the basis for explanations of the evidence.

Keywords: forensic genetics; full sibling; identity by state; binomial distribution; probability

基金项目: 国家自然科学基金面上资助项目(81671873)

作者简介: 赵琪(1990—), 女, 硕士, 主要从事法医遗传学研究; E-mail: zhaqiqiqi12@163.com

通信作者: 赵书民, 男, 副研究员, 主要从事法医遗传学研究; E-mail: zhaoshuminxl@hotmail.com

通信作者: 孙宏钰, 女, 教授, 主要从事法医遗传学研究; E-mail: sunhongyu2002@163.com

在亲缘关系鉴定中,原假设 H_0 (被鉴定人间为无关个体)与备择假设 H_1 (被鉴定人间存在所声称的亲缘关系)是一对有机整体。在司法部颁布实施的《生物学全同胞关系鉴定实施规范》(SF/Z JD0105002—2014,以下简称《规范》)中也规定,生物学全同胞(full sibling,FS)关系鉴定的参照系为无关个体^[1]。同时《规范》中规定采用状态一致性(identity by state, IBS)评分作为全同胞关系鉴定的判断参数(以下简称IBS评分)。目前已有学者^[2]仅依据短串联重复(short tandem repeat,STR)序列遗传标记等位基因的频率数据对IBS评分在无关个体对人群中的概率分布特征进行了描述,但作为逻辑推导过程的一个整体,仍需要对IBS评分在全同胞对人群中的概率分布特征进行推导,才可以形成完整的逻辑链条。

1 方法与结果

1.1 参数定义

比较一对生物学全同胞(C与D)在某一常染色体STR基因座的分型结果,存在3种相互排斥的结果:

有2个相同的等位基因、仅有1个等位基因相同、无相同等位基因,分别以 $a_2=1$ 、 $a_1=1$ 、 $a_0=1$ 表示。当对C与D完成 n 个相互独立的STR基因座分型后,则有 A_2 为 n 个STR基因座上 a_2 之和、 A_1 为 n 个STR基因座上 a_1 之和、 A_0 为 n 个STR基因座上 a_0 之和^[1-2]。以 ibs 表示C与D在单个STR基因座上相同的等位基因个数,依据 a_2 、 a_1 、 a_0 的概念有:

$$ibs=2 \times a_2 + 1 \times a_1 + 0 \times a_0, \quad (1)$$

$$A_2 + A_1 + A_0 = n, \quad (2)$$

$$IBS = 2 \times A_2 + 1 \times A_1 + 0 \times A_0. \quad (3)$$

1.2 生物学全同胞对中 a_2 、 a_1 、 a_0 的概率

依据人类现代婚配习俗,全同胞对C与D的生物学父母A与B为无关个体。在某常染色体STR基因座,A与B最少可检出1种等位基因(以P表示),最多可检出4种不同的等位基因(分别以P、Q、R、S表示)。A与B生育两名个体为两次独立事件,在不考虑STR基因座自发突变的情形下,依据孟德尔遗传规律,可得到A与B不同基因型组合对应C与D在该常染色体STR基因座 $a_2=1$ 、 $a_1=1$ 、 $a_0=1$ 的概率见表1。

表1 亲代不同基因型组合对应全同胞对 a_2 、 a_1 或 a_0 取值为1的概率

亲代 A 与 B 的基因型 组合	亲代 A 与 B 基因型组合 的概率	子代可能 的基因型	两名子代 C 与 D 可能的基因型组合及相应概率						
			$a_2=1$		$a_1=1$		$a_0=1$		
			HoHo	HeHe	HoHe	HeHe	HoHo	HoHe	HeHe
PP×PP	$p_2(\text{HoHo})$	PP	1×1=1	0	0	0	0	0	0
PQ×PQ	$p_2(\text{HeHe})$	PP/PQ/ QQ	$\frac{1}{4} \times \frac{1}{4} \times 2 = \frac{1}{8}$	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4} \times \frac{1}{2} \times 2 \times 2 = \frac{1}{2}$	0	$\frac{1}{4} \times \frac{1}{4} \times 2 = \frac{1}{8}$	0	0
PP×PQ	$p_1(\text{HoHe})$	PP/PQ	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$	0	0	0	0
PQ×QR	$p_1(\text{HeHe})$	PQ/PR/ QQ/QR	$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$	$\frac{1}{4} \times \frac{1}{4} \times 3 = \frac{3}{16}$	$\frac{1}{4} \times \frac{1}{4} \times 2 \times 2 = \frac{1}{4}$	$\frac{1}{4} \times \frac{1}{4} \times 2 \times 3 = \frac{3}{8}$	0	$\frac{1}{4} \times \frac{1}{4} \times 2 = \frac{1}{8}$	0
PP×QQ	$p_0(\text{HoHo})$	PQ	0	1×1=1	0	0	0	0	0
PP×QR	$p_0(\text{HoHe})$	PQ/PR	0	$\frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$	0	$\frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$	0	0	0
PQ×RS	$p_0(\text{HeHe})$	PR/PS/ QR/QS	0	$\frac{1}{4} \times \frac{1}{4} \times 4 = \frac{1}{4}$	0	$\frac{1}{4} \times \frac{1}{4} \times 2 \times 4 = \frac{1}{2}$	0	0	$\frac{1}{4} \times \frac{1}{4} \times 2 \times 2 = \frac{1}{4}$

注:亲代A与B可能的基因型概率计算公式见参考文献[2];HoHo表示均为纯合子,HoHe表示1个为纯合子、1个为杂合子,HeHe表示均为杂合子。

若以 p_{2FS} 、 p_{1FS} 、 p_{0FS} 分别表示全同胞对C与D间 a_2 、 a_1 、 a_0 取值为1的概率,则有:

$$p_{2FS} + p_{1FS} + p_{0FS} = 1, \quad (4)$$

即表1每一行对应的两名子代C与D不同基因型组合的概率和应为1。

依据表1可将 p_{2FS} 、 p_{1FS} 、 p_{0FS} 分解为:

$$p_{2FS} = p_{2FS(\text{HoHo})} + p_{2FS(\text{HeHe})}, \quad (5)$$

$$p_{1FS} = p_{1FS(\text{HoHe})} + p_{1FS(\text{HeHe})}, \quad (6)$$

$$p_{0FS} = p_{0FS(\text{HoHo})} + p_{0FS(\text{HoHe})} + p_{0FS(\text{HeHe})}. \quad (7)$$

1.3 生物学全同胞间不同参数概率计算公式的推导

设某STR基因座有 m 个等位基因,并以 f_i ($i=1,2,3,\dots,m$)表示该基因座上第 i 个等位基因的频率,则有:

$$\sum_{i=1}^m f_i = 1. \quad (8)$$

1.3.1 推导 p_{2FS} 计算公式

1.3.1.1 $p_{2FS(\text{HoHo})}$ 计算公式的推导

式(5)中 $p_{2FS(\text{HoHo})}$ 指C与D基因型相同且均为纯合子的概率,这一概率为表1中第4列各行概率值与相应A与B基因型组合概率乘积之和,即 $p_{2FS(\text{HoHo})}$ 可以写作:

$$p_{2FS(\text{HoHo})} = 1 \times p_2(\text{HoHo}) + \frac{1}{8} \times p_2(\text{HeHe}) + \frac{1}{4} \times p_1(\text{HoHe}) + \frac{1}{16} \times$$

$$p_1(\text{HeHe}). \quad (9)$$

依据参考文献[2]中无关个体对相应基因型概率,可以得到:

$$p_{2FS(\text{HoHo})} = \frac{1}{4} \times \left(\sum_{i=1}^m f_i^2 + 2 \sum_{i=1}^m f_i^3 + \sum_{i=1}^m f_i^4 \right). \quad (10)$$

1.3.1.2 $p_{2FS(HeHe)}$ 计算公式的推导

式(5)中 $p_{2FS(HeHe)}$ 指 C 与 D 基因型相同且均为杂合子的概率,这一概率为表 1 中第 5 列各行概率值与相应 A 与 B 基因型组合概率乘积之和,即 $p_{2FS(HeHe)}$ 可以写作:

$$p_{2FS(HeHe)} = \frac{1}{4} \times p_{2(HeHe)} + \frac{1}{4} \times p_{1(HoHe)} + \frac{3}{16} \times p_{1(HeHe)} + 1 \times p_{0(HoHo)} + \frac{1}{2} \times p_{0(HoHe)} + \frac{1}{4} \times p_{0(HeHe)} \quad (11)$$

依据参考文献[2]中无关个体对相应基因型概率,可以得到:

$$p_{2FS(HeHe)} = \frac{1}{4} \times [1 + \sum_{i=1}^m f_i^2 + 2(\sum_{i=1}^m f_i^2)^2 - 2 \sum_{i=1}^m f_i^3 - 2 \sum_{i=1}^m f_i^4] \quad (12)$$

所以,依据式(5)可以得到:

$$p_{2FS} = p_{2FS(HoHo)} + p_{2FS(HeHe)} = \frac{1}{4} \times [1 + 2 \sum_{i=1}^m f_i^2 + 2(\sum_{i=1}^m f_i^2)^2 - \sum_{i=1}^m f_i^4] \quad (13)$$

1.3.2 推导 p_{1FS} 计算公式1.3.2.1 $p_{1FS(HoHe)}$ 计算公式的推导

式(6)中 $p_{1FS(HoHe)}$ 指 C 与 D 在该基因座只有 1 个相同的等位基因且两者 1 个为纯合子、1 个为杂合子的概率,这一概率为表 1 中第 6 列各行概率值与相应 A 与 B 基因型组合概率乘积之和,即 $p_{1FS(HoHe)}$ 可以写作:

$$p_{1FS(HoHe)} = \frac{1}{2} \times p_{2(HeHe)} + \frac{1}{2} \times p_{1(HoHe)} + \frac{1}{4} \times p_{1(HeHe)} \quad (14)$$

依据参考文献[2]中无关个体对相应基因型概率,可以得到:

$$p_{1FS(HoHe)} = \sum_{i=1}^m f_i^2 - \sum_{i=1}^m f_i^4 \quad (15)$$

1.3.2.2 $p_{1FS(HeHe)}$ 计算公式的推导

式(6)中 $p_{1FS(HeHe)}$ 指 C 与 D 在该基因座只有 1 个相同的等位基因且两者均为杂合子的概率,这一概率为表 1 中第 7 列各行概率值与相应 A 与 B 基因型组合概率乘积之和,即 $p_{1FS(HeHe)}$ 可以写作:

$$p_{1FS(HeHe)} = \frac{3}{8} \times p_{1(HeHe)} + \frac{1}{2} \times p_{0(HoHe)} + \frac{1}{2} \times p_{0(HeHe)} \quad (16)$$

依据参考文献[2]中无关个体对相应基因型概率,可以得到:

$$p_{1FS(HeHe)} = \frac{1}{2} \times [1 - \sum_{i=1}^m f_i^2 - 2(\sum_{i=1}^m f_i^2)^2 - 2 \sum_{i=1}^m f_i^3 + 4 \sum_{i=1}^m f_i^4] \quad (17)$$

所以,依据式(6)可以得到:

$$p_{1FS} = p_{1FS(HoHe)} + p_{1FS(HeHe)} = \frac{1}{2} \times [1 + \sum_{i=1}^m f_i^2 - 2(\sum_{i=1}^m f_i^2)^2 - 2 \sum_{i=1}^m f_i^3 + 2 \sum_{i=1}^m f_i^4] \quad (18)$$

1.3.3 推导 p_{0FS} 计算公式1.3.3.1 $p_{0FS(HoHo)}$ 计算公式的推导

式(7)中 $p_{0FS(HoHo)}$ 指 C 与 D 无相同等位基因且两者均为纯合子的概率,这一概率为表 1 中第 8 列各行概率

值与相应 A 与 B 基因型组合概率乘积之和,即 $p_{0FS(HoHo)}$ 可以写作:

$$p_{0FS(HoHo)} = \frac{1}{8} \times p_{2(HeHe)} \quad (19)$$

依据参考文献[2]中无关个体对 $p_{2(HeHe)}$ 概率,可以得到:

$$p_{0FS(HoHo)} = \frac{1}{4} \times [(\sum_{i=1}^m f_i^2)^2 - \sum_{i=1}^m f_i^4] \quad (20)$$

1.3.3.2 $p_{0FS(HoHe)}$ 计算公式的推导

式(7)中 $p_{0FS(HoHe)}$ 指 C 与 D 无相同等位基因且两者 1 个为纯合子、1 个为杂合子的概率,这一概率为表 1 中第 9 列各行概率值与相应 A 与 B 基因型组合概率乘积之和,即 $p_{0FS(HoHe)}$ 可以写作:

$$p_{0FS(HoHe)} = \frac{1}{8} \times p_{1(HeHe)} \quad (21)$$

依据参考文献[2]中无关个体对 $p_{1(HeHe)}$ 概率,可以得到:

$$p_{0FS(HoHe)} = \frac{1}{2} \times [\sum_{i=1}^m f_i^2 - (\sum_{i=1}^m f_i^2)^2 - 2 \sum_{i=1}^m f_i^3 + 2 \sum_{i=1}^m f_i^4] \quad (22)$$

1.3.3.3 $p_{0FS(HeHe)}$ 计算公式的推导

式(7)中 $p_{0FS(HeHe)}$ 指 C 与 D 无相同等位基因且两者均为杂合子的概率,这一概率为表 1 中第 10 列各行概率值与相应 A 与 B 基因型组合概率乘积之和,即 $p_{0FS(HeHe)}$ 可以写作:

$$p_{0FS(HeHe)} = \frac{1}{4} \times p_{0(HeHe)} \quad (23)$$

依据参考文献[2]中无关个体对 $p_{0(HeHe)}$ 概率,可以得到:

$$p_{0FS(HeHe)} = \frac{1}{4} \times [1 - 6 \sum_{i=1}^m f_i^2 + 3(\sum_{i=1}^m f_i^2)^2 + 8 \sum_{i=1}^m f_i^3 - 6 \sum_{i=1}^m f_i^4] \quad (24)$$

所以,依据式(7)可以得到:

$$p_{0FS} = p_{0FS(HoHo)} + p_{0FS(HoHe)} + p_{0FS(HeHe)} = \frac{1}{4} \times [1 - 4 \sum_{i=1}^m f_i^2 + 2(\sum_{i=1}^m f_i^2)^2 + 4 \sum_{i=1}^m f_i^3 - 3 \sum_{i=1}^m f_i^4] \quad (25)$$

1.4 推导生物学全同胞对人群中 IBS 二项分布的特征参数

依据 1.1 节中关于 a_2 、 a_1 、 a_0 的定义可知,对生物学全同胞对 C 与 D 采用包含 n 个相互独立的 STR 基因座进行分型检测, $a_2=1$ 这样的事件所发生的次数(即 A_{2FS})服从总体率为 p_{2FS} 的二项分布,可记作 $A_{2FS} \sim B(n, p_{2FS})^{[3]}$ 。同样的, A_{1FS} 服从总体率为 p_{1FS} 的二项分布,记作 $A_{1FS} \sim B(n, p_{1FS})^{[3]}$ 。其中 p_{2FS} 为 1.3.1 节推导得到的各 STR 基因座 p_{2FSl} ($l=1, 2, \dots, n$) 的算术平均数, p_{1FS} 为 1.3.2 节推导到各 STR 基因座 p_{1FSl} ($l=1, 2, \dots, n$) 的算术平均数,即有:

$$p_{2FS} = \frac{1}{n} \sum_{l=1}^n p_{2FSl}, \quad (26)$$

$$p_{1FS} = \frac{1}{n} \sum_{i=1}^n p_{1FSi} \quad (27)$$

依据式(3)和二项分布的可加性^[3], *IBS* 亦为服从二项分布的变量。对于包含 *n* 个相互独立的 STR 基因座的分型系统而言, *IBS* 的最大值为 *2n*, 则 *IBS* 服从总体率为 π_1 的二项分布: $IBS \sim B(2n, \pi_1)$ 。

全同胞对人群中 *IBS* 的期望值 $E(IFS)$ 为:

$$E(IFS) = 2 \sum_{i=1}^n p_{2FSi} + \sum_{i=1}^n p_{1FSi} \quad (28)$$

全同胞对人群中 *IBS* 的总体率 π_1 为:

$$\pi_1 = \frac{E(IFS)}{2n} = \frac{1}{n} \sum_{i=1}^n p_{2FSi} + \frac{1}{2n} \sum_{i=1}^n p_{1FSi} \quad (29)$$

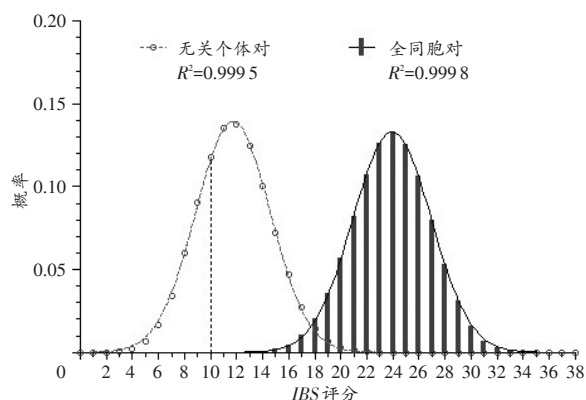
1.5 应用示例

1.5.1 *FGA* 等 19 个必检 STR 基因座对应各参数二项分布特征

采用参考文献[4-5]中对《规范》中 *FGA* 等 19 个必检 STR 基因座在华东汉族人群中的等位基因频率, 依据每一 STR 各自的等位基因频率和 1.3 节推导得到的计算公式, 可以分别计算出各个 STR 基因座对应的 p_{2FS} 、 p_{1FS} 、 p_{0FS} 值, 进而得到各参数在全同胞对人群中的分布特征(表 2)。依据 *IBS* 分布特征可得到相应参数在全同胞对人群中的分布曲线(图 1)。

表 2 华东汉族人群 *FGA* 等 19 个 STR 基因座对应各参数在不同研究人群中的二项分布特征

参数	无关个体对总体率	生物学全同胞对总体率
A_2	0.073 8	0.369 0
A_1	0.474 5	0.518 1
A_0	0.451 7	0.112 9
<i>IBS</i>	0.311 0	0.628 0



曲线为采用二项分布近似正态分布模拟得到, 非线性模拟采用 GraphPad Prism 5.0 软件包; 左侧曲线为 *IBS* 评分在无关个体对人群中的分布^[2]; 虚线高度为 *IBS*=10 时所检个体对为无关个体对的概率。

图 1 《规范》中 19 个必检基因座分型系统对应全同胞对与无关个体对的 *IBS* 二项分布

1.5.2 全同胞鉴定中遗传学证据价值评估示例

当两名被鉴定人采用 *FGA* 等 19 个《规范》中规定的必检 STR 基因座进行分型后, 依据表 1 可知, *IBS* 在

无关个体对人群中符合二项分布 $IBS \sim B(38, 0.311 0)$, 在全同胞对人群中符合二项分布 $IBS \sim B(38, 0.628 0)$ 。当全同胞对 C 与 D 经 *FGA* 等 19 个必检 STR 基因座分型后, 若 *IBS* 评分为 10, 采用 Excel 软件中的二项分布函数, 输入 BINOM.DIST(10, 38, 0.311 0, FALSE), 得到两者为无关个体的概率为 $p_{H0}=0.118 129 60$, 输入 BINOM.DIST(10, 38, 0.628 0, FALSE), 得到两者为全同胞对的概率为 $p_{H1}=0.000 004 26$ 。因此, 两者的生物学全同胞关系似然比(likelihood ratio, LR)为:

$$LR = \frac{p_{H1}}{p_{H0}} = \frac{0.000 004 26}{0.118 129 60} = 3.61 \times 10^{-5} \quad (30)$$

生物学全同胞相对概率为:

$$\frac{p_{H1}}{p_{H0} + p_{H1}} = \frac{LR}{1 + LR} = 3.61 \times 10^{-5} \quad (31)$$

即两者为全同胞的概率是两者为无关个体概率的 3.61×10^{-5} 倍(或两者为无关个体的概率约是两者为全同胞概率的 27 730 倍), 两者的全同胞相对概率为 0.000 036 06(或两者的无关个体相对概率为 0.999 963 94)。

1.5.3 对《规范》中 *FGA* 等 19 个必检 STR 基因座生物学全同胞检验系统效能的评估

依据《规范》中对检验 *FGA* 等 19 个必检 STR 基因座时的判断阈值, 采用 1.5.1 节中所示 *IBS* 在无关个体对人群与全同胞对人群中的二项分布特征可得到:

$IBS \leq 13$ 的真无关个体概率为: BINOMDIST(13, 38, 0.311 0, TRUE) $\approx 0.726 8$; 近似为图 1 所示无关个体对人群对应曲线的 $IBS \leq 13$ 的曲线下面积。

$IBS \geq 22$ 的真全同胞概率为: $1 - \text{BINOMDIST}(21, 38, 0.628 0, \text{TRUE}) \approx 0.787 6$; 近似为图 1 所示全同胞对人群对应曲线的 $IBS \geq 22$ 的曲线下面积。

$IBS \leq 13$ 的全同胞(误判为无关个体)概率为: BINOMDIST(13, 38, 0.628 0, TRUE) $\approx 0.000 3$; 近似为图 1 所示全同胞对人群对应曲线的 $IBS \leq 13$ 的曲线下面积。

$IBS \geq 22$ 的无关个体(误判为全同胞)概率为: $1 - \text{BINOMDIST}(21, 38, 0.311 0, \text{TRUE}) \approx 0.000 6$; 近似为图 1 所示无关个体对人群对应曲线的 $IBS \geq 22$ 的曲线下面积。

因此, 检验 *FGA* 等 19 个必检 STR 基因座时能够得到倾向性意见的概率(灵敏度, 即系统效能)约为: $(0.726 8 + 0.787 6)/2 = 0.757 2$; 得到的倾向性意见的平均误判率约为: $(0.000 3 + 0.000 6)/2 = 0.000 45$ 。

2 讨论

本研究依据 *IBS* 评分在无关个体对人群中的分布规律^[2]以及常染色体 STR 基因座的遗传符合孟德尔

遗传规律,推导得到了 IBS 评分在生物学全同胞对人群中的二项分布特征。之所以能够依据 IBS 评分在无关个体对人群中的二项分布规律来推导该参数在生物学全同胞对人群中的二项分布特征,是建立在一个相对稳固的假设前提之上,即:依照现代人类社会的婚配习俗,生物学全同胞的父母间应为无关个体。另外,对于一个足够大的群体(如我国汉族总人口)来说,可以认为该群体符合Hardy-Weinberg平衡。这两个相对稳固的假设前提构成了推导 IBS 评分在无关个体对人群与生物学全同胞对人群中的二项分布特征的基础^[6-11]。

从图1可以看到,当采用《规范》推荐的 FGA 等19个必检STR基因座时, IBS 评分在无关个体对人群与生物学全同胞对人群中的概率分布仍有不小的重叠区间,这是《规范》在给出生物学全同胞鉴定意见时,单独给出一个“无法给出倾向性意见”区间的原因。例如,采用 FGA 等19个必检STR基因座对一对个体进行检验后,当 $IBS=18$ 时,依据参考文献[2]和本文所给出的 IBS 评分在无关个体对人群和生物学全同胞对人群中的二项分布特征,采用Excel二项分布函数,可得到两者为无关个体的概率约为0.0145、为生物学全同胞的概率约为0.0200,两者为生物学全同胞的概率仅为两者为无关个体的概率的1.38倍。而当 $IBS=13$ 时,两者为无关个体的概率约为0.1245,为生物学全同胞的概率约为0.0002,两者为无关个体的概率是两者为生物学全同胞概率的622.5倍;当 $IBS=22$ 时,两者为无关个体的概率约为0.0004,为生物学全同胞的概率约为0.1074,两者为生物学全同胞的概率是两者为无关个体概率的268.5倍。这也是《规范》中当检验 FGA 等19个必检基因座时,分别设定 $IBS \leq 13$ (倾向于两名被鉴定人为无关个体)、 $IBS \geq 22$ (倾向于两名被鉴定人为生物学全同胞)两个阈值的原因,同时这也是《规范》中8.2条款的依据。

1.5.3节对《规范》中检验 FGA 等19个STR基因座对应的判断标准的系统效能进行了详细的解释。对于一个特定的分型系统(一组STR基因座组合)来说,对一个给定的判断标准均应对其判断的效能进行评估,包括这一判断标准的灵敏度和平均误判率等指标。但在具体个案中,因为已检出特定的 IBS 值,则可采用 IBS 值在相应人群中的二项分布直接计算两种假设对应的概率值,参照传统的法医学证据价值评估方法对个案中的遗传学证据价值进行评估,正如1.5.2节所示。

至此,依据参考文献[2]以及本文研究结果,已得

到生物学全同胞鉴定中原假设(H_0 ,被检验的两名个体为无关个体)与备择假设(H_1 ,被检验的两名个体为生物学全同胞)的概率计算工具。当检验任意符合《规范》要求的STR基因座组合时,依据该组计算工具以及STR基因座在相应人群中的等位基因频率,即可计算出任意 IBS 评分所对应的两名被鉴定人为无关个体和为生物学全同胞的概率,并能对相应的系统检验效能加以评估,从而实现《规范》的灵活应用。

参考文献:

- [1] 中华人民共和国司法部司法鉴定管理局. 生物学全同胞关系鉴定实施规范:SF/Z JD0105002—2014[S]. 2014.
- [2] 赵焕东,赵书民,陈玉祥,等. 推导 IBS 评分在无关个体对人群中概率分布的计算公式[J]. 法医学杂志, 2018, 34(4):370-374.
- [3] 郭祖超. 医学统计学[M]. 北京:人民军医出版社, 1999:17.
- [4] LI C, LI L, ZHAO Z, et al. Genetic polymorphism of 17 STR loci for forensic use in Chinese population from Shanghai in East China[J]. Forensic Sci Int Genet, 2009, 3(4):e117-e118.
- [5] 郭宏,林源,柳燕,等. $D6S1043$ 和 $D12S391$ 基因座在亲权鉴定中的应用[J]. 法医学杂志, 2007, 23(5):345-346.
- [6] GARCÍA D, CARLEOS C, PARRA D, et al. Sib-parentage testing using molecular markers when parents are unknown[J]. Anim Genet, 2002, 33(5):364-371.
- [7] ZHAO S, ZHANG S, LI C. Comparison study in determination of full sibling with Identifiler multiplex system between ITO method and identity by state scoring method[J]. Forensic Sci Int Genet Suppl Ser, 2011, 3(1):e335-e336.
- [8] AOKI Y, NAKAYAMA Y, SAIGUSA K, et al. Comparison of the likelihood ratio and identity-by-state scoring methods for analyzing sib-pair test cases: A study using computer simulation[J]. Tohoku J Exp Med, 2001, 194(4):241-250.
- [9] TZENG C H, LYOU J Y, CHEN Y R, et al. Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan[J]. Transfusion, 2000, 40(7):840-845.
- [10] 陈子翔,王福振,陆惠玲,等. 判别函数在同胞鉴定中的应用[J]. 中国法医学杂志, 2012, 27(2):129-132.
- [11] STEVENS E L, HECKENBERG G, ROBERSON E D, et al. Inference of relationships in population data using identity-by-descent and identity-by-state[J]. PLoS Genet, 2011, 7(9):e1002287.

(收稿日期:2018-11-16)

(本文编辑:张素华)