



Research paper

Evaluation of parentage testing accuracy of child trafficking cases: Combining the exclusion probability and likelihood ratio approaches

Kexin Yu, Wing Kam Fung*

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:

Child trafficking
DNA database
Parentage testing
Exclusion probability
Hardy-Weinberg equilibrium
Population substructure
Mutation
Likelihood ratio

ABSTRACT

The Chinese government established a national anti-trafficking DNA database in 2009 to help reunite trafficked children with their families. The database collects DNA information from missing children's parents, trafficked and homeless children, then conducts parentage testing using 18 or more loci to find matched pairs. This article evaluates the matching accuracy of parentage testing in child-trafficking cases, under both Hardy-Weinberg equilibrium and population substructure. Both one-parent and two-parent scenarios are considered, and mutations are taken into account. The number of random matches is first evaluated using exclusion probability (PE). It is found that there are a large number of single parent-child pairs that match at 18 loci, but the PE approach cannot tell which are the true positive ones. The likelihood ratio (LR) approach can help distinguish the true positive matches. So the second step is to obtain the true positive rate and false positive rate of matched pairs of single parent and child according to the LR approach. Based on the results of the two-step procedure, it is concluded that more than 18 loci should be used to ensure a correct match of single parent and child.

1. Introduction

Child trafficking, which refers to the criminal acts of kidnapping, abducting, robbing, stealing or transferring children for the purpose of selling, has been recognized as a serious social problem in modern China [1]. Trafficked children are usually traded through numerous hands before being sold for illegal adoption, in which the buyers are mostly ordinary families who cannot have their own children [2]. Other even less fortunate child victims are sold into forced labor, prostitution, or begging and stealing on street [2].

The crime of child trafficking not only brings immeasurable sorrow and misery to individuals and families, but also imposes risks on public security. Having recognized the severeness of this problem, the Chinese government have taken measures to fight against the crimes [1]. First of all, the criminals involved in child trafficking cases are always punished harshly, especially the leaders and repeat offenders [1,3]. In addition, a specialized anti-trafficking task force has been organized which 'carries out high-profile raids and liberates hundreds, sometimes thousands, of kidnapped children' every year [4]. Furthermore, a national anti-trafficking DNA database has been established by the Ministry of Public Security since the year 2009 [5]. The database collects DNA information using blood samples from parents who are looking for their missing kids, homeless children, as well as rescued ones from child-trafficking gangs, totally free of charge [5]. Using the technique of DNA profiling,

the database can help reunite separated families much quicker than ever, even for those long unsolved cases [6]. More than 4000 missing children have returned home with the help of the DNA database by the end of 2015, as announced by the State Council Information Office of China [7]. However, since China is a country of 1.3 billion people, and the numbers of parents and children in the database are quite large, it is not an easy task to find the matches accurately and efficiently. One of the main objectives of this article is to evaluate the effectiveness and accuracy of parentage testing utilized by the anti-trafficking DNA database.

The remainder of this article is organized as follows. Section 2 gives the exclusion probability for one-parent and two-parent cases, respectively, in a homogeneous population under Hardy-Weinberg equilibrium (HWE) and in a structured population. The formulae are then extended to allow for mutations. In Section 3, we assess the accuracy of parentage testing using 18 loci and develop a two-step method. The first step is to estimate the number of matched parents and children using the exclusion probability. Then in the second step, we employ the likelihood ratio approach to obtain the true positive rate and false positive rate. The results show that 18 loci is not adequate for single-parent case, and demonstrate the effectiveness of the method combining the exclusion probability and likelihood ratio approaches. Using the method developed in Section 3, Section 4 analyzes the accuracy of parentage testing for single parents with increasing number of loci.

* Corresponding author at: Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China.
E-mail addresses: violetyu@connect.hku.hk (K. Yu), wingfung@hku.hk (W.K. Fung).

Finally, Section 5 gives the concluding remarks.

2. Exclusion probability

2.1. Probability of excluding a random individual from parentage

Given the genetic information of one individual and one child, the individual can be excluded from parentage by inconsistency in at least one of the studied loci. The exclusion probability formula for this one-parent case under Hardy-Weinberg equilibrium has been provided by Garber and Morris [8]. However, the Hardy-Weinberg law is hardly exactly true in reality. To account for the population substructure [9], we denote the power or probability of excluding a random individual from parentage using R loci by $PE_{1,R}(\theta)$, where θ is the coancestry coefficient which can be regarded as the probability that two alleles are identical by descent due to recent shared ancestry. When $\theta = 0$, the population is under HWE. Note that $PE_{1,R}(\theta)$ corresponds to all possible parent-offspring combinations and is not limited to any particular case [10]. Consider a locus r having n alleles A_1, A_2, \dots, A_n with corresponding allele frequencies p_1, p_2, \dots, p_n . The probability of excluding a random man or woman from parentage at locus r , denoted by $Q_{1,r}(\theta)$, was given by Ayres [11] as

$$Q_{1,r}(\theta) = \frac{(1-\theta)}{(1+\theta)(1+2\theta)} \left[\sum_{i=1}^n p_i(\theta + (1-\theta)p_i)(1-p_i)(\theta + (1-\theta)(1-p_i)) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j (1-\theta)(1-p_i-p_j)(\theta + (1-\theta)(1-p_i-p_j)) \right]. \tag{1}$$

Suppose we study R autosomal loci in total for the purpose of parentage testing, then the overall exclusion probability for a random individual at all these R loci based on one or more loci inconsistency is taken as

$$PE_{1,R}(\theta) = 1 - \prod_{r=1}^R (1 - Q_{1,r}(\theta)). \tag{2}$$

It is noted that two equivalent formulae under HWE (i.e., $PE_{1,R}(0)$) were given by Garber and Morris [8] and Jamieson and Taylor [12].

2.2. Probability of excluding a random couple from parentage

In child trafficking cases, it is more than common that a couple are looking for their missing child together. Hence we need to consider the probability of excluding a random couple from parentage. Under the same setting as $Q_{1,r}(\theta)$, the probability of excluding a random couple from parentage at locus r is obtained as

$$Q_{2,r}(\theta) = \sum_{i=1}^n p_i^2 q + \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j p, \tag{3}$$

where

$$q = 1 - \frac{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)(1+3\theta)(1+4\theta)} \times [4(1+2\theta+2\theta^2) + (-4-3\theta+7\theta^2)p_i + (1-\theta)^2 p_i^2],$$

and

$$p = 1 - \frac{[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}{(1+\theta)(1+2\theta)(1+3\theta)(1+4\theta)} \times \{ [2\theta + (1-\theta)p_i][12 + 6\theta - 8(1-\theta)p_i - 9(1-\theta)p_j] + [2\theta + (1-\theta)p_j][12 + 6\theta - 8(1-\theta)p_j - 9(1-\theta)p_i] + [8(1-\theta)(1-p_i-p_j)][\theta + (1-\theta)(1-p_i-p_j)] \}.$$

The detailed derivation is given in Appendix I. Note that when $\theta = 0$, $Q_{2,r}(0)$ is the same as that given in Gundel and Reetz [13] and Jamieson and Taylor [12]. Suppose that totally we investigate R loci for parentage determination, then the overall exclusion probability for a random couple at all these R loci on the basis of one or more loci inconsistency is

$$PE_{2,R}(\theta) = 1 - \prod_{r=1}^R (1 - Q_{2,r}(\theta)). \tag{4}$$

2.3. Exclusion probability when allowing for mutation

Because of the high mutation rates of STR loci [14], it is a common practice to allow for one mismatched locus in parentage testing to avoid false exclusion. When one mutation is allowed, the overall exclusion probability based on two or more loci inconsistency for either a random individual ($PE'_{1,R}(\theta)$) or a random couple ($PE'_{2,R}(\theta)$) becomes

$$PE'_{i,R}(\theta) = 1 - \prod_{r=1}^R (1 - Q_{i,r}(\theta)) - \sum_{s=1}^R Q_{i,s}(\theta) \prod_{\substack{r=1 \\ r \neq s}}^R (1 - Q_{i,r}(\theta)), \quad i = 1, 2.$$

It is clear that the power of excluding a random individual or couple from parentage is reduced if we allow for one mismatched locus.

3. Parentage testing accuracy of child trafficking cases

3.1. The anti-trafficking DNA database

The Chinese government established an anti-trafficking DNA database in 2009, which extracts and collects genetic information from blood samples taken by the police from the following 5 groups of people (i) parents of missing children (ii) trafficked children or those suspected of having been trafficked (iii) homeless children (iv) children in social welfare institutions and (v) child beggars, according to the Ministry of Public Security [5]. The minimum number of STR loci to qualify for entry is 18, as required by the ministry [15]. Parentage testing can be conducted by more than 230 qualified DNA laboratories across the country having access to the database to find matched parents and offspring [5]. Since the establishment, the database has gathered genetic information of over $N_1 = 13,000$ single parents, over $N_2 = 65,000$ couples and over $N_c = 500,000$ children (*numbers obtained via personal communication*). The large and ever-increasing size of the database makes it a challenge to find matched parents and children efficiently and accurately.

3.2. Numbers of random matches

To evaluate the matching accuracy of parentage testing for child trafficking cases, we use 18 core STR loci that are commonly analyzed in parentage testing, namely, CSF1PO, D12S391, D13S317, D16S539, D18S51, D19S433, D1S1656, D21S11, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, FGA, TH01, TPOX and vWA. The exclusion probabilities are calculated using the allele frequencies of the Chinese Han population as reported by Wang et al. [16]. The results for one-parent and two-parent cases are listed in Tables 1 and 2, respectively.

Specifically, in Table 1, $PE_{1,18}(0) = 1 - 3.48 \times 10^{-5}$ means that under HWE, the overall probability of excluding a random individual from parentage of a child on the basis of one or more loci inconsistency among 18 loci is $1 - 3.48 \times 10^{-5}$, i.e., on average, among 100,000 random persons, about 3.48 persons are not excluded from being the true parent of a tested child as they match with the child's genotype at

Table 1

Exclusion probability for single parent when allowing for 0 and 1 mutation under HWE ($\theta = 0$) and population substructure ($\theta = 0.02$) using the allele frequencies of the 18 STR loci of the Chinese Han population.

Exclusion probability	$\theta = 0$	$\theta = 0.02$
No mutation	$PE_{1,18}(0) = 1 - 3.48E-5$	$PE_{1,18}(0.02) = 1 - 7.95E-5$
Allow for one mutation	$PE'_{1,18}(0) = 1 - 5.34E-4$	$PE'_{1,18}(0.02) = 1 - 1.10E-3$

Download English Version:

<https://daneshyari.com/en/article/6553262>

Download Persian Version:

<https://daneshyari.com/article/6553262>

[Daneshyari.com](https://daneshyari.com)