

EVALUATING THE STATISTICAL POWER OF DNA-BASED IDENTIFICATION, EXEMPLIFIED BY 'THE MISSING GRANDCHILDREN OF ARGENTINA'

Daniel Kling^{1,*}, Thore Egeland², Mariana Herrera Piñero³, Magnus Dehli Vigeland⁴

¹ Department of Forensic Services, Oslo University Hospital

² IKBM, Norwegian University of Life Sciences

³ Banco Nacional de Datos Genéticos

⁴ Department of Medical Genetics, Oslo University Hospital

Methods and implementations of DNA-based identification are well established in several forensic contexts. However, assessing the statistical power of these methods has been largely overlooked, except in the simplest cases. In this paper we outline general methods for such power evaluation, and apply them to a large set of family reunification cases, where the objective is to decide whether a person of interest (POI) is identical to the missing person (MP) in a family, based on the DNA profile of the POI and available family members. As such, this application closely resembles database searching and disaster victim identification (DVI).

If parents or children of the MP are available, they will typically provide sufficient statistical evidence to settle the case. However, if one must resort to more distant relatives, it is not *a priori* obvious that a reliable conclusion is likely to be reached. In these cases power evaluation can be highly valuable, for instance in the recruitment of additional family members.

To assess the power in an identification case, we advocate the combined use of two statistics: the *Probability of Exclusion*, and the *Probability of Exceedance*. The former is the probability that the genotypes of a random, unrelated person are incompatible with the available family data. If this is close to 1, it is likely that a conclusion will be achieved regarding general relatedness, but not necessarily the specific relationship. To evaluate the ability to recognize a true match, we use simulations to estimate exceedance probabilities, i.e. the probability that the likelihood ratio will exceed a given threshold, assuming that the POI is indeed the MP. All simulations are done conditionally on available family data.

Such conditional simulations have a long history in medical linkage analysis, but to our knowledge this is the first systematic forensic genetics application. Also, for forensic markers mutations cannot be ignored and therefore current models and implementations must be extended. All the tools are freely available in Familias (<http://www.familias.no>) empowered by the R library *paramlink*. The above approach is applied to a large and important data set: 'The missing grandchildren of Argentina'. We evaluate the power of 196 families from the DNA reference databank (Banco Nacional de Datos Genéticos, <http://www.bndg.gov.ar>). As a result we show that 58 of the families have poor statistical power and require additional genetic data to enable a positive identification.