



A172 Identifying Victim Remains From Uncertain Data

Mark W. Perlin, PhD, MD*, *Cybergenetics, 160 North Craig Street, Suite 210, Pittsburgh, PA 15213; and Alexander Sinelnikov, PhD, Erin Turo, BS, Matthew Legler, BA, and Meredith Clarke, MA, Cybergenetics, 160 North Craig Street, Pittsburgh, PA 15213*

After attending this presentation, attendees will understand how they can analyze uncertain mass disaster DNA data (victim remains, personal effects, family references) by inferring genotypes (up to probability) and matching them with a likelihood ratio statistic.

This presentation will impact the forensic community by describing how computer tools can help forensic analysts overcome mass disaster data uncertainty, and thereby make reliable identifications, in the context of the World Trade Center (WTC) STR data. A mass disaster can generate a vast amount of biological material. Such materials include the victim remains evidence at the disaster site, as well as the personal effects and family references of the missing people. These data may have considerable uncertainty, which occurs at many levels. Current reference sample expert systems do not address genotype uncertainty, so instead newer statistical computing methods are needed.

A mass disaster can generate a vast amount of biological material. Such materials include the victim remains (VR) evidence at the disaster site, as well as the personal effects (PE) and family references (FR) of the missing people. These data may have considerable uncertainty, which occurs at many levels.^[1] Current reference sample expert systems do not address genotype uncertainty, so instead newer statistical computing methods are needed. This paper describes how computer tools^[2] can help forensic analysts overcome mass disaster data uncertainty, and thereby make reliable identifications, in the context of the World Trade Center (WTC) STR data.

At the DNA level, VR samples are often degraded, mixed, burned, contain minimal DNA, or are compromised in other ways. PE samples are collected from contaminated environments, and may include mixed or degraded DNA. With FR samples, there can be uncertainty in the family kinship relationships. Human visual review of the STR data is needed to determine which data are good, mixed or unusable.

Indeed, the STR lanes (or injections) associated with a particular sample may be derived from different individuals. To disambiguate the data, we use a visual user interface software tool that lets an analyst view all of the provided lanes for a given sample, and then group compatible lanes together to form one or more genetic calculation requests for computer interpretation of that sample. In the WTC workflow, a visual user interface operator can inspect data and generate interpretation requests for 30 VR samples every hour.

When inferring a profile from uncertain mass disaster data, the resulting DNA profile may include more than one genotype. This profile uncertainty can be represented by assigning probabilities to genotypes.^[3] The forensic analyst can use a commercially available genetic calculator^[2] to infer genotypes in many common scenarios. For example, the calculator can infer profiles from mixtures having two or more unknown contributors, statistically combine uncertain data from multiple lanes, and infer a missing person's profile from kinship data.

To compare victim remains profiles with missing person profiles, the calculator's DNA match module^[2] computes a likelihood ratio that compares the probability of a victim remains match with a missing person to that with a random person. Every match comparison in the system, whether kinship or STR, has a numeric likelihood ratio score that measures the degree of DNA identity.^[3]

To help the forensic analyst integrate all of this profile and match information, the visual user interface provides visual representations at each locus of the genotype contributor probabilities and match likelihood ratios. The original EPG data can be seen and explored by the user at all times. A match-directed sample review takes only a few minutes, since the user can interactively focus on interesting questions about data, genotypes or matches.

A subset of the WTC data for 18,251 VRs and 2,386 PEs have been visually reviewed and statistically reanalyzed. Additionally 2,347 missing person profiles have been genetically reconstructed from 6,660 FRs. In a mass disaster project of this magnitude, forensic analyst expertise is essential for examining the data to ask the correct questions, and assess the calculated answers.^[1] Computer tools, including a visual user interface and a 24 processor genetic calculator supercomputer^[2] for statistical genotype inference and matching,^[3] can perform genetic calculations that help analysts apply their expertise to obtain reliable identifications with less effort.

References:

- ¹ Perlin, MW. "Mass casualty identification through DNA analysis: overview, problems and pitfalls." *Forensic Investigation and Management of Mass Disasters*. MI Okoye and CH Wecht. Tucson, AZ, Lawyers & Judges Publishing Co: 23-30, 2007.
- ² Perlin, MW. "Identifying human remains using TrueAllele® technology." *Forensic Investigation and Management of Mass Disasters*. MI Okoye and CH Wecht. Tucson, AZ, Lawyers & Judges Publishing Co: 31-38, 2007.



Criminalistics Section – 2009

- ³ Perlin MW, Kadane JB, Cotton RW. Forensic DNA Inference (Abstract). In: Seventh International Conference on Forensic Inference and Statistics; 2008 August; Lausanne, Switzerland.

Forensic DNA, Mass Disaster, Computer Software