

Forensic Genetics

Summer Institute in Statistical Genetics, July 20-22, 2016

David Balding: david.balding@unimelb.edu.au

Bruce Weir: bsweir@uw.edu

Bruce Weir Overview

Identifying Disaster Victims

“In most cases, say in a death involving a medical examiner’s office, the identification is pretty simple,” said Dr. Edward Kilbane, director of fatality management at the National Disaster Medical System, and arm of the federal Department of Health and Human Services. “But when you get a group of people together, who are unrelated, then that becomes a challenge. You really have to start applying more scientific methods to really make sure you don’t make any mistakes in the identification.”

New York Times, March 28, 2014. “Mountain that Roared Down Turns Into a Tomb of Debris.”

Legal v. Scientific Thinking

“The very goals of science and law differ. Science searches for the truth and seeks to increase knowledge by formulating and testing theories. Law seeks justice by resolving individual conflicts, although this search often coincides with one for truth.”

“Rules of decision that are not tailored to individual cases, such as those that turn on statistical reasoning, are often viewed as suspect.”

Feinberg et al., The Evolving Role of Statistical Assessments as Evidence in the Courts, 1989.

Forensic Science Approach

“The central problem of the criminal investigator is the establishment of personal identity – usually of the criminal, sometimes of the victim.”

P.L. Kirk, “Crime Investigation” 1974.

Need to distinguish between identity and individualization. *Identity* refers to unique existence – no two different things can be identical. The DNA profiles from a suspect and a crime scene are different things.

Individualization points to a specific person. A fingerprint from a crime scene is not identical to a suspect’s recorded fingerprint, but can be used to identify him and prove his individuality.

Uniqueness

“no two objects can ever be identical. They can and often do have properties that are not distinguishable. If enough of these properties exist ... *identity of source* is established.”

“The criminalist of the future may well be able to individualize the criminal directly through the hair he has dropped, the blood he has shed, or the semen he has deposited. All these things are unique to the individual, just as his fingerprints are unique to him.”

P.L. Kirk “Crime Investigation” 1974.

Relevant Forensic Science Question

Not: “Is this profile unique?” (it is).

Not: “Are these two profiles identical?” (they can’t be).

But: “Is there sufficient evidence to demonstrate that these two profiles originate from the same source?”

Statistical approach

C.R. Kingston, J Am Stat Assoc 60:70–80, 1028–1034, 1965.

Partial transfer evidence: physical material or impressions transferred from crime scene to perpetrator (or perpetrator's possessions), or vice versa.

PTE is characterized and assigned to an identity-set. Does a particular person (or their type) belong to the set? Does anyone else belong to the set?

“If it is *highly improbable* that another member could be found, we would be *reasonably sure* that the correct origin has been located. But if it is *quite probable* that other members exist, we would *not be so sure* that we have the correct origin.

Handwriting Evidence

Two signatures matched at 30 downstrokes. The probability of a coincidental match was estimated to be 1 in 5. The probability of 30 coincidences in one pair of signatures was “once in 2,666 millions of millions of millions.”

“This number far transcends human experience. So vast an improbability is practically an impossibility. Such evanescent shadows of probability cannot belong to actual life. They are unimaginably less than those least things which the law cares not for. The coincidence which has occurred here must have had its origin in an intention to produce it. It is utterly repugnant to sound reason to attribute this coincidence to any cause but design.”
(Mathematics professor Benjamin Pierce.)

Robinson v. Mandell, 20 F. Cas. 1027 (C.C.D. Mass. 1868)
(No. 11,959)

Bertillonage

Alphonse Bertillon (1853-1914), French anthropometrist. Son and brother of statisticians. Used 11 measurements:

1. Standing height
2. Arm reach
3. Sitting height
- 4.* Head length
- 5.* Head breadth
6. Length of right ear
7. Cheek width
- 8.* Length of left foot
- 9.* Length of left middle finger
10. Length of left little finger
11. Length of the left forearm and hand to the tip of extended middle finger

Bertillonage

The 11 characters had 3 subdivisions, plus 7 eye colors: $3^{11} \times 7 = 1,240,029$ possible profiles.

Searching was done on categories 4, 5, 8, 9. Each measurement divided into three subdivisions (large, medium, small) i.e. 81 possible combinations for a person. Filing cabinets with 81 drawers used for manual searching.

“Bertillon surmised that if a record could be made of eleven special measurements of the human body, then that record, when accompanied with a photograph, would establish unique, recordable, processable ID characteristics for every member of the human race.”

<http://www.deadmedia.org/notes/40/405.html>

Coincidental match

Two different men at Leavenworth in 1903 had very similar Bertillon dimensions (lengths in mm):

	Will West	William West
1	19.7	19.8
2	15.8	15.9
3	12.3	12.2
4	28.2	27.5
5	50.2	50.3
6	178.5	177.5
7	9.7	9.6
8	91.3	91.3
9	187.0	188.0
10	6.6	6.6
11	14.8	14.8

Fingerprints

“The arrangement of skin ridges is never duplicated in two persons.” J.C.A. Mayer, 1783.

J.E. Purkinje established categories of fingerprints in early 19th century.

W. Herschel, a British administrator, used fingerprints in India in 1850's.

H. Faulds, a British physician, used fingerprints in Japan.

Francis Galton wrote the book “Fingerprints” in 1892, and gave some probabilities for coincidental matches.

A good history is given in
“Probability Tales” Charles M. Grinstead, William P. Peterson,
J. Laurie Snell. American Mathematical Society, 2011.

Fingerprints

Galton considered that the chance that a random fingerprint would match a specified print was 2^{-36} . For a population of size 1.6×10^9 , the odds were about 1 to 40 that the print of any single finger would be exactly like the same finger of some other person.

“When two fingers of each of two persons are compared, and found to have the same minutiae, the improbability [of 1 in 2^{36}] becomes squared, and reaches a figure altogether beyond the range of the imagination.”

Francis Galton. 1892. Fingerprints.

Heritability of fingerprints

Galton looked at 105 sib-pairs:

		Second sib		
		Arches	Loops	Whorls
First sib	Arches	5	12	2
	Loops	4	42	15
	Whorls	1	14	10

Galton noticed that the diagonal counts of 5, 42, 10 are larger than those (2, 40, 6) expected if the sibs had independent fingerprints, but not as great as they could be (10, 61, 25). He did not have a formal statistical test available in 1892, but did conclude that there was an association.

He did not find racial differences.

Uniqueness of fingerprints

Probability arguments not now used. By 1924, textbooks would say “No two fingerprints are identical in pattern.” In 1939 J. Edgar Hoover wrote that fingerprints were “a certain and quick means of identification.”

Acceptance of uniqueness probably followed from “(i) striking visual appearance of fingerprints in court, (ii) a few dramatically successful cases, and (iii) a long period in which they were used without a single case being noted where two different individuals exhibited the same pattern.”

S.M. Stigler, 1995

Stigler anticipated the same growing acceptance of DNA profiles being unique.

DNA Evidence

DNA is extracted from bloodstains found at the scene of a crime. If the DNA profile does not match the profile found in a blood sample from a suspect, that person could not have been the contributor of the stains.

If the two DNA profiles do match, the suspect is not excluded as being a contributor.

How likely is the bloodstain to have that type if it came from the suspect?

How likely is the bloodstain to have that type if it came from some other person?

DNA Profiles - Serious Evidence

“For example, the three probes used in this case were pAC 255/locus D2S44, pAC 256/locus D17S79, and pAC 225/locus DXYS14. Probe pAC 255/locus D2S44 was tested on a data base comprised of 274 unrelated American blacks. Probe pAC 256/locus D17S79 was tested on 201 unrelated black individuals. The black data base for probe pAC 225/locus DXYS14 consisted of 220 unrelated individuals. These tests showed that the statistical likelihood of each probe’s matching DNA fragment appearing in the DNA of unrelated black individuals is 1/657, 1/1292, and 1/159, respectively. The overall likelihood of finding all three fragments in the same sample is determined by multiplying $1/657 \times 1/1292 \times 1/159$, resulting in an overall likelihood of 1/135,000,000. ”

SPENCER v. COM 384 S.E.2d 775 (Va. 1989)

Spencer became the first murderer in the United States to be convicted on the basis of DNA evidence.

CODIS Core Set of Genetic Markers

There is now a core set of 13 STR loci, on chromosomes 2, 3, 4, 5, 5, 7, 8, 11, 12, 13, 16, 18, 21. Each has at least 9 alleles, so there at least 45 possible genotypes per locus.

There are at least 6×10^{21} different 13-locus profiles, but there are only 6×10^9 people. Less than 1 profile in 1 trillion exists.

The question is not “Is the profile rare?” It is.

The question is “Is it possible that someone else has the profile?”

Statistical genetics approach

DNA profiles are genetic – they come with a structure and a history.

There is a dependence among profiles due to family relationships, and to a shared evolutionary history.

Presenting DNA Evidence

“The chance of a random person having the same DNA profile as found in the bloodstains on the rear gate at the Bundy residence is 1 in 57 billion.”

Marcia Clark, Prosecutor in O.J. Simpson case.

This number is an estimated probability, obtained by multiplying together “allele” frequencies, and refers to an infinite random mating population. It has nothing to do with the size of the world’s population.

Match Probabilities

Almost all the possible genotypes are not in the present population, and have expected frequencies that are very small. We don't *expect* that anyone living will have a particular profile – but of course we know that someone does if a defendant has that profile. The question is really whether we would see the profile in two people, given that we have already seen it in one person. This match (conditional) probability may be very low, but has nothing to do with the size of the population.

Extreme Numbers Were Not Allowed in Minnesota

“Schwartz contends that any probative value of statistical frequency evidence is outweighed by its prejudicial effect, as illustrated by the media exposure forensic DNA typing has received implying its infallibility. In dealing with complex technology, like DNA testing, we remain convinced that juries in criminal cases may give undue weight and deference to presented statistical evidence and are reluctant to take that risk.”

STATE v. SCHWARTZ No. C5-89-460.447 N.W.2d 422 (1989)

Refers to matching DNA profile with a frequency reported as 1 in 33 billion.

Quantifying Strength of Evidence

Before the evidence is introduced, there is some prior belief in guilt of the accused.

“Prior odds of guilt.”

This belief can be modified by the evidence, and becomes the “Posterior odds of guilt.”

Modification is by means of the “Likelihood Ratio.”

Posterior odds = Likelihood Ratio $\times$ Prior odds.

If the prior odds are only a million to one against guilt, but the DNA likelihood ratio is 57 billion, the posterior odds are 57 thousand to 1 for guilt.

Beyond Reasonable Doubt?

Samples of Judges, Jurors and Students were asked what probability corresponds to “beyond reasonable doubt.”

Probability	Judges	Jurors	Students
0–50%	0	5	3
50%	1	6	2
55%	2	2	1
60%	8	4	1
65%	2	1	0
70%	14	2	1
75%	23	2	1
80%	58	8	9
85%	21	2	3
90%	68	9	20
95%	44	3	17
100%	106	25	30
Total	347	69	88

Which Database?

Mr. Passino's paternal grandparents were Italian.

His maternal grandfather was Native American, his maternal grandmother was half French and half Native American.

"The defendant has proffered evidence that he belongs to an ethnic group whose genotypic frequencies may occur more frequently than the FBI's estimate."

Judge R.F. Kilburn.

Which Database?

The court's order excluding the DNA evidence expressly acknowledged that defendant contested only the FBI's probability calculations, which purportedly ruled out a coincidence in the match between defendant and semen found on the victim's body. The discussion was confined to the composition of the comparison database and the assumptions made about defendant's ancestry, two critical components of an accurate probability assessment. The court found the FBI probability analysis flawed, and held that "[b]ecause the probability estimates are such an integral part of the FBI's DNA profiling, the test results in this case must be suppressed," and added that results showing a match are not admissible without reliable statistics.

STATE V PASSINO.92-078; 161 Vt. 515; 640 A.2d 547

The Central Question

At a trial the trier of fact is asked to decide between two alternative explanations of some event.

The event E may be that the DNA profiles of the suspect S and the crime scene stain (left by the perpetrator P) match. Two possible explanations are

H_p : the crime scene stain is from the suspect

H_d : the crime scene stain is not from the suspect

The likelihood ratio is the probability of the evidence if H_p is true divided by the probability of the evidence if H_d is true.

$$L = \frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

Database Searches

If a profile has a frequency of P , then NP copies are expected in a database of size N . It has been suggested that the product NP , instead of P , is reported as a match probability. If a single hit is found in a database of one million profiles and the matching profile has an estimated frequency of 1 in a million, only the product of 1 would be reported.

The likelihood ratio for assessing the strength of the match still involves the probability of the match if the person identified is not the source of the evidence profile, and this has not been changed by the fact of a database search.

The strength of the evidence has actually been increased by the exclusion of 999,999 possible contributors (Balding and Donnelly, 1996).

People v. Johnson: The Evidence

Victim alleged rape by two men. DNA profiles of vaginal swab and of suspect Kelvin Johnson were determined by California Department of Justice laboratory (Keith Inman). The swab profile was a mixture: 3 or 4 alleles at each locus:

Locus	Swab	K. Johnson
D1S7	a,b,c	a,b
D2S44	a,b,c	a,b
D4S139	a,b,c,d	a,b
D5S110	a,b,c	a,b
D10S28	a,b,c	a,b
D17S79	a,b,c	a,b

People v. Johnson: Relatives

The fact that five of the six loci had only three alleles suggests allele-sharing by two contributors. This, in turn, suggests relatives.

Search of California offender database for relatives of Kelvin Johnson found his half-brother George.

Locus	Swab	Kelvin	George
D1S7	a,b,c	a,b	a,c
D2S44	a,b,c	a,b	a,c
D4S139	a,b,c,d	a,b	c,d
D5S110	a,b,c	a,b	a,c
D10S28	a,b,c	a,b	a,c
D17S79	a,b,c	a,b	a,c

People v. Johnson: Relatives

The prosecutor believed that George and Kelvin were the contributors. The defense did not offer an alternative set of contributors.

Needed to consider pairs of explanations that assumed Kelvin either was or was not a contributor, allowing for the two contributors to be related.

At the DNA admissibility hearing, the defense pointed out that Kelvin has another brother George. This man is Kelvin's full brother.

R v. Deen: Transposed Conditional

“The probability of this blood typing evidence if Mr. Deen is innocent is 1 in 700,000.

“Therefore the probability that Mr. Deen is innocent is 1 in 700,000.”

Use of Offender Database

Anthony Ruark thought he had got away with murder for 17 years until a combination of scientific expertise and a minor crime helped to bring him to justice.

The Forensic Science Service played a crucial part in helping the Metropolitan Police find the killer of Jacqueline Poole, who was raped and murdered in her own home in London in 1983.

Ruark, found guilty of murder at the Old Bailey today, was linked to the crime by his DNA profile. It was obtained after painstaking and unorthodox work by FSS scientists using the latest DNA profiling techniques.

When the resulting DNA profile was checked against the National DNA Database last April it matched with Ruark whose DNA profile had been placed on the database just three months previously for a minor theft.

Non-human Genetic Evidence

In October of 1994, a 32-year-old woman disappeared from her home in Richmond, Prince Edward Island, Canada. Three weeks later a man's leather jacket stained with blood was found in the woods 8 km from her home, and in May of 1995 her body was uncovered from a shallow grave. DNA testing established a match between the victim and the blood on the jacket. How could the jacket be linked to the suspect, the woman's estranged common-law husband? The Royal Canadian Mounted Police found white domestic cat hairs in the jacket lining, and also found that the suspect lived with his parents and Snowball, a white American shorthair cat. Could DNA tie Snowball to the jacket, and therefore tie the suspect to the crime?

Non-human Genetic Evidence

The Laboratory of Genomic Diversity of the National Cancer Institute, under the direction of Dr. Stephen O'Brien, is well-known for its feline studies and Dr. Marilyn Menotti-Raymond there has developed a panel of 400 simple tandem repeat (STR) genetic markers for the domestic cat. She also maintains a three-generation pedigree that can be used to establish genotyping protocols and establish linkage relationships for these markers. Upon being approached by the RCMP, Drs. Menotti-Raymond, O'Brien and Victor David, extracted DNA from the root of one the hairs off the jacket and from a blood sample from Snowball. They found a 10-locus STR match between these two DNA samples, and the suspect was convicted of second-degree murder in 1996.