Forensic Statistics

15th International Symposium on Human Identification

From the ground up…

John V. Planz, Ph.D.
UNT Health Science Center at Fort Worth

Why so much attention to statistics?

Exclusions don’t require numbers

Matches do require statistics

Problem of verbal expression of numbers

Transfer evidence

Laboratory result

1. Non-match - **exclusion**
2. Inconclusive- **no decision**
3. Match - **estimate frequency**

Statistical Analysis

Focus on the question being asked…

About "Q" sample

"K" matches "Q"

Who else could match "Q"

partial profile, mixtures

Match – estimate frequency of:

Match to forensic evidence

NOT suspect DNA profile

Who is in suspect population?

So, what are we really after?

Quantitative statement that expresses the rarity of the DNA profile

Estimate genotype frequency

1. Frequency at each locus

   Hardy-Weinberg Equilibrium

2. Frequency across all loci

   Linkage Equilibrium

Terminology

Genetic marker variant = allele

DNA profile = genotype

Database = table that provides frequency of alleles in a population

Where Do We Get These Numbers?

Coin Toss

Probability of heads with a penny 1/2

Probability of heads with a nickel 1/2

Probability of both coins as heads? tails? heads/tails?

Human Beings

23 different chromosomes

2 sets of chromosomes (from mom and dad) – two copies of each marker

Each genetic marker on different chromosome

Thus, each marker treated like coin toss – two possibilities

Anchor principle

- Analysis of genetic makeup in individuals is based on the *Genotype* at the locus being queried
- To remove "individual variation" so that we can focus on population-wide variation we must meld all the genotypes into a pool…separated as alleles

Alleles in populations –

The Hardy-Weinberg Theory

Basis: Allele frequencies are inherited in a Mendelian fashion and frequencies of occurrence follow a predictable pattern of probability

Hardy - Weinberg Equilibrium

A_1A_1	A_1A_2	A_2A_2
p_1^2	$2p_1p_2$	p_2^2

$\text{freq}(A_1) = p_1$

$\text{freq}(A_2) = p_2$

	A_1	A_2
A_1	p_1^2 A_1A_1	p_1p_2 A_1A_2
A_2	p_1p_2 A_1A_2	p_2^2 A_2A_2

$$(p_1 + p_2)^2 = p_1^2 + 2p_1p_2 + p_2^2$$

A Hardy-Weinberg Population

LARGE POPULATION

NO NATURAL SELECTION

NO MUTATION

NO IMMIGRATION / EMIGRATION

RANDOM MATING

A Hardy-Weinberg Population

We don't care these about criteria!

Only concerned about alleles…

Why "complicated" statistical tests?

- to estimate frequencies
- genotypes are rare
- many not seen in population sample
- HW equilibrium ?
- Linkage equilibrium ?
- need to use product rule

Estimate genotype frequency:

1. Frequency at each locus

2. Frequency across all loci

Product Rule

Product Rule

The frequency of a multi-locus STR profile is the product of the genotype frequencies at the individual loci

$$f\ \text{locus}_1 \times f\ \text{locus}_2 \times f\ \text{locus}_n = f_{\text{combined}}$$

Criteria for Use of Product Rule

Inheritance of alleles at one locus have no effect on alleles inherited at other loci

Linkage Equilibrium

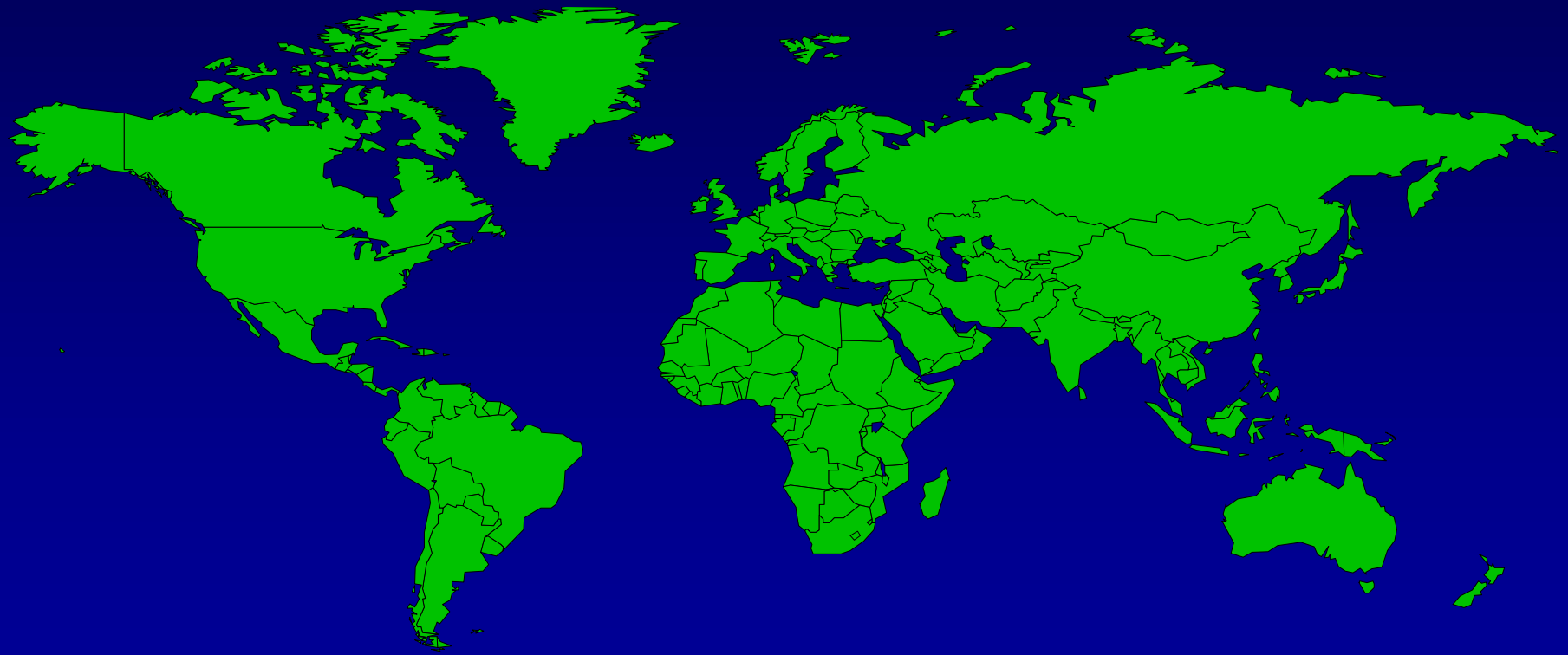
Condition in which genomes are composed of a random association of gametes

Linkage disequilibrium between two loci means that knowledge of a genotype at one locus gives at least a statistical clue as to the genotype at the other locus.

You can see that this is the common scenario we have discussed with regard to Y chromosome loci and also mtDNA sequence data or SNPs.

Linkage disequilibrium can exist either because of population substructure or because of physical linkage.

POPULATION DATA and Statistics



DNA databases are needed for placing statistical weight on DNA profiles

Population database

Look up how often each allele occurs at the locus in a population (or populations)

AKA looking up the “allele” frequency

TECHNICAL NOTE

Bruce Budowle,[1] Ph.D.; Tamyra R. Moretti,[1] Ph.D.; Anne L. Baumstark,[1] B.S.; Debra A. Defenbaugh,[1] B.S.; and Kathleen M. Keys,[1] B.S.

Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians*

REFERENCE: Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44(6):1277–1286.

markers are required, and all laboratories that contribute to the database should use the same genetic loci. Short tandem repeat (STR) loci are the most informative PCR-based genetic markers available to date for attempting to individualize biological material (2–5). The 13 STR loci CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818,

Bruce Budowle,[1] *Ph.D.; Brendan Shea,*[2] *M.S.; Stephen Niezgoda,*[2] *M.B.A.; and Ranajit Chakraborty,*[3] *Ph.D.*

CODIS STR Loci Data from 41 Sample Populations*

REFERENCE: Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. J Forensic Sci 2001:46;(3):453–489.

Materials and Methods

Samples

ProfIler Plus

Item	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
Q1	16,16	15,17	21,22	13,13	29,30	16,20	8,12	12,12	8,11

CoFIler

Item	D3S1358	D16S539	TH01	TPOX	CSF1P0	D7S820
Q1	16,16	10,12	8,9.3	9,10	12,12	8,11

D3S1358 = 16, 16 (homozygote)

Frequency of 16 allele = ??

TABLE 1—*Observed allele distributions (as %) for 13 STR loci in six population groups.*

D3S1358	African American (N=210)	Bahamian (N=157)	Jamaican (N=194)	Trinidad (N=80)	Caucasian (N=203)	Hispanic (N=209)
<12	0.476	0.000	0.000	0.000	0.000	0.000
12	0.238	0.000	0.515	0.000	0.000	0.000
13	1.190	0.000	1.546	0.000	0.246	0.239
14	12.143	7.643	6.701	5.625	14.039	7.895
15	29.048	31.847	33.763	31.250	24.631	42.584
15.2	0.000	0.318	0.258	0.000	0.000	0.000
16	30.714	33.758	30.670	31.875	23.153	26.555
17	20.000	19.745	21.134	20.000	21.182	12.679
18	5.476	6.369	4.639	11.250	16.256	8.373
19	0.476	0.318	0.773	0.000	0.493	1.435
>19	0.238	0.000	0.000	0.000	0.000	0.239
Homozygosity (Obs.)	21.4%	25.5%	27.8%	16.3%	19.2%	26.3%
Homozygosity (Exp.)	23.5%	26.2%	25.8%	25.0%	20.3%	28.0%
(p)	0.482	0.838	0.513	0.070	0.691	0.595
Exact Test	0.797	0.758	0.270	0.222	0.084	0.333
PD	0.903	0.885	0.886	0.878	0.920	0.880
PE	0.543	0.499	0.508	0.511	0.589	0.492

D3S1358 = 16, 16 (homozygote)

Frequency of 16 allele = 0.3071

When same allele:

Frequency = genotype frequency (p^2)
(for now!)

Genotype freq = 0.3071 x 0.3071 = 0.0943

ProfIler Plus

Item	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
Q1	16,16	15,17	21,22	13,13	29,30	16,20	8,12	12,12	8,11

CoFIler

Item	D3S1358	D16S539	TH01	TPOX	CSF1P0	D7S820
Q1	16,16	10,12	8,9.3	9,10	12,12	8,11

VWA = 15, 17 (heterozygote)

Frequency of 15 allele = ??

Frequency of 17 allele = ??

VWA	African American (N=180)	Bahamian (N=162)	Jamaican (N=244)	Trinidad (N=85)	Caucasian (N=196)	Hispanic (N=203)
11	0.278	0.926	0.410	0.588	0.000	0.246
13	0.556	2.778	0.820	0.588	0.510	0.000
14	6.667	6.173	7.377	8.824	10.204	6.158
15	23.611	15.123	22.746	14.118	11.224	7.635
16	26.944	26.235	29.098	29.412	20.153	35.961
17	18.333	20.679	18.238	26.471	26.276	22.167
18	13.611	18.210	13.115	13.529	22.194	19.458
19	7.222	7.099	5.328	4.706	8.418	7.143
20	2.778	2.778	2.254	1.765	1.020	1.232
21	0.000	0.000	0.615	0.000	0.000	0.000
Homozygosity (Obs.)	11.7%	17.3%	20.9%	20.0%	22.4%	24.6%
Homozygosity (Exp.)	18.9%	17.6%	19.4%	20.0%	18.7%	22.9%
(p)	0.014	0.928	0.557	0.991	0.179	0.564
Exact Test	0.328	0.790	0.655	0.229	0.063	0.928
PD	0.926	0.942	0.933	0.917	0.932	0.914
PE	0.624	0.648	0.617	0.602	0.625	0.563

VWA = 15, 17 (heterozygote)

Frequency of 15 allele = 0.2361

Frequency of 17 allele = 0.1833

When heterozygous:

Frequency = 2 X allele 1 freq X allele 2 freq

(*2pq*)

Genotype freq = 2 x 0.2361 x 0.18331 = 0.0866

Overall profile frequency =

Frequency D3S1358 X Frequency vWA

0.0943 x 0.0866 = 0.00817

This is basically what Popstats does for us in it's simplest task

What tools were used?

Population database

Some math equations

Steps – Single Sample Target Profile

- enter alleles of target profile
- look up allele frequencies at all loci for all populations
- determine if homozygous or heterozygous at each locus
- calculate genotype frequency at each locus
- calculate profile frequency with product rule

But this doesn't address all of the issues!

What if…

We encounter alleles not represented in the population database…

…or alleles that are extremely rare in the database???

Ideally, we should know the frequency of every genotype that might be encountered

Do we?

How many genotypes at a locus?

k alleles, so there are:

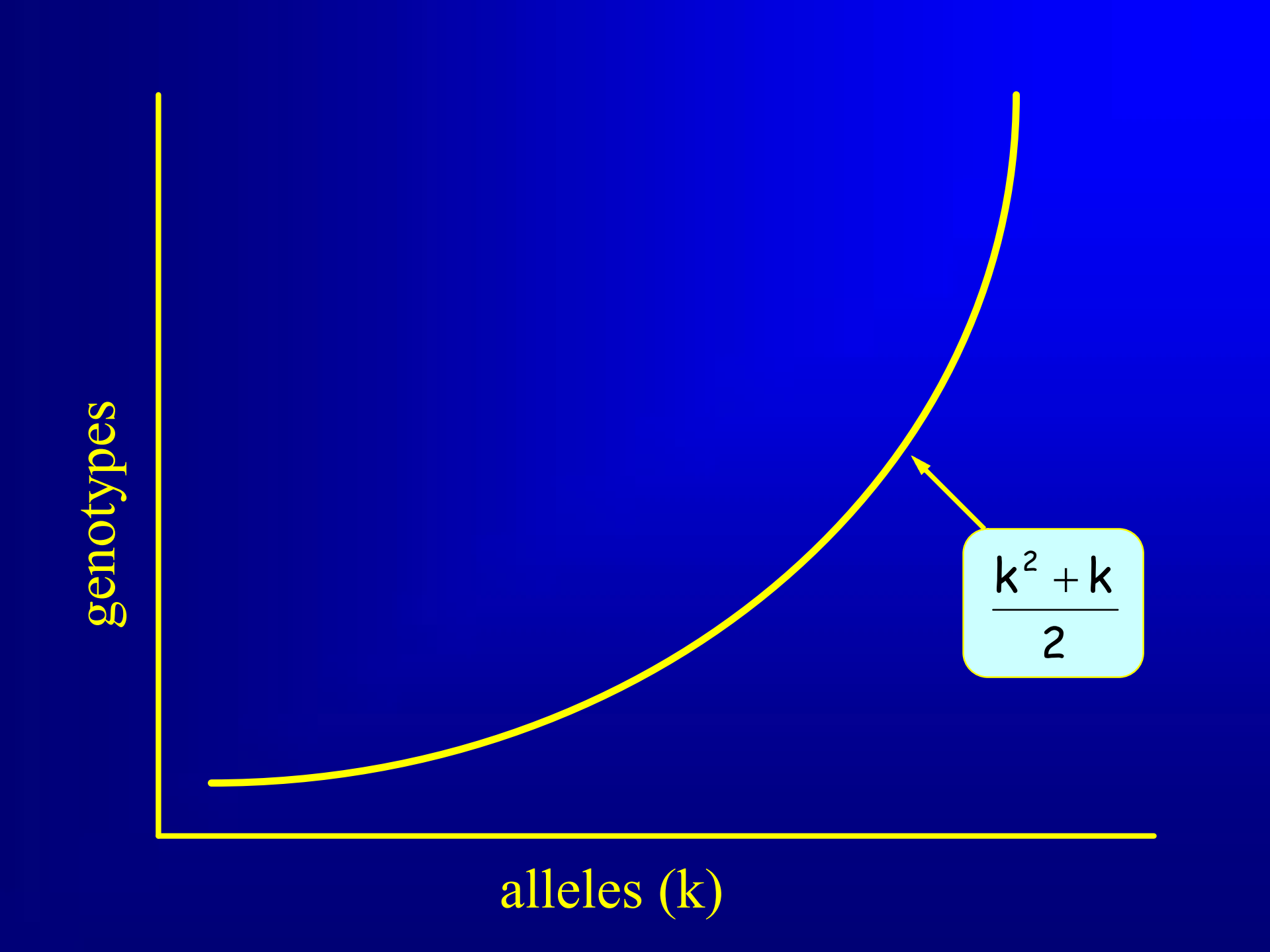
k homozygotes

k x (k-1)/2 heterozygotes

k + (k x (k-1))/2 = $k(k+1)/2$

Number of genotypes $= \frac{k(k+1)}{2} = \frac{k^2 + k}{2}$

k	$\frac{k^2 + k}{2}$
2	3
3	6
4	10
5	15
6	21


genotypes
alleles (k)
$\frac{k^2 + k}{2}$

Caucasian Database for Locus vWA

$N = 196$ Individuals

	11	12	13	14	15	16	17	18	19	20	21
11											
12											
13						1			1		
14				1	4	10	10	10	4		
15					3	7	14	8	4	1	
16						11	27	7	3	1	
17							11	23	8	1	
18								16	6		
19									3	1	
20											
21											

66 Possible Genotypes $(N)(N+1)/2$

27 Genotypes Seen In Caucasians

Profiler Plus - 9 loci:

locus	alleles	genotypes
D3S1358	8	36
vWa	8	36
FGA	12	78
D8S1179	9	45
D21S11	14	105
D18S51	12	78
D5S818	6	21
D13S317	7	28
D7S820	8	36

Number of genotypes detectable is

$$36 \times 36 \times 78 \times 45 \times 105 \times 78 \times 21 \times 28 \times 36 =$$

$$\mathbf{7.89\ X\ 10^{14}}$$

But you will never see all of them!!!

Discriminatory Power

number of alleles

"evenness" of frequencies

heterozygosity is a measure of discrimination

$$\text{Homozygosity} = \sum_{i=1,k} p_i^2$$

$$\text{Heterozygosity} = 1 - \text{Homozygosity}$$

$$\text{Heterozygosity} = 1 - \sum_{i=1,k} p_i^2$$

Example: 4 alleles

pop # 1	pop # 2	pop # 3
.25	.20	.10
.25	.20	.10
.25	.20	.10
.25	.40	.70
het = ?	?	?

$$\text{Heterozygosity} = 1 - \sum_{i=1,k} p_i^2$$

pop# 1: $1 - [\,(.25)^2+(.25)^2+(.25)^2+(.25)^2\,] = .75$

pop# 2: $1 - [\,(.20)^2+(.20)^2+(.20)^2+(.40)^2\,] = .72$

pop# 3: $1 - [\,(.10)^2+(.10)^2+(.10)^2+(.70)^2\,] = .48$

Example: 4 alleles

	pop # 1	pop # 2	pop # 3
	.25	.20	.10
	.25	.20	.10
	.25	.20	.10
	.25	.40	.70
het =	.75	.72	.48

Well…unfortunately, the Power of Discrimination and Power of Exclusion are a bit more involved.

Power of Discrimination is related to the what has been called the random match probability…

… the probability that two randomly selected individuals have identical phenotypes/genotypes by chance alone

Here, however, *p* = the phenotype frequency

Now, Power of Discrimination is simply:

$$PD = 1 - P_i$$ **for one locus, or**

$$PD = (P_1 P_2 P_3 \ldots P_n)$$ **for a panel of loci**

While we're at it lets cover Power of Exclusion

Where the random match probability is the sum of the squares of the observed phenotype/genotype frequencies in a database,

The Power of Exclusion of a genetic locus is based on the 1 – the sum of squares of all the expected phenotypes/genotypes!

These measures tell us two things about our markers and databases:

Power of Discrimination
– how powerful our loci are at individualizing

Power of Exclusion
– how powerful our marker panel is at excluding particular genotypes

VWA	African American (N=180)	Bahamian (N=162)	Jamaican (N=244)	Trinidad (N=85)	Caucasian (N=196)	Hispanic (N=203)
11	0.278	0.926	0.410	0.588	0.000	0.246
13	0.556	2.778	0.820	0.588	0.510	0.000
14	6.667	6.173	7.377	8.824	10.204	6.158
15	23.611	15.123	22.746	14.118	11.224	7.635
16	26.944	26.235	29.098	29.412	20.153	35.961
17	18.333	20.679	18.238	26.471	26.276	22.167
18	13.611	18.210	13.115	13.529	22.194	19.458
19	7.222	7.099	5.328	4.706	8.418	7.143
20	2.778	2.778	2.254	1.765	1.020	1.232
21	0.000	0.000	0.615	0.000	0.000	0.000
Homozygosity (Obs.)	11.7%	17.3%	20.9%	20.0%	22.4%	24.6%
Homozygosity (Exp.)	18.9%	17.6%	19.4%	20.0%	18.7%	22.9%
(p)	0.014	0.928	0.557	0.991	0.179	0.564
Exact Test	0.328	0.790	0.655	0.229	0.063	0.928
PD	0.926	0.942	0.933	0.917	0.932	0.914
PE	0.624	0.648	0.617	0.602	0.625	0.563

So, what we need to consider now is "How good are our databases?"

We know we don't have full representation of all of the genotypes possible…

We must consider then that we don't have an accurate representation of some of the rarer alleles either!

Minimum allele frequency

The first NRC report proposed a minimum allele frequency based on NO empirical data and without any statistical basis!

10 % or 0.1

What…you are surprised??

Ceiling Principle

Minimum allele frequency

Weir, B.S. 1992. minfreq = $1 - \alpha^{1/2N}$

Budowle, B., K. Monson, R. Chakraborty, 1996. minfreq = $1 - [\ 1 - (\ 1 - \alpha)^{1/C}\ ]^{1/2N}$

NRC II, 1996, pg. 148. minfreq = 5/2N

Minimum allele frequency

This method requires a minimum of 5 copies of an allele before the allele frequency can be used for calculation of genotype frequency

5

Total number of alleles at locus

For the 13 allele at vWA:
Actual Freq = 2 / 392 = 0.0051
Minimal Freq = 5 / 392 = 0.0128

Conservatism & also addresses some substructure effects

This estimate is strictly driven by database size:

N	min allele freq
100	2.50 % (0.025)
150	1.67 % (0.0167)
200	1.25 % (0.0125)
250	1.00 % (0.01)
300	0.83 % (0.0083)

Where N is the number of individuals in database

So the only real thing left to consider regarding the NRC concerns is population subdivision.

Population Structure

- **Racial, ethnic subgroups**
- **Excess of homozygotes**
- **What is "theta" ⑧**
- **Why modify just homozygous calculation?**
- **NRC Formula 4.1 vs 4.4 vs 4.10**

Population Subdivision

We've always surmised…

Racial / ethnic group composed of distinct sub-groups within the sample population

Only a concern if sub-groups differ substantially at allele frequencies at the loci

Human Genetic Variation

between populations within racial groups …

between racial groups ……

within populations within racial groups ……

- Barbujani, Magagni, Minch, Cavalli-Sforza. 1997. An apportionment of human DNA diversity. *PNAS* 94:4516-4519.

Problems created by population subdivision

Genotype frequencies calculated from population average allele frequencies could lead to:

Wrong estimates!

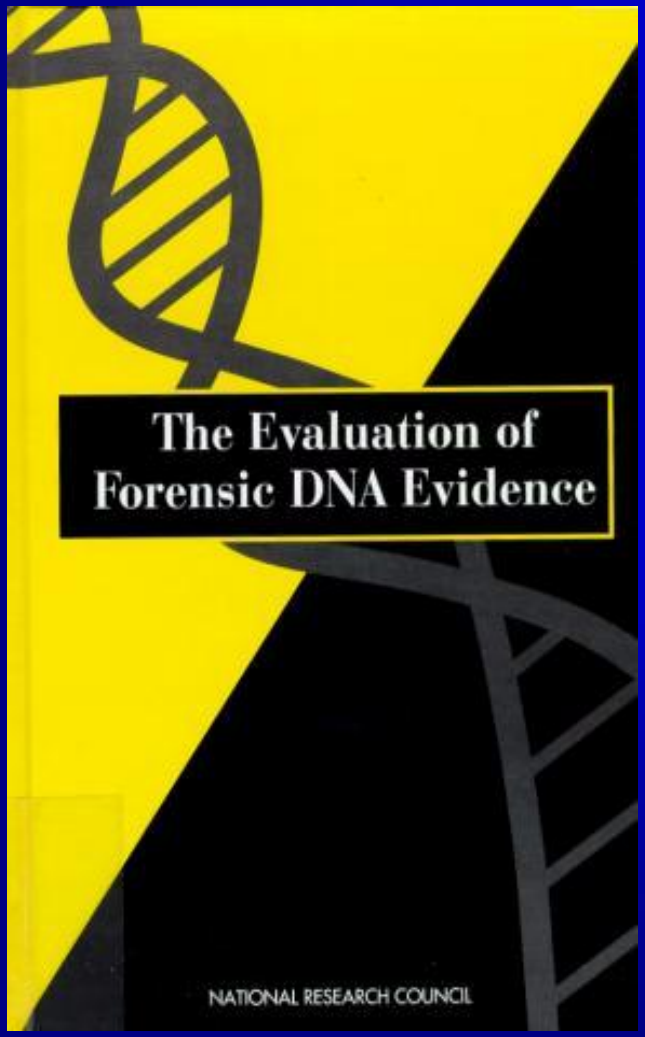
Employ a Theta (θ) Correction

θ is used as a measure of the effects of population subdivision (inbreeding)

How many Great, Great, Great, Great, Great, Great, Great… Grandparents do you have?

⑧ is equivalent to F_{ST} and G_{ST}

National Research Council Report II

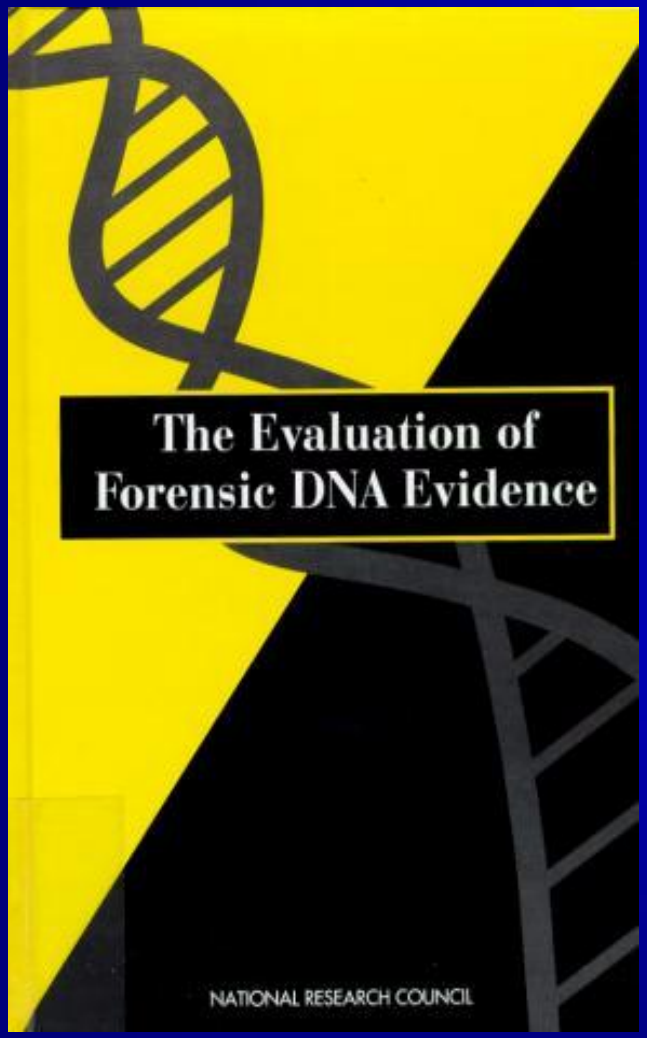


National Academy
of Sciences

Data support the recommendation that F_{ST} of 0.01 is conservative

Issued in May 1996

National Research Council Report II



The significance of this F_{ST} is

That some Hardy-Weinberg expectations do not have to be met

TABLE 6—*F_{ST} values for the thirteen CODIS core STR loci.*

Locus	African American	Caucasian	Hispanic	Asian	Native American
CSF1PO	−0.0009	−0.0007	−0.0003	−0.0012	0.0244
D3S1358	−0.0005	−0.0009	0.0014	0.0035	0.0764
D5S818	0.0010	−0.0001	0.0010	0.0028	0.0656
D7S820	0.0000	−0.0005	0.0010	0.0039	0.0201
D8S1179	−0.0001	0.0000	0.0005	0.0025	0.0125
D13S317	0.0029	−0.0008	0.0047	0.0071	0.0157
D16S539	−0.0013	−0.0005	0.0067	0.0017	0.0132
D18S51	0.0012	0.0001	0.0011	0.0046	0.0268
D21S11	0.0005	0.0008	0.0013	0.0056	0.0371
FGA	0.0004	−0.0004	0.0008	0.0029	0.0168
TH01	0.0015	−0.0012	0.0041	0.0058	0.0356
TPOX	0.0021	−0.0015	0.0024	0.0100	0.0164
vWA	0.0011	−0.0011	0.0029	0.0027	0.0172
F_{ST} over all loci	0.0006	−0.0005	0.0021	0.0039	0.0282

Modifying the pro

Use correction factor f

P^2 ☞ $P(1-P)$⑧

⑧☞ 0.01

Intermediate to the F_{ST} that you would find in populations with 1st and 2nd cousin matings

0.03 for Native populations

use $2p_ip_j$ for heterozygotes (ie: no correction)

Really, this is more than ten fold more conservative

Modifying the product rule

- Formula 4.1 - HW
- Formula 4.4 - Simple subdivision
- Formula 4.10 - assumption of population
- Conditional vs Unconditional Probability

HWE: p^2

NRC II, 4.4a: $p^2 + p(1 - p)\theta$

NRC II, 4.10a: $\dfrac{[2\theta + (1 - \theta)p_i][3\theta + (1 - \theta)p_i]}{(1 + \theta)(1 + 2\theta)}$

This last formula addresses a conditional probability of the suspect genotype, given that of the perpetrator, $P(A_iA_i \mid A_iA_i)$, considering the person contributing the evidence and the suspect are from the same subgroup.

When and why should we consider this??

Takes into account the assumption that the person contributing the evidence and the suspect are from the same subgroup

What it gives us is a conditional probability of the suspect genotype given that we have already seen that genotype in the perpetrator.

Example… use if the suspect and all possible perpetrators are from the same small isolated town i.e. religious sects, native communities

Although we **CAN** correct the heterozygote genotype estimate…it is **not** generally necessary.

HWE: $2pq$

NRC II, 4.4a: $2pq(1 - \theta)$

NRC II, 4.10b:

$$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

$$P(A_iA_j \mid A_iA_j)$$

So,

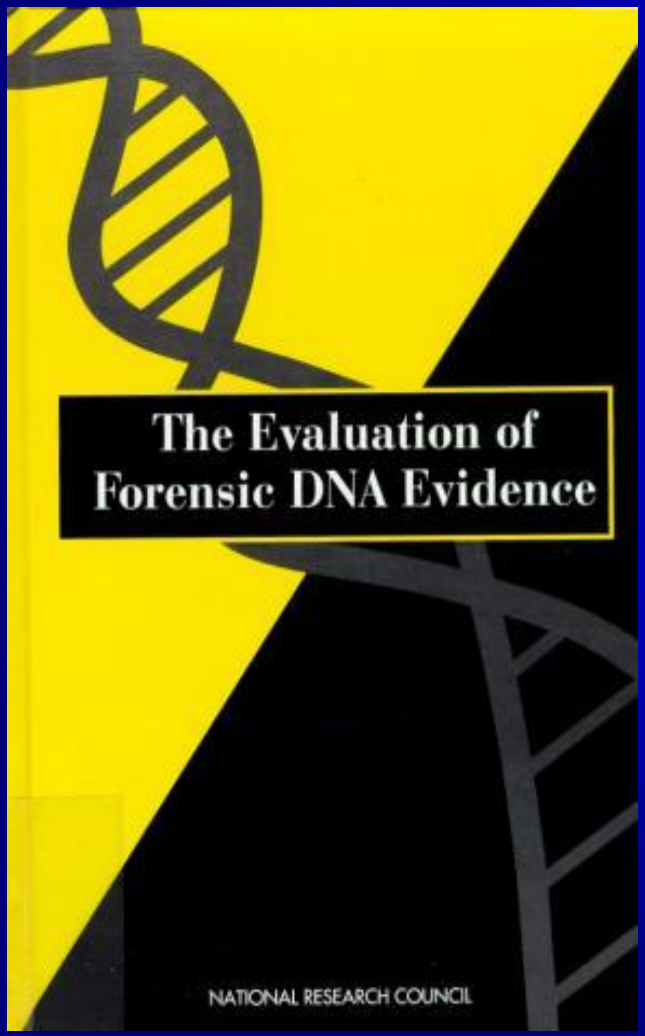
We've calculated these nice frequency estimates that we desired…

What do we do with them???

Well,

We report them of course!

But we should consider what we are reporting and the information we are conveying in our "statistics"



NRC II May 1996

"...that profile might be said to be unique if it is so rare that it becomes unreasonable to suppose that a second person in the population might have the same profile."

Source Attribution

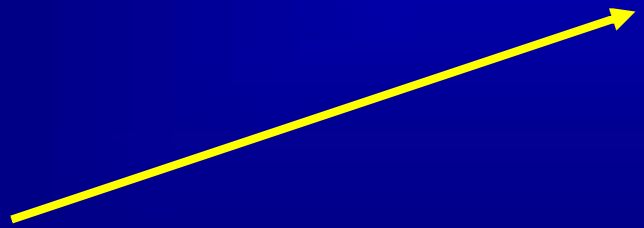
Hot topic for statistical debate

With the current panel of genetic markers available to forensic testing, it is not uncommon for the reciprocal of the random match probability determined for a genetic profile to exceed the worlds population several fold.

So, how do you want to express this fact in your reports and testimony?

What do these *numbers* mean to you?

the prosecutor?
the defense?
the judge?
the jury?



This is what really matters!!!

Big Number Names:

1,000,000	million
1,000,000,000	billion
1,000,000,000,000	trillion
1×10^{15}	quadrillion
1×10^{18}	quintillion
1×10^{21}	sextillion
1×10^{24}	septillion
1×10^{27}	octillion
1×10^{30}	nonillion
1×10^{33}	decillion

Even Bigger Number Names:

1×10^{36}	**undecillion**
1×10^{39}	**duodecillion**
1×10^{42}	**tredecillion**
1×10^{45}	**quattordecillion**
1×10^{48}	**quindecillion**
1×10^{51}	**sexdecillion**
1×10^{54}	**septendecillion**
1×10^{57}	**octodecillion**
1×10^{60}	**novemdecillion**
1×10^{63}	**vigintillion**

To address uniqueness we are back to the same old question…
population sample size

Here the population size differs from what we discussed when calculating allele frequencies…

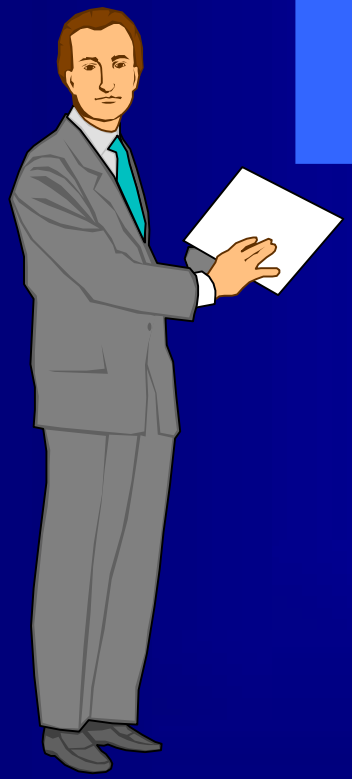
The relevant population is at issue here

Define the Question

(or at least make sure you know what question you are answering)

Define the Question

Estimates of the Rarity of
a DNA Profile:



1 in 130 million

Based on
unrelated individuals

1 in 128

Based on
brothers

Uniqueness / Source Attribution
Webster's Definitions

only one

Unusual

Some [circumstance] that is the only one of it kind

Source Attribution

Attribution evaluated within context of case

Rarely is the world's population the appropriate context

Thus, a circumstance that is the only one of its kind is appropriate context

Uniqueness ?

A profile that exists in one person and no other (excluding identical twins)

Context?

- **Population of the world…maybe**
- **Population of the US….there is a thought!**
- **Population with access to a crime scene…**

A profile that exists in one person and no other (excluding identical twins)

Actually we are interested in source attribution, not whether the profile is unique in the world

Is it reasonable to consider the profile to be so rare that one can opine about the source of the evidence?

Let the RMP of a given evidentiary profile X be p_x
(Calculate using NRC II Report Recommendations)

Then $(1-p_x)^N$

is the probability of not observing the profile
in a population of N unrelated individuals

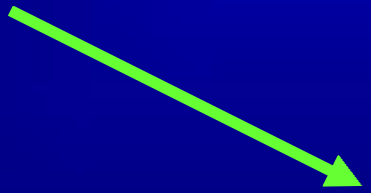
This probability should be greater than
or equal to a 1- α confidence level

$$(1-p_x)^N \geq 1-\alpha$$

$$p_x \leq 1-(1-\alpha)^{1/N}$$

Source Attribution

- Specify (1- α) confidence level of 95% or 99% (uses an α of 0.05 or 0.01, respectively)
- Determine RMP threshold to assert with a specific degree of confidence that the particular evidence profile is unique with a population of N unrelated individuals



What population????

Source Attribution Values

Calculate p for major population groups

θ = 0.01or 0.03

Take the most common value for p
Increase p by factor of 10
Determine if $p \leq 1 - (1 - a)^{1/N}$

What N ??

The standard basis that is used here in the US is an estimate of US population of approximately 260 million people

So, taking this and if we accept an α of 0.01 (99% confidence level) with

$$p_x \leq 1-(1-\alpha)^{1/N}$$

A random match probability less than 3.9×10^{-11} would convey at least 99% confidence that the evidentiary profile is unique in the population

RMP thresholds for source attribution at various population sizes and confidence levels

SAMPLE SIZE N	CONFIDENCE LEVELS 0.90	0.95	0.99	0.999
2	5.1×10^{-2}	2.5×10^{-2}	5.0×10^{-3}	5.0×10^{-4}
3	3.5×10^{-2}	1.7×10^{-2}	3.3×10^{-3}	3.3×10^{-4}
4	2.6×10^{-2}	1.3×10^{-2}	2.5×10^{-3}	2.5×10^{-4}
5	2.1×10^{-2}	1.0×10^{-2}	2.0×10^{-3}	2.0×10^{-4}
6	1.7×10^{-2}	8.5×10^{-3}	1.7×10^{-3}	1.7×10^{-4}
7	1.5×10^{-2}	7.3×10^{-3}	1.4×10^{-3}	1.4×10^{-4}
8	1.3×10^{-2}	6.4×10^{-3}	1.3×10^{-3}	1.3×10^{-4}
9	1.2×10^{-2}	5.7×10^{-3}	1.1×10^{-3}	1.1×10^{-4}
10	1.1×10^{-2}	5.1×10^{-3}	1.0×10^{-3}	1.0×10^{-4}
25	4.2×10^{-3}	2.1×10^{-3}	4.0×10^{-4}	4.0×10^{-5}
50	2.1×10^{-3}	1.0×10^{-3}	2.0×10^{-4}	2.0×10^{-5}
100	1.1×10^{-3}	5.1×10^{-4}	1.0×10^{-4}	1.0×10^{-5}
1×10^{3}	1.1×10^{-4}	5.1×10^{-5}	1.0×10^{-5}	1.0×10^{-6}
1×10^{5}	1.1×10^{-6}	5.1×10^{-7}	1.0×10^{-7}	1.0×10^{-8}
1×10^{6}	1.1×10^{-7}	5.1×10^{-8}	1.0×10^{-8}	1.0×10^{-9}
1×10^{7}	1.1×10^{-8}	5.1×10^{-9}	1.0×10^{-9}	1.0×10^{-10}
5×10^{7}	2.1×10^{-9}	1.0×10^{-9}	2.0×10^{-10}	2.0×10^{-11}
2.6×10^{8}	4.1×10^{-10}	2.0×10^{-10}	3.9×10^{-11}	3.9×10^{-12}
1×10^{9}	1.1×10^{-10}	5.1×10^{-11}	1.0×10^{-11}	1.0×10^{-12}
5×10^{9}	2.1×10^{-11}	1.0×10^{-11}	2.0×10^{-12}	2.0×10^{-13}

So with typical results obtained for a ⑧ = 0.01

Locus	CAU	BLK	SEH	SWH
D3S1358	4.6529E-02	4.1600E-02	2.7701E-02	1.7185E-02
VWA	4.1106E-02	3.5938E-02	3.6987E-02	4.4303E-02
FGA	4.6005E-02	1.8050E-02	3.5152E-02	2.0049E-02
D8S1179	7.4442E-02	9.5057E-02	8.5725E-02	7.5293E-02
D21S11	3.6039E-02	2.8637E-02	4.0358E-02	3.5239E-02
D18S51	2.3427E-02	1.5228E-02	1.4971E-02	1.0950E-02
D5S818	2.9041E-01	1.8569E-01	2.4943E-01	2.4480E-01
D13S317	6.1431E-02	3.5080E-02	5.2533E-02	2.8834E-02
D7S820	6.5690E-02	7.7793E-02	6.4360E-02	5.6800E-02
CSF1PO	4.6424E-02	3.2880E-02	4.9027E-02	5.0685E-02
TPOX	1.3412E-01	1.3395E-01	8.4350E-02	3.7185E-02
TH01	1.0530E-01	9.2329E-02	1.1869E-01	1.6298E-01
D16S539	1.1082E-01	[illegible]	[illegible]	5.9165E-02

	CAU	BLK	SEH	SWH
Total	4.709E-16	1.420E-17	6.379E-17	3.231E-18

3.9×10^{-11}

Profile frequency is less than 99% threshold

Source Attribution

- Method is simple
- Conservative because N is so large (260,000,000)
- If N = 260,000,000, then RMP threshold is 3.9×10^{-11}
- Most of the time the RMP is far less, so confidence is greater than 0.99

Source Attribution

- N can be configured to context of the case
- Two individuals to entire town, state, or whatever
- Laboratory policy to set N

"To a reasonable degree of scientific certainty, _____ is the source of the DNA in specimen Q2."

"I have a high degree of confidence, _____ is the source of the DNA in specimen Q2."

Assignment of DNA origin as a frequency (random match probability)

We are not stating that _____ is the only person to possess that profile. We are stating that we would not expect to find it in a population of N individuals.

Presenting Statistics

- Keep it simple
- Make it vivid
- Understand the data
- Know your audience
- Credibility

Random match probability is NOT

Chance that someone else is guilty

Chance that someone else left the bloodstain

Chance of defendant not being guilty

"We are neither hunters nor gatherers. We are statisticians."