NIST
National Institute of Standards and Technology
U.S. Department of Commerce

FORENSIC SCIENCE ERROR MANAGEMENT
INTERNATIONAL FORENSICS SYMPOSIUM
JULY 20-24, 2015 • WASHINGTON, DC

http://www.nist.gov/director/international_forensics_home.cfm

DNA Error Rates

John M. Butler

NIST Fellow & Special Assistant to the Director for Forensic Science

National Institute of Standards and Technology

July 23, 2015

Disclaimers

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial equipment, instru
materials are identified in orde
experimental procedures
In no case does s
recommend
Institu
i
eq
ava

I am from the U.S. government and I am here to help you…

Full Title of My Presentation

To Err is Human, but **How Might We Measure Error Rates in Forensic DNA Testing** and What Would These Error Rates Really Mean?

Daubert 1993 Ruling Mentions Error Rates

https://en.wikipedia.org/wiki/Daubert_standard

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993) is a United States Supreme Court case determining the standard for admitting expert testimony in federal courts. The *Daubert* Court held that the enactment of the Federal Rules of Evidence implicitly overturned the Frye standard; the standard that the Court articulated is referred to as the Daubert standard.

…

Factors relevant: The Court defined "scientific methodology" as the process of formulating hypotheses and then conducting experiments to prove or falsify the hypothesis, and provided a nondispositive, nonexclusive, "flexible" set of "general observations" (i.e. not a "test") that it considered relevant for establishing the "validity" of scientific testimony:

1. **Empirical testing**: whether the theory or technique is falsifiable, refutable, and/or testable.
2. Whether it has been **subjected to peer review and publication**.
3. **The known or potential error rate.**
4. The **existence and maintenance of standards and controls** concerning its operation.
5. The degree to which the theory and technique is **generally accepted by a relevant scientific community**.

Daubert Ruling on Error Rates

SUPREME COURT OF THE UNITED STATES

No. 92-102

WILLIAM DAUBERT, et ux., etc., et al., PETITIONERS *v.* MERRELL DOW PHARMACEUTICALS, INC.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

[June 28, 1993]

Justice Blackmun delivered the opinion of the Court.

Justice Harry Blackmun wrote in the majority opinion:

"Additionally, in the case of a particular scientific technique, **the court ordinarily should consider the known or potential rate of error**, see, *e. g.*, *United States* v. *Smith*, 869 F. 2d 348, 353-354 (CA7 1989) (surveying studies of the error rate of spectrographic voice identification technique) …"

https://www.law.cornell.edu/supct/html/92-102.ZO.html

Calls for Using Proficiency Test Data to Estimate Error Rates

Professor Jay Koehler, Northwestern University School of Law, has been the most vocal advocate for this topic

2011

Proficiency Tests to Estimate Error Rates in the Forensic Sciences

Jonathan Koehler
Northwestern University School of Law, jay.koel

Law, Probability and Risk (2013) **12**, 89–98 doi:10.1093/lpr/mgs013
Advance Access publication on September 3, 2012

Proficiency tests to estimate error rates in the forensic sciences

JONATHAN J. KOEHLER*
Northwestern University School of Law, Chicago, IL 60611, USA

[Received on 19 July 2011; accepted on 14 May 2012]

Proficiency Testing Data Does Exist – **but what does it really mean?**

Collaborative Testing Services, Inc
FORENSIC TESTING PROGRAM

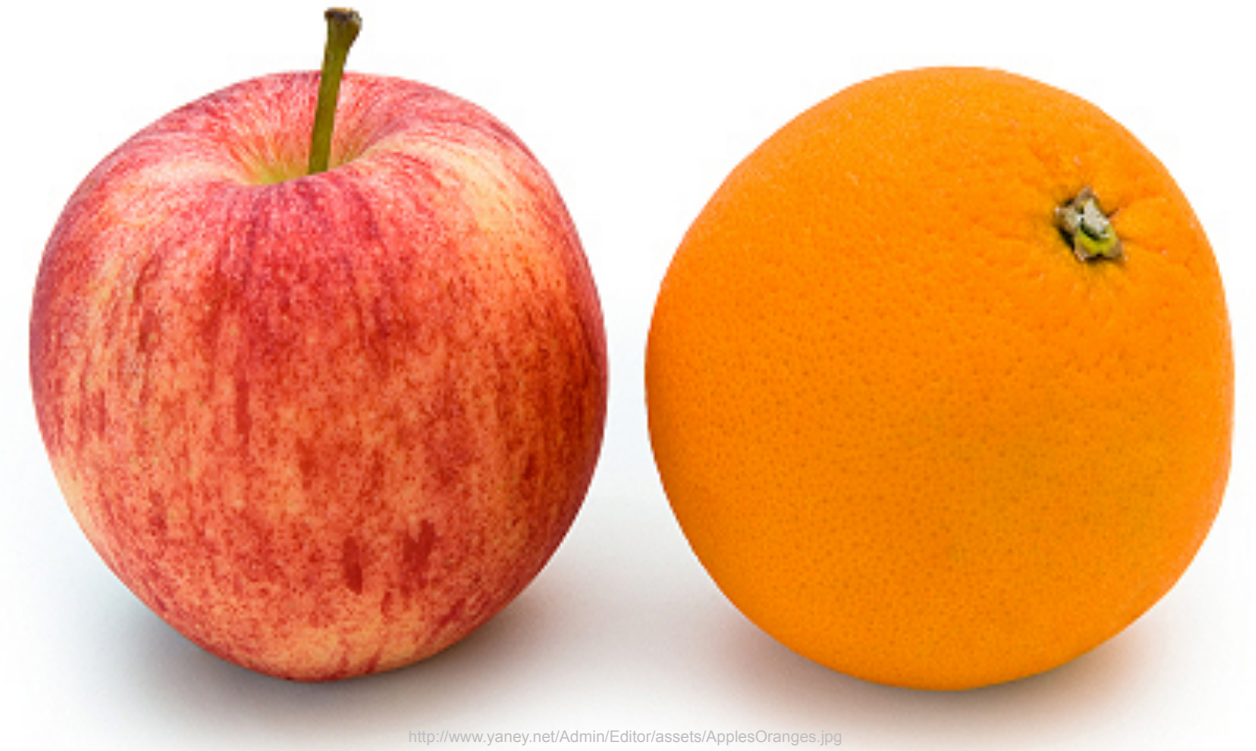
DNA - Mixture Test No. 15-581 Summary Report

This proficiency test was sent to 153 participants. Each participant received a sample pack consisting of two known bloodstains and two questioned stains which they were requested to analyze using their existing protocols. Data were returned from 145 participants (95% response rate) and are compiled into the following tables:

http://www.ctsforensics.com/assets/news/3581_Web1.pdf

"This report contains the data received from the participants in this test. Since these participants are located in many countries around the world, and **it is their option how the samples are to be used (e.g., training exercise, known or blind proficiency testing, research and development of new techniques, etc.)**, **the results compiled in the Summary Report are not intended to be an overview of the quality of work performed in the profession** and cannot be interpreted as such. The Summary Comments are included for the benefit of participants to assist with maintaining or enhancing the quality of their results. **These comments are not intended to reflect the general state of the art within the profession.**"

Proficiency Test “Error Rates” and Casework Case-Specific Error Rates



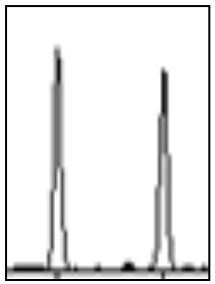
http://www.yaney.net/Admin/Editor/assets/ApplesOranges.jpg

Math Analogy to DNA Evidence

$$\int_{x=0}^{\infty} f(x)\,dx$$

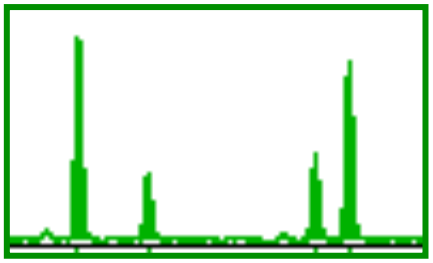
$2 + 2 = 4$ $2x^2 + x = 10$

Basic Arithmetic



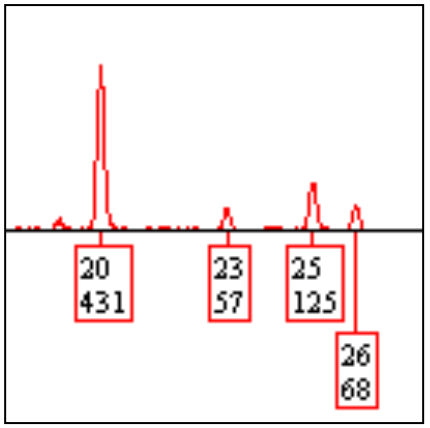
Single-Source DNA Profile
(DNA databasing)

Algebra



Sexual Assault Evidence
(2-person mixture with high-levels of DNA)

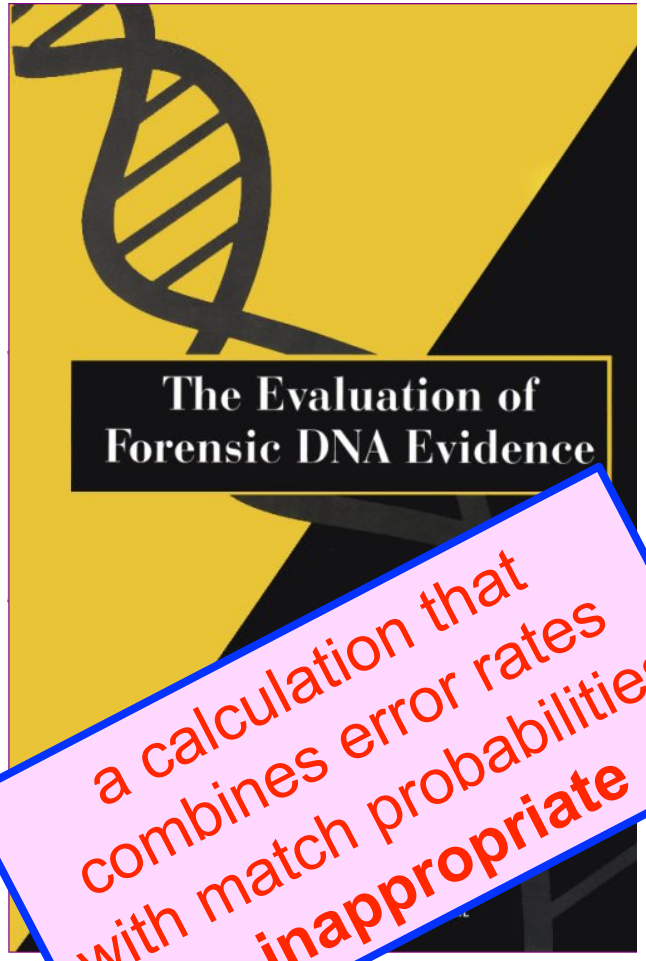
Calculus



Touch Evidence
(>2-person, low-level, complex mixtures perhaps involving relatives)

http://www.cstl.nist.gov/strbase/pub_pres/Butler-DNA-interpretation-AAFS2015.pdf

The Second National Research Council Report (**NRC II**) Published in 1996



a calculation that combines error rates with match probabilities **is inappropriate**

Pages 85-87

Should an Error Rate Be Included in Calculations?

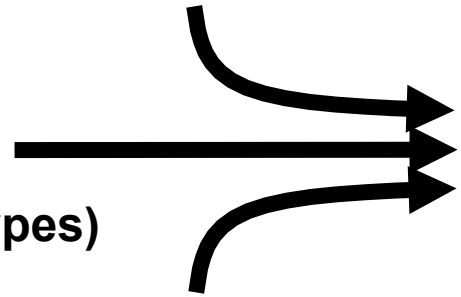
- "The question to be decided is not the general error rate for a laboratory or laboratories over time but rather whether the laboratory doing DNA testing in this particular case made a critical error."
- "To estimate accurately, from proficiency test results, the overall rate at which a laboratory declares nonmatching samples to match, as has been suggested, **would require a laboratory to undergo an unrealistically large number of proficiency trials**."
- "The pooling of proficiency-test results across laboratories has been suggested as a means of estimating an "industry-wide" error rate (Koehler et al. 1995). But that could penalize the better laboratories…"

Information that goes into a DNA rarity estimate (i.e., where errors can occur)

2 Estimates are derived from testing a small subsection of a population

Population allele frequencies

1 **Evidentiary DNA Profile**
(with specific alleles/genotypes)



Rarity estimate of DNA profile
(e.g., RMP or LR)

Genetic formulas and assumptions made

The risk of error goes up with complexity of the DNA profile (e.g., >2 person mixture or low-quality, low-template DNA sample)

3 "All models are wrong – but some are useful" (George Box, 1979)

Recent FBI Erratum on Allele Frequencies **Errors Made in 1999**

JOURNAL OF FORENSIC SCIENCES

July 2015 issue of the *Journal of Forensic Sciences*

ERRATUM

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, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44(6):1277–86.

Since the development in the late 1990s of the original short tandem repeat (STR) typing systems that included the 13 CODIS core loci, new amplification kits that expand the number of loci to 24 in a multiplex reaction are now commercially available. To establish allele distributions for the additional loci, population samples that were originally genotyped using AmpFlSTR Profiler Plus, COfiler, Identifiler (Thermo Fisher Scientific, South San

using the original and corrected data is expected to be less than a factor of two in a full profile. The actual minimum ratio that we could obtain for a constructed profile in the direction of the profile probability being more rare in the original as compared to the amended data was for a highly homozygous partial profile in the Jamaica dataset. It was 0.76, which is well within the factor of 10 suggested by previous studies and the National Research Council (7–10). See Fig. 1 and Table 2. Amended data will be available at fbi.gov and through FBI PopStats. The authors are of the view that these discrepancies require acknowledgment but are unlikely to materially affect any assessment of evidential value.

- **Genotyping errors were made in 27 samples, affecting the reported frequencies of 51 alleles** **In Table 1, 255 allele frequencies are impacted**
- For alleles requiring a frequency correction, the magnitude of the change in frequencies ranged from 0.000012 to 0.018 (average 0.0020 ± 0.0025)
- "The authors are of the view that these discrepancies require acknowledgment but are **unlikely to materially affect any assessment of evidential value**"

Original NIST Identifiler 2003 Dataset
(almost all typed by John Butler with very little second review)

Butler, J.M., et al. (2003) *J. Forensic Sci.* 48(4): 908-911

J Forensic Sci, July 2003, Vol. 48, No. 4
Paper ID JFS2003045_484
Published 19 May 2003
Available online at: www.astm.org

FOR THE RECORD

John M. Butler,[1] Ph.D.; Richard Schoske,[1] M.A.; Peter M. Vallone,[1] Ph.D.; Janette W. Redman[1]; and Margaret C. Kline,[1] M.S.

Allele Frequencies for 15 Autosomal STR Loci on U.S. Caucasian, African American, and Hispanic Populations*

A total of **700 unique STR profiles** were evaluated: 302 Caucasian, 258 African American, and 140 Hispanic. There were 660 males and 40 females.

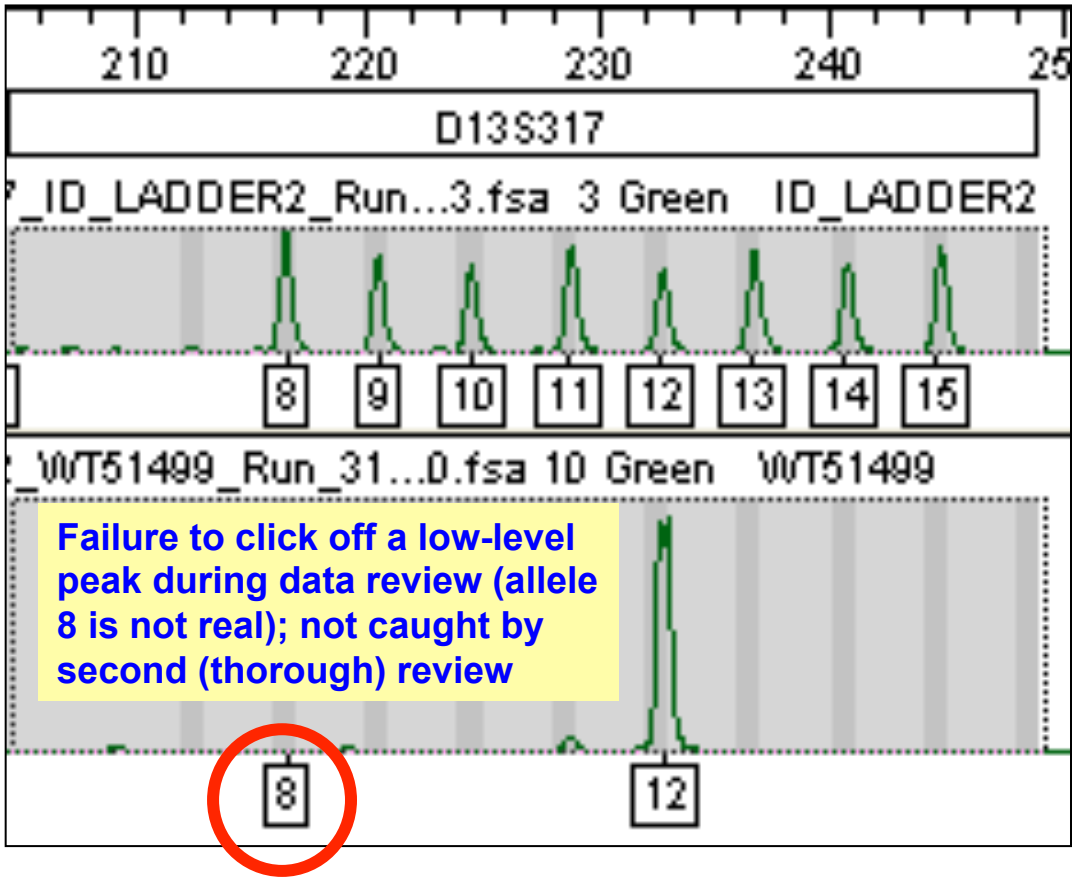
- 700 samples x 16 loci = **11,200 genotypes** or **22,400 alleles**
- **7 errors later found** (presence or absence of single alleles missed)
- **John Butler error rate in 2003**: 7/22400 = 0.00031 = 0.031%
  (if genotypes, then 7/11200 = 0.000625 = **0.063% or ~1 in 1600**)

Post-publication review of NIST 2003 Identifiler dataset revealed 7 errors in the published genotypes

#	ERROR and CORRECTION	REASON
1	OT05588 (African American) is missing a third allele for TPOX; correct type is 9,10,11	**Genotyper table import of only 2 alleles**
2	OT05576 (African American) has an extra allele 11 for D13S317; correct type is 12,12	**stutter peak not removed**
3	GT38066 (Caucasian) is missing allele 18 at VWA; correct type is 17,18	hard to distinguish if allele 18 is real due to potential bleed-through from TH01 *(comments: confirmed with PP16 and Identifiler repeat)*
4	GT36864 (Caucasian) is missing allele 25 at D2S1338; correct type is 19,25	**allele clicked off accidently in initial data review**
5	UT57289 (Caucasian) has an extra allele 12 at D7S820; correct type is 10,10	**bleed-through (green to blue) from high signal** D16S539 allele 12
6	GT36886 (Caucasian) is missing a third allele for TPOX; correct type is 8,10,11	Genotyper table import of only 2 alleles
7	WT51499 (African American) has an extra allele 8 at D13S317; correct type is 12,12	**Failure to click off a low-level peak during data review**

http://www.cstl.nist.gov/strbase/NISTpopdata/JFS2003IDresults.xls (Errata tab)

Uncovering Previous Mistakes During Data Review with Expert System



- Data included in Butler *et al.* (2003) *J. Forensic Sci.* 48(4): 908-911

- D13S317 African American **allele 8 frequency** changes from **0.03295** to **0.03101**

Minor impact

Correct call should be 12,12 for D13S317

(Discovered while reviewing FSS-i^3 "discordant" calls)

NIST Population Datasets

http://www.cstl.nist.gov/strbase/NISTpop.htm

Population Studies Conducted by the NIST Forensics/Human Identity Project Team

During the summer of 2002, our project team obtained over 600 anonymous male samples from Interstate Blood Bank (Memphis, TN) in the form of liquid blood. (Ft. Lauderdale, FL). These samples, which come from U.S. Caucasian, African American, and Hispanic males (self-identified), were subjected to a bulk extracti individual. In 2007, a set of anonymous 800 father/son samples from U.S. Caucasian, African American, Hispanic, and Asian individuals were provided to NIST with these samples to examine a number of DNA markers that are used or may be used in the future for human identity testing applications. DNA typing informatio

DNA Data [Autosomal Markers] [Y-Chromosome Markers] [Mitochondrial DNA]

NIST 1036 U.S. Population Dataset - 29 autosomal STR loci and 23 Y-STR loci NEW

- covers all STR loci present in current commercially available STR kits from Life Technologies and Promega Corporation
- Butler, J.M., Hill, C.R., Coble, M.D. (2012) Variability of new STR loci and kits in U.S. population groups. *Profiles in DNA*. Available at http://www.pror str-loci-and-kits-in-us-population-groups/
- Data as Excel file - includes results from 23 Y-STR haplotypes generated using PowerPlex Y23 reported in Coble, M.D., Hill, C.R., Butler J.M. (2013) I groups. *Forensic Sci. Int. Genet.* 7: e66-e68 and 29 autosomal STRs reported in Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U e82-e83.
- Allele frequencies from autosomal STRs as Excel file - from Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. populat (Supplemental Material Table 2).

2012 NIST 1036 (29 locus) dataset

- Extensive re-testing of samples with multiple STR kits containing different primer sets
- Many people carefully reviewing the dataset

Autosomal Markers

Autosomal STRs - 15 Loci and amelogenin using the Identifiler kit (Applied Biosystems)

- allele frequencies published in *J. Forensic Sci.* July 2003; 48(4):908-911
- Raw Data as Excel file

2003 Identifiler dataset

- Single STR kit primer set used
- Primarily just JB reviewing the dataset

Netherlands Forensic Institute (NFI) Article on Forensic DNA Error Rates

Forensic Science International: Genetics 12 (2014) 77–85

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Error rates in forensic DNA analysis: Definition, numbers, impact and communication

Ate Kloosterman [a,b,c,*], Marjan Sjerps [b,d], Astrid Quak [a]

[a] Department of Human Biological Traces (HBS), Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands
[b] Department of Science, Interdisciplinary Research, Statistics and Knowledge Management (WISK), Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands
[c] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[d] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

Kloosterman et al. (2014) Error rates in forensic DNA analysis: definition, numbers, impact and communication. *FSI Genetics* 12: 77-85

Reported DNA Error Rates

Medical Laboratory Literature Data

	year	# tests	# errors	1 in	%
Plebani & Carraro [33]	1997 (3 mo.)	40,490	189	**214**	0.47%
Carraro & Plebani [36]	2007 (3 mo.)	51,746	160	**323**	0.31%
Stahl et al. [34]	1998 (3 yr.)	676,564	4,135	**164**	0.61%
Hofgärtner & Tait [35]	1999 (1 yr.)	88,394	293	**302**	0.33%

Not all quality issue notifications (aka “errors”) are equal

Table 3

Types of quality issue notifications (QINs) at the NFI in the years 2008–2012. In 2011 it was decided to no longer incorporate the type c QIN: opportunities for improvement ($n = 2$ in 2011 and $n = 10$ in 2012) in the yearly totals of this overview.

	2008	2009	2010	2011	2012
a. External origin	23	10	23	54	100
b. External contamination	3	0	5	24	22
c. Room for improvement	11	6	3	(2)	(10)
d. Positive response	19	9	11	6	17
e. Clerical (no adverse outcome)	29	25	92	77	82
f. Not related to case work	13	9	20	10	5
g. Other (NFI related)	230	270	281	355	346
Total	328	329	435	526	572

Kloosterman et al. (2014) Error rates in forensic DNA analysis: definition, numbers, impact and communication. *FSI Genetics* 12: 77-85

Checks and Controls on Forensic DNA Results

Community	FBI DNA Advisory Board's Quality Assurance Standards *(also interlaboratory studies)*
Laboratory	ASCLD/LAB, ANAB, A2LA Audits and Accreditation
Analyst	Proficiency Tests & Continuing Education
Method/Instrument	**Validation of Analytical Performance** *(with aid of **traceable reference materials**)*
Protocol	Standard Operating Procedure is followed
Data Sets	Allelic ladders, positive and negative amplification controls, and reagent blanks are used
Individual Sample	Internal size standard present in every sample
Interpretation of Result	**Second review by qualified analyst/supervisor**
Court Presentation of Evidence	Defense attorneys and experts with power of discovery requests

Wisdom of **Wilmer Souder**

National Bureau of Standards (1911-1913, 1917-1954)

"The honest expert never looks upon the outcome of his work as a result of luck, the reward of a game, or victory in a battle of wits. He has built his qualifications through hard work. He establishes his conclusions through exacting procedures; he presents his testimony in the face of keen opposition and asks no favor beyond an honest consideration of the facts disclosed. Having done so, he has fulfilled the high obligations of his profession.

"**Justice is sometimes pictured as blindfolded. However, scientific evidence usually pierces the mask**."

- **Wilmer Souder**, "Effective Testimony for Scientific Witnesses", *Science* (1954) 119: 819-822

National Commission on Forensic Science (NCFS):
www.justice.gov/ncfs

Organization of Scientific Area Committees (OSAC):
www.nist.gov/forensics/osac/index.cfm

NIST FORENSIC SCIENCES

www.nist.gov/forensics

301-975-4049 john.butler@nist.gov

Where can DNA errors occur?

Sampling

Analysis

Sample Limitations

- Contamination
  - At point of collection, lab analysis, etc.
- Data processing
  - Incorrect allele/genotype calls made due to data artifacts or sensitivity
- Statistical calculations
  - allele frequency mistakes
  - Incorrect model (theta correction) applied
- Stochastic effects in low template DNA samples
  - limiting appropriate representation of the original source
- Allele stacking and stochastic effects in mixtures
  - limiting the correct deconvolution of components and interpretation of results