Can the Validation Process in Forensic DNA Typing Be Standardized?

John M. Butler[1], Christine S. Tomsey[2], Margaret C. Kline[1]

[1]National Institute of Standards and Technology
[2]Pennsylvania State Police DNA Laboratory

15th International Symposium on Human Identification
Phoenix, AZ
October 6, 2004

International Symposiums on Human Identification
and the Topic of Validation

#	Year	Validation in Title	Total Talks	Validation in Title	Total Posters	%
1	1989	1	10	--	--	10.0
2	1991	0	21	0	14	0
3						
4						
5						
6						
7	1996	2	30	1	77	2.8
8	1997	3	34	11	81	12.1
9	1998	3	25	14	80	16.2
10	1999	0	44	7	70	6.1
11	2000	8	33	11	107	13.6
12	2001	4	30	7	76	10.4
13	2002	2	27	8	78	9.5
14	2003	4	26	17	86	**18.8**
15	2004					
	TOTAL	34	**384**	91	**836**	**10.2**

~10% out of 1,220 presentations have "validation" in the title

Statement of Project Purpose

- Review validation practices currently in use and available standards and guidelines
- Refine general philosophy of validation and steps involved with goal to see if these steps can be standardized
- Attempt to define a minimum number of samples that could be recommended for various validation scenarios
  - Is there a consensus in the community (or can there ever be)?

Conventional forensic DNA typing methods are now widely used and accepted in courts of law. However, new technologies, software, or instrumentation will continue to be developed and therefore need to be validated in laboratories prior to use in casework.

Can we learn from the past as we move into the future?

Validation Definitions

ISO 17025

5.4.5.1 Validation is the **confirmation by examination** and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled

DAB Quality Assurance Standards for Forensic DNA Testing Laboratories

2 (ff) Validation is a **process by which a procedure is evaluated** to determine its efficacy and reliability for forensic casework analysis and includes:

To demonstrate that a method is suitable for its intended purpose…

DAB Quality Assurance Standards for Forensic DNA Testing Laboratories

Manufacturer

(1) **Developmental validation** is the acquisition of test data and determination of conditions and limitations of a new or novel DNA methodology for use on forensic samples.

(2) **Internal validation** is an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform as expected in the laboratory.

Forensic Lab

SWGDAM Revised Validation Guidelines

Section 1.1 Validation is the process by which the scientific community acquires the necessary information to

(a) Assess the ability of a procedure to obtain reliable results.

(b) Determine the conditions under which such results can be obtained.

(c) Define the limitations of the procedure.

The validation process identifies aspects of a procedure that are critical and must be carefully controlled and monitored.

Reliability, Reproducibility, Robustness, Range

Presentation Outline

- Summary of Findings (Community Consensus?)
  - Literature review
  - Interviews with labs
  - **Validation questionnaire**
- Steps Involved in Going "On-Line"
- Resources Under Development to Aid Future Validation Efforts

PubMed Literature Search

http://www.ncbi.nlm.nih.gov/PubMed

Search Results with term "validation" (9/8/04)

- *J. Forensic Sci.* - 71 references
- *Int. J. Legal Med.* - 21 references
- *Forensic Sci. Int.* - 47 references
- *Electrophoresis* – 62 references (12 on DNA)

- All of PubMed - **28,035 references**

Review of Promega conference proceedings:
125 with "validation" in title of talk or poster

Total number of papers examined: 64

Contacting the Community

- Validation Standardization Questionnaire handed out at NIJ DNA Grantees meeting (June 28-30, 2004)
- Emails sent to >200 scientists (July-Aug 2004)
  - Attendees from the NIJ DNA Grantees meeting
  - Participants in NIST interlaboratory studies
  - Contacts through STRBase website
- Responses from **52 scientists** were compiled
  - Covering 27 states + Puerto Rico, 4 companies, 2 outside US
- Specific interviews were conducted to gain perspectives from a small lab, a large lab, a private lab, and court testimony experience

Representative Labs Interviewed

- **Montgomery County Crime Lab** – small lab, 3 analysts, ~180 cases/year; using PP16 and ABI 310
- **Orchid Cellmark** – private contract lab, 40 analysts and technicians, ~5,000 cases/year; Profiler Plus/ COfiler and Identifiler with ABI 310 and ABI 3100; extensive court experience
- **AFDIL** – large federal lab, ~120 analysts/technicians, remains identification rather than strictly forensic cases, >1,000 cases/year (mtDNA & STRs); Profiler Plus/COfiler and PP16 with ABI 377 and ABI 3100

Information from interviews is included in the written report of this project…

Validation Standardization Questionnaire (conducted June-August 2004)

Review of Survey Questions

- What is validation?
- How do you know when you are finished validating a kit, instrument, software, or procedure?
- What steps are needed in internal validation and how many samples should be run at a minimum?
- **How many total samples do you think it takes to internally "validate" a new forensic kit?**
- How many different sets of samples are needed? Over what time period?
- Where do you look for guidance currently in terms of validation?
- **What are some kits, software, instruments that you are considering for validation in the next year?**
- How are validation, training, and proficiency testing related to one another?
- Do you think that the process of validation can be standardized?
- If a standard protocol or set of guidelines existed for validation, would you use it?
- If a standard set of samples existed for performing validation testing, would you use them?

Used to help define specific examples …

How I felt after taking on this project…

Me

Literature, Validation Data, Survey Responses

Validation Standardization Questionnaire (conducted June-August 2004)

How do you know when you are finished with a validation study? (1)

- "When you have demonstrated that it works as expected over a range of samples that is representative of what is seen in casework"
- "When repeat performance gave the same result"
- "When you pull the toothpick out and it is dry?... Meet at least minimum expectations and DAB guidelines"
- "You are very comfortable that you know how it works and your documentation will convince a reviewer you have put the kit thru a rigorous review/test."

Validation Standardization Questionnaire (conducted June-August 2004)

How do you know when you are finished with a validation study? (2)

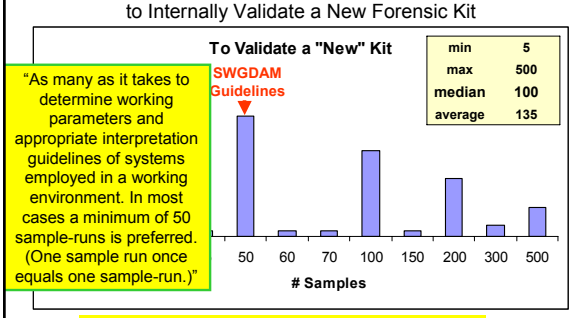
- "Once a reasonable body of data has been assembled and analyzed, quirks have been revealed, and the upper and lower limits of the system have been challenged using a range of samples that one could expect to encounter in the everyday operation of the system"
- "When you achieve accuracy and precision to the desired statistical level of certainty"
- "You can never know…but it is always nice to have more samples!"
- "Validation is never complete"

Validation Standardization Questionnaire (conducted June-August 2004)

Survey Summary for Recommended **Total Number of Samples**

to Internally Validate a New Forensic Kit

"As many as it takes to determine working parameters and appropriate interpretation guidelines of systems employed in a working environment. In most cases a minimum of 50 sample-runs is preferred. (One sample run once equals one sample-run.)"



Choices in survey were: **10, 50, 500, or other ____**

Validation Standardization Questionnaire (conducted June-August 2004)

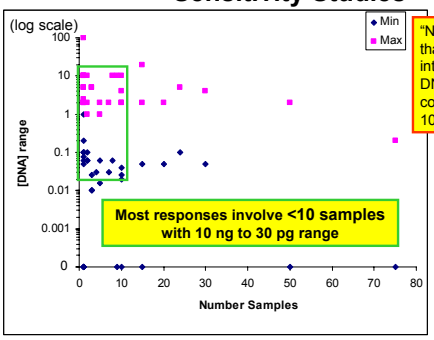
Survey Summary for Recommended **Precision Studies**

A few of the responses:

- "100 allelic ladder injections"
- "1 allelic ladder with 10 injections"
- "Depends upon the system being tested. For a databanking system, 50-100 runs of 50-100 specimens. Again, stats tell you when you've processed enough specimens to understand the system."
- "Minimum: Run one sample at least 8 times. Recommended: Run at least two samples plus allelic ladder at least 8 times." (24 sample-runs)

Validation Standardization Questionnaire (conducted June-August 2004)

Survey Summary for Recommended **Sensitivity Studies**

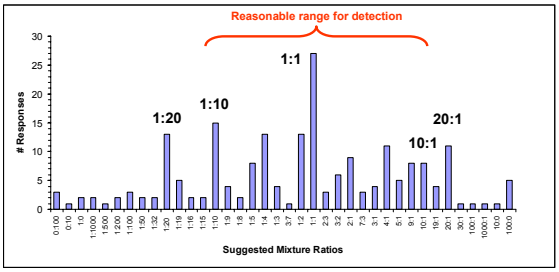


"Need to run samples that challenge interpretation at high DNA and low DNA concentrations—e.g., 10 ng and <0.2 ng"

Most responses involve <10 samples with 10 ng to 30 pg range

Validation Standardization Questionnaire (conducted June-August 2004)

Survey Summary for Recommended **Mixture Studies**



Some Recommended Numbers of Samples:
5 different 2-person mixtures
50 amplifications from at least 10 different mixtures
1 set of samples (ranging from 1:10 to 10:1)

Validation Standardization Questionnaire (conducted June-August 2004)

Survey Summary for Recommended **Non-Human Cases**

A few of the responses:

- "10-20 food animals, companion animals, local wildlife, ferrets"
- "I don't believe this is necessary in internal validation if external results are published. This would not be expected to vary in different analysts' hands."
- "I've trusted system manufacturers to handle this. Should I have?"
- "Minimum: Include information from developmental studies. If performing developmental studies, include at least bacterial and yeast/fungal example, plus mammalian and non-mammalian examples."

Validation Standardization Questionnaire (conducted June-August 2004)

Survey Summary for Recommended **Non-Probative Cases**

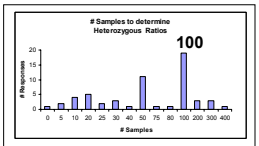
A few of the responses:

- Most responses were between 5-10 cases (range 3-25)
- "More important that the number of cases is the range of forensic samples that are typed during validation."
- "Complete cases are not required to test a system. Recommended: **Run at least 8 mock non-probative samples**. Note: Non-probative samples are not guaranteed to provide complete profiles. They are needed only to show that false results are not generated. Lack of results or incomplete results do not affect the validity of a validation."

Validation Standardization Questionnaire (conducted June-August 2004)

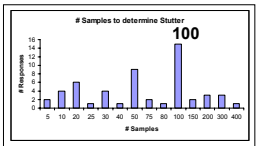
Survey Summary for Recommended **Numbers of Samples**

to Determine Heterozygote Peak Height Ratios and Stutter Values



min	0
max	400
median	50
average	85

Heterozygote Peak Height Ratios



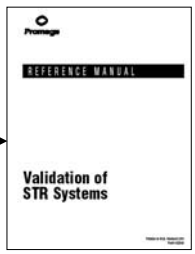
min	5
max	400
median	63
average	88

Stutter Values

Validation Standardization Questionnaire (conducted June-August 2004)

Where do you look for guidance currently in validation?

- SWGDAM
- DAB standards and ISO 17025
- Other scientists
- Literature publications
- Presentations at meetings
- Promega's validation guide
- FBI studies and publications
- NIST studies and publications
- Previous scientific training
- Common sense



Published in March 2001

Validation Standardization Questionnaire (conducted June-August 2004)

Can Validation be Standardized?

Statements from survey responders…

Over 86% (45/52) said yes

Those who responded "no" said

- "to some degree it can be, however, validation is specific to the platform, kits, …",
- "a start-up lab should do much more than an experienced lab…",
- "validation builds on previous work by lab or published data",
- "parts of it can be standardized; I don't think the non-probative cases could be", and
- "only in a general way, as with the SWGDAM guidelines. The uniqueness of each new procedure would make standardization difficult."

Our Conclusion…

to a certain extent it can…but everyone will always have a different comfort level…and **inflexible, absolute numbers for defined studies will not likely be widely accepted**

Validation Standardization Questionnaire (conducted June-August 2004)

If a Standard Protocol or Set of Guidelines Existed for Validation, Would You Use It?

90% (47/52) said yes

Some responses

- "No-I would reference them. I may not completely abide by them but I would certainly review them",
- "No-but it would be taken into consideration",
- "Yes-we would have to or there would be problems in court",
- "Yes-as long as they remain updated, relevant and feasible guidelines and do not become dogma",
- "Yes-if it would pass an audit for validation", and
- "Yes-unless they were far less stringent than current practice."

Validation Standardization Questionnaire (conducted June-August 2004)

If a Standard Set of Samples Existed for Performing Validation Testing, Would You Use Them?

90% (47/52) said yes

Some responses

- "Yes-would love to have something like that available; we are always eager to have benchmarks for assessment",
- "Yes-these types of samples would cut down on time for validation. It would be efficient if they were ready for the particular type of validation…",
- "Yes-as long as they are readily available at a reasonable price",
- "No-this approach is not recommended. It is most important that systems work with the materials available in individual laboratories. Laboratories should be allowed, even encouraged, to select their own preferred materials. Choices for such selection of standard materials for within laboratory analyses and cross-laboratory comparison already exist from a variety of government and commercial entities."

Summary of Literature Examined

Reported Developmental Validation Efforts

Numbers of Samples Run in Developmental Validation Studi…

Kit	Reference	Sensitivity	Precision	Stutter	Mixture		…ases
PP16							
Profiler Plus							
Cofiler							
Identifiler							
SGM Plus							
PP1.1							
PP2.1							
PP 16 BIO							
PP…							
Se…							
Powe…							
Y-PLE…							
Y-PLEX…							
Y-PLEX 12							
Yfiler							

Information will be posted on new STRBase Validation Homepage http://www.cstl.nist.gov/biotech/validation.htm

A total of 64 papers examined

Full list of forensic DNA literature reviewed is available on STRBase

There are Different Opinions…

in Who Should Perform Validation

Development of New STRs for Forensic Casework: Criteria for Selection, Sequencing & Population Data and Forensic Validation

Angel Carracedo and M.V. Lareu
Institute of Legal Medicine. University of Santiago de Compostela, Spain

http://www.promega.com/geneticidproc/ussymp9proc/content/21.pdf

Validation studies following similar parameters to those recommended by TWGDAM were carried out. These include robustness, stability, mixtures, non-human studies, mutation rate and checking for independence with other loci. In our opinion the final validation of a system cannot be carried out by individual groups and companies and should always be performed by an internationally established validation group. In Europe a final assessment and intercomparison exercises are usually performed by the EDNAP group, a working group of the ISFH.

Abstract from talk presented at Promega meeting in 1998

Validation Section of the DNA Advisory Board Standards

issued July 1998 (and April 1999); published in *Forensic Sci. Comm.* July 2000

STANDARD 8.1 The laboratory shall use validated methods and procedures for forensic casework analyses *(DNA analyses)*.

8.1.1 Developmental validation that is conducted **shall be appropriately documented**.

8.1.3 Internal validation **shall be performed and documented by the laboratory**.

FORENSIC SCIENCE COMMUNICATIONS JULY 2000 VOLUME 2 NUMBER 3

Revised SWGDAM Validation Guidelines

(July 2004)

http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm

Forensic Science Communications July 2004 – Volume 6 – Number 3

Standards and Guidelines

Revised Validation Guidelines

Table of Contents
Back Issues
Search

Editors
About FS…
Instructio…
Authors

Scientific Working Group on DNA Analysis Methods (SWGDAM)

3. Internal Validation
…a total of **at least 50 samples**
(some studies may not be necessary…)

Program for DNA Analysis by the Technical Working Group on DNA Analysis Methods (*Crime Laboratory Digest* 1995;22(2):21-43) has been revised due to increased laboratory experience, the advent of new technologies, and the issuance of the Quality Assurance Standards for Forensic DNA Testing Laboratories by the Director of the FBI (*Forensic Science Communications* available: www.fbi.gov/hq/lab/fsc/backissu/july2000/codis2a.htm)

The document provides validation guidelines and definitions approved by SWGDAM July 10, 2003.

A Thoughtful Comment from One Interviewee

Before a set of validation experiments is performed…

- The question should be asked "Do we already know the answer to this question from the literature or a previous study performed in-house?"
- If the answer is "yes" and we document how we know this answer, **then there is no need to perform that set of validation experiments**.

A good example of this scenario is non-human DNA studies.

Common Perceptions of Validation

The goal is not to experience every possible scenario during validation…

"You cannot mimic casework because every case is different."

Effort

Lots of experiments are required

Many labs are examining far too many samples in validation and thus delaying application of casework and contributing to backlogs…

Significant time is required to perform studies

Time

How an Assay Evolves

NIJ-funded project or company efforts

Performed by manufacturer

- Research
- Development
- Optimization

Performed by forensic lab

- Pre-Validation — Learning what questions to ask
- Validation — Performance Check (Kit QC or Following Instrument Repair)
- Re-Validation
- Implementation — Writing SOP, Training Others and Going "On-Line"

Steps Surrounding "Validation" in a Forensic Lab

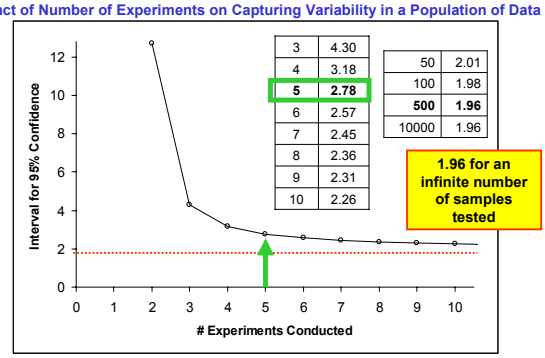
Effort to Bring a Procedure "On-Line"

This is what takes the time…

- **Installation** – purchase of equipment, ordering supplies, setting up in lab
- **Learning** – efforts made to understand technique and gain experience troubleshooting; can take place through direct experience in the lab or vicariously through the literature or hearing talks at meetings
- **Validation of Analytical Procedure** – tests conducted in one's lab to verify range of reliability and reproducibility for procedure
- **SOP Development** – creating interpretation guidelines based on lab experience
- **QC of Materials** – performance check of newly received reagents
- **Training** – passing information on to others in the lab
- **Qualifying Test** – demonstrating knowledge of procedure enabling start of casework
- **Proficiency Testing** – verifying that trained analysts are performing procedure properly over time

A Comment on Minimum Numbers of Samples for Validation Studies…

Impact of Number of Experiments on Capturing Variability in a Population of Data



3	4.30
4	3.18
5	**2.78**
6	2.57
7	2.45
8	2.36
9	2.31
10	2.26

50	2.01
100	1.98
500	**1.96**
10000	1.96

1.96 for an infinite number of samples tested

From ***The HitchHiker's Guide to the Galaxy***

http://www.bbc.co.uk/dna/h2g2/

The Answer to the Ultimate Question Of Life, The Universe, And Everything

(and the Minimum Number of Samples for Internal Validation?)

➢**42**

Validation Standardization Questionnaire (conducted June-August 2004)

Survey Summary of Planned Near-term "Validation"

Commercial Kits

Extraction
- DNA IQ
- Qiagen
- Biomek 2000

DNA Quant
- **Quantifiler**

STR Amp Kits
- Identifiler
- PowerPlex Y
- Yfiler
- PowerPlex 16
- ProPlus/COfiler reduced volume

Software
- **GeneMapper*ID***
- GeneScan/ Genotyper NT
- TrueAllele
- SQL*LIMS and Forensic Solution

Analysis Instruments
- **ABI 3100 Avant**
- ABI 3100
- FMBIO III+
- MegaBACE

For RT-PCR
- **ABI 7000**
- Stratagene RT-PCR

Internal Validation

Example: PowerPlex 16

- Switch from ProfilerPlus/COfiler kits to PowerPlex 16
- Retaining same instrument platform of ABI 310

Recommendations:

- Concordance study (somewhat, but better to review literature to see impact across a larger number of samples and which loci would be expected to exhibit allele dropout-e.g., D5S818)
- Stutter quantities, heterozygote peak height ratio
- Some sensitivity studies and mixture ratios
- **Do not need precision studies to evaluate instrument reproducibility**

Internal Validation

Example: ABI 3100*Avant*

- Evaluation of a new ABI 3100*Avant* when a laboratory already has experience with ABI 310
- STR kits used in lab will remain the same

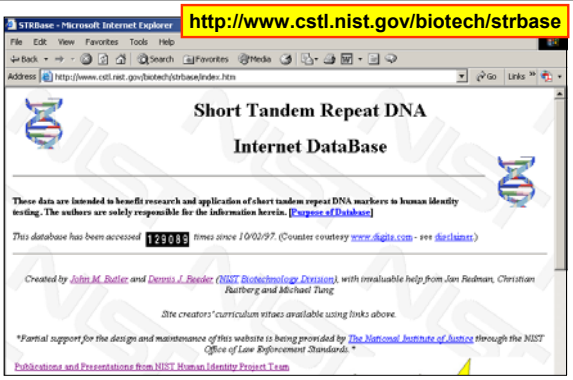
Recommendations:

- Precision studies to evaluate instrument reproducibility
- Sensitivity studies
- **Do not need new stutter, mixture ratio, peak height ratio, etc. (these relate to dynamics of the the kit used)**

Resources to Aid Future Validation Studies

- STRBase Validation Website
  - **http://www.cstl.nist.gov/biotech/strbase/validation.htm**
  - Examples with recommended minimum numbers
  - Validation summary sheets
- NIST Calibration Data Set
  - **set of ~200 sample data files** that can be used to evaluate common STR typing "artifacts" such as stutter, non-template addition, spikes, peak imbalance, tri-allelic patterns, variant alleles, single base resolution
  - will help meet NDIS Appendix B requirements for Expert Systems evaluation
- Quality Control Program (Dave Duewer, NIST)
  - Software to monitor STR electropherogram performance (resolution, sensitivity) over time

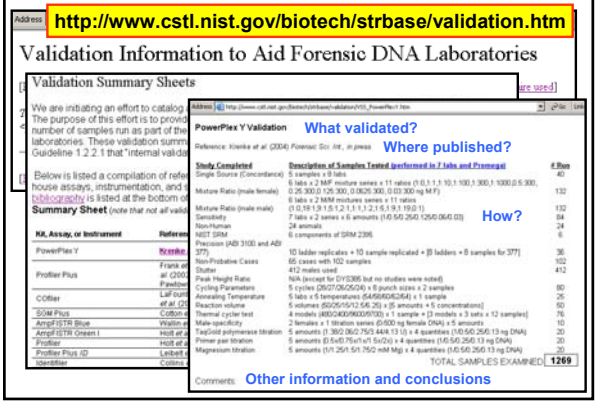
http://www.cstl.nist.gov/biotech/strbase



A Human Identity Testing Community Resource…

New Validation Homepage on STRBase

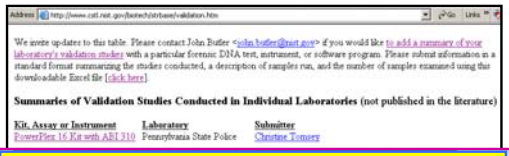
http://www.cstl.nist.gov/biotech/strbase/validation.htm



Validation Summary Sheet for PowerPlex Y

Study Completed (17 studies done)	Description of Samples Tested (performed in 7 labs and Promega)	# Run
Single Source (Concordance)	5 samples x 8 labs	40
Mixture Ratio (male:female)	6 labs x 2 M/F mixture series x 11 ratios (1:0,1:1,1:10,1:100,1:300,1:1000,0.5:300, 0.25:300,0.125:300, 0.0625:300, 0.03:300 ng M:F)	132
Mixture Ratio (male:male)	6 labs x 2 M/M mixtures series x 11 ratios	132
Sensitivity	7 labs x 2 series x 6 amounts (1/0.5/0.25/0.125/0.06/0.03)	84
Non-Human	24 animals	24
NIST SRM	6 components of SRM 2395	6
Precision (ABI 3100 and ABI 377)	10 ladder replicates + 10 sample replicated + [8 ladders + 8 samples for 377]	36
Non-Probative Cases	65 cases with 102 samples	102
Stutter	412 males used	412
Peak Height Ratio	N/A (except for DYS385 but no studies were noted)	
Cycling Parameters	5 cycles (28/27/26/25/24) x 8 punch sizes x 2 samples	80
Annealing Temperature	5 labs x 5 temperatures (54/58/60/62/64) x 1 sample	25
Reaction volume	5 volumes (50/25/15/12.5/6.25) x [5 amounts + 5 concentrations]	50
Thermal cycler test	4 models (480/2400/9600/9700) x 1 sample + [3 models x 3 sets x 12 samples]	76
Male-specificity	2 females x 1 titration series (0-500 ng female DNA) x 5 amounts each	10
TaqGold polymerase titration	5 amounts (1.38/2.06/2.75/3.44/4.13 U) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Primer pair titration	5 amounts (0.5x/0.75x/1x/1.5x/2x) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Magnesium titration	5 amounts (1/1.25/1.5/1.75/2 mM Mg) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Krenke *et al.* (2004) *Forensic Sci. Int., in press*	TOTAL SAMPLES EXAMINED	1269

Laboratory Internal Validation Summaries



We invite updates to this table. Please contact John Butler <john.butler@nist.gov> if you would like to add a summary of your laboratory's validation studies with a particular forensic DNA test, instrument, or software program. Please submit information in a standard format summarizing the studies conducted, a description of samples run, and the number of samples examined using this downloadable Excel file [click here].

Summaries of Validation Studies Conducted in Individual Laboratories (not published in the literature)

Kit, Assay or Instrument	Laboratory	Submitter
PowerPlex 16 Kit with ABI 310	Pennsylvania State Police	Christine Tomsey

Soliciting Information on Studies Performed by the Community

Study Conducted	Description of Samples Run with PowerPlex 16 Validation	# Male	# Female Mix
Single Source (Concordance)	8 samples (Promega concordance) + 200 samples (part of population concordance study)	208	100
Mixtures	45	45	10
Mixture Ratio	1 sample x 11 ratios (1:0,1:91,1:41,1:21,1:11,1:2,1:4,1:9,1:19,0:1) x 2 injections (5/10 seconds)	22	33
Sensitivity	5 samples x 6 amounts (50/0.6:50:250:1250:0.06/0.03 ng) + [5 samples x 3 points (ab/above/below dropout)]	55	33
Non-Human	11 animals	11	0
NIST SRM 2391b	12 components	12	12
Precision (ABI 310)	(5 samples x 10 injections each) + 10 injections of allelic ladders	60	60
Non-Probative Cases	5 cases x 4 samples each (evidence EF/EF/Victim/suspect)	20	20
Stutter	200 samples (data used from population samples)	-	-
Peak Height Ratio	200 samples (data used from population samples)	-	-
Cycling Parameters	14 samples x 2 different cycle numbers (30/32) x 2 injection times (3/5 seconds)	56	-
Annealing Temperature	3 samples x 4 concentrations (2.0/1.0/0.5/0.25 ng) x 5 temperatures (56/58/60/62/64)	60	0
Proficiency	9 sets x 4 samples per set	36	12
Substrate	9 common substrates x 1 sample each	9	0
Environment	5 conditions (outside/60oC/60oC/40oC/RT) x 6 time points (3/5/1 2/24/48/85 days)	30	0
Various tissues	Bone, hair_teeth, semen, perspiration, urine, blood, semen, vaginal swab (minimum of one sample each)	9	0
	TOTAL SAMPLES RUN	633	280

Acknowledgments

National Institute of Justice
The Research, Development, and Evaluation Agency of the U.S. Department of Justice

- **NIJ Funding** for NIST Project Team through NIST Office of Law Enforcement Standards
- **Co-Authors: Chris Tomsey and Margaret Kline**
- Dave Duewer (NIST)
- Kari Tontarski (Montgomery County Crime Lab)
- Robin Cotton (Orchid Cellmark)
- Tim McMahon (AFDIL)

- **Many members of forensic DNA typing community for their valuable input on our validation questionnaire**

Further Information

- Final version of this talk will be available:
  - http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
- See also new **STRBase Validation Homepage**
  - http://www.cstl.nist.gov/biotech/strbase/validation.htm
- My email address: **john.butler@nist.gov**