Can the Validation Process in Forensic DNA Typing Be Standardized?

John M. Butler¹, Christine S. Tomsey², Margaret C. Kline¹

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

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

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

(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.

International Symposiums on Human Identification and the Topic of Validation

<table>
<thead>
<tr>
<th>Year</th>
<th>Validation in Title</th>
<th>Total Talks</th>
<th>Validation in Title</th>
<th>Total Posters</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>1</td>
<td>40</td>
<td>0</td>
<td>14</td>
<td>10.0%</td>
</tr>
<tr>
<td>2000</td>
<td>2</td>
<td>30</td>
<td>1</td>
<td>77</td>
<td>2.3%</td>
</tr>
<tr>
<td>2001</td>
<td>3</td>
<td>24</td>
<td>11</td>
<td>81</td>
<td>12.1%</td>
</tr>
<tr>
<td>2002</td>
<td>4</td>
<td>25</td>
<td>7</td>
<td>75</td>
<td>10.5%</td>
</tr>
<tr>
<td>2003</td>
<td>5</td>
<td>44</td>
<td>11</td>
<td>107</td>
<td>13.6%</td>
</tr>
<tr>
<td>2004</td>
<td>6</td>
<td>50</td>
<td>7</td>
<td>76</td>
<td>10.4%</td>
</tr>
<tr>
<td>2005</td>
<td>7</td>
<td>32</td>
<td>8</td>
<td>75</td>
<td>11.5%</td>
</tr>
<tr>
<td>2006</td>
<td>8</td>
<td>28</td>
<td>17</td>
<td>86</td>
<td>16.8%</td>
</tr>
<tr>
<td>2007</td>
<td>9</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td>10.0%</td>
</tr>
<tr>
<td>2008</td>
<td>10</td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>10.0%</td>
</tr>
<tr>
<td>2009</td>
<td>11</td>
<td>3</td>
<td>12</td>
<td>15</td>
<td>10.0%</td>
</tr>
<tr>
<td>2010</td>
<td>12</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>10.0%</td>
</tr>
<tr>
<td>2011</td>
<td>13</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>10.0%</td>
</tr>
<tr>
<td>2012</td>
<td>14</td>
<td>4</td>
<td>17</td>
<td>8</td>
<td>10.0%</td>
</tr>
<tr>
<td>2013</td>
<td>15</td>
<td>2</td>
<td>17</td>
<td>9</td>
<td>10.0%</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>364</td>
<td>11</td>
<td>835</td>
<td>10.2%</td>
</tr>
</tbody>
</table>

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

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

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

PubMed Literature Search


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

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 Identifier 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.

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

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…

Litetature, Validation Data, Survey Responses

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.”

Survey Summary for Recommended
Total Number of Samples

to Internally Validate a New Forensic Kit

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)

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
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."

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."

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

Can Validation be Standardized?

Statements from survey responders...

Published in March 2001

Survey Summary for Recommended Numbers of Samples to Determine Heterozygote Peak Height Ratios and Stutter Values
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-as long as they remain updated, relevant and feasible guidelines and do not become dogma."
- "Yes-if it would pass an audit for validation."
- "Yes-unless they were far less stringent than current practice."
- "Yes-if they could show a benefit of standardized validation."
- "Yes-as long as reviews are made and the guidelines are kept up to date."
- "Yes-if everyone is using them."
- "Yes-unless they are ambiguous or too vague."
- "Yes-but it would be a starting point."
- "Yes-if they were science based."
- "Yes-as an industry standard, not a dogma."
- "Yes-as long as it would allow for a consistent practice."
- "Yes-if it would pass an audit for validation."

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."
- "Yes-as long as ongoing testing validates that the material is superior to current practice."
- "Yes-if the samples were relevant to the work that is actually being performed."
- "No-this is just another way to tax the system."
- "Yes-as long as there is a clear, neutral authority funding and controlling the material source."
- "Yes-unless they are too restrictive."

Summary of Literature Examined

Reported Developmental Validation Efforts

<table>
<thead>
<tr>
<th>Kit</th>
<th>Reference</th>
<th>Number of Samples Run in Developmental Validation Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yfiler</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y-PLEX 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y-PLEX 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y-PLEX 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PowerPlex Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sefiler</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP ES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP 16 BIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGM Plus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifiler</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cofiler</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profiler Plus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A total of 64 papers examined. Full list of forensic DNA literature reviewed is available on STRBase.

Validation Section of the DNA Advisory Board Standards


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.

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


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 ENNAP group, a working group of the ISFH.

Abstract from talk presented at Promega meeting in 1998

Revised SWGDAM Validation Guidelines

(July 2004)


3. Internal Validation

...a total of at least 50 samples

(some studies may not be necessary...)
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...

- Lots of experiments are required
- “You cannot mimic casework because every case is different.”
- 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

Steps Surrounding “Validation” in a Forensic Lab

- Effort to Bring a Procedure “On-Line”
- 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
- Proficiency Testing – verifying that trained analysts are performing procedure properly over time

How an Assay Evolves

NIJ-funded project or company efforts

Performed by manufacturer

Performed by forensic lab

Research → Development → Optimization → Pre-Validation → Validation → Implementation

Learning what questions to ask → Performance Check (QA/QC or Following Instrument Repair)

Steps Surrounding “Validation” in a Forensic Lab

- Effort to Bring a Procedure “On-Line”
- 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
- 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

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
Survey Summary of Planned Near-term “Validation”

<table>
<thead>
<tr>
<th>Commercial Kits</th>
<th>Software</th>
<th>Analysis Instruments</th>
<th>Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA IQ</td>
<td>GeneMapperID, GeneScan/Genotyper NT, TrueAllele, SQL*LIMS and Forensic Solution</td>
<td>ABI 3100, ABI 3100, FMBIO III+, MegaBACE</td>
<td>DNA IQ</td>
</tr>
<tr>
<td>Qiagen</td>
<td></td>
<td></td>
<td>Qiagen</td>
</tr>
<tr>
<td>Biomek 2000 DNA Quant</td>
<td></td>
<td></td>
<td>Biomek 2000 DNA Quant</td>
</tr>
<tr>
<td>Quantifiler</td>
<td></td>
<td></td>
<td>Quantifiler</td>
</tr>
</tbody>
</table>

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

Example: PowerPlex 16
- Switch from ProfilerPlus/COfiller 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., DSS818)
- Stutter quantities, heterozygote peak height ratio
- Some sensitivity studies and mixture ratios
- Do not need precision studies to evaluate instrument reproducibility

Example: ABI 3100Avant
- Evaluation of a new ABI 3100Avant 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 kit used)

Resources to Aid Future Validation Studies
- STRBase Validation Website
  - 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

New Validation Homepage on STRBase
http://www.cstl.nist.gov/biotech/strbase/validation.htm
Validation Summary Sheet for PowerPlex Y

<table>
<thead>
<tr>
<th>Study</th>
<th>Description of Samples Tested</th>
<th>No.</th>
<th>Study</th>
<th>Description of Samples Tested</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Source (Concentrations)</td>
<td>5 samples x 2 dites</td>
<td>40</td>
<td>Male Specificity</td>
<td>10 ladders x 10 samples + 10 sample replicated + [3 ladders x 8 samples for 377]</td>
<td>58</td>
</tr>
<tr>
<td>Mutant Ratio (male/female)</td>
<td>6 tests x 2 MIF mixture series x 11 dites</td>
<td>132</td>
<td>Non-Probative Cases</td>
<td>45 cases with 132 samples</td>
<td>102</td>
</tr>
<tr>
<td>Mutant Ratio (male/male)</td>
<td>6 tests x 2 MIF mixture series x 11 dites</td>
<td>132</td>
<td>Solute</td>
<td>412</td>
<td>4 males used</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>7 tests x 2 series x 6 amounts (0.50, 0.25, 0.125, 0.0625 ng DNA)</td>
<td>84</td>
<td>Peak Height Ratio</td>
<td>N/A (except for DYS385 but no studies were noted)</td>
<td>50</td>
</tr>
<tr>
<td>Hair Human</td>
<td>34 analyses</td>
<td>24</td>
<td>Cycling Parameters</td>
<td>5 cycles (28/27/26/25/24) x 2 punch sizes x 2 samples</td>
<td>50</td>
</tr>
<tr>
<td>NIST STR</td>
<td>6 components of STRM 2395</td>
<td>6</td>
<td>Annealing Temperature</td>
<td>5 tests x 5 temperatures (45/50/55/60/65°C) x 1 sample</td>
<td>25</td>
</tr>
<tr>
<td>Precision (ABI 3100 and ABI 377)</td>
<td>10 ladder replicates x 10 sample replicated + [8 ladders x 8 samples for 377]</td>
<td>58</td>
<td>Reaction Volume</td>
<td>6 volumes (50/25/15/12.5/6.2/3.1 mL) x 6 concentrations</td>
<td>50</td>
</tr>
<tr>
<td>Non-Probative Cases</td>
<td>45 cases with 132 samples</td>
<td>102</td>
<td>Thermal Cycle Test</td>
<td>5 replicates x 2 series x 12 sampling</td>
<td>78</td>
</tr>
<tr>
<td>Water</td>
<td>412 males used</td>
<td>412</td>
<td>Melting Specificity</td>
<td>2 fragments x 2 fragment series (5.0/8.0 ng female DNA) x 6 amounts each</td>
<td>15</td>
</tr>
<tr>
<td>Reaction Volume</td>
<td>6 volumes (50/25/15/12.5/6.2/3.1 mL) x 6 concentrations</td>
<td>50</td>
<td>TaqGold Polymorpherase Titration</td>
<td>5 amounts (1.0, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.025, 0.0125, 0.00625 ng DNA)</td>
<td>25</td>
</tr>
<tr>
<td>Primer pair titration</td>
<td>5 amounts (0.75, 0.5, 0.25, 0.125, 0.0625 ng DNA) x 4 quantities</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TaqGold Polymerase Titration</td>
<td>5 amounts (1.0, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.025, 0.0125, 0.00625 ng DNA)</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnesium Titration</td>
<td>5 amounts (1.0, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.025, 0.0125, 0.00625 ng DNA)</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


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:
- See also new STRBase Validation Homepage
  - http://www.cstl.nist.gov/strbase/validation.htm
- My email address: john.butler@nist.gov