VALIDATION

Debunking Some Urban Legends Surrounding Validation Within the Forensic DNA Community

By John Butler
National Institute of Standards and Technology, Gaithersburg, Maryland, USA

EDITOR'S NOTE: Dr. Butler has more than a dozen years of experience validating new DNA assays and instruments in both government and private-sector laboratories. He was a member of the SWGDAM Validation Subcommittee that produced the 2004 Revised Validation Guidelines. For more information on his work with validation, see: **www.cstl.nist.gov/biotech/strbase/validation.htm**

Avoiding the urban legends described here should simplify the internal validation process and enable new techniques to be brought on-line in a timely fashion so that crimes involving biological evidence can be solved more readily.

INTRODUCTION

Urban legends are funny (or sometimes horrifying) stories that spread quickly, often via e-mail (**http://en.wikipedia.org/wiki/Urban_legend**; see also **www.snopes.com**). While they are seldom based in reality, urban legends often reflect the paranoia of the population that perpetuates them. Similarly, in recent years a number of misconceptions have arisen within the forensic DNA community surrounding the process and philosophy of validation.

The EURACHEM Guide *"The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics"* (see also ISO 17025 section 5.4.5.1) defines validation as "confirmation by examination and provision of objective evidence that the particular requirements for a specified intended use are fulfilled" (1). The FBI's DNA Advisory Board Quality Assurance Standards (2), and more recently the SWGDAM Revised Validation Guidelines (3) describe developmental validation as being performed by a manufacturer of a new procedure or instrument, while each individual forensic DNA laboratory conducts internal validation.

This article is focused on laboratories performing internal validation. Once the developer of a particular measurement technique demonstrates that it is robust, reliable and reproducible, validating the technique for use in your lab just requires establishing that it is working properly (4). Unfortunately, some forensic DNA labs, often because they are driven by fear of auditors, are taking far too long or running far too many samples as part of their "validation studies". This over-validation can contribute to backlogs in already overburdened DNA laboratories, as it delays the initiation of forensic casework with a new technique or instrument that in many cases would help speed data collection or interpretation. Since technology will continue to advance, validation of new methodologies will always play an important role in forensic DNA laboratories. In short, there will always be something to validate, and it is increasingly important to accomplish this task quickly and reliably.

Validation is an essential part of the overall quality assurance program in a laboratory. A quality assurance program helps ensure that: (a) consistent and correct results are obtained from samples given suitable quality and quantity of material and (b) failure to obtain a result from a tested sample is due to insufficient material rather than an invalid measurement procedure. Unreliable data could (and should) be contested in court. Validation is the essence behind the reliable analytical data required by the *Frye* and *Daubert* rulings that impact court admissibility of scientific results.

VALIDATION

URBAN LEGEND #1: HUNDREDS OR THOUSANDS OF SAMPLES ARE REQUIRED TO FULLY VALIDATE AN INSTRUMENT OR METHOD

I recently reviewed a validation study where 960 samples were examined to verify the precision of STR allele sizing on an ABI PRISM® 3100 Genetic Analyzer. The +/–3 standard deviation for these results was less than 0.5bp, suggesting that reliable STR genotyping could be performed on this laboratory's instrument using the default allele bin sizes used by the analysis software. While it is admirable that so many data points were included in this study, the same conclusion could have been reached with far fewer experiments. Besides reducing the workload on DNA analysts performing the validation studies, a direct benefit of running fewer validation experiments is that casework can be initiated sooner with a new instrument—bringing greater capacity to the laboratory.

In 1908, William Sealy Gosset, writing under his pen name "Student", introduced the t-test to help place confidence levels on judgments made from small sample sets in comparision to a potential larger population of data. The Student's t-test can be helpful in defining potential sample numbers for validation experiments. After running 5–10 replicate samples for a particular experiment, there are diminishing returns to adding additional results. The number five is already in use throughout the forensic DNA community. The 1996 National Research Council report *The Evaluation of Forensic DNA Evidence* requires at least five observations of an allele when establishing a minimum allele frequency (5), and the National DNA Index System (NDIS) expert system validation requirements involve the observation of at least five challenge events for each issue such as stutter, spikes, etc. (6).

When conducting an internal validation, the SWGDAM Revised Validation Guidelines recommend running a total of at least 50 samples—not 50 samples per experiment. Typical internal validation studies include concordance testing with known and nonprobative evidence samples, examining precision, reproducibility and sensitivity, and assessing stochastic effects and the detectable range of mixtures and contamination.

When conducting an internal validation, the SWGDAM Revised Validation Guidelines recommend running a total of at least 50 samples—not 50 samples per experiment.

For example, to verify that a new STR kit works appropriately in your laboratory, the following samples could be run: (a) standard samples with known types (e.g., the kit positive control, DNA from staff members and NIST SRM samples) along with 5–10 nonprobative casework samples previously examined with other kits, (b) 5–10 injections of allelic ladders to define system precision across typical time and environmental conditions used in running a batch of samples, (c) sensitivity samples covering the dynamic range of the STR kit (e.g., two sets of samples with the following dilution series: 5ng, 2ng, 1ng, 0.5ng, 0.2ng, 0.1ng, and 0.05ng), and (d) mixtures covering a range of ratios (e.g., two sets of samples with different allele combinations and the following ratios: 10:1, 3:1, 1:1, 1:3, and 1:10). Such a set of experiments would meet the requirement of 50 total samples for internal validation and could be performed in a relatively short period of time (less than a week by a single scientist).

URBAN LEGEND #2: VALIDATION IS UNIFORMLY PERFORMED THROUGHOUT THE COMMUNITY

A common perception of people outside the forensic community looking into the field—and even many forensic scientists and lab managers—is that validation experiments and operational protocols are fairly uniform across laboratories. A survey conducted in 2004 of over 50 forensic DNA analysts revealed that a wide range of comfort levels exists throughout the community in terms of validation requirements (7). The range of responses reflects different perspectives of what validation should entail and makes it challenging to develop community-wide consensus on minimum sample numbers recommended for various studies. Auditors need to realize that variability can exist among validation studies.

URBAN LEGEND #3: EACH COMPONENT OF A DNA TEST OR PROCESS MUST BE VALIDATED SEPARATELY

The entire process of extraction, quantitation, amplification, separation/detection and data interpretation impact the final DNA typing results. Since the final outcome of a DNA test is what matters, multiple steps in a process may be validated together. Samples used in testing amplification reproducibility can also be used in verifying data interpretation software. However, for troubleshooting purposes (i.e., solving a problem with an instrument or assay), the isolation or decoupling of variables is essential. Remember though that validation is not the same thing as troubleshooting.

VALIDATION

URBAN LEGEND #4: VALIDATION SHOULD SEEK TO UNDERSTAND EVERYTHING THAT COULD POTENTIALLY GO WRONG WITH AN INSTRUMENT OR TECHNIQUE

It is impossible to mimic everything that might be seen in casework or in samples processed through a laboratory in the future. Remember that validation simply confirms that the STR kit, instrument or software is "fit-for-purpose" and works within the range of conditions defined by the validation experiments conducted.

The objectives of validation studies should be decided—and agreed upon by management—from the beginning of the task to avoid the temptation to toddle down tangential topics uncovered during validation experiments. Without a "stop point" validation experiments can easily lose their focus. Discussions with other labs that have previously validated the same kit or instrument can be valuable.

URBAN LEGEND #5: LEARNING THE TECHNIQUE AND TRAINING OTHER ANALYSTS ARE PART OF VALIDATION

This urban legend is probably behind the slow implementation of new techniques in many forensic laboratories. Too often validation is not distinguished from training (7). While learning and training play an important role in bringing a new technique "on-line", validation is the demonstration that a particular measurement process works properly when it is performed by appropriately trained personnel. Remember that instruments are calibrated, methods are validated and people are trained.

URBAN LEGEND #6: VALIDATION IS BORING AND SHOULD BE PERFORMED BY SUMMER INTERNS SINCE IT IS BENEATH THE DIGNITY OF A QUALIFIED ANALYST

Validation experiments can be tedious, but these studies are tremendously important and should be performed by experienced and qualified analysts. As noted above, validation should not require large numbers of samples to confirm that an instrument or method is working properly in your laboratory.

Unfortunately, some forensic DNA labs, often because they are driven by fear of auditors, are taking far too long or running far too many samples as part of their "validation studies".

To make valid analytical measurements, it is assumed that the staff making the measurements is both qualified and competent to undertake the task. Validation should not be viewed as an opportunity for a summer intern to become familiar with a particular technique. Since these interns are not permanent staff, the knowledge they gain in performing experiments leaves with them.

Everyone within a laboratory using a particular technique should be familiar with the validation results in order to understand the limitations of the method. Perhaps part of the qualifying test performed by analysts being trained in a new technique or instrument should be a quiz regarding details of a laboratory's internal validation.

URBAN LEGEND #7: DOCUMENTING VALIDATION IS DIFFICULT AND SHOULD BE EXTENSIVE

Documentation of validation results is required by the DAB Quality Assurance Standards section 8 (2). As noted by the EURACHEM Guide section 9.2 (1), appropriate documentation will help ensure that application of the method from one occasion to the next is consistent. However, documentation is not difficult nor does it have to be extensive. An example of a simple documentation format is provided by Angelo Della Manna of the Alabama Department of Forensic Sciences in their validation of the ABI PRISM® 7000 Sequence Detection System and Quantifiler® Human DNA Quantification Kit available at: **www.cstl.nist.gov/biotech/strbase/validation/ADFS-BH_7000val.pdf**

URBAN LEGEND #8: ONCE A VALIDATION STUDY IS COMPLETED YOU NEVER HAVE TO REVISIT IT

In a certain sense, validation is never complete because it is part of a good quality assurance program. Instruments (e.g., laser power, CCD camera response) may change slightly over time and impact sensitivity of DNA tests. Environmental conditions that impact STR allele-sizing precision, such as room temperature, may also change over time. Ongoing monitoring (essentially "re-validation") should be performed regularly to verify that results are within the expected range.

Most laboratory protocols involving the ABI PRISM® 310 or 3100 Genetic Analyzer use more than one allelic ladder in a sample batch in case one of the ladders fails to inject properly. A simple way to conduct an ongoing performance check/system re-validation is to examine the

VALIDATION

additional allelic ladder(s) in a batch as samples against the first allelic ladder. If all alleles are called in the "sample" allelic ladder(s), then this provides confidence that the environmental conditions are stable and sufficiently precise to reliably genotype the samples in the batch. Likewise, sensitivity of an analytical system can be monitored over time by noting the signal observed from the positive control. If the relative fluorescence units change dramatically for the positive control DNA, then troubleshooting measures are required.

CONCLUSIONS

Use common sense in your approach to validation studies. First, establish concordance with previous results. The same typing results should be obtained with the new technique compared with results for Certified Reference Materials (e.g., NIST SRMs) and "real samples" previously analyzed by an established method. Second, through constant monitoring of performance, such as checking multiple allelic ladders in a batch against one another, demonstrate consistency in results over time.

Resources are available to aid in current and future validation studies. The STRBase web site contains a validation section with helpful information: **www.cstl.nist.gov/biotech/strbase/validation.htm**. The contribution of additional internal validation studies from members of the community to this web site will aid other forensic laboratories in their validation studies. Promega has updated their validation reference manual (8), and Applied Biosystems is producing a validation software tool.

Due to the success of forensic DNA typing in solving crimes and its continued value to the criminal justice community, laboratories involved in DNA analysis will be subjected to increasing requests for sample analysis in the future. Avoiding the urban legends described here should simplify the internal validation process and enable new techniques to be brought on-line in a timely fashion so that crimes involving biological evidence can be solved more readily.

ACKNOWLEDGMENTS AND DISCLAIMER

The assistance of a supportive human identity project research team at NIST including Margaret Kline, David Duewer and Peter Vallone is greatly appreciated. Robyn Ragsdale (FDLE) and Chris Tomsey (PASP) have provided helpful input from a forensic casework perspective. The author is supported in part by funding from the National Institute of Justice through interagency agreement 2003-IJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology; nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

REFERENCES

1. EURACHEM Guide (1998) The fitness for purpose of analytical methods: A laboratory guide to method validation and related topics. This can be viewed online at: **www.eurachem.ul.pt/guides/valid.pdf**
2. DNA Advisory Board (2000) Quality assurance standards for forensic DNA testing laboratories and for convicted offender DNA databasing laboratories. *Forensic Sci. Comm.* **2(3). This can be viewed online at: www.fbi.gov/hq/lab/fsc/backissu/july2000/codispre.htm**
3. Scientific Working Group on DNA Analysis Methods (SWGDAM) revised validation guidelines. (2004) *Forensic Sci. Comm.* **6**(3). This can be viewed online at: **www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm**
4. Butler, J.M. (2005) *Forensic DNA typing: Biology, technology, and genetics of STR markers*, 2nd edition, Chapter 16 "Laboratory Validation", 389–412.
5. National Research Council (1996) *The evaluation of forensic DNA evidence.* National Academy Press, Washington, D.C., 148.
6. National DNA Index System (NDIS) DNA data acceptance standards operational procedures (2005) Appendix B guidelines for submitting requests for approval of an expert system for review of offender samples.
7. Butler, J.M., Tomsey, C.S. and Kline, M.C. (2004) Can the validation process in forensic DNA typing be standardized? *Proceedings of the 15th International Symposium on Human Identification*. This can be viewed online at: **www.promega.com/geneticidproc/ussymp15proc/oralpresentations/butler.pdf**; presentation available at: **www.cstl.nist.gov/biotech/strbase/pub_pres/PromegaTalkOct2004.pdf**
8. *Internal Validation of STR Systems*, Promega Corporation. This can be viewed online at: **www.promega.com/techserv/apps/hmnid/referenceinformation/powerplex/ValidationManual.pdf**