Episode 3: Reading List

Validation Essentials

ANZPAA
Australia New Zealand
Policing Advisory Agency

NIFS
NATIONAL INSTITUTE OF FORENSIC SCIENCE AUSTRALIA NEW ZEALAND

Reading List

Instrumental-Based Methods

Statistical Tools and Approaches to Validate Analytical Methods: Performances and Limits

Belouafa, S., Habti, F., Benhar, S., Belafkih, B., Tayane, S., Hamdouch, S., Bennamara, A., Abourriche, A. (2017)

DOI: 10.1051/ijmqe/2016030

This review paper explains the statistical concepts and tests commonly used in analytical method validation, including measures of accuracy, precision, confidence intervals, linearity, detection limits, and robustness. The authors integrate international validation requirements (ICH, FDA, USP) with worked examples from pharmaceutical and forensic analyses to demonstrate practical application of these techniques. The paper emphasises that a sound understanding of statistical principles is essential for demonstrating method reliability, estimating uncertainty, and ensuring analytical results are fit for their intended purpose.

Guideline for Internal Validation/Verification of Various Aspects of the DNA Profiling Process

ENFSI DNA Working Group (2023)

ENFSI-Validation-Guideline-04012024.pdf

This ENFSI guideline provides recommendations for the internal validation and verification of DNA profiling methods, instruments, kits, and software used in forensic laboratories. It recommends that laboratories prepare a validation plan outlining experimental design, sample selection, acceptance criteria, risk assessment, and data evaluation before testing begins. The document provides a structured framework for determining what evidence must be generated during validation and ensures that studies are designed to demonstrate fitness for purpose in accordance with ISO/IEC 17025 requirements.

Guidance for the Validation of Analytical Methodology and Calibration of Equipment Used for Testing of Illicit Drugs in Seized Materials and Biological Specimens

UNDOC (2009)

https://www.unodc.org/documents/scientific/validation_E.pdf

This guidance is particularly valuable during the planning phase of validation testing because it provides a structured framework for deciding what validation experiments are required before any laboratory work begins. The document emphasises defining the intended purpose of the method, selecting appropriate validation parameters (e.g., specificity, precision, accuracy, LOD, LOQ, recovery, uncertainty), and developing a documented validation plan and SOP. It also provides recommended sample numbers, concentration levels, performance criteria, and acceptance limits for different analytical techniques, helping laboratories design validation studies that are fit for purpose and aligned with quality assurance requirements.

Human- Based Methods

Performance claims in Forensic Science Expert Opinion Evidence

Smith, C., Thompson, M. (2019)

DOI: 10.38127/uqlj.v38i2

This paper argues that forensic pattern-matching disciplines often make claims about examiner abilities without clearly defining those claims in a way that can be scientifically tested. The authors reviewed professional literature across multiple forensic disciplines and found that most claims lack explicit statements about what examiners can do, under what conditions, and with what expected level of accuracy. They propose a simple framework—"I can do X, under

conditions Y, with a success rate of Z"—to help forensic sciences develop empirically testable claims and strengthen the evidence base for expert testimony.

Scientific Guidelines for Evaluating the Validity of Forensic Feature-Comparison Methods

Scurich, N., Faigman, D. L., Albright, T. D. (2023)

DOI: 10.1073/pnas.2301843120

This paper provides empirically grounded scientific guidelines for designing validation studies of forensic feature-comparison (human-based) methods. It covers study design elements including open-set design, blind testing, independent researchers, representative sample construction, and the handling of inconclusive outcomes. It specifically advocates including by design ambiguous comparisons for which inconclusive is the correct response and discusses statistical techniques for analysing performance data when multiple outcome categories (including inconclusive) exist.

Considerations when designing human performance tests in the forensic sciences

Martire, K. A., Kemp, R. I (2016)

DOI: 10.1080/00450618.2016.1229815

This paper provides practical guidance for designing rigorous studies that evaluate the performance of forensic examiners on tasks involving human judgment and comparison. The authors explain key research design principles, including clearly defining the claim being tested, establishing objective measures of accuracy, selecting representative stimuli, using appropriate participant samples, controlling for bias, and ensuring ethical conduct. Their central message is that collaboration between forensic scientists and cognitive scientists is essential to produce valid, reliable performance testing that can support claims of forensic expertise

Shining a Light on Forensic Black-Box Studies

Khan, K., Carriquiry, A. L. (2023)

DOI: 10.1080/2330443X.2023.2216748

This paper critically evaluates the design flaws in existing forensic black-box validation studies, focusing on non-representative examiner sampling and high rates of missing data (non-response). It demonstrates that reliance on self-selected examiner participants creates bias in error rate estimates and that current studies cannot support discipline-wide error rate claims. The paper outlines what constitutes appropriately designed studies per PCAST (2016): open-set design, blind testing, and independent researcher oversight.

Sample Size

Sample Size Determination and Power Analysis Using the G*Power Software

Kang, H. (2021)

DOI: 10.3352/jeehp.2021.18.17

This paper explains how to use the freely available G*Power software for sample size and power calculations across a range of statistical methods (F, t, chi-squared, z, and exact tests). It reviews the conceptual foundations of effect size, power, and alpha, and provides step-by-step examples. It argues that inadequate sample sizes prevent accurate estimates of therapeutic or measurement effects, making evidence-based decisions difficult — while excessively large samples can produce misleading statistical significance for trivial differences.