Best Practice Manual for the internal validation of probabilistic software to undertake DNA mixture interpretation

DOCUMENT TYPE :	REF. CODE:	ISSUE NO:	ISSUE DATE:
TEMPLATE	ENFSI-BPM-DNA-01	001	17 05 2017

TITLE

Best Practice Manual for the internal validation of software for DNA mixture interpretation

CONTENTS

1. AIMS

This Best Practice Manual (BPM) aims to provide a framework for procedures, quality principles, training processes and approaches to the forensic examination. This BPM can be used by Member laboratories of ENFSI and other forensic science laboratories to establish and maintain working practices in the field of forensic genetics that will deliver reliable results, maximize the quality of the information obtained and produce robust evidence. The use of consistent methodology and the production of more comparable results will facilitate interchange of data between laboratories.

The term BPM is used to reflect the scientifically accepted practices at the time of creating. The term BPM does not imply that the practices laid out in this manual are the only good practices used in the forensic field. In this series of ENFSI Practice Manuals the term BPM has been maintained for reasons of continuity and recognition.

2. SCOPE

This BPM is aimed at experts in the field and assumes prior knowledge in the discipline. It is not a standard operating procedure and addresses the requirements of the judicial systems in general terms only.

A number of different software solutions are now available with respect to mixture interpretation. In recognition of the challenges that laboratories face to implement new software, the new ISFG DNA commission [1] has recently reported a comprehensive discussion between users and developers of software in order to reach a consensus of understanding. This DNA commission was intended to cover all kinds of software for biostatistical calculations in forensic genetics. However, it is inevitable that most users will be interested in mixture interpretation of complex DNA profiles, and this guideline is written accordingly. It is also intended to be entirely consistent with the ISFG recommendations which should be read concurrent with this document.

2.1 External validation

This guideline does not cover the external validation required by a developer of software. It is assumed that this validation has been accomplished and would usually be demonstrated by peer reviewed publications (recommendations 1 and 2 of the ISFG DNA Commission). It will be a decision for the laboratory to be satisfied that the external validation is 'fit-for-purpose' within the scope of its intended use.

3. DEFINITIONS AND TERMS

For the purposes of this Best Practice Manual (BPM), the relevant terms and definitions given in ENFSI documents, the ILAC G19 "Modules in Forensic science Process", as in standards like ISO 9000, ISO 17000 and 17020 apply.

For recommendations on the correct terminology to use in relation to the kinds of probabilistic software available see the Appendix.

4. RESOURCES

Only field specific quality advice relating to the best practice manual should be outlined.

4.1 Personnel

Once a laboratory has made a decision to adopt software with the intention of implementing in casework, a person(s) should be nominated to be responsible to act as the 'local expert' with the broadest knowledge about the software. This contact point will act as the primary lead to be used by colleagues to resolve any difficulties, or questions. He/she is responsible for assessing the competence of users and to act as a conduit between the software developer and the laboratory, monitoring updates and informing users of any changes or problems that may arise (ISFG DNA Commission recommendation 8).

4.2 User manual

The software developer should create instructions on how to validate and configure software within the laboratory (ISFG DNA Commission recommendation 4) and will supply a user manual (ISFG DNA Commission recommendation 5) for end users.

.

5. Methods

Not applicable

6. VALIDATION

6.1 Validation plan

The laboratory will need to develop a documented validation plan (ISFG DNA recommendation 10) to 'scope' the kinds of samples that they wish to analyse.. This scope will take into account the different processing, analytical methods and equipment used by the laboratory where relevant to the interpretation method employed. The user should have access to examples from the developer, used in the external validation exercise. However, it is recommended that an internal validation exercise is carried out using samples that are representative of casework within the laboratory (ISFG DNA commission recommendation 11-12). In the first instance, a laboratory may prefer to restrict the scope (for example to 3 person mixtures), later expanding e.g. to 4-5 person mixtures, subject to additional internal validation. These 'mock casework' samples can be prepared from within the laboratory, utilising its standard test methods. Mock casework is essential because there is certainty about the ground truth. Real casework samples can also be used for testing, but cannot provide the sole basis of validation. Some caution is needed because strictly speaking, the ground truth is unknown with these samples.

6.2 Materials used in the validation testing

Samples chosen for use in the competency exercise should span the kinds of samples that are routinely tested within that laboratory, for example 2 person or 3 person mixtures, with and without drop-out, relatives or unrelated people. In addition, known non-contributors to a mock crime-sample should be tested (the likelihood ratio should be less than one for these examples, but there may be a false positive rate which should be characterised as a limitation of the software). The laboratory will need to make clear the limitation criteria in their standard operating procedures (SOPs).

6.3 Evaluation of results

1) Recommendation 13 of the ISFG DNA Commission states: "*The laboratory should determine whether the results produced by the software are consistent with the laboratory's previously validated interpretation procedure if the data and/or method exist."* However, if the method is completely new to a laboratory, then there may be nothing to compare – the question is to establish that the results obtained are *reasonable*, i.e. do not represent unexpected findings: for example, the likelihood ratio of a mixture cannot be greater than 1/p(G) where p(G) is the probability of the questioned reference sample of the defendant.

6.4 Establish limitations of testing

The laboratory will establish a series of criteria that define the limitations of testing. For example the 2006 ISFG DNA Commission [2] recommendation 8 stated:

"If the alleles of certain loci in the DNA profile are at a level that is dominated by background noise, then a biostatistical interpretation for these alleles should not be attempted."

If the profile of interest is predominantly below some defined level or a specified number of alleles have dropped-out (under the prosecution hypothesis) then interpretation may be deemed to fall below the specification required by the laboratory. It is important that users have a clear understanding on the limitations. To facilitate this, users must be presented with examples considered unsuitable for testing.

6.5 Probability of drop-in

The laboratory will determine the probability of drop-in according to their own analytical process(es). Drop in probability, p(C), without considering peak height, is calculated as $p(C) = \left(\frac{x}{n \times L}\right)^1$ where *x* is the number of drop-in events observed in *n* negative controls, with *L* loci tested in the multiplex (excluding amelogenin). Drop-in alleles are considered to be independent from different sources, rather than from the same individual.
For older multiplexes such as SGM plus, the original advice was to restrict the number of accepted drop-in alleles to no more than two per DNA profile [3]. Newer multiplexes have many more loci, hence the accepted number of drop-in alleles is increased. Taylor [4], fig 6, describes a method using Poisson distribution to calculate the accepted maximum number of drop-in alleles that may be incorporated into a model. A rough guideline is one drop-in peak allowed per 5 markers, hence for a 15 locus multiplex, the maximum allowed is three, whereas four or more are considered to be gross contamination.

7. PROFICIENCY TESTING

Proficiency tests should be used to test and assure the quality of mixture interpretation. Once the SOPs have been written, based on the internal validation (ISFG DNA commission recommendation 14), users will be required to undergo competency exercises (ISFG DNA commission recommendation 15), where a series of samples will be analysed that encompass the scope of the software and the kinds of samples that are to be routinely processed by the laboratory. To accommodate the requirement for external proficiency testing, it is recommended that this can be achieved by expansion of GEDNAP exercises, or any other PT provider. This would be a useful development.

8. HANDLING ITEMS

Not applicable

9. INITIAL ASSESSMENT

Not applicable

10. PRIORITISATION AND SEQUENCE OF EXAMINATIONS

Not applicable

[1] Note that the formula in [3] is given as P(C)=x/n, but the correct formula to use is per locus, as provided here.

11. RECONSTRUCTION OF EVENTS

Not applicable

12. EVALUATION AND INTERPRETATION

12.1 Training Policy

The laboratory must develop a policy to support training and competency testing, to demonstrate that persons who are involved in the reporting of complex mixtures using software are proficient in its use (ISFG DNA Commission recommendations 6 and 15). To use such software they must also be aware of the relevant theory. In particular, they must have received training with respect to:

a. Likelihood ratio theory
b. Case pre-assessment, the formulation of propositions and their limitations
c. The theory that underpins the software
d. The method of using the software
e. The limitations of the software
f. The use of (for example) non-contributor analysis, and/or other tests, in order to test the robustness of the results generated by the software
g. Statement writing

13. PRESENTATION OF EVIDENCE

1) The overriding duty of those providing expert testimony is to the court and to the administration of justice. As such, evidence should be provided with honesty, integrity, objectivity and impartiality.

2) Hypothesis building is fundamental to using probabilistic models, and users must understand the concepts. Therefore, guidance must be provided in the form of documentation supported by training.

3) The 2012 ISFG DNA commission [3] stated: *"Software tools used for casework implementation must be evaluated with known samples and each laboratory will have to establish reporting guidelines and testimony training to properly present the results to courts."* i.e each jurisdiction will follow accreditation standards, which will differ between them.

4) Evidence can be presented to the court either orally or in writing. Only information which is supported by the examinations carried out should be presented. Presentation of evidence should clearly state the results of any evaluation and interpretation of the examination.

5) Written reports should include all the relevant information in a clear, concise, structured and unambiguous manner as required by the relevant legal process. Written reports must be peer reviewed.

6) Expert- witnesses should resist responding to questions that take them outside their field of expertise unless specifically directed by the court, and even then a declaration as to the limitations of their expertise should be made.

14. HEALTH AND SAFETY

Not applicable

15. REFERENCES

[1] Coble MD, Buckleton J, Butler JM, Egeland T, Fimmers R, Gill P, et al. DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Sci Int Genet. 2016;25:191-7.
[2] Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int. 2006;160:90-101.
[3] Gill P, Gusmao L, Haned H, Mayr WR, Morling N, Parson W, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Sci Int Genet. 2012;6:679-88.
[4] Taylor D, Bright JA, McGoven C, Hefford C, Kalafut T, Buckleton J. Validating multiplexes for use in conjunction with modern interpretation strategies. Forensic Sci Int Genet. 2016;20:6-19.

16. APPENDIX

16.1 Terminology for probabilistic mixture models

The nomenclature used to describe probabilistic models can be confusing. We need *generic* terms to describe probabilistic models to interpret mixtures, as well as *specific* terms to be clear what the model does.
All models are continuous, but none is *fully* continuous. The latter implies that the model takes account of all of the variation/ characteristics in the electropherogram but this is never true. Models will take account of one or more of the following characteristics of the DNA profile (fig 1), ordered in importance of the relative effects on the likelihood ratio.

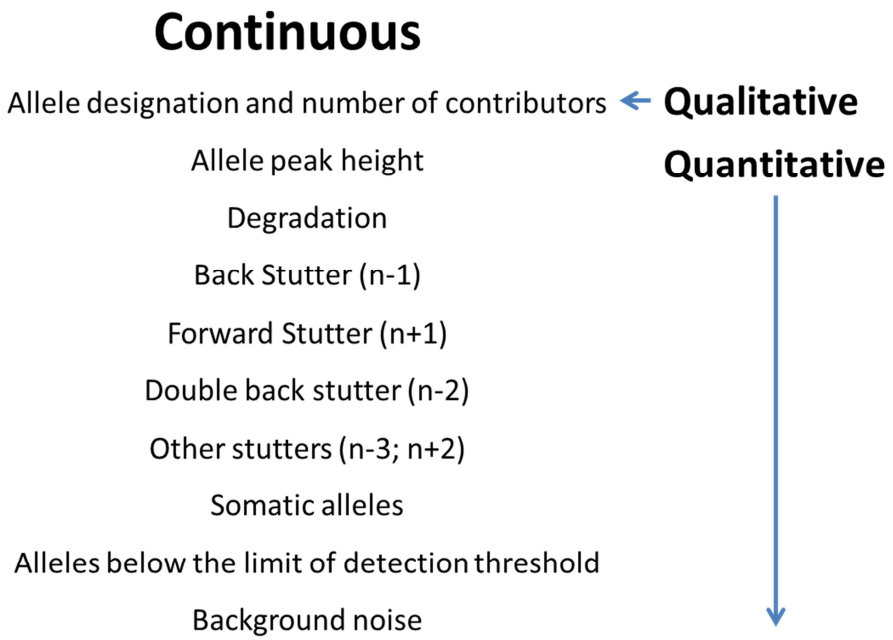

Fig 1: List of DNA profile attributes that can be incorporated to affect the likelihood ratio
Qualitative continuous model

To describe the model used to interpret, the most basic of these simply considers allele designation and the number of contributors. Possible stutters are also included in the list of alleles, which may affect the number of contributors (since ignoring stutters would violate the ISFG 2006 recommendations). This kind of model is called *qualitative continuous.*

Quantitative continuous model

The *quantitative continuous* models take account of the qualitative aspects of allele designation and will also include allele peak height as a minimum. Some models take account of one or more other aspects listed in fig 1, but none take account of all. In order to be more specific about the attributes of a particular model it can be further qualified, for example, as *quantitative continuous* with *peak height, back stutter and degradation*, which could be shortened if required to *pk.Ht; BSt;deg*.

Generic Term

Since all models are continuous, but some are more continuous than others, the generic term is simply "*Continuous*" to describe both *qualitative* and *quantitative* models

###