Ballot Name: Approval of ANSI/ASB Standard 018
Ballot URL: https://workspace.aafs.org/higherlogic/ws/groups/DNA_CB/ballots/ballot?id=61
Document Standard 018
Document Title: Validation Standards for Probabilistic Genotyping Systems

Note: a specific Proposed Resolution must accompany each comment or it cannot be considered.

#	Section	Type of Comment (E-Editorial, T-Technical)	Comments	Proposed Resolution	Final Resolution
39	Foreword	T	The following text was omitted from the OSAC document: "With multi-laboratory systems using a common protocol, internal validation may be shared by all locations." This topic needs to be addressed in the ASB document.	Reinsert the text.	Reject. Out of scope.
40	Foreword	E	"of" is missing between "current as" and "the publication" on last line of Foreword	add "of" between "current as" and "the publication" on last line of Foreword	Accept
101	General	T	The standard is generally lacking in specificity that makes is subject to abuse such that labs could conduct studies that fufill the standard but are wholly inadequate		Non-actionable comment. No specifics provided.
41	Keywords	E	DNA should be listed	add "DNA"	Partial accept: added DNA Standards
22	List of draft working group members	E	My name should be Susan A. Greenspoon, Ph.D.		No longer relevant. ASB decision to no longer include Acknowledgments in any ASB documents.
85	Overall	T	This proposed standard lacks critical requirements (see below).		Addressed by specific suggestions. See in order of comments. Column A.
24	TOC	E	Annex A-Principles is misspelled	Correct spelling "Principles"	Accept
42	Abstract	E	add "DNA" to abstract	"...the interpretation of autosomal *DNA* short tandem repeat...." or "...short tandem repeat *DNA* analysis"	Accept
62	1	T	The purpose and scope of this standard require greater clarity. The document seems to be a mixture of a standard addressing software performance and a standard addressing internal validation studies, but doesn't provide clear guidance in either direction. Based on the stated definition of probabilistic genotyping system, this standard's intent is to address validation of hardware and software that can encompass issues regarding software settings, coding, modifications, functionality, and performance checks. Throughout the standard and Annex B, the stated requirements address hardware and software. Where the document becomes confusing is that there is mention of developmental and internal validation studies, where very little detail is provided explaining the importance of such studies and how they should be conducted. If the purpose of this standard is to offer guidance to individuals familiar and unfamiliar with study design and quality assurance, then more detail on the process for conducting developmental and internal validation should be provided. The standard is not totally encompassing and explanatory of the two concepts being conveyed.	Clarify the scope/purpose of the standard and the testing that must be done to ensure laboratories are able to use the software.	See comment 3
3	1.2	E/T	The standards should be made to apply to <u>all</u> labaratories, and should apply retroactivly. To do otherwise would lead to inconsistent validation levels of software systems between labaratories, where one lab's version of the software was validated properly according to the standards set forth, while another was not, merely based on the date the software system was acquired by the lab. This poses issues of fundamental fairness to criminal defendants, whose right to have reliable evidence presented against them presumably is part of the purpose of these standards in the first place. There is no countervailing reason for not requiring retroactive validation according to these standards-- these proposals are not hugely costly or cumbersome and, even if they were, surely are worthwhile in order to endeavor to ensure reliable results. Indeed, if labaratories do not follow proper validation standards and, as a result, produce unreliable results, the entire field of probabilistic genotyping may be unfairly criticized or called into question.	Change: These standards are not meant to be applied to probabilistic genotyping systems which have been previously validated. However, laboratories are advised to review their previous validation relative to these standards. " To: "These standards are meant to be applied to probabilistic genotyping systems which have been previously validated. Laboratories are advised to review their previous validation relative to these standards. "	Accept with revision: Revise 1.2 to read: Laboratories are advised to review validation for compliance with these standards, supplement validation where necessary, and modify existing protocols accordingly.
18	1.2	T	By reading this section, it seems like the standard is not meant retroactively. The ASB standard 020 for validation of mixture interpretation protocols sounds like it IS retroactive. This contrast is a bit confusing.	The standards should take the same stance on whether the documents are retroactive or not. Make wording similar between documents to increase clarity of intent.	See # 3
55	1.2	T	The statement "These standards are not meant to be applied to probabilistic genotyping systems which have been previously validated. However, laboratories are advised to review their previous validation relative to these standards" will allow substandard validation to stand for the significant percentage of developers and laboratories which have already purported to validate these programs. The standard should be retroactive.	Make the standards retroactive.	See # 3
80	1.2	Technical	Reference to retrospective actions may stray from scientific arena to legal one. There is no scientific reason why validation studies used to report past data should not be reviewed and consequent impact assessed. Validation is a continual process.	Remove first sentence in this section. Commence with "Laboratories are advised to review....	See # 3
86	1.2	T	Standard should be retroactive. Labs across the country have completed validations of probabilistic genotyping systems already. A lack of retroactivity will mean that thousands of cases will be interpreted and reported without assurances of properly validated probablistic genotyping systems.	Add requirement that the standards apply retroactively.	See # 3
102	1.2	t	The standard should be retroactive. Many labs have begun and some have completed validations of probabilistic genotyping systems already. If this standard represents best practices, then labs should be required to conform their previous valdiations and protocols to the new standard for probablistic genotyping systems.	Add requirement that the standards apply retroactively.	See # 3
43	1.2	E	not necessary or appropriate	delete 1.2; accreditating bodies can determine how and when these standards should be applied; could move last sentence to Annex, but it seems that this statement is implied without additional relevant discussion	Partial accept. This section has been modified.See # 3
63	2	T	No software standards or practices are included in the references.	Clarify software standards and requirements.	Accept: Included one reference (#7) to software standards in Annex B Bibliography.
82	2	E	This sentence doesn't make sense. Break up? Eg. There are no normative reference documents. Annex C Bibliography contains informative references.		Accept with revision - Staff to revise.
25	3	E	Should be two sentences	Replace comma after "documents" with a period	See # 82 (applies to section 2 not section 3)
53	3	E	replace period with colon after "apply"	replace punctuation	Reject based on ASB Style Guidelines

64	3	T	SWGDAM has published clear and thorough definitions of all of the terms mentioned in this standard in the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. It seems very practical to use the terminology as defined by SWGDAM rather than create new definitions for these terms. Using the preexisting definitions promotes uniformity in terminology and meaning across documents referenced by the forensic DNA community and avoids confusion for practitioners and non-practitioners alike.	Utilize the terms and definitions as stated in the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems.	Reject. The consensus body feels that it is important that the definitions that are specific to this document be contained herein.
90	3	T	Material modification is not defined (an example or two are given).	Define material modification	Reject with revision: "material modification" removed and definition reworded for clarification.
6	3.1	E	Suggest adding a comma within sentence: "A statistical model and accompanying method that evaluates DNA profiles by assigning weights for the observed data assessing the presence or absence of allelic peaks for different contributor genotypes."	"A statistical model and accompanying method that evaluates DNA profiles by assigning weights for the observed data, assessing the presence or absence of allelic peaks for different contributor genotypes."	Accept. Modification made.
27	3.1	E	3.10, semi-continuous model-this term does not appear anywhere in the standards and therefore no definition is necessary	Delete 3.10	Accept. Definition deleted
44	3.1	E	"performed" is used twice in second sentence	revise to read "…these are studies for establishing that…"	Accept with revision. 2nd "performed" changed to "made"
45	3.1	T	Ambiguous mixtures is undefined, and its meaning may be unclear.	Suggest saying "Profiles from mixtures of 3 or more contributors are not suitable for accuracy studies."	Revised: "However, profile results where the ground truth is not known are not suitable for accuracy studies."
65	3.1	T	Greater clarity is needed for the phrase "ambiguous mixture profiles" within the definition for accuracy studies.	Provide a definition for "ambiguous mixture profiles" or provide greater detail of the samples that should not be used in an accuracy study within the text. State the limitations of the accuracy studies, particularly with respect to generalizing beyond the specific types of samples and conditions tested.	Reject. See #45
19	3.2	E	This definition of "case-type profiles" is different from "case-type samples" defined in ASB Standard 020.	Since both standards reflect regulations on validations, definitions about the same concept should be the same. Combine or alter the definitions to be more consistent between the two standards.	Reject. "Case-type profiles" is the appropriate term in this document, and the defintion is correct for the uses within the text.
46	3.2	T	Need specification that case-type profiles are generated from samples with known composition (e.g., known contributors with known genotypes, known number of contributors and mixture ratios).	add necessary specification to the definition	Reject, The specifications are contained in 4.1.2 and 4.1.3
4	3.3	T	Statement as written is that developmental validation is the accumulation of test data **within the laboratory**…. However, the Foreword to the document statest that "[d]evelopmental validation may be conducted outside the laboratory planning to use it…" Also 4.1.1 states that developmental validation may be conducted by manufacturer, etc.	move phrase "within the laboratory" from section 3.3 (developmental validation) to section 3.5 (internal validation)	Accept. phrase moved.
47	3.3	T	"within the laboratory" suggests that the laboratory must do the work rather than the developers of the software; seems to confuse developmental and internal validation	Use the original OSAC approved definition; critical to keep requirement for establishing some limitations of the system	Partial Accept. Added some text from SWGDAM definition. See comment 66/line 33 SWGDAM text reference.
66	3.3	T	The purpose of this standard appears to address internal validation so laboratories can use probabilistic genotyping software. In the event a laboratory creates its own software, developmental validation is mentioned in the standard to cover all bases. However, the term as defined and used in this document does not address all of the concerns and issues that must be addressed in developmental validation. For example is a developmental validation an evaluation of a novel technique/method? How is a laboratory accumulating test data if a developmental validation is performed by an outside manufacturer? The use of the phrase "expected values" in this definition also leads to numerous questions that are not answered in the standard. It would seem that developmental validation is where the expectation of performance (parameters and limitations) is established and an internal validation is where the performance is confirmed. Greater clarity is needed to explain the definition and the requirements in studies for developmental validation.	The definition for developmental validation from the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems should be used.	Partial Accept: Deleted text " and that the information/results/data obtained is correct and consistent with expected values." Following text added to end of definition: "Developmental validation should also demonstrate any known limitations of the system. Developmental validation may be conducted outside the laboratory planning to use it (i.e., by the manufacturer, developer, or other testing laboratory) and will precede any internal validations."
87	3.3	T	Determining the limits of a system is one of the goals of validation (and is adequately expressed elsewhere in the definitions but should be as well here).	add "determination of limitations of system"	Accept with modifications. See comment 47
88	3.3 & 3.5	T	Acquisition is used in the internal validation definition, and accumulation is used for developmental validation. Unclear why different wording.	explain why different terminology is used or make consistent	Accept: See # 5
5	3.3 and 3.5	E	Is there a difference between the "accumulation" (section 3.3) and the "acquisition" (section 3.5) of test data?	define any difference, or use one word in both instances if no difference exists	Accept: Revised with other comments on 3.3
35	3.3 and 3.5	T	The definitions of developmental and internal validation have been switched when compared to the original OSAC document.	Switch them back.	Reject. Additional public comments have required clarification of definitions from OSAC version.
26	3.4	E	3.4, fully-continuous model-this term does not appear anywhere in the standards and therefore no definition is necessary	Delete 3.4	Accepted
81	3.4 and 3.10	Technical	Uncomfortable with reference to 'fully continuous models' as many so denoted have thresholds such as limit of detection and stutter thresholds. Also explain rationale for probabilistic reasoning and why continuous and semi continuous models are useful for ambiguous DNA profiles where the PrE/Hp in the likelihood ratio evaluation cannot be 1 or 0 but is a continuous variable between 0 and 1.	Replace 3.4 with 'continuous' models and description. Consider rationale for probabilistic systems (in Foreword?) and inserting description of binary models (extra clause in section 3) and add how binary systems are unsuited for DNA profiles with uncertainty such as drop out and drop in because probabilities cannot be 1 or zero.	Reject. These definitions have been removed due to other comments.
20	3.5	E	This definition of "internal validation" is different from the same item defined in ASB Standard 020.	Since both standards reflect regulations on validations, definitions about the same concept should be the same. Combine or alter the definitions to be more consistent between the two standards.	Rejected. It is allowed to have a definition specific to a standard. This definition reflects how the term is used in this standard.
48	3.5	T	needs "by and within the laboratory" to differentiate developmental from internal validation	insert "by and within the laboratory" after "test data"	Accepted as 'within the laboratory'
67	3.5	E	For consistency purposes "within the laboratory" should be added after "data" in the definition for internal validation. The standard mentions several times that laboratories must perform internal validation; this concept should be reflected in the term. The results of an internal validation should be measured against results of a developmental validation to assess how well the software is working. Is it possible that the definitions for developmental and internal validation were switched in this document?	Change the beginning of the first sentence of this definition to: "The acquisition of test data within the laboratory to verify the functionality of the system….." The definition should include language that reflects comparison of results to parameters established in a developmental validation.	Accepted as 'within the laboratory'
89	3.5	T	Missing "in the laboratory"	Add "in the laboratory"	Accepted as 'within the laboratory'
36	3.6	T	Using "a change in the input parameters such as the number of MCMC iterations" as an example of something triggering a performance check is too broad. Is the consensus group suggesting that every increase in the number of MCMC iterations above a default and validated minimum would require a performance check?	Delete the text.	Accepted.
68	3.7	E	If the probabilistic genotyping software may not inherently produce the same statistical calculation from repeated analysis, the standard should provide some guidance on what precision studies should entail since exactness is not expected.	More clarity on precision studies should be provided in the definition.	Accepted See 103
91	3.7	T	Insufficient regarding how a lab should validate the range of values produced by the system to determine the limits of variability that will be accepted in casework.	Add back in language deleted from prior draft "studies should demonstrate the range of values that can be expected from multiple aalyses of the same data"	Accepted See 103
92	3.7	T	How much variation is okay? No concrete guidance is provided.	Prior draft contained language "..and are the basis for establishing an acceptable amount of variation in the statistical calculation" AND provide guidance about what acceptable is	Accepted See 103
103	3.7	T	Insufficient regarding how a lab should validate the range of values produced by the system to determine the limits of variability that will be accepted in casework.	Add back in language deleted from prior draft "studies should demonstrate the range of values that can be expected from multiple analyses of the same data."	Accepted

7	3.12	E	Statement as written: "Studies performed to assess the ability of the probabilistic gentoyping system to support true non-contributors. True non-contributors would correctly indicate the absence of an individual who is known not to contribute." Awkward phrasing in 1st sentence - 3.11 states how the Prob Gen system supports *the presence* of a known contributor. 3.12 is not written as supporting the *absence* of a true non-contributor, but is written as supporting a true non-contributor. 3.12 second sentence then states that true non-contributors would correctly indicate the absence.... This is awkward phrasing. A small LR might indicate the absence of a true non-contributing individual	Rethink the use of "true non-contributor" as anything other than this phrase being a person. Suggested rewrite: "Studies performed to assess the ability of the probabilistic gentoyping system to support the absence of true non-contributors. True non-contributors are those who are known not to contribute."	Accept with revision. Text modified to clarify definition.
21	3.13	E	This definition of "validation" differs in puncutation from the one in ASB Standard 020.	Since both standards reflect regulations on validations, definitions about the same concept should be the same. Combine or alter the definitions to be more consistent between the two standards.	Accept. Definition modified to match the one in Std 020. Semicolon added to match the definition in Std 20.
69	4	E	Annex B is listed as normative, yet there is no mention that it should be referenced in addition to the listed requirements.	Add a comment requiring reference to Annex B.	Accepted: Include a note in Requirements/section 4. (See note in Std 40 to ensure consistency between Std 18 and Std 40)
70	4	T	The standard relies heavily on the reader knowing all of the defined terms and what is meant by the defined terms in order to conduct a validation study. However, many of the terms' definitions are vague and confusing. Additionally, there is limited information on the quality, quantity, and variety of the samples used to generate the data that will be put into the software. The requirements are not explicit in requiring the preparation of these samples to be stressful to the software in order to find the true limitations of the system. The samples used in a validation of mixtures interpretation, which probabilistic genotyping is designed to address, should include: (1) a variety of samples with multiple contributors based on the number of contributors the lab intends to interpret; (2) a pool of participants that demonstrate the diversity of the United States; (3) mixtures created from related individuals; (4) mixtures created both from individuals that are of different ethnicities and from individuals of the same ethnicity; (4) a range of mixture ratios; and (5) degraded samples. All validation samples should be run in replicate to evaluate stochastic effects between amplifications and varied likelihood ratios calculated. The evaluation of multiple mixed samples from related individuals, degraded samples, and mixtures from the same and different ethnicities would ensure well informed mixture interpretation protocol and understanding of the values generated by the software. Lastly, the standard provides little guidance the interpretation protocol for parameters not tested during validation. For example, during validation if the lowest total amount of DNA for a three-person mixture is 1ng, is the software capable of evaluating a three-person mixture with samples lower than 1 ng?	Provide greater detail and examples of the sample preparation process that results in the data entered into the software. The Standards for Validation of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol should be referenced in section 4 or a statement should be added to Annex B detailing the samples that should be used for developmental and internal validation studies. Additionally, more guidance should be provided about the interpretation of samples beyond the limitations tested during developmental and internal validation, (or, more specifically, cautioning against the interpretation of these samples).	Partial Accept. Ammended 4.1.4 and added 4.1.5, in relation to proposition between related individuals. 4.1.3 requires the range of actual case type samples intended for analysis, 4.1.5 requires that the lab demonstrate the limitation and reliability of the softare. Further requirements on how to meet these points would be overly proscriptive.
93	4	T	There is no requirement for a separate verification of the software, i.e. the need to validate the software component of a probabilistic genotyping system separately as software or that it must meet software engineering and governance standards as comprehensively established by another body such as the Institute for Electrical and Electronics Engineers (IEEE). IEEE was specifically referenced in the 2016 ISFG probabilistic genotyping guidance document included in Annex C. This is also recognized in Haned, H., et al, *Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations* , Science & Justice, 56 (2016) 104-108; IEEE Standard for System and Software Verification and Validation, IEEE Std 1012-2012 (or latest version available).	Make separate software verification part of the standard.	Reject with revision. The verification and validation of the software is being done during the developmental validation which makes the IEEE document not necessary and should therefore not be required to implement the requirements in this Standard. IEEE Std 1012-2012 has been added to bibliography for informational purposes.
8	4.1	E	Check spacing between sentences. Document uses single space between sentences. Likely have two spaces between sentences here.	delete extra space	Accept
56	4.1	T	The President's Council of Advisors on Science and Technology listed four fundamental questions to be answered by those validating probabilistic genotyping standards in "Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Models" (Sept. 2016). This proposed standard fails to adequately ensure that a lab or developer performing a validation answers them.	These should be incorporated into the requirements of Standard 18: "1) How well does the method perform as a function of the number of contributors to the mixture? How well does it perform when the number of contributors to the mixture is unknown? 2) How does the method perform as a function of the number of alleles shared among individuals in the mixture? Relatedly, how does it perform when the mixtures include related individuals; 3) How well does the method perform--and how does accuracy degrade--as a function of the absolute and relative amounts of DNA from various contributors?....4) Under what circumstances--and why--does the method produce results (random inclusion probabilities) that differ substantially from those produced by other methods?" PCAST Report, p.79-80.	Reject: These points are covered in the document although they are not called out specifically in the terminology and order of the PCAST report.
57	4.1	T	Consideration of the number of contributors is included in 4.1.2; 4.1.3; and 4.1.5; however, there is no explicit breakdown of how this should be done, i.e. n +1, n+2.	Add requirement.	Reject. 4.1.3 requires the range of actual case type samples intended for analysis, 4.1.5 requires that the lab demonstrate the limitation and reliability of the softare. Further requirements on how to meet these points would be overly proscriptive.
58	4.1	T	Allele sharing and relatedness are not adequately specified. Although mentioned under the definition for "case-type" profiles, it should be made a specific requirement that mixtures with various degrees of allele sharing, as well as relatives, are tested.	Add requirement.	Reject. See comment 57.
59	4.1	T	Because there have been at least several examples of cases in real-life criminal trials where different probabilistic genotyping programs have generated hugely discordant results, there must be a requirement that the program be compared to others. *See* PCAST at 80; Garafano, P, D. Caneparo, et al., *An alternative application of the consensus method to DNA typing interpretation for Low Template-DNA mixtures* , Forensic Sci. Int'l: Genetics Supp. Series, 5 (2015), e422-424.	Add requirement.	Reject. It is not the place of this document to require a facility to acquire multiple programs.
60	4.1	T	Independent experts must be provided access to source code of a probabilistic genotyping. This is a fundamental component of any comprehensive review of a program without which the question: Is the program actually implementing the models/algorithms intended? cannot be answered satisfactorily.	Require that source code be disclosed to the defense in a criminal case.	Reject. This is outside of the scope of the document. It is not the place of a lab to provide proprietary source code, this would potentially violate IP laws.
61	4.1	T	Should require separate software validation/verification. This is a requirement in the standards of the the Institute for Electrical and Electronics Engineers (IEEE). It is critical that software verification be required independently from validation of probabilistic genotyping particularly given the lack of ground truth answers with complex DNA Mixtures (see Steele C.D. and DJ Balding, Statistical evaluation fo forensic DNA profile evidence" Annu. Rev. Stat. Appl. 1, 2014, 361-84.	Cross reference IEEE standards.	Reject. See comment 93
71	4.1	E	The standards specify only one requirement of the people designing and evaluating the validation studies, i.e., "knowledge in the calculation and explanation of likelihood ratios." This does not fully represent the necessary skills and knowledge required for this work.	Provide a more comprehensive list of necessary skills and knowledge, , for example, including study design methodology and quality assurance procedures.	Partial Accept. A comprehensive list cannot be fully specified, therefore, the last sentence in 4.1 has been modified to read: "The individuals designing and evaluating the validation studies should possess, at a minimum, the appropriate foundational knowledge in the calculation and explanation of likelihood ratios."
96	4.1	T	PCAST at p.80 in report stresses that a program's results should be compared to other programs. This is absolutely critical given actual casework examples where there is substantial variation among results generated by different programs. This should be part of the standard.	include requirement that program must be compared with other probabilistic genotyping programs.	Reject. See comment 59.

1	4.1.1	E/T	Validation must be performed by an expert or lab that is fully independent of the developer and the requistioning lab. This independent validation is consistent with the recommendations of the President's Council of Advisors on Science and Technology (PCAST). In its 2016 Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Models," PCAST recommended that probabilistic genotyping systems be independently reviewed before being accepted as "foundationally valid" or "valid as applied,"*see* Report, at pp. 80-82. Internationally recognized software engineering groups, including the IEEE, also recommend independent review of software programs, particularly where, as here, the software performs critical functions. See IEEE Standard for System and Software Verification and Validation, inEEE Std. 1012-2004, May 12, 2012, Appd'x B and C. The International Society of Forensic Genetics, in 2016, recognized this IEEE standard as appropriate for use in validating probabilistic genotyping software systems. See Coble M.D., J. Buckleton, J.M. Butler, T. Egeland, R Fimmers, P. Gill, et al. "DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing bio statistical calculations for forensic genetics applications," Forensic Sci. Int.: Genetics 25, 2016., n. 16. Drs. Haned and Gill propose another, also widely used, procedure for software code review. This procedure also calls for independent code review, along with comparison with other, similar software programs, and review of the "concept," "operation," and "software code itself," through "visual inspection and re-coding." See Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations, Science & Justice, 56 (2016) 104-108.	Change: " Developmental validation may be conducted by the manufacturer/developer of the application or another laboratory/agency." To: " Developmental validation may be conducted by the manufacturer/developer of the application or another laboratory/agency, as long as it also is conducted by another labatory/institute/agency that has full independence from the developer and the requestioning labaratory. "	Comment Accepted
37	4.1.2 (commenter stated 3.12, that was the incorrect section)	T	The consensus group added the following text: "that represent (in terms of number of contributors, mixture ratios, and total DNA template quantities) the range of scenarios that would likely be encountered in casework. Studies shall not be limited to pristine DNA samples but shall also include compromised DNA samples (e.g., low template, degraded, and inhibited samples)." It is my recollection that the OSAC committee specifically avoided putting this language in the developmental validation studies section, because the range of sample types tested will vary considerably between different end users. For example, one laboratory may only test samples with up to three contributors using the multiplex manufacturer's standard settings, while another laboratory may test up to six-person mixtures using enhanced detection approaches. The concern during the OSAC discussions was that the manufacturer of the probabilistic genotyping system should not have to validate all possible profile types all laboratories may wish to test. This is why that language appears only in the internal validation studies section of the OSAC document.	Delete the text.	Partial accept. Text in 4.1.1 states "Validations shall include both developmental and internal studies. Developmental validation may be conducted by the manufacturer/developer of the application or another laboratory/agency." And revision made to 4.1.3 "The internal validation shall not exceed the scope of the conditions tested in the developmental validation. Case type profiles that fall outside the range of conditions explored in developmental validation shall require additional developmental validation studies."
38	4.1.2	T	The consensus group added the following text: "that represent (in terms of number of contributors, mixture ratios, and total DNA template quantities) the range of scenarios that would likely be encountered in casework. Studies shall not be limited to pristine DNA samples but shall also include compromised DNA samples (e.g., low template, degraded, and inhibited samples)." It is my recollection that the OSAC committee specifically avoided putting this language in the developmental validation studies section, because the range of sample types tested will vary considerably between different end users. For example, one laboratory may only test samples with up to three contributors using the multiplex manufacturer's standard settings, while another laboratory may test up to six-person mixtures using enhanced detection approaches. The concern during the OSAC discussions was that the manufacturer of the probabilistic genotyping system should not have to validate all possible profile types all laboratories may wish to test. This is why that language appears only in the internal validation studies section of the OSAC document.	Delete the text.	Partial accept. Text in 4.1.1 states "Validations shall include both developmental and internal studies. Developmental validation may be conducted by the manufacturer/developer of the application or another laboratory/agency." And revision made to 4.1.3 "The internal validation shall not exceed the scope of the conditions tested in the developmental validation. Case type profiles that fall outside the range of conditions explored in developmental validation shall require additional developmental validation studies."
72	4.1.2	E/T	The standard should provide guidance on the type of testing that warrants developmental validation. Is a developmental validation a requirement of a novel method or is it a process necessary for a preexisting method where a lab or developer is attempting to establish new settings or ranges for testing? For example, does a laboratory wanting to evaluate 10 person mixtures with software already in use need to perform a developmental validation for such testing?	An explanation as to when a developmental validation should take place can be added to section 4.1.2 or it could be added to the definition of a developmental validation in section 3. Additionally, an explanation of the degree to which a developmental validation study can be extrapolated to cover conditions not tested during validation should be added to this section.	Partial Accept. Text added to 4.1.3 to address comment: "The internal validation shall be a subset of the range of cases used in the developmental validation. Cases that fall outside the range explored in developmental validation shall require additional developmental validation."
95	4.1.2 and 4.1.3	T	Effect of various degrees of allele sharing and accounting for relatedness is not specified as a requirement during both developmental and internal validation studies. Relatedness is simply a reality in casework. *See* PCAST at 79.	Include requirement.	Partial Accept: Section 4.1.5 added to address multiple propositions to include relatedness.
99	4.1.2 and 4.1.3	T	There is no requirement concerning the # of samples to be used in developmental or internal validation. While this may vary among laboratories according to intended use, there should be a minimum standard. *See* PCAST at 81 (discussing use of hundreds of distinct samples in experimental validation of important diagnostic methods in human molecular genetics).	Add standard concerning a minimum number of samples to be used in testing or how a lab must determine whether it has used a sufficient number of samples in validation.	Partial Accept. It is not possible to establish a minimum number of samples to test as this will be dependent on the DNA test used, the testing parameters used by the laboratory, the types of samples tested, etc. For example, a laboratory may choose to test only a small number of four person contributor mixtures to decide that they do not intend to interpret profiles likely to contain four or more contributors, whereas laboratories planning to interpret profiles from four or more contributors should have a large sample set of multi-contributor mixtures in their validation studies sufficient to develop and appropriately verify robust protocols. Added requirement to 4.1.3 in Annex A to clarify why a minimum number of samples cannot be used.
94	4.1.2; 4.1.4	T	The field still has not figured out how to accurately determine the number of contributors to a mixture. The true number of contributors to a complex mixture in casework is unknown. Although there is a mention of "alternate hypotheses testing", there should be an explicit requirement to test N +1 and N+ 2 contributors that the lab intends to interpret in casework. The effect of underestimating and overestimating the number of contributors on the LR generated must be part of both developmental and internal validations.*See* President's Council of Advisors on Science and Technology, "Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Models" (Sept. 2016) p.79.	Add requirement.	Accept: Language added to 4.1.5.
9	4.1.3	E	Check spacing between sentences. Document uses single space between sentences. Likely have two spaces between sentences here.	delete extra space between "laboratories." and "Studies"	Accept.
73	4.1.3	E/T	The standard does not offer guidance on the testing or interpretation of samples that fall outside of the range of those tested during the internal validation.	An explanation of the degree to which an internal validation study can be extrapolated to cover conditions not tested during validation should be added to this section.	Reject: Outside of scope of document. This comment would apply to an inperpretation protocol, not a validation protocol.
10	4.1.4	E	Check spacing between sentences. Document uses single space between sentences. Likely have two spaces between sentences here.	delete extra space between "software." and "Therefore"	Accept.
74	4.1.4	E/T	The requirement addresses alternate hypothesis testing, but the meaning of this phrase is not clear. An assumption can be made that it is meant to address mixtures where the data looks like a two person mixture but there are really three contributors, or it could mean the addition of a known contributor to the conditioning of the hypotheses. Assuming the number of contributors in a mixture is a known problem within the forensic DNA community and testing conditions where three person mixtures may present like two person mixtures or five person mixtures may present like three person mixtures should be a requirement explicitly stated.	Alternate hypothesis testing can be defined in section 3 of the standard or text can be added 4.1.4 to explain what it meant by alternative hypothesis testing. The expectation for mixtures with known numbers of contributors to be tested with alternate number of contributor hypotheses and testing with known contributor conditioning should be clearly communicated in the standard.	Accept. See comment 94

75	4.1.5	T	Greater clarity is needed for the reader to understand the expectation of this of this requirement. Is there a set or recommended number of appropriate samples that need to be tested in order to demonstrate the potential limitations and reliability of the software?	Add text clarifying the number of samples needed for the intended analysis (accuracy, sensitivity, specificity, and precision tests).	Partial Accept. See comment 99
2	4.2	E/T	In additional to publising the "mathematical basis" for the software program, the source code also shoud be either publically available or freely open to inspection upon request without any restriction. Several experts, including Drs. Balding and Steele, recommend the probabilistic genotyping software be either"open source" or "open to scrutiny." " ***Open source software is highly desirable in the court environment because openness to scrutiny by any interested party is an invaluable source of bug reports and suggestions for improvement." Steele, C.D. and Balding, D.J., Statistical evaluation of forensic DNA profile evidence, Annu. Rev. Stat. Its Appl., vol. 1, pp. 361–384, 2014.***	Change: "The underlying scientific principle(s) of the probabilistic gen otyping model and associative method and software including the mathematical basis and underlying algorithms shall be published for publication in peer-reviewed scientific journal(s)" To: "The underlying scientific principle(s) of the probabilistic genotyping model and associative method and software including the mathematical basis and underlying algorithms shall be published for publication in peer-reviewed scientific journal(s). Additionally, the software source code should be made either open source or freely avaialbe for public inspection without restriction."	Reject. See comment 60.
11	4.2	E	Awkward phrasing: "... shall be published for publication..."	"... shall be published..."	Accept: "...shall be published in peer-reviewed scientific journal(s)."
28	4.2	E	4.2-Incorrect wording at end-"published for publication"	I am assuming that the intent is that the items are published, so I suggest removing "for publication" from the sentence.	Accept. See comment 11.
49	4.2	E	extra words in last line ("for publication")	Delete "for publication" or state "...shall be published or accepted for publication in peer-reviewed..."	Accept. See comment 11.
54	4.2	E	remove "for publication"	delete additional wording	Accept. See comment 11.
97	4.2	E	Grammar error (shall be published for publication)	delete "for publication"	Accept. See comment 11.
98	4.2	T	There should be a requirement that source code and a copy of the program be made available to the defense in a case at a minimum, and to others (eg interested scientists) upon request. The developer of the program should be required to provide the developmental validation studies to the defense and others.	Add standard concerning making the source code and copies of program for each version of the program available for inspection.	Reject. See comment 60.
104	4.2	T	There should be a requirement that the source code and a copy of the program must be made available to the defense upon request. The developer of the program should be required to provide its developmental validation as well.	Add standard concerning making the source code and copies of program available for inspection.	Reject. See comment 60.
50	4.3	T	Guidelines may not provide adequate specificity and necessity; whereas protocols do.	substitute "protocols" for "guidelines" in the first line	Accept.
76	4.4	E	The standard could provide greater clarity as to when a validation should be performed instead of a performance check.	An example should be given in the requirement similar to the example given where neither a validation or performance check is needed after a software modification. Additionally, the standard should note that the source code should be open to inspection when modifications impact the analytical process, interpretation, or reported results, and documentation explaining the nature and reasons for the modifications should be retained by laboratory.	Reject. See comment 60.
77	4.5	E	Labs involved in developmental validations should also document and retain developmental validation studies.	The sentence should be changed to say all developmental validation, internal validation, and performance check studies shall be documented and retained by the laboratory.	Partial Accept. "All validation and performance check studies conducted by the laboratory shall be documented and retained by the laboratory."
23	4.6	E	This is written almost exclusively for one program, STRMix.	References to specifics should be removed, such as "random seed number". The intent of the requirement can be made without specifics to one software program.	NOTE: Believe that this is referring to text in Annex B section referring back to Requirment 4.6. Accept. Example added for other probabalaistic genotyping systems.
29	4.6	E	4.6-This appears to be a requirement intended to be addressed with each use of the system in casework. As such, it goes beyond the scope of this document. If it is meant to apply for each validation run, it belongs in section 4.1.	Either delete 4.6 or move it into the validation procedures in section 4.1.	Reject. Section is intended to state that settings are to be recorded with each validation.
51	4.7	T	"a different data set" implies only one profile or limited set of profiles is sufficient	make plural - "...utilizing different data sets than were originally used..."	Accept.
30	5	E	As these are potentially international standards, not every lab may have a "technical leader"	Add "or other appropriate personnel"	Partial Accept: Text revised to "... (or equivalent)."
52	5	E	not all laboratories have DNA technical leaders and other individuals may also be required to review and approve the documentation.	substitute "...approved by the appropriate laboratoriy authorities, and will be..."	Partial Accept. See #30
79	5	E	The standard should offer clarity as to whom is meant by assessors seeking review.	Documentation demonstrating conformance with the standard should be made readily avialable to all parties inquiring.	Partial Accept. "by the assessor" has been removed.
100	Add requirement	T	Should specify that any change to instruments or processes, etc. in the lab which can affect DNA interpretation requires re-validation. For instance, if the lab switches from 3130 CE machines to the more sensitive 3500s, there would need to be a new validation. This should be made explicit in the standard.	Add requirement that any change in laboratory testing/machinery/instrumentation/processes that could impact DNA interpretation be subject to revalidation.	Accept: modifications made to section 4.4 regarding the upstream analytical process.
12	Annex A	E	"The validation of computer software... for ... [prob gen] is a critical component of the validation process any caseworking laboratory undergoes." If a laboratory is not using prob gen, then the validation of prob gen software certainly is not a critical component of that lab's validation process.	"The validation of computer software systems used for probabilistic evaluation and interpretation of genetic information from forensic casework is a critical component of the validation process for any caseworking laboratory using such software."	Accept. Modification made.
13	Annex A	E	"Validation plans of such systems provide the study results and conclusions ..." The *plans* do not provide any results or conclusions. The plans are made prior to any validation testing actually being performed. Prior to the testing being performed there are no results or conclusions.	"Validation of such systems provide the results and conclusions..." OR "Validation of such systems should provide the results and conclusions..." (since we don't know if the validation will actually provide confidence in the prob gen system)	Accept. "plans" removed.
14	Annex A	E	"... confidence in the evidence provided." I typically think of "evidence provided" as the swab of blood, clothing, etc. Using such a definition, there is no confidence needed in the evidence provided. There is need to have confidence in the interpretation of the DNA results, and confidence in the conclusions drawn from using the prob gen software.	"Validation of such systems provide the results and conclusions..." OR "Validation of such systems should provide the results and conclusions..." "...necessary for customers of forensic science service providers to have confidence in the system."	Reject. The consensus body does not feel this wording change adds any value to the document.
31	Annex B	E	The last sentence under Requirement 4.4 is a shall statement. Should this be listed as an actual requirement under 4.4 in the standard?	Consider adding the last sentence to section 4.4	Reject. The Annex B is Normative. It is a requirement.
32	Annex B	Requirement 4.6	While this is important, as written it does not appear to apply to the validation procedure. If it does, it should be moved under 4.1 and then would not appear in this annex.	Delete this paragraph	Reject. Setting the parameters is an important part of the validation process. Annex B is normative.
33	Annex B	Requirement 4.6	If this paragraph is not deleted, add a space following "run specific parameters" in the 4th sentence.	Add missing space prior to parentheses.	Accept
15	Annex B, requirement 4.4	E	"It is impossible, for example to base a requirement..." Comma needed	"It is impossible, for example, to base a requirement..."	Accept

16	Annex B, requirement 4.4	T	"It is impossible, for example to base a requirement on changes to software version numbers or build numbers." If a lab has a requirement that absolutely any change to software version number or build number will require a complete validation, I don't see how this would be in violation of the intent of this document. It may not be the most time-effective policy, but it would meet the requirements of these standards.	"It is not recommended to base a requirement simply on changes to software version numbers or build numbers. A requirement shall be based on the list of documented changes..."	Accept with modification: A laboratory need not base a requirement for revalidation solely upon changes to software version numbers or build numbers.
17	Annex B, requirement 4.5	E	Period needed at end of paragraph	Add period at end of paragraph	Accept
83	Annex B, Requirement 4.5	E	Period missing at the end of the section (after "documented")		Accept
34	Annex C	Footnote	Reference 12 is properly listed as "latest version" of the QA Standards; however, footnote 1 would lead the reader to a specific version of the Standards, which will be obsolete next year.	Either remove footnote, or change the link to the FBI page which would list the most current version of the Standards, rather than the link to the 9-1-11 version only.	Accept. New link used.
84			Any substantive changes to the operating systems and hardware that result in changes to the probabilistic genotyping software functions should also be subject to validation.	(comment posted on Ballot, no resolution recommended)	Accept. Modifications made to include computing platform in 4.4
105	Annex A, 4.4	E	Chenge the word requirement in last 2 sentences of requirement 4.4 to another word to avoid confusion.	Change to "A laboratory does not need to perform additional validation based solely upon changes to software version numbers or build numbers. Additional validation or a performance check shall be based on the list of documented changes provided by the developer that accompany each updated version of the software installed in the laboratory."	Accept.