JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2010, Vol. 55, No. 1
doi: 10.1111/j.1556-4029.2009.01257.x
Available online at: interscience.wiley.com

Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54(4):810–21.

Sir,

We welcome the Budowle et al. (1) initiative to clarify some of the issues in relation to the interpretation of mixed DNA profiles using random man not excluded (RMNE) reporting. Whereas the ISFG recommendations (2) were based mainly on likelihood ratio principles, the Budowle et al. paper concentrates primarily on the RMNE approach. Consequently, the emphasis of the two papers is different. Although both approaches are complementary to each other and are accepted by all authors, important differences nevertheless exist that we highlight below. In particular, the statistical framework that applies to the RMNE approach does not necessarily apply to the likelihood ratio (*LR*) approach (and *vice-versa*).

There would certainly be significant advantages for the two sets of recommendations to align with each other, and we see no reason why this should not happen. To achieve this goal, we call for a constructive dialogue between all major commentators, to include ISFG, ENFSI, and SWGDAM.

Recommendations that are at variance even in minor details can only lead to confusion within the scientific community, and in court.

Definition of LCN versus Conventional DNA Profiling

There has been much discussion of characterizing DNA profiles according to whether they are deemed to be *low-copy-number* (LCN) or *conventional*. Previous authors have attempted to provide definitions relative to modified techniques (e.g., elevated cycle number). We do not believe that it is helpful to attempt delineation between two theoretical categories. This is because there is no possible natural delineator that can be used (3,4). The transition of the two "states" is gradual rather than sudden and is independent of the methodology utilized. Rather, we prefer to work towards a single integrated approach that can be applied universally across all techniques that is independent of PCR cycle number, etc. Many of the difficulties associated with interpretation are universally observed across all DNA profiling technologies (5). In its deliberations, the ISFG DNA commission (2) dealt extensively with the phenomenon of dropout, which produces the partial profile, and is the main defining feature of the low template (LT) sample. In this paper, we reiterate principles published elsewhere in order to provide guidance on the interpretation of DNA profiles that are partial.

Our main difference from Budowle et al.'s recommendations arises exactly in the area where drop-out is possible. We contend that it is the philosophy and the rigor of the likelihood ratio approach that allows the court to determine the strength of evidence, and its relevance in the context of the specific case-work circumstances. We also note with concern, that literal implementation of the Budowle et al. guidelines would result in cases unnecessarily or even wrongly being discarded as unreportable or *inconclusive*. Whereas it is tempting to believe that the Budowle guidelines would always act in favor of the defense, this is by no means a sound proposition. It is important to emphasize that many cases currently considered to favor the defense would themselves be reported as inconclusive or would inadvertently support the prosecution if the *cannot be excluded* phraseology is used.

The reality is that most real casework samples are compromised in some way and there is no perfect solution. Our purpose is to provide robust guidance for experts, that allows interpretation of real casework samples in a way that safeguards the rights of a defendant but does not unnecessarily waste evidence.

Differences Between the RMNE and LR

The RMNE statistic is not *suspect-anchored* in the same way as the *LR*. With the former, a statistic can be calculated without knowledge of the suspect's profile, but there is no indication of the strength of the evidence until there is a declaration of match with the suspect's reference sample. With the *LR* method it is necessary to condition the statistic on the suspect under the prosecution hypothesis. To summarize, the RMNE calculation is a two-step *consecutive* process: (i) is the suspect included as a contributor to the profile? (ii) Given that an inclusion is declared, what fraction of people in a defined population would also be included? (6). The *LR* method assesses the strength of the evidence relative to two alternative hypotheses. If the *LR* is less than one then the defense hypothesis is favored, whereas if it is greater than one then the prosecution hypothesis is favored. The need to designate an inclusion/exclusion is neatly avoided because the conditioning that is used in the calculation obviates the requirement for a two-step approach. Proponents of the RMNE approach see the need to condition on a suspect's profile as potentially hazardous. But this process must not be confused with the essentially mechanical aspects of the calculation of an *LR*. A scientist operating under the *LR* approach may be "blind" to information of the suspect's profile, and indeed we recommend that he/she is "blind," until such a time as the statistic is calculated (7). There is no mathematical or legal dilemma implied by such an approach.

In relation to step 1 of the RMNE approach, the following definitions are used by Budowle et al.:

- "An exclusion is declared when the reference sample has alleles that are not observed in the evidence and these unobserved alleles cannot be because of degradation within the evidence sample."
- "An inclusion is declared when the genetic results obtained from a mixture is such that the reference sample(s) can not be excluded as a part contributor of the mixed profile."

There is circularity in the definitions used above. Despite the fact that these two definitions appear exhaustive there is a third category: "inconclusive" which is defined by Budowle et al. as follows:

- "An inconclusive call can be divided into two categories: (i) those profiles that are unsuitable for comparison (other than for exculpatory purposes); and (ii) an interpretation where the profile or portion of a profile is not used for statistical purposes such as for any locus of an indistinguishable mixture when any potentially attributable allele to a single contributor(s) is below the empirically established MIT."

Our views differ somewhat from Budowle et al. on this point. We have shown that if alleles corresponding with the reference sample are not *all* present in the crime-stain evidence, then treating the evidence as an *inclusion* of the suspect or as *inconclusive* can,

in some, but not all circumstances, be seriously misleading (8). This is true even if there is a valid reason that alleles may be missing in the crime stain. Consequently, we are unable to support the use of words such as *match*, *included*, or even some interpretations of the word *inconclusive*, when describing the comparison between such a reference and evidence stain.

Our approach to interpretation follows a somewhat different, and we claim more rigorous and powerful, philosophy which is summarized below:

- Using *LR*s, there is no need to define a sample in terms of an inclusion/exclusion.
- Conversely, exclusion probabilities are usually defined in terms of *inclusion.* Typically, "If all of the alleles present in the probative reference sample are present in the crime stain then the suspect is included as a potential contributor."
- Once we consider the possibility of *drop-out*, the evidence can occur even if the contributor has alleles outside the set visualized in the mixture. Clearly, if we consider the possibility that the contributor alleles can both *drop-out* and *drop-in* then no reference sample can be excluded as a potential contributor. Consequently, we have some real difficulty to define profiles in terms of *inclusion* and *exclusion*.
- This ambiguity in declaring an *inclusion* or *exclusion* is what drives the suggestion that the comparison at that locus is *inconclusive* and cannot be reported. However, such a declaration may occur when the evidence is not neutral (8). We would be especially concerned if a locus was deemed *inconclusive* and yet it held considerable exculpatory value.
- If one locus is declared to be *inconclusive* because of potential drop-out then it follows that *drop-out* may be possible at other loci, especially at higher molecular weight. If the logic is that any locus where *drop-out* is possible is to be declared *inconclusive* then this would suggest that the whole evidential profile is potentially *inconclusive*, at least within the RMNE framework.
- Our own work suggests that the risk areas in declaring loci *inconclusive* are those where it is necessary to invoke *drop-out* in order to explain the evidential profile (C) with the suspect (S) as a donor (e.g., $S = AB$ and $C = A$). This has happened even by experienced interpretation specialists in casework (9).
- The likelihood ratio follows a very different route. A statistic cannot be calculated in the absence of the suspect's profile—two alternative propositions are considered *concurrently*. Typical examples are: (i) the crime-stain came from the suspect versus (ii) the crime-stain came from an unknown (unrelated) individual. In addition, we typically need to postulate the number of contributors to condition the calculation. In general, the most parsimonious explanations (the fewest unknown contributors needed to explain the evidence) are usually the ones that maximize the respective likelihoods under *Hp* and *Hd* (2) (caveats are given in this reference). If necessary, a range of options may be calculated. There is no reason why exploratory calculations may not be made in order to take account of multiple sets of defense and prosecution propositions. There is no single definitive statistic that can be provided. There is no definitive requirement to know the absolute number of contributors under the *LR* method because a valid/conservative statistic can be formulated based on n or more contributors.[1]
- There is little difficulty in the *LR* framework in dealing with the situation where the suspect has alleles not visualized in the mixture. If we consider a set of alleles for a mixture, C, and a suggestion, Hp, that a person, S, is a contributor, then the probability Pr($C|H$p) varies between 0 and 1. When the value ≈ 0 it strongly favors the defense hypothesis of exclusion.
- Note that the *LR* framework uses a phraseology that is itself different from the RMNE method. With the RMNE method the onus is on the scientist to make a definitive statement of *inclusion*, *exclusion*, or *inconclusive*. In the *LR* framework, we talk about evidence that either *favors the defense hypothesis of exclusion* or evidence that *favors the prosecution hypothesis of inclusion*.
- The ISFG document stated: "the RMNE approach is restricted to profiles where the profile is unambiguous." This specifically means that the suspect alleles are fully represented within the evidential crime stain and that drop-out is not possible given the heights of the represented alleles.
- There is no accepted formal probabilistic method that includes drop-out within the RMNE framework,[2] but this is relatively straightforward within the *LR* framework. The development of a mathematical model that incorporated a probabilistic treatment of drop-out and drop-in was originally published in 2000 (5), and was the foundation stone that underpinned the introduction of low-level (formerly LCN) DNA profiling. The approach has gained broad support (2). Central to this concept is the probability of dropout, Pr(D).
- Consider the situation where $S = ab$, $V = cd$, and $C = acd$ (suspect = S, victim = V, and crime-stain = C; a, b, c, and d are different alleles). Using the RMNE calculation, Budowle et al. suggest treating this locus as *inconclusive* or suggests using the $2p$ rule if the a allele was below the stochastic threshold MIT but above the PAT (using Budowle et al.'s nomenclature for T and limit of detection [*LOD*], respectively).[3] Buckleton and Triggs (8) demonstrated that the traditional $2p$ calculation can be highly anti-conservative in restricted circumstances: if the probability of allele drop-out (Pr(D)) is small (especially if it is close to zero). Of particular concern is that assigning "neutral evidence" to *inconclusive loci* can also be highly nonconservative. This extends to some mixtures where there are no minor alleles visible at all at a locus (9).
- The use of the stochastic threshold (*MIT* in the Budowle et al. paper) provides a means to estimate the level where Pr(D) ≈ 0 (see Gill et al. [12] for an experimental design and discussion on the estimation of this threshold [T] using logistical regression). If $S = ab$ and $C = a$, and $a > T$ then this would not be a reasonable proposition unless a primer binding site mutation was demonstrated (for example). See the International Society of Forensic Genetics (ISFG) DNA commission paper (2) for examples of how the *LR* varies with Pr(D) for different genotype scenarios. In Fig. 1, there are three simple examples to show how reliable reporting can proceed, relative to T, even if the allele peak heights are less than T.

[1]With the LR method, multiple pairs of propositions can be evaluated. We would like to encourage software development that facilitates rapid formulation of multiple contributor scenarios since this will undoubtedly assist the role of "scientist as facilitator" in court (see Gill et al. [10] for an example).

[2]Van Nieuwerburgh et al. (11) have attempted a solution to this problem, but their solution has the unworkable condition that you know the number of alleles that have dropped out and suffers from the restriction that it can never produce a statistic that favors the defense even when such a statistic is justified.

[3]There is an urgent need to use standardized nomenclature to avoid confusion. We suggest original terms should be used wherever possible.

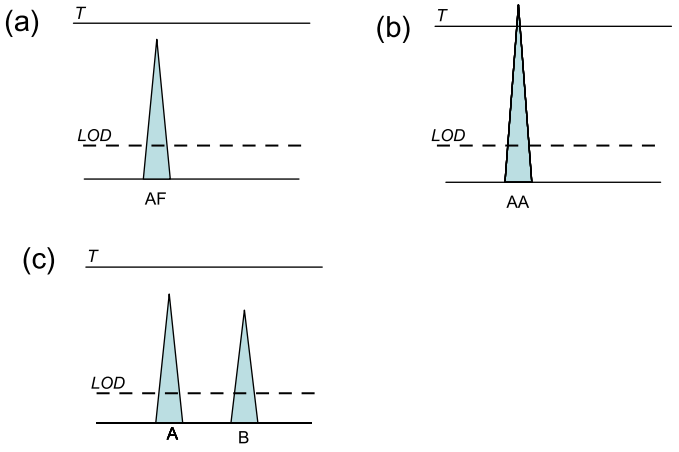

FIG. 1—*Under our scheme, the allelic designation is dependent upon the conditioning hypothesis. T is the threshold corresponding to a risk defined by Pr(D) that is chosen to be close to zero (but cannot be zero). In this example, we condition on the suspect S: (a) The 2p calculation is applied to the AF designation. This can be anti-conservative, but only if the suspect S = AB (8). If the suspect is homozygote AA, then the 2p rule is never anti-conservative. (b) If the suspect is AB, and A>T, then Pr(D) ≈ 0, which results in a very low LR favoring Hd (this could be defined as an exclusion). If the suspect is AA then the calculation is straightforward. (c) If the suspect is AB then the calculation can proceed as per normal.*

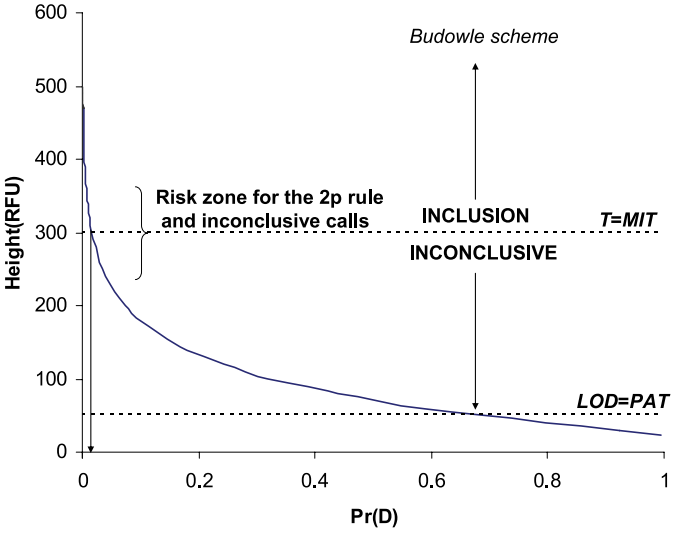

FIG. 2—*Stochastic threshold T = MIT is denoted relative to some low level of Pr(D). Note the T cannot be set to Pr(D) = 0 because* $T_{Pr(D)=0}$ *= infinity. It is inevitable, therefore, to be of any practical use, that there must be instances when an AB heterozygote will be wrongly assigned as a homozygote because (a) drop-out has occurred, (b) it cannot be guaranteed with certainty that the survivor allele is always less than T (rfu).*

Reasons for Threshold Limits

Biologists and applied scientists in many fields use thresholds to delineate between two states. Thresholds are always applied for "convenience," because the underlying models are always continuous, which means that the transition between two states is gradual. For this reason, any attempt to apply a strict threshold will always fail, because there will always be examples that will fall into the "wrong" category. Thresholds are commonly used by forensic biologists in order to delineate signal background noise from the allelic information (*LOD*, renamed and partly redefined as *PAT* by Budowle et al.). A second limit known as the stochastic threshold (*T*), renamed as *MIT* by Budowle et al., is used to delineate the level where allele drop-out may occur relative to the height of a single surviving allele. This limit needs to be assessed so that a homozygote *AA* can be distinguished from a genetic heterozygote (*AB*) where the latter is visualized as an apparent homozygote because one of the alleles has dropped out.

On the Threshold of a Dilemma

Although most labs use thresholds of some description, this philosophy has always been problematic because there is an inherent illogicality which we call the *falling off the cliff effect.* If *T* = an arbitrary level (e.g., 150 rfu), an allele of 149 rfu is subject to a different set of guidelines compared with one that is 150 rfu even though they differ by just 1 rfu (Fig. 1). Recently, Gill et al. (12) defined *T* in relation to the height of the surviving allele of a heterozygote given that *drop-out* had occurred.

Because it is not possible to provide a definitive threshold, this means that it must be set in relation to some predetermined level of acceptable risk of mis-designation. This is an unavoidable consequence that applies to all thresholds where the underlying model is continuous. Note that many labs still use the 150 rfu threshold. This level is historical and was assessed empirically in the early 1990s in relation to flat-bed gels. The sensitivity of CE instrumentation probably necessitates upward revision of this level (12) to ensure that it is consistent with Pr(*D*) ≈ 0 (Fig. 2).

The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of probabilistic models to circumvent the requirement for a threshold and to safeguard the legitimate interests of defendants. In particular it was demonstrated how Pr(*D*) could be directly incorporated into the *LR* equation itself (2) (see appendices of this reference). The ISFG DNA commission did not make specific recommendations about *how* to estimate Pr(*D*) (a new DNA commission is now under way in order to resolve this). Since publication, significant progress has been made, e.g., by Tvedebrink et al. (13), but these solutions are only possible within the *LR* framework. We cannot see how to apply such principles within the RMNE framework.

Finally, we note that Budowle et al. continue to use the LCN definition that we regard as redundant and even detrimental. They define LCN as profiles with sub-200 pg starting DNA. Most *conventional* processes will readily obtain interpretable DNA profiles at and below this level (5).[4] Our view is that it is problematic to attempt a "blanket" definition of LCN; we prefer to refer to low template (Lt-DNA) or *low-level-DNA*, based on a "loose" quantification value (3) without attempting to define a strict delineator that leads to illogical consequences.

As a simple instance, if major/minor mixtures are analyzed, the minor component will often be sub-200 pg while the total DNA concentration may be much greater. However, more generally, quantification is known to be quite inaccurate at low levels and the key matter to consider is whether stochastic effects are to be expected or not. The stochastic threshold, *T*, is used as an indicator that is set on an essentially continuous scale. But the central matter is to assess the risk associated with any particular strategy, which can be defined by a stochastic threshold (12) for any process.

[4]Perhaps it is important to point out that when the LCN term was first "coined" by Gill et al. (5), this work utilized flat-bed technology. The development of CE instrumentation, along with routine methods to concentrate DNA samples obviates the need to increase sensitivity by increasing the number of PCR cycles as the sole method. Consequently, we claim all laboratories have to deal with the "effects" of low-level DNA.

If a protocol, typically associated with Lt-DNA is used, *and* all of the evidential alleles are well-represented (above *T* evaluated for that system) then there is no need to classify the result as Lt-DNA as stochastic effects are not expected.

The stochastic *effects* of low-level DNA profiling are manifested with all protocols in current use. There is no special requirement to use a protocol with enhanced numbers of cycles (for example).

Interpretation methods are required, and are available, to take account of the effects of low-level DNA profiling that includes allele drop-out, allele drop-in, often in admixture. The characteristics of *any* system (*heterozygote balance*, *drop-out*, *drop-in* rates, etc.) can be defined with proper experimentation. These models have been published for nonmixtures and for mixtures (5,10). Consequently, in common with Budowle et al., we strongly urge due caution whenever there is the possibility of allele *drop-out* (or allele *drop-in*), regardless of the process used to obtain the result, but most especially if RMNE is to be used as the interpretation tool. However, this advice does not preclude careful interpretation of (*low-level*) results using probabilistic *LR* methods that are informed by experimental characterization of the process in use (5,10,12,14–16).

Conclusion

Although heuristic models that include the use of thresholds are unavoidable at present, we would urge continued research to determine: (i) limitations of the heuristic approach, (ii) proper experimental designs to measure the various parameters, and (iii) to design new models that incorporate a continuous probabilistic approach—where these models can be used: (i) to validate/modify existing heuristic models, (ii) to replace heuristic models, either in part or entirely. It is unlikely that the RMNE method can be developed further to accommodate additional probabilistic theory, but there is no such limitation to the *LR* approach. It is quite important that interpretation limitations diagnosed using the RMNE approach are not improperly applied to the more powerful and rigorous *LR* approach.

Acknowledgments

We acknowledge the comments of Sue Vintiner and Jo-Anne Bright that have greatly improved this letter.

References

1. Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, et al. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54(4):810–21.
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, Brown R, Fairley M, Lee L, Smyth M, Simpson N, et al. National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. Forensic Sci Int Genet 2008;2(1):76–82.
4. Buckleton J. Validation issues around DNA typing of low level DNA. Forensic Sci Int Genet 2009;3(4):255–60.
5. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less that 100 pg of DNA. Forensic Sci Int 2000;112(1):17–40.
6. Buckleton J, Curran J. A discussion of the merits of random man not excluded and likelihood ratios. Forensic Sci Int Genet 2008;2(4):343–8.
7. Clayton TM, Whitaker JP, Sparkes R, Gill P. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci Int 1998;91:55–70.
8. Buckleton J, Triggs CM. Is the 2p rule always conservative? Forensic Sci Int 2006;159:206–9.
9. Balding DJ, Buckleton JS. Interpreting low template DNA profiles. Forensic Sci Int Genet. In press.
10. Gill P, Kirkham A, Curran J. LoComatioN: a software tool for the analysis of low copy number DNA profiles. Forensic Sci Int 2007;166(2–3):128–38.
11. Van Nieuwerburgh F, Goetghebeur E, Vandewoestyne M, Deforce D. Impact of allelic drop-out on evidential value of forensic DNA profiles using RMNE. Bioinformatics 2009;25(2):225–9.
12. Gill P, Puch-Solis R, Curran J. The low-template DNA (stochastic) threshold—its determination relative to risk analysis for national DNA databases. Forensic Sci Int Genet 2009;3(2):104–11.
13. Tvedebrink T, Eriksen PS, Mogensen HS, Morling N. Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Sci Int Genet 2009;3(4):222–6.
14. Curran JM, Gill P, Bill MR. Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. Forensic Sci Int 2005;160:47–53.
15. Gill P, Curran J, Neumann C, Kirkham A, Clayton T, Whitaker J, et al. Interpretation of complex DNA profiles using empirical models and a method to measure their robustness. Forensic Sci Int Genet 2008;2(2):91–103.
16. Gill P, Curran J, Neumann C. Interpretation of complex DNA profiles using Tippett plots. Forensic Sci Int Genet Suppl Ser 2008;1(1):646–8.

Peter Gill,[1] Ph.D. and John Buckleton,[2] Ph.D.
[1]University of Strathclyde, Glasgow, UK
[2]ESR, P.B. 92021, Auckland, New Zealand
E-mail: peter.gill@strath.ac.uk