JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2010, Vol. 55, No. 1
doi: 10.1111/j.1556-4029.2009.01258.x
Available online at: interscience.wiley.com

Authors' Response

Sir,

The Gill and Buckleton letter (1) focuses on low copy number (LCN) typing and recommends use of the likelihood ratio (LR) for LCN mixture interpretation over the random man not excluded (RMNE) approach. In contrast, the Budowle et al. (2) paper describes guidelines for interpreting mixtures for robust and reliable results from profiles where allele drop-out does not occur or where the probability of drop-out is so low such that it can be generally ignored. Gill and Buckleton (1) ignore this and deal with interpretation of mixed profiles from LCN samples producing peaks below the match interpretation threshold. Budowle et al. (2) describe the use of thresholds to identify profiles (and partial profiles) that are usually reliably reproduced. Particularly, Budowle et al. ([2], on page 821) stated "we strongly urge caution with mixture interpretation with any LCN typing. The interpretation guidelines described above do not apply to LCN typing." Gill and Buckleton (1) question the value of thresholds that define the regions where the probability of drop-out is exceedingly unlikely even while acknowledging "the use of thresholds is unavoidable at present." They do not address the issues or merits of non-LCN mixture profile interpretation guidelines and thus their discussion of the Budowle et al. (2) paper is neither germane nor appropriate.

Gill and Buckleton (1) state that "the Budowle et al. paper concentrates primarily on the RMNE approach." Even a cursory reading of the Budowle et al. (2) paper will show this to be incorrect and will also show that most of the paper is dedicated to interpretation and not statistical weight assessment. Moreover, both RMNE and LR approaches are discussed and supported. Budowle et al. (2) stated

"There are two approaches available for rendering an estimate."

"We support that forensic scientists should be trained to calculate either statistical approach; but do not support that the LR is a preferred method that must be captured in the notes. It is clear that the significance of some mixtures may not be easily calculated using the LR, such as some mixtures with three or more contributors. Instead, we support the position of the DAB: 'Rarely is there only one statistical approach to interpret and explain the evidence. The choice of approach is affected by the philosophy and experience of the user, the legal system, the practicality of the approach, the question(s) posed, available data, and/or assumptions. For forensic applications, it is important that the statistical conclusions be conveyed meaningfully. Simplistic or less rigorous approaches are often sought. Frequently, calculations such as the random match probability and probability of exclusion convey to the trier of fact the probative value of the evidence in a straightforward fashion. Simplified approaches are appropriate, as long as the analysis is conservative or does not provide false inferences. LR approaches compare mutually exclusive hypotheses and can be quite useful for evaluating the data. However, some LR calculations and interpretations can be complicated, and their significance to the case may not be apparent to the practitioner and the trier of fact.' Also the DAB stated 'The DAB finds either one or both PE or LR calculations acceptable and strongly recommends that one or both calculations be carried out whenever feasible and a mixture is indicated.' This is a more balanced position and is more practical for addressing the various mixture profiles that may be encountered. It is better to use what is best determined to be meaningful for assessment and/or for communication by a laboratory. However, whatever is used must be clearly documented in the standard operating protocol and any assumptions impacting the calculation should be recorded."

Clearly, Budowle et al. (2) do not advocate the sole or primary use of the RMNE method. In a more balanced approach, the merits of both RMNE and LR are presented.

Gill and Buckleton (1) attack the use of thresholds (a concept used by many who perform STR typing), and the use of terms, such as inconclusive, match, and LCN even though they agree that the use of "heuristic models that include use of thresholds are unavoidable at present." The tone of their letter seems to imply that the LR they advocate is the only way to apply a statistical analysis. In contrast, we agree with alternate comments made by Gill recently when he testified in *NY v Megnath* (3) as follows:

On page 507

Q Does the fact that there is a diversity in the practice and one lab may not follow the recommendations of the ISFG, undermine the reliability of a particular technique that the lab has validated?

A No. That would be unfair. You have to bear in mind that techniques are constantly changing. That is the nature of science, and forensic science is a science. And therefore by definition it is constantly changing. By definition there will be a variety of viewpoints. But I like to think of these viewpoints as being generic rather than specific. So the guidelines that we introduce tend to be generic guidelines. When you have generic guidelines then you can expect a diversity and you can justify diversity of practice.

The forensic DNA community should strive to improve. We welcome a better approach, if it is objective, reliable, and robust. It is noteworthy that Gill and Buckleton (1) do not describe how interpretation is carried out currently in their respective laboratories (the FSS formerly for Gill and the ESR for Buckleton). It would be beneficial to the forensic community if these protocols for interpretation and statistical weight analysis were made available.

Gill and Buckleton (1) advocate the elimination of thresholds because there is no absolute cut-off for a threshold. It is true that the point where the probability of allele drop-out increases above 0 cannot be precisely defined. However, a lack of an absolute value does not support the concept that the use of thresholds is wrong. Many scientific and diagnostic procedures, along with most forensic DNA protocols, employ thresholds. Indeed, thresholds are used in many other forensic applications.

It is equally important for DNA typing to use thresholds. Yet, under the Gill and Buckleton (1) approach there is no threshold and it is left to the discretion of the analyst. This concept could be taken to an extreme. Consider a mixture profile with peaks with heights between 5 and 10 rfu. It is conceivable under the no threshold approach that an analyst could choose those peaks that appear slightly above noise as true alleles and if the suspect has alleles that are at those positions, he might be included. At the other loci in the mixture profile where the suspect's alleles are not observed, he could still be a potential contributor because the analyst asserts the missing alleles are lost in the noise and the other observed alleles are from other contributors. While this is an extreme example, it makes the point that a non-threshold approach can be problematic particularly if there is an inclination or bias towards fitting the suspect profile to the evidence. Such a profile should not be interpreted.

Budowle et al. (2) define practices that avoid situations where allele drop-out can occur so that a robust approach can be

developed. Interpretation is essential to a sound forensic DNA typing system. Instead, Gill and Buckleton (1) state "it is not helpful to attempt to delineate between two theoretical categories." This is not a theoretical issue. In their continuum, the opposing ends behave very differently. One end is highly reliable and reproducible with little evidence for allele drop-out and the other end of the continuum experiences all the vagaries of LCN typing. A threshold can be readily drawn where the former is essentially guaranteed and this provides a robust environment for interpretation. Although it may result in more inconclusive calls, it does minimize false inclusions. This "bias" is in keeping with the premise of our judicial systems.

Gill and Buckleton (1) suggest that a threshold is equivalent to the fall off the cliff syndrome. They state "the literal implementation of the Budowle et al. guidelines would result in cases unnecessarily or even wrongly being discarded as unreportable or *inconclusive*." Although in some cases a defendant *might* have been included but by having a threshold the interpretation was deemed inconclusive, the threshold approach is in keeping with an important premise of our justice system, namely keeping innocent persons out of jail at the expense of allowing guilty ones to escape punishment. Their fallacious argument asserts that some results that we would call inconclusive (which *they* have deemed should be neutral) could actually exclude a defendant and these would be ignored. Gill and Buckleton (1) apparently ignored the guidelines of Budowle et al. (2) where it is conveyed that interpretation is not symmetrical; for example on p. 818 Budowle et al. (2) describe a scenario where the interpretation would be inconclusive but clearly state that "the alleles could be used only for exclusionary purposes." An interpretation of exclusion is allowable even if the locus is deemed inconclusive for inculpatory purposes. We have always supported that the DNA profile or its loci should still be considered for exculpatory value even if the initial interpretation was deemed inconclusive. Therefore, the Gill and Buckleton (1) literal approach does not directly support their position because exculpatory evidence would not be lost with the threshold approach described in Budowle et al. (2). Misleading and unsubstantiated statements by Gill and Buckleton (1) such as "any attempt to apply a strict threshold will always fail, because there will always be examples that will fall into the 'wrong' category" are of little value. They do not describe what constitutes a failure under the Budowle et al. (2) guidelines. However, one could argue that without a threshold a failure, namely false inclusion of a suspect, is more likely to occur. Yet Gill and Buckleton (1) do not seem to consider that point. We repeat that we are willing to discard some profiles as inconclusive to ensure confidence in the interpretation of a failure to exclude instead of pushing the envelope and potentially falsely including a suspect or overstating the strength of the evidence in the unreliable zone where LCN typing results would be assessed. Until a robust approach is developed we again stress the need to exercise caution.

Clearly, the term "inconclusive" is a difficult concept for Gill and Buckleton (1). They state "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." In the Budowle et al. (2) guidelines a locus is deemed inconclusive prior to the comparison, unlike the approach that seems to be advocated by Gill and Buckleton. It is vital to avoid interpretation bias based on the suspect's profile. The Gill and Buckleton (1) opinion on use of the term inconclusive conflicts with recent testimony by Gill in *NY v Meganth* (3) where these same words and concepts are discussed:

On page 475

A… profile, you either don't get one, or the one that you get is so partial as not to be usable or horrendous mixture which you can't possibly interpret or evaluate.

Q And again does that impact upon the reliability of the science when a profile is in fact obtained?

A Not in my opinion, no, it's just a fact of life.

On page 553

A… The question before the Court is, does the DNA profile match the suspect. The second question is, how did it get there.

On page 658

A… the statistic will become small and really a very, very limited value. And we would probably call those kinds of profiles inconclusive (note the transcript contains "inclusive" but it is obviously a transcription error and should be inconclusive).

The transcript is an example of contradiction by Gill of his own recommendations presented in the letter. Perhaps such terms are useful for communicating interpretations. Instead of arguing over semantics which is not particularly helpful herein, the terminology used should be defined as recommended by Budowle et al. (4). This would better reduce the inconsistencies of word usage displayed by Gill and Buckleton (1).

To help clarify the misunderstanding of the use of inconclusive, a short discussion is in order. Inconclusive is a broadly used interpretation to cover a variety of situations where the result obtained does not fall within established criteria for rendering a conclusive interpretation. The term inconclusive is not exclusive to DNA typing; it is used widely in the forensic sciences and also by the broader scientific diagnostic arena. Human clinical diagnostics, veterinary diagnostics, food testing, and environmental testing are all applications of science that use the term inconclusive in reporting unsuccessful laboratory results (for examples see [5,6]). The general meaning of the term is that the result does not meet laboratory established criteria, based on validation studies, for rendering a conclusive result. Our justice system, however, works to avoid wrongly associating an individual with an evidence sample. The interpretation of forensic evidence can therefore also be weighted in a similar fashion. When an inconclusive result yields DNA data below accepted interpretation criteria (but above the peak amplitude threshold), it is imperative that a suspect be excluded if the data support such an interpretation. In other words, exclusionary evidence should be declared even if below the MIT (2).

Gill and Buckleton (1) suggest the use of 2p for situations of potential allele drop-out may not be conservative. Their argument is fallacious since the 2p rule would not be applied as they infer. More importantly, the locus has been reviewed *a priori* and determined that drop-out may have occurred (as opposed to the approach of asserting allele drop-out after observing the suspect's alleles are missing at the locus). In fact, if the locus *a priori* was deemed to have all alleles present (so there is no evidence of allele drop-out) and the suspect did not have those alleles, the suspect should be excluded. This is the advantage (for confidence) of interpreting profiles where drop-out is unlikely. In such scenarios $2p > 2p_1p_2$ and $2p > p^2$. Moreover, if allele drop-out is essentially 0, then there would be no need to invoke use of the 2p formula. We also note that Gill et al. (7) advocated the 2p approach for the LR for single source LCN samples.

Interestingly, the Gill and Buckleton (1) letter seems to focus on mixture samples, and yet Fig. 1 only illustrates single source profiles, which often are not as problematic to interpret. They show no examples of how to accommodate mixtures with a high probability

of allele drop-out, peak imbalance, and exaggerated stutter with 3, 4, 5, 6, 7, or more alleles per locus (i.e., complex mixtures that will have additive effects to consider). The problems arise with mixed samples and particularly those interpreted in the LCN area. It is allele drop-out and the inability to use quantitative data with LCN typing that are difficult to address.

The loci where missing alleles can occur should be designated before considering the suspect's profile. This is an important difference that Gill and Buckleton (1) fail to consider. In reality addressing the probability of allele drop-out is a serious problem that has yet to be adequately developed for LCN typing protocols. The hypotheses of the LR may not have been well considered given that allele drop-out may be highly probable in LCN typings, especially for mixed samples. Mixed samples are comprised of two or more individuals and therefore the individual contributors on average are at much lower levels of DNA than would be generally for single source samples of similar template quantities. So allele drop-out is likely in most if not all mixture cases with LCN typing. These considerations also imply that the possibilities of allele drop-out are not independent across loci. With these in mind, it is not clear how Gill and Buckleton would actually carry out an interpretation and statistical analysis; they did not give any real case examples let alone any hypothetical examples. Interestingly, a recent communication by Keith Simpson from the Crown Prosecutor Service in a response dated (8–19–09) to a request for discovery materials in a FSS case (*Queen v Reed and Reed*) stated "The FSS do not rely on probability data for allele drop-out/allele drop-in and stutters in the statistical analysis." So we are in a quandary as to what are the acceptable interpretation approaches in the LCN typing laboratories. However, it is clear that Gill and Buckleton must rely on the suspect's profile to select the loci where allele-drop-out may have occurred. One could surmise that to make their approach work they might perform the following. An LCN evidence profile is generated. There are no thresholds, so the analyst decides what peaks represent true alleles. Then the suspect profile is reviewed. There is little or no *a priori* assessment that states which loci of the mixture profile will have experienced allele drop-out. Allele drop-out is only determined as a possibility after observing the suspect profile. Consider in one case a 10-locus STR kit is used and alleles are present and scored at all loci. All the suspect's alleles are observed (along with other alleles) at eight of the loci and two loci (in this case D2S1338 and D19S433) do not display any of the suspect's alleles. In the next case, a similar scenario occurs; however the D8S1179 and FGA loci do not display the suspect's alleles. In another case it could be another two loci where the suspect's alleles are not observed. The prosecution hypothesis in each of these scenarios could be that the suspect is the source of the evidence and the missing alleles are due to allele drop-out. The probability of observing the evidence under this hypothesis might use the probability of allele drop-out (of which to date valid values have yet to be generated) as suggested by Gill and Buckleton (1) and would be applied at the two missing allele loci. This is where bias creeps into the assessment. It is possible in this scenario that any loci can experience allele drop-out when using LCN typing and not until the suspect is typed are the loci selected for some compensation. This "sliding window allele drop-out phenomenon" is a result of reverse conditioning on the suspect's profile and is problematic.

The situation would be exacerbated if there were three suspects all compared with the same evidence and each of their alleles are observed in the evidence (i.e., part contributors of the mixture) at all but two loci and the two loci differ for each suspect. The evidence does not support that all three are contributors but that one or two might possibly be. The problem is that there are now six loci where allele drop-out must be considered. Lastly, if allele drop-out is suspected (or observed) at one locus, then there is a high probability that allele drop-out will occur at another locus. Drop-out is not independent; it is impacted by the quality of the evidence. One cannot assume that allele drop-out is independent across loci; that assumption would be erroneous. Gill and Buckleton (1) recognize the dilemma when using LCN typing; they state "Clearly, if we consider the possibility that the contributor alleles can 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 is the real dilemma for LCN typing interpretation, i.e., it cannot be used for exculpatory purposes. They do not describe what to do. If the loci with drop-out can change with each suspect instead of based on the quality of the evidence, then all loci should have a compensation for allele drop-out and not just those that do not have alleles shared with the suspect (or some other appropriate compensatory approach could be exercised that does not apply reverse conditioning). We note that there is no defined number of loci espoused by Gill and Buckleton (1) where missing alleles can be tolerated. It is conceivable it could range from one to almost all loci. At this time it is left to the judgment of the reporting scientist. The defense hypothesis could be: the loci with the alleles that are shared with the suspect are mixtures and those two alleles in common with the suspect are from two unknown individuals with their partner alleles dropped-out. The defense also could assert that the number of contributors is much greater than proposed by the laboratory-generated defense hypothesis and if not considered the LR is inflated. We do not know if this bias is addressed or even considered by practitioners of the LCN technology; Gill and Buckleton (1) have not described it in their letter.

Gill and Buckleton (1) state "Note that many labs use the 150 rfu threshold. This level is historical and was assessed empirically in relation to flat-bed gels." This is another misleading statement by Gill and Buckleton (1). Gill and Buckleton seem unaware that there is likely a good portion of laboratories that did not implement STR typing until capillary electrophoresis instruments were in place. Therefore, their thresholds were not based on flat bed gel electrophoresis. Indeed, the FBI lab set its threshold based on capillary electrophoresis systems. So these thresholds can be quite relevant. Regardless, it is immaterial since laboratories are required to do internal validations to set these levels. A major technology change (e.g., from flat bed gel to CE) would require establishment of new thresholds.

It seems that the Gill and Buckleton (2) letter is more about a forum for LCN typing as opposed to whether mixture interpretations above a threshold are reliable. They state "we note that Budowle et al. continue to use the LCN definition that we regard as redundant and detrimental. They define LCN as profiles with sub-200 pg starting DNA… we prefer to refer to low template (Lt-DNA) or low-level DNA, based on a 'loose' quantification level." This is an odd and trivial statement, but worth noting its misdirection. Budowle et al. (2) hardly mention LCN typing and only stress that their guidelines do not apply to LCN typing. The emphasis of the Budowle et al. (2) paper is guidelines for the robust regions of DNA typing and to proceed with extreme caution when interpreting profiles where stochastic affects are exacerbated. Indeed, it was Gill et al. (7) who coined the term LCN typing. Budowle et al. (8) used the same term many years ago because they wanted to convey that there are methods in addition to increasing cycle number that can increase sensitivity of detection and enable visualization of profiles with exacerbated stochastic effects from a very limited amount of template DNA. Budowle

et al. (2) simply are using the same terminology for communication and reference purposes. So, we wonder how the issue of the term LCN as being redundant or detrimental adds anything to the discussion on reliably interpreting mixed profiles. Moreover, all should be aware that the 200 pg level was described by Caddy et al. (9) in their recent commissioned review of the current LCN practices after the *Queen v Hoey* case (10). Also interesting is that Gill et al. (11) describing quantitative models as useful (relating similarly to the above threshold approach described in Budowle et al. [2]) use the same threshold of template DNA stating "The quantitative models have primarily been developed for profiles where there is a significant amount (>200 pg) of DNA present." Had Gill and Buckleton cited Budowle et al. (12) they would have conveyed that we already have addressed this issue—"These quantitative threshold values are based on an amount of template DNA where peak height imbalance becomes exaggerated and are relative to specific assays, kits, and methodologies. The value will change with technology and genetic markers typed and the 200 pg threshold therefore will not necessarily apply to all systems. More likely a heterozygote peak height imbalance ratio may be a better criterion for a stochastic threshold."

Conclusions

The Gill and Buckleton (1) letter does not actually point to any problems with the Budowle et al. (2) guidelines. Indeed, they say that thresholds are unavoidable but suggest working towards an integrated approach that accommodates highly reliable as well as highly unreliable data. This integrated approach is not readily accomplished and to date no one has offered a robust solution. We welcome any methods that will enable a robust and unbiased (i.e., not based on suspect's profile) approach. Until then it would be prudent to continue with the practical and reliable approach that most forensic analysts use.

We also caution cognizance of the misleading and incorrect statements made by Gill and Buckleton (2); they are neither helpful nor constructive. For example, they state "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." This statement cannot be justified. It is inconsistent with what Gill and Buckleton (1) stressed throughout their letter. They do not advocate use of a threshold; so there can be no T. Yet, they invoke one. However, more troubling is when using LCN typing conditions (where allele drop-out is prevalent, quantitative analysis is invalid, the amount of template DNA is low, and a 34-cycle approach advocated by Gill et al. [7] was used) how can Gill and Buckleton (1) ever say "all the evidential alleles are well represented?" The sample is an unknown. They can say all the alleles that the suspect has are observed in the profile but they cannot say all the alleles are present under LCN typing conditions. Again a bias in interpretation is illustrated.

Lastly, it would make for a better discussion if the policies and protocols for LCN mixed sample interpretations and statistical assessments were made available. We encourage all those who use LCN typing in forensic analyses to make their protocols publicly available so a discussion of the merits and limitations of actual practices can ensue. LCN typing procedures' reliability is intimately related to the interpretation and statistical analysis employed (more so because the molecular biology analytical results are not reproducible).

References

1. Gill P, Buckleton J. 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 2010;55:265–268.
2. 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;54:810–21.
3. Gill P. *Transcript, NY v Megnath*, Indictment No. 917-07. New York: Frye Hearing, December 18–19, 2008.
4. Budowle B, Bottrell MC, Bunch SG, Fram R, Harrison D, Meagher S, et al. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J Forensic Sci 2009; 54:798–809.
5. Wax TDCA, Johnston WW. Clinical significance of an inconclusive cytopathologic diagnosis: a five-year experience at Duke University Medical Center. II. Analysis and follow-up of specimens from the respiratory tract. Acta Cytol 1997;41(4):1053–7.
6. Mathieu IMS, Willemart B, Destine M, Mazy G, Lonneux M. Inconclusive triple diagnosis in breast cancer imaging: is there a place for scintimammography? J Nucl Med 2005;46(10):1574–81.
7. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 2000;112:17–40.
8. Budowle B, Hobson DL, Smerick JB, Smith JAL. Low copy number—consideration and caution. Twelfth International Symposium on Human Identification; 2001 Oct 9-12; Madison, WI: Promega Corporation, 2001, http://www.promega.com/ussymp12proc/default.htm.
9. Caddy B, Taylor DR, Lincare AMT. A review of the science of low template DNA analysis. 2008, http://police.homeoffice.gov.uk/publications/operational-policing/Review_of_Low_Template_DNA_1.pdf?view=Binary.
10. *The Queen v Sean Hoey*, Neutral Citation Number [2007]. NICC;49. http://www.xproexperts.co.uk/newsletters/feb08/R%20v%20Hoey.pdf.
11. Gill P, Kirkham A, Curran J. LoComatioN: a software tool for the analysis of low copy number DNA profiles. Forensic Sci Int 2007;3:128–38.
12. Budowle B, Eisenberg AJ, van Daal A. Validity of low copy number typing and applications to forensic science. Croatian Med J 2009;50(3): 207–17.

Bruce Budowle,[1] Ph.D.; Ranajit Chakraborty,[2] Ph.D.; and Angela van Daal,[3] Ph.D.

[1]Department of Forensic and Investigative Genetics, and Institute of Investigative Genetics, University of North Texas Health Science Center, Ft Worth, TX.

[2]Center for Genome Information, Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH.

[3]Faculty of Health Science & Medicine, Bond University, Gold Coast, Qld, Australia.

E-mail: bbudowle@hsc.unt.edu