Forensic science, reliability and scientific validity: Advice from America

Tony Ward,[∞] Gary Edmond[†], Kristy A. Martire[§] and Natalie Wortley[¶]

1. Introduction

In this article we review an important report produced by the President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (the PCAST report).[1] The PCAST report builds on an earlier report prepared by the National Research Council, *Strengthening Forensic Science in the United States: The Path Forward* published in 2009 (the NRC report).[2] These reports are focused on the organisation, funding and practice of the forensic sciences in the United States. In their deliberate and unflinching concern with probative value, particularly the validity and reliability of procedures used by forensic scientists and the way opinions are expressed in expert reports and testimony, both have application to England and Wales. Both reports speak directly to forensic scientists, law enforcement, lawyers and courts. Forensic scientists, advocates, judges and legislators must respond to criticisms and recommendations if we hope to place the forensic sciences on firm scientific foundations.[3]

PCAST 'is an advisory group of the Nation's leading scientists and engineers, appointed by the President' to advise on 'the full range of issues where understandings from the domains of science, technology, and innovation bear potentially on the policy choices before the President.'[4] The PCAST report is primarily oriented toward forensic feature-comparison procedures. Used routinely in the US, as well as England and Wales, these procedures (or methods) 'aim to determine whether an evidentiary sample (e.g., from a crime scene) is or is not associated with a potential source sample (e.g., from a suspect) based on the presence of similar patterns, impressions, features, or characteristics in the sample and the source.'[5] Typically they involve comparison and some kind of 'matching'. They include 'the analysis of: DNA, hair, latent fingerprints, firearms and spent ammunition, tool and toolmarks, shoeprints and tire tracks, bitemarks, and handwriting' as well as voices and images.[6] Following on from concerns about the level of scientific support for forensic

---

[∞] Professor of Law, Law School, Northumbria University.

[†] Professor, School of Law, UNSW and Research Professor, Law School, Northumbria University. This research was supported by the Australian Research Council (LP16010000).

[§] ARC DECRA Fellow, School of Psychology, UNSW. This research was supported by the Australian Research Council (DE140100183).

[¶] Principal Lecturer in Law, Law School, Northumbria University.

[1] Executive Office of the President of the United States, 2016, available at https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcastforensicsciencereportfinal.pdf.

[2] The NRC report is also known as the National Academy of Sciences (or NAS) report. The PCAST report is, like the NRC report, critical in tone. PCAST was assessing reforms, following the NRC report, to assist institutions such as the newly established National Commission on Forensic Science.

[3] The reports recognise that some techniques (e.g. some chemical analyses and basic DNA profiling) are robust. Concern and criticism was primarily directed at the research base, along with (uncritical) legal acceptance and reliance.

[4] PCAST report, iv. For a list of the Council, the working group and advisors, see PCAST report v-ix.

[5] PCAST report, 23, Box 1.

[6] PCAST report, 23, Box 1, and 27ff. On voices and images, see David Ormerod, 'Sounds Familiar? Voice Identification Evidence' [2001] *Criminal Law Review* 595; David Ormerod, 'Sounding Out Expert Voice Identification Evidence' [2002] *Criminal Law Review* 771; Gary Edmond, Kristy Martire and Mehera San Roque, 'Unsound law: Issues with ("expert") voice comparison evidence' (2011) 35 *Melbourne University Law Review* 52; Geoffrey Morrison, 'Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, and approaches to forensic voice comparison' (2014) 54 *Science & Justice* 245; Geoffrey Morrison and William Thompson, 'Assessing the Admissibility of a New Generation of Forensic Voice Comparison Testimony' (2017) *Columbia Science & Tech LR* (forthcoming); Gary Edmond and Natalie

science practices expressed in the NRC report, PCAST sought to fill a conspicuous gap by examining 'the fundamental scientific validity and reliability of many forensic methods used every day in courts.'[7]

Before moving to review the findings and recommendations in the PCAST report, it is useful briefly to summarise the findings from the earlier inquiry conducted by the National Research Council (NRC), the research arm of the US National Academy of Sciences (NAS).

2. Building on the NRC report (and ongoing responses)

In 2009 the NRC published the outcome of a multi-year review of the forensic sciences in the United States. The inquiry was prompted by mistakes exposed through innocence projects and a high-profile error by FBI fingerprint examiners in the mis-identification of Brandon Mayfield.[8] The resulting report – *Strengthening the forensic sciences in the United States: The path forward* – was unprecedented in its critical assessment.[9] The NRC report was summarised by PCAST in the following terms:

> The 2009 report described a disturbing pattern of deficiencies common to many of the forensic methods routinely used in the criminal justice system, most importantly a lack of rigorous and appropriate studies establishing their scientific validity, concluding that 'much forensic evidence—including, for example, bitemarks and firearm and toolmark identifications—is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline.'[10]
>
> The 2009 NRC study concluded that many of these difficulties with forensic science may stem from the historical reality that many methods were devised as rough heuristics to aid criminal investigations and were not grounded in the validation practices of scientific research. … the report found the problems plaguing the forensic science community are systemic and pervasive—the result of factors including a high degree of fragmentation (including disparate and often inadequate training and educational requirements, resources, and capacities of laboratories); a lack of standardization of the disciplines, insufficient high-quality research and education; and a dearth of peer-reviewed studies establishing the scientific basis and validity of many routinely used forensic methods.[11]

Like the NRC report, the PCAST report displays an unremitting commitment to the need for scientific research. Indeed, the report is oriented toward addressing 'A Critical Gap: Scientific Validity'.[12]

We now turn to describe the findings and recommendations presented in the PCAST report.

Wortley, 'Interpreting image evidence: Facial mapping, police familiars and super-recognisers in England and Australia' (2016) 3 *Journal of International and Comparative Law* 473.

[7] PCAST report, 39.

[8] See e.g. Brandon Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Harvard University Press, Cambridge MA, 2011); Office of the Inspector General, *A review of the FBI's progress in responding to the recommendations in the Office of the Inspector General Report on the fingerprint misidentification in the Brandon Mayfield Case* (U.S. Department of Justice, 2011).

[9] For a detailed review of the NRC report, see Gary Edmond, 'What Lawyers should know about the forensic "sciences"' (2015) 36 *Adelaide Law Review* 33.

[10] PCAST report, 22. See NRC report, pp. 107-8.

[11] PCAST report, 32-3. See also NRC report, 128. On the misguided interest in 'uniqueness', see PCAST report, 61-62.

[12] PCAST report, 39.

3. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*

The PCAST report builds on previous reviews of the forensic sciences, explains the role of scientific validity (for courts), and describes scientific criteria for validity and reliability of forensic feature-comparison methods. It then evaluates the scientific validity of: (i) DNA analysis of single-source and simple-mixture samples; (ii) DNA analysis of complex-mixture samples; (iii) bitemark analysis; (iv) latent fingerprint analysis; (v) firearms analysis; (vi) footwear analysis; and (vii) hair analysis. The report concludes by making recommendations to US domestic institutions, such as the National Institute of Standards and Technology (NIST), the Office of Science and Technology Policy (OSTP), the Federal Bureau of Investigation (FBI) Laboratory, the Attorney General, and finally the judiciary. Key themes of the report and representative excerpts are reproduced below.

A. Scientific validity and reliability must be demonstrated through empirical studies

PCAST draws on the scientific discipline of metrology to define scientific validity and reliability:[13]

> For a metrological method to be scientifically valid and reliable, the procedures that comprise it must be shown, based on empirical studies, to be *repeatable, reproducible,* and *accurate,* at levels that have been measured and are appropriate to the intended application.[14]

This selection was conditioned by the fact that the 'feature-comparison methods … all belong to the same broad scientific discipline, *metrology*, which is "the science of measurement and its application," in this case to measuring and comparing features.'[15]

The key constructs that comprise scientific *reliability* are defined:

> By 'repeatable,' we mean that, with known probability, an examiner obtains the same results, when analyzing samples from the same sources.
>
> By 'reproducible,' we mean that, with known probability, different examiners obtain the same result, when analyzing the same samples.
>
> By 'accurate,' we mean that, with known probability, an examiner obtains correct results both (1) for samples from the same source (true positives) and (2) for samples from different sources (true negatives).[16]

The PCAST report relies on two distinct types of validity: *foundational validity* and *validity as applied:*

> *Foundational validity* … means that a method can, *in principle,* be reliable.[17]
>
> *Validity as applied* means that the method has been reliably applied *in practice.*[18]

In other words, foundational validity refers to the extent to which a method is able to do what it is intended to do (e.g., link a sample to its source) in a repeatable, reproducible and accurate manner. Validity as applied incorporates the proficiency of the practitioner as well as the practical constraints associated with using the method under casework conditions. A method that is repeatable,

[13] Metrology is the science of measurement and its application, PCAST report, 44.
[14] PCAST report, 47.
[15] PCAST report, 23.
[16] PCAST report, 47, Box 2.
[17] PCAST report, 4.
[18] PCAST report, 5.

reproducible and accurate under controlled laboratory conditions may not be valid in specific applications – e.g., a particular analyst may not be sufficiently proficient in the application of the technique or a sample may be inadequate or excessively degraded.

The report then states two fundamental principles. First, foundational validity cannot be assumed:

> (1) methods must be presumed to be unreliable until their foundational validity has been established based on empirical evidence and (2) even then, scientific questioning and review of methods must continue on an ongoing basis.[19]

Secondly, empirical testing is essential:

> … valid scientific knowledge can *only* be gained through *empirical* testing of specific propositions.[20]

B. Error rate information must be reported (to the court)

PCAST expressed concern about expert witnesses misrepresenting the value of their evidence in reports and testimony:

> … reviews have found that expert witnesses have often overstated the probative value of their evidence, going far beyond what the relevant science can justify. Examiners have sometimes testified, for example, that their conclusions are "100 percent certain;" or have "zero," "essentially zero," or "negligible," error rate. As many reviews—including the highly regarded 2009 National Research Council study—have noted, however, such statements are not scientifically defensible: *all laboratory tests and feature-comparison analyses have non-zero error rates*.[21]

The PCAST report explains that appropriate, empirically-derived, estimates of error are essential for the assessment of the probative value (and weight) of evidence and so must be determined and reported:[22]

> An empirical measurement of error rates is not simply a desirable feature; it is essential for determining whether a method is foundationally valid.[23]

Moreover, the provision of error rates and other limitations are vital for decision makers in legal settings:

> Without appropriate estimates of accuracy, an examiner's statement that two samples are similar—or even indistinguishable—is scientifically meaningless: *it has no probative value,* and considerable potential for prejudicial impact.[24]

---

[19] PCAST report, 32.
[20] PCAST report, 46.
[21] PCAST report, 3 (emphasis added). See also NRC report, 142.
[22] The term 'error' is defined in the PCAST report, 51, Box 3. See also Bryan Found et al, 'Reporting on the comparison and interpretation of pattern evidence: recommendations for forensic specialists' (2012) 44 *Australian Journal of Forensic Science* 193.
[23] PCAST report, 53, 50: 'It is necessary to have appropriate empirical measurements of a method's false positive rate and the method's sensitivity. As explained … it is necessary to know these two measures to assess the probative value of a method.'
[24] PCAST report, 27 (emphasis added). To say that such evidence has *no* probative value is not, however, strictly correct, if probative value is thought of in orthodox legal terms. Such evidence may have *some* probative value, i.e. some tendency to

Unless opinions are expressed in empirically-predicated terms, that include limitations, they are not susceptible to rational evaluation.[25]

PCAST also makes clear that appropriately designed 'black box' studies are required to determine the relevant error rate:[26]

> Importantly, error rates cannot be inferred from casework, but rather must be determined based on samples where the correct answer is known.[27]

Unless procedures and practitioners are tested in circumstances where the correct answer (or ground truth) is known, it is not possible to be confident that the method is valid or the performance proficient.

C. Forensic feature-comparison methods pose difficulties for jurors

PCAST expressed particular concerns about the way evidence derived from forensic feature-comparison procedures is expressed in reports and testimony, because

> [t]he vast majority of jurors have no independent ability to interpret the probative value of results based on the detection, comparison, and frequency of scientific evidence. … The potential prejudicial impact is unusually high, because jurors are likely to overestimate the probative value of a 'match' between samples.[28]

Without insight into limitations, accuracy and error, as well as the frequency with which certain features appear (or are interrelated), jurors (and judges) are likely to over-estimate the value of forensic science evidence.[29] They are vulnerable to concluding that similar traces (whether shoe prints or striations on shell casing and so on) were produced by the same person or object.

D. Experience, training and professional practices cannot substitute for empirically demonstrated validity and reliability

The report explains that it is inappropriate to infer validity or reliability on the basis of a forensic practitioner's professional experience, casework, confidence or a variety of institutional practices (including accreditation and proficiency testing):

> Notably, some forensic practitioners espouse the notion that extensive 'experience' in casework can substitute for empirical studies of scientific validity. Casework is not scientifically valid research, and experience alone cannot establish scientific validity. In particular, one cannot reliably estimate error rates from casework because one typically does not have independent knowledge of the 'ground truth' or 'right answer'.[30]

---

make the existence of a fact in issue more or less likely, but its *degree* of probative value is unknown. See Michael S. Pardo, 'Evidence Theory and the NAS Report on Forensic Science' [2010] Utah L Rev 367.

[25] See Gary Edmond, 'The conditions for rational (jury) evaluation of forensic science evidence' (2015) 39 *Melbourne University Law Review* 1.

[26] Characteristics of 'black-box' trials to establish foundational validity are described in the PCAST report, 52, Box 4.

[27] PCAST report, 53.

[28] PCAST report, 45. See also Gary Edmond, Matt Thompson & Jason Tangen, 'A guide to interpreting forensic testimony: Scientific approaches to fingerprint evidence' (2013) 12 *Law, Probability & Risk* 1.

[29] PCAST report, 9, 149. Exaggeration and over-estimation are forms of unfair prejudice to the defendant.

[30] PCAST report, 32.

> Importantly, good professional practices—such as the existence of professional societies, certification programs, accreditation programs, peer-reviewed articles, standardized protocols, proficiency testing, and codes of ethics—cannot substitute for actual evidence of scientific validity and reliability.[31]

These conclusions have serious implications for conventional admissibility and procedural rules and the considerable confidence invested in adversarial procedures – frequently operating in the absence of validations studies and indicative error rates.

E. Assessment of the validity of seven feature-comparison methods

PCAST undertook one of the first independent reviews of forensic feature-comparison methods focusing on empirical evidence for *foundational validity* and *validity as applied*. We summarise the conclusions for the seven methods, noting that where there is foundational validity, validity as applied must also be considered.

DNA analysis of single-source and simple-mixture samples:

DNA analysis of samples from a single individual or from a 'simple' mixture of two individuals, one of whom is known, is found to be 'an objective method in which the laboratory protocols are precisely defined and the interpretation involves little or no human judgment.'[32] It meets all the criteria for foundational validity, while 'the probability of a match arising by chance in the population by chance can be estimated directly from appropriate genetic databases and is extremely low'.[33]

DNA analysis of complex-mixture samples:

By contrast, the 'analysis of complex mixtures of biological samples from multiple unknown individuals in unknown proportions', while it uses similar laboratory techniques, involves a much greater degree of human judgment in the interpretation of the resulting DNA profile.[34] PCAST finds that such 'subjective analysis … has not been established to be foundationally valid and is not a reliable methodology.'[35]

Bitemark analysis:

PCAST's view of the comparisons of marks on a victim or object with dental impressions taken from a suspect is particularly damning:

> available scientific evidence strongly suggests that examiners not only cannot identify the source of [a] bitemark with reasonable accuracy, they cannot even consistently agree on whether an injury is a human bitemark. For these reasons, PCAST finds that bitemark analysis is far from meeting the scientific standards for foundational validity.[36]

Latent fingerprint analysis:

Latent fingerprint analysis, i.e. the comparison of impressions made or developed on an item with prints taken from a known subject, or with another latent print, is found to be:

---

[31] PCAST report, 55.
[32] PCAST report, 7.
[33] PCAST report, 7, 69-75. See also NRC report, 128-133.
[34] PCAST report, 7.
[35] PCAST report, 8, 75-83. Contrast *Dlugosz* [2013] EWCA Crim 2, where the court allowed the forensic practitioner to express a subjective impression.
[36] PCAST report, 9, 83-87. See also NRC report, 173-176.

> a foundationally valid subjective methodology—albeit with a false positive rate that is substantial and is likely to be higher than expected by many jurors based on longstanding claims about the infallibility of fingerprint analysis. The false-positive rate could be as high as 1 error in 306 cases based on the FBI study and 1 error in 18 cases based on a study by another crime laboratory. In reporting results of latent-fingerprint examination, it is important to state the false-positive rates based on properly designed validation studies.[37]

Firearms analysis:

In discussing the analysis of 'toolmarks' on ammunition to identify the weapon from which ammunition was fired, PCAST acknowledges that there has been significant progress since the NRC report in 2009, in that one substantial 'black box' study has been conducted, the results of which are available online.[38] However,

> [t]he scientific criteria for foundational validity require that there be more than one such study, to demonstrate reproducibility, and that studies should ideally be published in the peer-reviewed scientific literature. Accordingly, the current evidence still falls short of the scientific criteria for foundational validity.[39]

Footwear analysis:

PCAST does not consider the validity of identifications from shoeprints of 'class characteristics' such as make and size of shoe, but points to a lack of appropriate studies or analyses of 'identifying characteristics' which purportedly link an impression to a specific shoe: 'Such associations are unsupported by any meaningful evidence or estimates of their accuracy and thus are not scientifically valid'.[40]

Hair analysis:

PCAST did not examine the comparison of microscopic features of hair in depth, but its review of the papers cited in documentation submitted by the Department of Justice found 'that these studies do not establish the foundational validity and reliability of hair analysis.'[41]

F. Recommendations for the Attorney General.

PCAST made two sets of recommendations to the US Attorney General. The first addresses the use of feature-comparison methods in federal prosecutions, and the second concerns guidelines on expert testimony prepared by the Department of Justice (DOJ). The most radical recommendations are the following:

> The Attorney General should direct attorneys appearing on behalf of the Department of Justice (DOJ) to ensure expert testimony in court about forensic feature-comparison methods meets the scientific standards for scientific validity.[42]

---

[37] PCAST report, 9-10, 87-104. See also NRC report, 136-145; National Institute of Standards and Technology, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (US Department of Commerce, 2012); Lord Anthony Campbell, *The Fingerprint Inquiry Report* (APS Group Scotland, 2011).

[38] David P. Baldwin, Stanley J. Bajic, Max Morris, and Daniel Zamzow, *A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons* (2014), available at https://www.ncjrs.gov/pdffiles1/nij/249874.pdf

[39] PCAST report, 11, 104-114. See also NRC report, 150-155.

[40] PCAST report, 13, 114-117. See also NRC report, 145-150 and *R v T* [2010] EWCA Crim 2439.

[41] PCAST report, 13, 118-122. See also NRC report, 155-161.

> Where there are not adequate empirical studies and/or statistical models to provide meaningful information about the accuracy of a forensic feature-comparison method, DOJ attorneys and examiners should *not offer testimony based on the method*.[43]

Given the findings reviewed above, these recommendations would dramatically curtail the use of forensic science evidence in federal prosecutions. Other recommendations deal with encouraging research and revised guidelines on the reporting of forensic science evidence; stressing the need to include information about error rates when these are known, and to point out the possibility of error in every case.

G. Recommendations for the Judiciary

Finding 'lack of rigor in the assessment of the scientific validity of forensic evidence' to be 'a real and significant weakness in the judicial system',[44] PCAST made one recommendation to the US judiciary regarding the use of scientific validity as a foundation for expert testimony. Recommendation 8 states:

> (A) When deciding the admissibility of expert testimony, Federal judges should take into account the appropriate scientific criteria for assessing scientific validity including: … (i) foundational validity ... and (ii) validity as applied. ...
>
> (B) Federal judges, when permitting an expert to testify about a foundationally valid feature-comparison method, should ensure that testimony about the accuracy of the method and the probative value of proposed identifications is scientifically valid in that it is limited to what the empirical evidence supports.[45]

Finally, PCAST recommends the preparation of a best practices manual (8C) and advisory note relating to the admissibility of expert testimony based on forensic feature-comparison methods, as well as scientific education programs relevant to those methods (8D).

While PCAST's discussion consciously builds on US admissibility rules (notably *Daubert v Merrell Dow Pharmaceuticals, Inc.* and the consequent revision of r702 of the US Federal Rules of Evidence), PCAST's advice transcends jurisdictional practices.[46] The PCAST report endeavours to capture what is required of any (adversarial) system interested in drawing upon and rationally assessing scientific and technical forms of evidence.

4. Getting PCAST (and its concerns) before the courts of England and Wales

The PCAST report's conclusions are stark. It recommends that certain types of forensic feature-comparison evidence should not be adduced in criminal cases because they do not meet scientific standards for foundational validity. In relation to some categories of expert evidence, the report's conclusions will not surprise English readers. Bitemark comparison evidence, for example, is rarely seen in English courts. But English legal professionals may be rather more surprised to discover that some footwear comparison, firearms analysis, and DNA analysis of complex mixtures using

---

[42] PCAST report , 18: Recommendation 6(A).
[43] PCAST report, 19: Recommendation 7(B)(ii). (italics added)
[44] PCAST report, 22.
[45] PCAST report, 19, 145.
[46] *Daubert v Merrell Dow Pharmaceuticals Inc*. 509 US 579 (1993).

Combined Probability of Inclusion-based methods, were found to lack foundational validity.[47] There are serious and, as yet, unanswered questions about the value of these procedures and how courts operating in a rational tradition should respond. The PCAST report has the potential to support challenges to such evidence. However, the mechanisms by which it (or the recommendations of the Forensic Science Regulator, discussed below) might be adduced, either before the tribunal of law in relation to preliminary admissibility decision-making, or before the tribunal of fact so that it can be taken into account when determining weight, are constrained by rules of evidence and procedure, by traditions of practice as well as by widespread ignorance

There are three main ways in which the PCAST report might be used in English courts. First, it could support a submission that a particular expert's opinion should be excluded on the ground that it does not have a 'sufficiently reliable scientific basis'.[48] If the method the expert has employed lacks foundational validity (or validity as applied), it may be insufficiently reliable to be admitted. The case law prior to what is now Criminal Practice Direction (CPD) 19A (originally issued in 2014) suggested that the threshold of reliability was a low one, but the Practice Direction could form the basis of a 'more rigorous' approach, as promised by Leveson LJ in *Stephen H v R*.[49]

Secondly, even if feature-comparison evidence is deemed to meet the threshold of 'sufficient reliability' under CPD 19A, the opinions expressed by the report's authors could lend weight to a submission that the evidence ought to be excluded under either s.78 of the Police and Criminal Evidence Act 1984 or at common law.[50] Discretionary exclusion, at least in theory, constitutes a second hurdle that expert evidence has to cross once it has cleared the (traditionally very low) threshold of reliability required for admissibility.[51] In light of the Practice Direction, however, and also the Court of Appeal's view as set out in *Atkins,*[52] the possibility of using s.78 or the common law discretion to get a second bite at the cherry of unreliability is probably theoretical rather than real. A judge who accepted that the probative value of some piece of unvalidated scientific evidence was insufficient to outweigh the prejudicial effect of presenting it to the jury as 'science' would have ample grounds to deem the evidence 'insufficiently reliable to be admitted'.

Thirdly, if a judge declines to exclude the evidence, the report could potentially be used to challenge the forensic practitioner's testimony before the tribunal of fact via cross-examination. Though, unless the practitioner accepts the authority of PCAST and particular findings, the status of the report remains uncertain. In criminal proceedings, an expert providing testimony derived using a procedure that has been questioned by PCAST ought to refer to the PCAST report as part of standard disclosure.[53] In addition to general disclosure obligations that arise, particularly in relation to

[47] We appreciate that there is work being undertaken in these areas, particularly in relation to mixed and complex DNA samples.

[48] *Criminal Practice Directions 2015 Division V* (Oct, 2015) (hereafter CPD) accessed from the internet 02/01/2017 https://www.justice.gov.uk/courts/procedure-rules/criminal/practice-direction/2015/crim-practice-directions-V-evidence-2015.pdf, 19A.4.

[49] [2014] EWCA Crim 1555, [44]. See, for example, Law Commission of England and Wales, *Expert Evidence in Criminal Proceedings in England and Wales*, HMSO, London, 2011, [1.8], [1.17], [1.21], [2.16], [3.3], [3.4], [6.10].

[50] Section 82(3) of PACE preserves the court's common law discretion to exclude evidence on the basis that its prejudicial effect outweighs its probative value (*R v Sang* [1980] AC 425).

[51] *R v Luttrell* [2004] 2 Cr App R 31, [28], [38]; *R v Dlugosz* [2013] 1 Cr App R 32, [27].

[52] *R v Atkins* [2010] 1 Cr App R 8, [9].

[53] See *ACPO/CPS Guidance Booklet for Experts - Disclosure: Experts' Evidence, Case Management and Unused Material* (May 2010). Of course, disclosure of validation results or their absence should extend beyond the procedures discussed by PCAST to all scientific, medical and technical procedures grounding opinion evidence.

prosecution evidence,[54] Criminal Procedure Rule (CrimPR) 19.4(h) imposes obligations on all experts to include in their reports such information as the court may need to decide whether evidence is sufficiently reliable to be admitted.[55]

If PCAST were to be referenced by a forensic practitioner as part of their general disclosure obligations, the court would be able to take the report into account in evaluating the evidence. The PCAST report could then form part of the material to be weighed in the balance by the judge in determining whether expert evidence is sufficiently reliable, or whether to exclude the evidence. If the report is not referred to by the forensic practitioner, unlike the situation in civil cases, it cannot simply be adduced by the opposing advocate.

In the civil courts a report of this type could be served and relied upon directly in court. In *Rogers v Hoyle*[56] the Court of Appeal considered the admissibility of a report that had been prepared independently of the proceedings by a statutory body. Insofar as the report contained statements of fact, it was admissible, the rule against hearsay having been abolished by s.1 of the Civil Evidence Act 1995. The expressions of opinion in the report were also admissible, since the authors were 'qualified experts on subjects involving special expertise'.[57] Furthermore, the Civil Procedure Rules (CPR) Part 35, which gives civil courts certain powers and duties to restrict expert evidence, did not apply because CPR 35.2 defines an 'expert' as 'a person who has been instructed to give or prepare expert evidence for the purpose of proceedings'. As the authors had not been instructed to prepare the report for the proceedings, the claimants did not require the permission of the court to adduce it. Academic research appears to be admissible in civil proceedings on the same basis.[58]

Similarly, the Criminal Procedure Rules define an 'expert' as someone who has been 'instructed' for the purposes of the proceedings,[59] and thus it could be argued that a report such as PCAST's does not have to comply with CrimPR 19. Criminal courts, however, remain subject to the rule against hearsay, and the PCAST report is undoubtedly hearsay. The only specific exception to the hearsay rule which looks remotely relevant is rule 1(a) of the common law rules preserved by the Criminal Justice Act 2003, s. 118(1), which covers 'scientific works … as evidence of facts of a public nature'. The common law rule in question, however, applies to works that are referred to in the evidence of an expert witness.[60] The only way in which the PCAST report could be admitted as evidence in itself is 'in the interests of justice' under s.114(1)(d) of the Act. While a court might not admit a public report from a different jurisdiction in the exercise of its inclusionary discretion, it is not outlandish to suggest that it might be in the interests of justice to do so. In the Canadian case of *Bornyk,*[61] the report of the Scottish Fingerprint Inquiry,[62] among other sources, was consulted (on the judge's own

[54] There is a general obligation in criminal cases under the Criminal Procedure and Investigations Act 1996, s. 3 to disclose evidence that reasonably might be considered capable of undermining the prosecution case or assisting the case for the defence.

[55] See generally Gary Edmond, 'Expert evidence and the professional responsibilities of prosecutors' in P. Roberts *et al* (eds), *Integrity in the Criminal Process* (Oxford, Hart, 2016) 225-245; Evidence-based Forensics Initiative, 'Model forensic science' (2016) 39 *Australian Journal of Forensic Sciences* 174.

[56] [2013] EWCA Civ 257.

[57] *Ibid.,* [65].

[58] *Interflora* v. *Marks and Spencer plc* [2014] EWCA Civ 1403, [161]; Hodge Malek et al (eds), *Phipson on Evidence*, 18th ed., 1st Supp. (Sweet & Maxwell, 2015), para. 33-25;

[59] CrimPR 19.2.

[60] *R v Abadom* (1983) 76 Cr App R 48.

[61] *R v Bornyk* 2013 BCSC 1927.

[62] *The Fingerprint Inquiry Report* (2011) available at: http://www.webarchive.org.uk/wayback/archive/20150428160022/http://www.thefingerprintinquiryscotland.org.uk/inquiry/3127-2.html.

initiative) as evidence of the limitations of fingerprint evidence. Unfortunately, the judge consulted these materials only after counsel had made their submissions. The British Colombia Court of Appeal[63] was critical of their being used in this way, pointing out that the material on which the judge relied was beyond the scope of judicial notice.[64] Had the judge instead encouraged counsel to raise the issue at an earlier stage, this material might have been introduced in a way more consistent with the adversarial process.[65]

There seems to be no reason why an opposing expert could not rely on the PCAST report as the basis for an opinion that, for example, a footwear examiner who purports to match a footwear impression to a shoe is using a method that lacks foundational validity and the examiner's evidence may, in consequence, be unreliable or insufficiently reliable. As Davies and Piasecki note, the earlier NRC report has been used in that way in at least two cases – though without moving the Court of Appeal to act on it.[66] Once the PCAST report has been referred to in this way by the opposing expert, the judge may take the report's contents into account in evaluating the forensic practitioner's conclusions. In deciding whether the evidence is 'sufficiently reliable' to be admitted, the PCAST report may be directly relevant as, under CPD 19A.5(a), the court may take into account the validity of the methods by which the forensic practitioner's data was obtained. In addition, CPD 19A.6 provides that a court 'should be astute to identify potential flaws in [an expert's] opinion which detract from its reliability', including 'being based on a hypothesis which has not been subjected to sufficient scrutiny (including, where appropriate, experimental or other testing)'.[67]

Where PCAST's report is directly relevant to the matters dealt with by the prosecution expert, it ought to be mentioned in the expert's own report, either as part of the 'range of opinion' that the expert is required to 'summarise' under CrimPR 19.4(f)(i), or, if the expert accepts PCAST's conclusions, as a necessary 'qualification' of her opinion under para. 19.4(g). An expert who does *not* accept PCAST's conclusion that a particular technique is scientifically invalid should explain why.[68] If the expert's report does not comply with these requirements, the expert cannot give evidence unless the parties agree or the court so directs.[69] Making the admissibility of expert evidence conditional on the expert drawing attention to adverse findings – to criticisms and recommendations produced by authoritative organisations such as PCAST – would be consistent with a series of Court of Appeal judgements stressing the need for judges to control the terms in which expert evidence is given.[70] Conscientious scientists should, however, raise problems about validity themselves, rather than leave it to defence lawyers who may be unaware of the issues.[71]

Where the expert's evidence is admitted, an advocate who *is* aware of the issues could use the PCAST report to inform the cross-examination of the forensic practitioner about the reliability of the method used (and perhaps the competence and impartiality of the witness where PCAST is apposite

[63] *R v Bornyk* 2014 BCCA 450. Discussed in Gary Edmond, David Hamer and Emma Cunliffe, 'A little ignorance is a dangerous thing: Engaging with exogenous knowledge not adduced by the parties' (2016) 25 *Griffith Law Review* 383.
[64] The leading English case on the limited judicial notice that may be taken of scientific works is *McQuaker v Goddard* [1940] 1 KB 387, 400-1.
[65] This occurred at re-trial in January 2017.
[66] *Otway v R* [2011] EWCA Crim 3; *R v Ferdinand* [2014] EWCA Crim 1243. Gemma Davies and Emma Piasecki, 'No More Laissez Faire?' (2016) 80 J Crim L 327, 337.
[67] CPD 19A.6(a).
[68] CrimPR 19.4(f)(ii).
[69] This is the combined effect of CrimPR 19.3(3)(a) and 19.3(4)(a).
[70] *R v Atkins* [2010] 1 Cr App R 8, [23], [31]; *R v Reed* [2010] 1 Cr App R 23, [122]; *R v C* [2011] 3 All ER 509, [32], *R v Dlugosz* [2013] 1 Cr App R 32.
[71] Davies and Piasecki, n. 66 above.

yet not disclosed).[72] The report may then become evidence the tribunal of fact can take into account in deciding what weight to assign to the evidence, although weighing evidence whose true probative value is unknown is a difficult exercise, and juries may fail to appreciate how fundamental validation is to the production of reliable results. Here we might note that English judges have not uniformly recognised the centrality of foundational validity.

We should also recognise that most cases involving expert evidence never go to trial. We have recently seen the introduction of streamlined reports, which (controversially) do not include information about validity or error.[73] When prosecutors rely on expert opinion that is based upon a method that is not foundationally valid, legal practitioners need to appreciate this as an issue and, where appropriate, challenge the evidence (or demand more detail from investigators, forensic practitioners and prosecutors). This will require raising awareness of PCAST (and codes and guidelines produced by the Forensic Science Regulator, discussed below) among practitioners if the report is to have any impact in England and Wales. The apparent unwillingness of practitioners to engage with the domestic Practice Direction to date is not promising in this regard.

5. 'Sufficient reliability' and common law principles

Because of the lack of case law on CPD 19A, the precise nature of the 'sufficient reliability' test remains unclear. CPD 19A re-enacts an earlier Direction, which came into effect in October 2014,[74] and which partially implemented the recommendations of the Law Commission Report on Expert Evidence.[75] What the Practice Direction could not implement was the Law Commission's proposed statutory test of the admissibility of expert evidence. Since the Government declined to introduce the proposed legislation, the guidance which the Practice Direction adopts from the Law Commission report has to be applied in determining whether the common law criteria of admissibility are satisfied.[76]

There are two possible interpretations of the relationship between CPD 19A and the common law. On one account, set out by one of us elsewhere, 'sufficiently reliable to be admitted' means sufficiently reliable to satisfy the three limbs of the common-law admissibility test: that the evidence is relevant; necessary to assist the jury ('helpful'); and will be given by a competent witness.[77] The second view, advocated by Stockdale and Jackson,[78] is that 'sufficiently reliable to be admitted' is an independent fourth limb of the common-law test, introduced by *Dlugosz* and other cases where the Court of Appeal adopted the phrase used in the Law Commission report.[79] The difficulty with this is that 'sufficient reliability' as a free-standing test is vacuous: it tells us nothing about what degree of reliability is 'sufficient'. In our view the test is that the reliability of expert evidence must be sufficient to establish its relevance, to assist the jury, and to establish that the individual expert is

[72] See Evidence-Based Forensics Initiative, 'How to cross-examine forensic scientists: A guide for lawyers' (2014) 39 *Australian Bar Review* 174.
[73] See Oriola Sallavaci, 'Streamlined reporting of forensic evidence in England and Wales: Is it the way forward? (2016) 20(3) IJEP 235.
[74] Practice Direction: Criminal Proceedings (Various Changes) [2014] EWCA Crim 1569.
[75] Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (Law Com No 325, 2011).
[76] CPD, n.48 above, 19A.3.
[77] Tony Ward, '"A New and More Rigorous Approach" to Expert Evidence in England and Wales?' (2015) 19 *E&P* 228.
[78] Michael Stockdale and Adam Jackson, 'Expert Evidence in Criminal Proceedings: Current Challenges and Opportunities' (2016) 80 *J Crim L* 344, 348.
[79] [2013] EWCA Crim 2.

competent.[80] Most scientific, technical and medical forms of evidence should include information about validity and reliability.

First on the list of factors which judges 'may take into account' in assessing admissibility is the 'validity of the methods by which' the data supporting the expert's opinion were obtained.[81] PCAST's concepts of 'foundational validity' and 'validity as applied' are consonant with the common law principles that ultimately determine admissibility, and could therefore provide a useful gloss to the concept of validity under the Practice Direction.

What PCAST calls 'foundational validity' may bear on the relevance of the evidence but is also important in determining whether the evidence will assist the court, while PCAST's 'validity as applied' relates both to the common-law test of competence and to the issues highlighted in the Practice Direction as to whether the limitations of the forensic practitioner's methods and data have been properly taken into account.[82] The test of assistance to the court or 'helpfulness' is rather too briefly summarised in the Practice Direction as whether the evidence 'is needed to provide the court with information likely to be outside the court's own knowledge and experience'.[83] As the Court of Appeal explained in *Gilfoyle,* quoting the leading Scottish case of *Davie v Magistrates of Edinburgh,* it is not enough that the information is outside the court's knowledge. In addition, 'expert witnesses must furnish the court "with the necessary scientific criteria for testing the accuracy of their conclusions, so as to enable the judge or jury to form their own independent judgment by the application of these criteria to the facts proved in evidence"'.[84] Without such criteria, the information provided by a forensic practitioner cannot be rationally evaluated by the court and is therefore not 'helpful' or 'needed'.

The 'scientific criteria' alluded to in *Davie* must be criteria that are grounded in science but capable of application by non-scientists. What PCAST very helpfully does is to propose a set of scientific concepts – repeatability, reproducibility and accuracy – which together, in the context of feature-comparison evidence, can be treated as equivalent to the legal concept of 'reliability,' and which are reasonably easy for non-scientists to understand and apply. Assessing the degree of reliability in the context of the case as a whole is primarily a matter for the jury[85] but in order to be 'sufficiently reliable to be admitted'[86] the evidence must, we submit, provide sufficient indicia of reliability to make such an assessment possible. Crucially, as PCAST urges repeatedly, the jury must be given some well-founded estimate of the risk of a 'false positive'. This provides a common-law basis on which to mount an argument that evidence lacking in 'foundational validity' is 'insufficiently reliable to be admitted'. So is evidence unsupported by indications of 'validity as applied', such as rigorous proficiency tests passed by the individual practitioner, because the evidence fails to establish that that the witness is competent.[87]

The PCAST report also provides an opportunity to reassess judicial guidance on directing the jury in Crown Court trials involving expert evidence. The new *Crown Court Compendium*, to which Judges and Recorders should have regard when summing up, contains a summary of the law relating to

---

[80] Dale Nance, 'Reliability and the Admissibility of Experts'(2003) 34 *Seton Hall L Rev* 191; Ward, n 77 above
[81] *Ibid.,* 19A.5.
[82] CPD, 19A.5(b), (c), (f), (g), (h); PCAST report, 56.
[83] *Ibid.,* 19A.1.
[84] *Gilfoyle* [2001]2 Cr App R 5 at 67*, Davie v. Edinburgh Magistrates* 1953 S.C. 34 at 40.
[85] *Luttrell* [2004] 2 Cr App R 31, [36], again quoting *Davie,* above.
[86] *Dlugosz* [2013] EWCA Crim 2, CPD 19.4.
[87] PCAST report, 57-59. The Council expressed deep concerns about proficiency tests in use being too easy; addressing the need for accreditation but without helping to test abilities or enhance training.

expert evidence, guidance as to how to direct the jury in cases involving experts, and an example of an appropriate direction for a case involving a handwriting expert.[88]

The legal summary in the Compendium suggests that, where an expert testifies as to a degree of confidence in a match:

> (e.g. "no support, limited support, moderate support, support, strong support, powerful support") it may help the jury to explain that these terms are no more than labels which the witness has applied to his opinion of the significance of his findings and that because such opinion is entirely subjective different experts may not attach the same label to the same degree of comparability: *Atkins*.[89]

This direction fails to distinguish clearly between relative judgments which have a scientific basis (and are used as a purportedly user-friendly substitute for a likelihood ratio) and those which – as in *Atkins* – are what PCAST would call 'scientifically meaningless'.[90] The latter can fairly be described as 'entirely subjective' and the important point to make here is that while the jury might attach some weight to a personal opinion based on experience, they should be clear that it is not a scientific finding[91] and should treat it with caution.[92]

The Compendium also indicates that juries should be directed to take into account an expert's 'qualifications/practical experience/methodology/source material/quality of analysis/objectivity of the experts, and the impression they make when giving evidence'.[93] It is concerning that the experts' methodology – and the validity or otherwise of that methodology – is placed on a par with 'the impression they make'. Juries should not be left 'to flounder in the formation of a general impression,'[94] but should be told to focus on whether or not the evidence is based on a method that has been validated, and if not (assuming that the evidence is nevertheless deemed 'sufficiently reliable to be admitted') to treat it with caution. The example direction in the Compendium relates to handwriting evidence, which is not discussed by PCAST, but raises similar issues.[95] The direction does not refer to validity or reliability, nor give jurors any indication of how these might be assessed in order that they might evaluate the weight of the evidence. Handwriting comparison evidence is likely to remain admissible on the basis that even if it is not science, it is better than unaided lay judgment, but we submit that, at the very least, clearer and more cautious guidance is required as to its evaluation.[96]

6. PCAST, validation and the Forensic Science Regulator

[88] Judicial College, *Crown Court Compendium,* https://www.judiciary.gov.uk/wp-content/uploads/2016/05/crown-court-compendium-part-i-jury-and-trial-management-and-summing-up.pdf <accessed 6th February 2016> at 10-3.
[89] 10-3, para 5. See also Gary Edmond et al, '*Atkins v The Emperor*: The "Cautious" use of Unreliable "Expert" Opinion' (2010) 14 IJEP 146.
[90] See above, n 24 and NRC report, 185.
[91] On the need to avoid calling such evidence 'scientific', see *R v T (Footwear Mark Evidence)* [2011] 1 Cr App R 9, [96].
[92] See *R v Luttrell* [2004] 2 Cr App R 31, [41-2].
[93] 10-3 para 13(2)(b). See also para 7.
[94] *R v Henderson* [2010] 2 Cr App R 34, [218].
[95] See e.g. D. Michael Risinger with Michael J. Saks, 'Science and Nonscience in the Courts: *Daubert* meets Forensic Handwriting Expertise' (1996) 82 Iowa L. Rev. 21.
[96] See Jennifer R. Mnookin, 'Scripting Expertise: The History of Handwriting Identification Evidence and the Judicial Construction of Reliability' (2001) 71 Va. L. Rev. 1723. For the purposes of this article we are neutral as to whether such evidence *should* be admitted.

The PCAST report and its legal significance (beyond the United States) find support in the activities of the Forensic Science Regulator (UK).[97] The office of the Forensic Science Regulator ('the Regulator') was established in 2007 as a result of the House of Commons Science and Technology Committee report identifying that '[j]udges are not well placed to determine scientific validity without input from scientists.'[98] The Regulator considers that the Criminal Practice Directions clarify the expectation 'that all methods routinely employed within the Criminal Justice System (CJS), whether for intelligence or evidential use, will be validated prior to their use on live casework material'.[99] Using this definition as the basis of her *Guidance FSR-G-201: Validation* (hereafter *Guidance*) on the subject,[100] the manner and content of the Regulators' engagement with issues relating to validity is largely consistent with that put forward by PCAST.

The Regulator is responsible for ensuring that the provision of forensic science services across the criminal justice system is subject to an appropriate regime of scientific quality standards.[101] The Regulator publishes Codes of Practice and Conduct specifying the required scientific quality standards for providers of laboratory-based forensic science services.[102] In addition to the general Codes, there are specific Codes for certain forensic disciplines, including some types of feature-comparison evidence.[103] The Regulator also issues guidelines on specific subjects, including contamination,[104] cognitive bias[105] and validation.[106] Although all commercial and police providers of forensic science evidence are expected to conform to the Regulator's published standards, neither the Codes nor the *Guidance* have statutory force.[107] The Government has recently indicated that it 'will develop proposals to give the Forensic Science Regulator statutory powers, to place the current remit and the associated Codes of Practice on a statutory basis and enable the Forensic Science Regulator to investigate non-compliance where necessary'.[108]

---

[97] Kent Roach 'Forensic Science and Miscarriages of Justice: Some Lessons from Comparative Experience'

[98] Following the closure of the FSS and the HoC Select Committee Report Forensic Science on Trial – consequent need for independent regulation.

[99] Forensic Science Regulator, *Guidance– Validation* (*FSR-G-201*, 2014) (hereafter Guidance) available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375285/FSR-G-201_Validation_guidance_November_2014.pdf, accessed 02/01/2017.

[100] *Ibid*.

[101] The Forensic Science Service dominated the provision of forensic science services until its closure in March 2012, but it was never the sole provider of forensic expertise to the police.

[102] Forensic Science Regulator, *The Regulator's Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System* (2015) (hereafter The Codes) available at https://www.gov.uk/government/collections/forensic-science-providers-codes-of-practice-and-conduct, para 1.1. Accessed 02/01/2017.

[103] See Forensic Science Regulator: *Code of Practice and Conduct 2015: fingerprint comparison* available at https://www.gov.uk/government/publications/forensic-science-providers-codes-of-practice-and-conduct-2015*; DNA analysis: codes of practice and conduct* (2014) available at https://www.gov.uk/government/publications/dna-analysis-codes-of-practice-and-conduct; *Video analysis: codes of practice for forensic service providers* (2014) available at https://www.gov.uk/government/publications/video-analysis-codes-of-practice-for-forensic-service-providers. Accessed 02/01/2017.

[104] Forensic Science Regulator: *Crime scene DNA: anti-contamination guidance* (2016) available at https://www.gov.uk/government/publications/crime-scene-dna-anti-contamination-guidance; *Laboratory DNA: anti-contamination guidance* (2015) available at https://www.gov.uk/government/publications/laboratory-dna-anti-contamination-guidance. Accessed 02/01/2017.

[105] Cognitive bias effects relevant to forensic science examinations. See also Gary Edmond, Rachel Searston, Jason Tangen and Itiel Dror, 'Contextual bias and cross-contamination in the forensic sciences: The corrosive implications for investigations, plea bargains, trials and appeals' (2014) 13 Law, Probability & Risk 1-25.

[106] Guidance, n 99 above.

[107] Annual Report Forensic Science Regulator, *Annual Report 2015* (2015) available at https://www.gov.uk/government/publications/forensic-science-regulator-annual-report-2015, accessed 02/01/2017

[108] Forensic Science Strategy (Cm 9217) March 2016, para. 44.

The Codes of Practice and Conduct for forensic science providers and practitioners in the criminal justice system confirm that it is the duty of individual providers to ensure that all methods and procedures that are used are validated.[109] Validation 'may be performed by the provider, manufacturer or another provider'.[110] The Codes state that validation should ensure that 'results are consistent, reliable, accurate, robust and with an uncertainty measurement', echoing PCAST's concerns that results should be repeatable, reproducible and accurate.[111]

Both the Regulator and PCAST are advocates for the central role of validity as a precursor to the forensic application of any scientific method or technique. Their definitions are similar and consistent[112] in holding that validation entails a demonstration of reliability.[113] The *Guidance* is emphatic regarding the importance of conducting validation studies. The Code requires documentation to state that validation has been completed[114] given that 'forensic science is science':[115]

> [T]he requirement to demonstrate the reliability of scientific evidence clearly extends to any methodology when the operation has an impact in the result obtained, wherever it is used. Therefore validation should be the norm.[116]

Like PCAST, the Regulator is sensitive[117] to the distinction between foundational validity and validity as applied, referring to them as 'external/developmental validation' and 'internal validation' respectively.[118] The issue of proficiency testing raised by PCAST is also addressed by the Regulator as 'technical competence'.[119] Here the Regulator engages in a more expansive consideration of validation opportunities (and requirements) associated with applied science than considered by PCAST in the forensic feature-comparison context. For example, the Regulator suggest that the following specifications should be articulated for any method implementing a scientific model or theory:

i. The validity of the model/theory;
ii. The validity of the application of the model/theory in the method;
…
iv. The validity of the assumptions and any limits on the application of the assumptions;[120]

The importance of identifying and communicating error and uncertainty logically flows from considerations of validity and is explicitly articulated by both the Regulator and PCAST. For the Regulator, the specification of the application of a scientific model or theory should entail a consideration of 'any issues or limitations'[121] or 'any caveats that might apply, e.g., error rates'.[122]

---

[109] The Codes, 20.1.1.
[110] *Ibid*., 20.2.1.
[111] *Ibid*., 20.8.2 (l).
[112] Guidance, 2.2.2 and PCAST p 4-5 (1).
[113] *Ibid*., 1.1.1: 'can be shown to be reliable'; PCAST, 4-5(1): 'can, *in principle,* be reliable."
[114] Ibid., 1.1.9.
[115] *Ibid*., 1.1.1.
[116] *Ibid*., 3.3.3.
[117] *Ibid*., 1.1.5 – 1.1.6.
[118] *Ibid*., 4.2.6 - 4.2.7.
[119] *Ibid*., Section 4.3.
[120] *Ibid*., 4.4.10.
[121] *Ibid*., 4.4.11.
[122] *Ibid*., 3.4.2.

Along with PCAST, the Regulator considers this information to be vital and requires it to be communicated clearly.[123] The Regulator even suggests a possible model for how error rates might be reported: 'e.g., an x% error rate'.[124] Issues associated with the uncertainty of measurement are also given detailed consideration by the Regulator; with an entire section devoted to the concepts of accuracy and precision.[125] Both PCAST and the Regulator insist that it is not enough to merely complete the requisite validation work. It is, in addition, vital for validation results to be peer reviewed, published and made accessible to the public (particularly attentive scientists): 'the Regulator is very clear that "secret science" is not in the interests of transparent justice.'[126]

As explained in Sections 4 and 5, the Codes could be used as a basis for asking a forensic practitioner to demonstrate the foundational validity of her method. Currently this is not a feature of legal practice in England and Wales. As with the CPD, there are few reported cases where advocates or judges refer to the Regulator's Codes or *Guidance* in support of a submission that expert evidence should be excluded (either on the basis that it is non-compliant and therefore insufficiently reliable or under s.78 of PACE), or even as part of the cross-examination of a forensic practitioner in a review of the weight of the evidence.[127] The PCAST report could inform the way Codes and *Guidance* are used to mount challenges to the reports and testimony of forensic practitioners in feature-comparison cases.

The existence of an independent framework for regulation does not yet appear to be ensuring the validity and reliability of forensic science methods in practice. For, as the Regulator's *Guidance* on validity makes clear, '[t]he responsibility for validation rests with the forensic science provider'.[128] This assumes that commercial providers can be relied upon to commit resources to developing and conducting studies of the type and on the scale PCAST recommends. Yet, there are few incentives for an individual provider to conduct the sort of research that might be capable of establishing foundational validity, particularly where this is not expressly requested (or required by courts) and there is a risk that a method – relied upon as a source of income – might be found to be invalid. There is little evidence that the black-box studies envisaged by PCAST are being conducted or that indicative error rates are being measured and reported as a matter of routine. Simultaneously, resource pressures and the drive for efficiency – exemplified in the emergence of streamlined forensic reports – are working to reduce the amount of information available to decision-makers (including the defence). Such orientations tend to elide the lack of validation and uncertainties around the prevalence of errors, and therefore reduce the prospects of identifying or challenging evidence that is speculative or unreliable. They discourage transparency and accountability and in

[123] *Ibid*., 3.4.3.
[124] *Ibid*., 3.4.3. See also Guide, n 28 above.
[125] *Ibid*., section 7.
[126] *Ibid*., 4.4.5. The contention that proprietary interests should override the interests or ability of criminal justice actors to obtain access to information has been criticised by Lord Thomas, CJ, 'Expert evidence and the future of forensic science in criminal trials'. Speech presented at the Criminal Bar Association Kalisher Lecture; 2014, 14 October.
[127] The only references to the Forensic Science Regulator, revealingly relate to DNA challenges and one fingerprint appeal. Notwithstanding more serious question marks hanging over other forensic comparison methods, lawyers and judges have focused almost exclusively on DNA evidence. This is ironic, because, without wanting to suggest that DNA techniques are without their problems, they are some of the few techniques that came from science and have been largely validated. Many techniques that remain unvalidated are yet to be challenged by (supposedly attentive and critical) defence counsel. See, for example, *Smith v R* [2011] EWCA Crim 1296 at [10], [62] (fingerprints); *Broughton v R* [2010] EWCA Crim 549 at [33]-[34] (Low template DNA); *Dlugosz* [2013] EWCA Crim 2 at [23], [29] (Low template DNA – mixed profile without random match probabilities); *Reed and Reed v R* [2009] EWCA Crim 2698 at [5], [72]-[73], [111], [116] (Low template DNA).
[128] Guidance, 8.1.1.

the process do not to convey what is actually known about procedures and the evidence they generate.

Significantly, there is no group in the United Kingdom undertaking the kind of evidence review (or meta-analysis) performed by PCAST for the various forensic feature-comparison methods. It is one thing for the Forensic Science Regulator to develop guidelines and policy documents.[129] It is quite another for them to be adopted or enforced. But, to actually review the available evidence to inform practice is necessary if we hope to provide decision-makers – whether investigators, prosecutors, defence lawyers, judges and juries – with the materials to make informed and defensible decisions.

By virtue of her engagement with validation for all, rather than just the forensic feature-comparison, methods, the Forensic Science Regulator is concerned with the institutional context and the management of risk in a way that PCAST is not. For example, although PCAST clearly considers admissibility for the purpose of minimising wrongful convictions and improving the performance of criminal justice systems, the Regulator devotes a subsection of the *Guidance* to articulating and exploring a risk assessment framework to be informed by validation results.[130] Explicit consideration is given not only to wrongful convictions, but also the potential for wrongful acquittals and obstructing and delaying investigations.[131] In addition, the risk assessment process is intended to aid the development of validation studies, to ensure 'that the validation study correctly assesses whether the risk mitigation put in place works'.[132] It also proposes that the process of risk assessment should take into account the likely weight of the evidence as well as additional information about the method (for example the accuracy of its output, the extent to which its output can be influenced by context, the meaning of absent information, and the potential for delays in the application of the method). These considerations are both complex and worthy of attention. Their explicit consideration by the Regulator is valuable, if only to illustrate that validation information ultimately serves to inform a complex decision environment.

Despite similarities with the PCAST report the Regulator's activities are not without limitation. PCAST purposefully focused on a subset of forensic science practices (i.e. feature-comparison methods) in order to make concrete recommendations specific to those areas. The President's Council proposed a research design considered optimal for the purpose of producing relevant validation data (i.e., 'appropriately designed black-box studies).[133] By contrast the Regulator is obliged to develop materials that apply across the forensic sciences. An unfortunate consequence is that the diversity in forensic science practice (from analytical chemistry to visual feature-comparison) has understandably limited the extent to which the Regulator can offer discipline specific recommendations on validation. While the *Guidance* is unapologetically 'descriptive rather than prescriptive in style',[134] its generality may have resulted in the provision of advice that may be problematic in some instances. The Regulator's position on the use of case materials for validation purposes – at least as applied to the feature-comparison disciplines – provides one example.

The Regulator suggests that:

---

[129] The consistencies are more important than the differences. Both stress the primacy of validation. While we accept that there may be scope to disagree with or qualify aspects of PCAST, the NRC report and even the FSR's advice and guidelines, there is little scope to question the need to validate most procedures used in forensic science and medicine – especially those in routine use. Moreover, there is a need to direct attention to limitations, error and cognitive bias.

[130] Guidance, 5.4.

[131] *Ibid.,* 5.4.3.a-c.

[132] *Ibid*., 5.4.2.

[133] PCAST report, 46.

[134] Guidance, 2.3.2.

> the use of casework material [for validation] is valuable in many areas because it may be difficult to generate a set of test data that adequately reflect the range, quality and complexity of the material submitted in case work.[135]

However, this suggestion is in direct contrast to the advice offered by PCAST:

> … one cannot reliably estimate error rates from casework because one typically does not have independent knowledge of the 'ground truth' or 'right answer'.[136]

The extent to which the Regulator is conscious of this fundamental concern is unclear. For example the *Guidance* suggests the there is a real risk that 'different results may be obtained in the validations study' compared to the casework, giving rise to the possibility that an error had been made in the case.[137] However, without ground truth, if the original casework and validation conclusions differ the source of the error is unknown. Interestingly, the Regulator's apparent reticence with the use of casework material for validation purposes is primarily concerned with the need for permission rather than guaranteeing that the materials are fit for purpose.[138]

Despite this significant departure from the recommendations in the PCAST report, the Regulator and PCAST are in lockstep both about the fundamental importance of validity and the requirement for it to be demonstrated and communicated if we hope to improve the quality of forensic science evidence.

Conclusion

The PCAST report both highlights and extends the important work of the Forensic Science Regulator. Not only does the PCAST report, like the NRC report before it, lend very considerable weight to the need for validation of procedures used in forensic science (and medicine), PCAST went further and actually reviewed the evidence available for a range of procedures. The result of that review, by independent scientists, might come as a surprise to English lawyers and judges. Five of seven feature-comparison methods were found to lack evidence of foundational validity. While the methods selected might not be representative, they include some of the most widely used techniques in the US and UK. In parallel to developments in the US, the Forensic Science Regulator has been actively developing advisory codes and guidelines to inform and improve practice. However, the Regulator is not in a position to review experts' reports or the evidence base for particular procedures or even 'fields' to ascertain whether codes and guidance have been followed. The lawyers and courts practically responsible for regulating expert reports have been slow to engage with new procedural rules and materials published by the Regulator.

A second point to make in conclusion is that there are few grounds for believing that English criminal justice institutions, traditions and practices can somehow circumvent the force of the general scientific critique offered by PCAST. According to the NRC, PCAST and the Forensic Science Regulator, 'nothing can substitute' for validation: '[f]oundational validity is a *sine qua non*, which can only be shown through empirical studies.'[139] Cross-examination, judicial warnings, long experience,

---

[135] Guidance, 6.1.2. See also Forensic Science Regulator, *Protocol: Validation of Casework Material* (FSR-P-300, 2016) available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/510850/Protocol_on_Casework_Material_Issue_1.pdf, accessed 02/02/2017.

[136] PCAST report, 33.

[137] Guidance, 6.1.6.

[138] *Ibid.,* 6.1.8

[139] PCAST report, 147.

other evidence, historical reliance, apparent plausibility, the existence of standards and accreditation, proficiency testing (and so on) cannot substitute for validation. None of these tell us whether a procedure works, in what conditions and how well. If, however, juries are to be left to evaluate purportedly scientific evidence in the absence of such information, they should at least be informed in the clearest possible terms of the importance of the information that is missing.

Thirdly, if we accept that validation is fundamental, and cannot be replaced by other activities (legal or scientific), then the review of evidence undertaken by a group of scientists in the United States suggesting that some procedures do not have the required support would appear to have direct application to the feature-comparison methods in use in the United Kingdom. An unlikely exception might arise if English forensic scientists could point to credible validation studies not considered by PCAST. For those who would argue that the PCAST report's findings to do not apply to the forensic feature-comparison methods used in England and Wales, we would inquire as to the results of rigorous validation studies. To the contention that things are very different in England and Wales (from many jurisdictions in the US) we would agree, but nevertheless once again ask for evidence that any particular difference or set of differences matter. In the absence of empirical evidence, we cannot be confident about English (or Canadian or Australian and so forth) practices, training and traditions in forensic science and medicine or law making a meaningful difference to performance or reliability.

English lawyers and judges should engage with their procedural rules as well as the kinds of insights provided by the NRC, the Forensic Science Regulator and now PCAST. The latter infuse the rules and practice directions with meaning and might help to reform admissibility practices. The PCAST report identifies some of the dangers with the traditional 'laissez faire' approach to admissibility as well as the threat posed to both truth and fairness where procedures are not formally evaluated in ways that would satisfy independent scientists.