JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, November 2009, Vol. 54, No. 6
doi: 10.1111/j.1556-4029.2009.01176.x
Available online at: interscience.wiley.com

Angi M. Christensen,[1] *Ph.D. and Christian M. Crowder,*[2] *Ph.D.*

Evidentiary Standards for Forensic Anthropology*

ABSTRACT: As issues of professional standards and error rates continue to be addressed in the courts, forensic anthropologists should be proactive by developing and adhering to professional standards of best practice. There has been recent increased awareness and interest in critically assessing some of the techniques used by forensic anthropologists, but issues such as validation, error rates, and professional standards have seldom been addressed. Here we explore the legal impetus for this trend and identify areas where we can improve regarding these issues. We also discuss the recent formation of a Scientific Working Group for Forensic Anthropology (SWGANTH), which was created with the purposes of encouraging discourse among anthropologists and developing and disseminating consensus guidelines for the practice of forensic anthropology. We believe it is possible and advisable for anthropologists to seek and espouse research and methodological techniques that meet higher standards to ensure quality and consistency in our field.

KEYWORDS: forensic science, forensic anthropology, *Daubert,* Scientific Working Group, SWGANTH, professional standards, error rate, validation

Admissibility criteria for expert testimony have been established and clarified through three United States Supreme Court decisions. The decisions from *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, *General Electric Co. v. Joiner*, and *Kumho Tire Co. v. Carmichael* were intended to ensure the reliability and relevance of scientific or technical testimony admitted as evidence in federal courts. Since the 1993 *Daubert* ruling, many forensic disciplines including anthropology have determined that there is a need to critically re-evaluate some of the techniques and methods used in their examinations, as well as the validity of the underlying scientific theories. Disciplines like forensic anthropology may be problematic in the eyes of the courts since they employ a combination of traditional scientific methodologies and less rigorous observational methodologies coupled with case study evaluations or casework experience. Several recent papers have advocated more earnest consideration of the *Daubert* guidelines when conducting research and preparing testimony in forensic anthropology (1–5). This has likely contributed to the increased awareness and interest in evaluating some of the techniques most often used by forensic anthropologists. Issues related to quality assurance, however, have been less often and less aggressively addressed.

At present there are no professionally agreed upon standards for the application of forensic anthropological methods regarding the recovery and analysis of remains, which leaves the individual or forensic institution to develop their own guidelines and standards. Some organizations such as the Department of Defense's (DOD) Joint POW/MIA Accounting Command (JPAC) Central Identification Laboratory (CIL) and the Federal Bureau of Investigation (FBI) Laboratory are accredited by the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB), but ASCLD/LAB does not specifically recognize anthropology as an independently accreditable discipline. The American Board of Forensic Anthropology (ABFA) was created to examine and certify forensic anthropologists and set standards for their individual proficiency, but this organization does not (nor does any other) provide protocols to ensure consistency and reliability in the application of forensic anthropological methods.

Creating and maintaining professional and scientific standards is achieved in many disciplines by Scientific Working Groups (SWGs). A SWG consists of a group of experts in a particular scientific discipline that meets periodically to formulate and review standards (for both examination protocols and validation testing) applied in their respective fields, and standards set by SWGs are increasingly recognized and considered by the courts. As the courts continue to raise the bar regarding scientific standards, forensic anthropology must be committed to providing analyses that are of the highest quality and reliability, and we believe that creating and adhering to recognized standards will facilitate achieving this objective.

The purpose of this paper is multi-faceted; it reviews the three major court decisions involving admissibility of expert testimony, examines the current interplay between legal and scientific culture within the forensic disciplines, specifically identifies areas where the field of forensic anthropology can improve regarding issues of method validation and quality assurance, and introduces the current discourse within the field regarding best practice protocols for practitioners.

Daubert v. Merrell Dow Pharmaceuticals

The parents of Jason Daubert and Eric Schuller filed suit against the pharmaceutical company Merrell Dow claiming that their children's birth defects were the result of the mothers' prenatal ingestion of the prescription drug Bendectin, which is designed to alleviate morning sickness. Crucial scientific evidence showing a

[1]Federal Bureau of Investigation, Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135.
[2]Office of Chief Medical Examiner, 520 First Avenue, New York, NY 10016.
*Presented at the 60th Annual Meeting of the American Academy of Forensic Sciences, February 18–23, 2008, in Washington, DC.

Received 25 Sept. 2008; and in revised form 5 Dec. 2008; accepted 6 Dec. 2008.

causal relationship between the use of the drug during pregnancy and birth defects was deemed inadmissible because it did not meet the general acceptance criterion established in *Frye v. United States* (6). The Supreme Court granted certiorari, "in light of sharp divisions among the courts regarding the proper standard for the admission of expert testimony" (7). The Supreme Court stated that the 1975 Federal Rules of Evidence (FRE) Rule 702 superseded the *Frye* test, thus determining the standard for relevance and reliability of expert testimony (8). Furthermore, the Supreme Court ruled that it is the trial judge who must ensure that all scientific testimony is relevant and reliable. To assist the trial judges, the Court provided the following general guidelines for assessing the admissibility of scientific testimony known as the *Daubert* standard:

- Has the theory or technique been tested?
- What is the known or potential rate of error?
- Do standards exist for the control of the technique's operation?
- Has the theory or technique been subjected to peer review and publication?
- Has the theory or technique been generally accepted within the relevant scientific community?

It is important to understand that the *Daubert* opinion instructs the judge to focus on principles and methods and not the conclusions that they generate.

The 1993 Supreme Court decision *Daubert v. Merrell Dow Pharmaceuticals, Inc*. was intended to ensure the reliability and relevance of the scientific testimony, instructing the judge to be the "gatekeeper" in keeping "junk science" out of the courtroom or restrict evidence with less than a direct fit to the issues of the trial (9). One of the goals of this decision was to curtail court cases from becoming a battle of the experts, which in effect may cause the trial to be decided by the experts rather than the courts (8). In the wake of *Daubert*, flexibility in the admissibility criteria for expert testimony has been called for in two other landmark United States Supreme Court cases which supplemented and clarified the decision in *Daubert*. Confronted with some of the more difficult issues regarding expert testimony, the Supreme Court's decisions in *General Electric Co. v. Joiner* and *Kumho Tire Co. v. Carmichael* were explanatory of *Daubert*. These landmark cases coupled with *Daubert* form the trilogy of Supreme Court decisions that set the legal standard for evaluating the admissibility of expert testimony.

General Electric Co. v. Joiner

Robert Joiner was an electrician in the Water & Light Department of Thomasville, Georgia who filed suit in state court regarding his claim that his small-cell lung cancer had been caused by exposure to polychlorinated biphenyls (PCBs) on the job. The court ruled the experts' testimony failed to demonstrate the link between PCB exposure and small-cell lung cancer and, therefore, was inadmissible because "it did not rise above subjective belief or unsupported speculation" (10). According to the court, there is nothing stated in the FRE or *Daubert* ruling that requires a district court to accept expert testimony which is connected to existing data only by the unproven assertion of the expert (10,11). Furthermore, as indicated by *Joiner*, the court may conclude that there is too great an analytical gap between the data and the opinion proffered. This questions whether existing scientific evidence can be generalized to address specific causal relationships. As professionals, experts should be cautioned in over extending analyses; however, admissibility of this type of testimony would be assessed on a case-by-case basis by the courts.

The decision put forth from *Joiner* produced two significant results. First, the Supreme Court ruled that the appellate court erred in overturning the admissibility decision of the trial court. The Supreme Court explained that the appellate courts should only overturn admissibility decisions when the trial court abuses its discretion, which is a failure to take into proper consideration the facts and law relating to a particular matter. Second, it was argued that methodology and conclusions are not completely distinct from each other as stated in *Daubert* (8). According to *Joiner*, the two are linked together and an expert's conclusion should be excluded in the event that valid reasoning does not support it. Overall, the *Joiner* decision questioned and clarified the language in *Daubert* and solidified the burden of admissibility for the trial courts.

Kuhmo Tire Co. v. Carmichael

In July of 1993, eight members of the Carmichael family were involved in a motor vehicle accident when one of the tires of their vehicle failed. The plaintiffs' expert determined that the tire did not fail from abuse, thus the tire was defective. The trial judge of the district court ruled that the expert's testimony did not satisfy the admissibility criteria outlined by the *Daubert* standard. The District Court determined that the expert testimony "fell outside the range where experts might reasonably differ, and where the jury must decide among the conflicting views of different experts, even though the evidence is 'shaky'" (12). The decision was appealed under the grounds that the expert was not a scientist, thus his testimony not subject to *Daubert*. The appellate courts determined that the district court applied the *Daubert* criteria incorrectly, being that the expert's testimony should be considered non-scientific and not applicable under *Daubert*. The Supreme Court resolved the disagreement between the two courts and provided reasons why *Daubert*'s general reliability requirement applies to all expert testimony (12,13).

The *Daubert* "gatekeeping" obligation applies not only to "scientific" testimony, but to all expert testimony. Rule 702 does not distinguish between "scientific" knowledge and "technical" or "other specialized" knowledge, but makes clear that any such knowledge might become the subject of expert testimony. *Daubert* referred only to "scientific" knowledge because that was the nature of the expertise there at issue. Rules 702 and 703 grant all expert witnesses, not just "scientific" ones, testimonial latitude unavailable to other witnesses on the assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline. Finally, it would prove difficult, if not impossible, for judges to administer evidentiary rules under which a "gatekeeping" obligation depended upon a distinction between "scientific" knowledge and "technical" or "other specialized" knowledge, since there is no clear line dividing the one from the others and no convincing need to make such distinctions (12).

In the delivering of the Court's opinion, Justice Breyer relays the flexibility of the *Daubert* factors and states that all of the *Daubert* criteria do not necessarily apply in all circumstances even when scientific evidence is being reviewed. The Court explained that the judicial goal is always to ensure the reliability and relevancy of expert testimony (12). The Court referred to their decision in *Joiner* and reiterated "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* [he himself said it] of the expert" (10). In other words, expert testimony may not be admissible if the data is not sufficiently connected to the conclusions.

The outcome of the Supreme Court's decision in *Kumho v. Carmichael* produced significant results regarding the Court's approach to the admissibility of expert testimony. First, the Court reaffirmed the gatekeeper role of the judge and held that both scientific and non-scientific expert witness testimony must be reliable and relevant to be admissible. With that said, the Court realized that there must be a flexible approach for assessing expert testimony considering how the type of evidence may vary across disciplines. This should not be interpreted to mean that a trial court might ignore *Daubert* factors in non-scientific cases (13). The judge should consider that not all of the *Daubert* criteria may be applicable to the expert testimony and those that do apply should be utilized to evaluate admissibility. The *Kuhmo* decision clarified that the expert witnesses should use "the same level of intellectual rigor that characterizes the practice of an expert in the relevant field" (12). This statement essentially tethers admissibility standards with professional standards (11). Finally, the decision indicates that the courts need to recognize that expert testimony may employ a combination of scientific methodology and less rigorous case study and observational methodologies (14).

Aftermath of the Trilogy

While not all states have adopted the *Daubert* standard, research has shown that in both *Frye* and *Daubert* jurisdictions 94% of state court judges contend that they find *Daubert* valuable to their decision-making (15). The results of the *Daubert* trilogy outlined the importance of judicial gate-keeping and emphasized the need for judges and legislators to set the standards for admitting testimony. Under the trilogy rulings, the admissibility standard has become more encompassing and will decrease the introduction of "soft" science and speculation-based experience into the courtroom (16). Subsequent amendments to the FRE following these landmark cases further clarify the legal language and eliminate loopholes within the FRE regarding admissibility criteria for expert testimony.

Considering how law and science continue to converge, the science of anthropology (as well as other forensic disciplines) must be conceived under the rubric of evidentiary examination and methods need to be based on a sound scientific foundation with justifiable protocols. It is extremely important to state that the admissibility rules set forth from the courts do not instruct scientists how to perform scientific research. *Daubert* does not and will not dictate science, but the ruling did lift the proverbial blinders from many forensic disciplines causing the realization that scientific rigor may be lacking. Furthermore, scientists and other professionals may greatly improve the admissibility potential of their testimony by providing guidance to the courts on how to "assess causality at the level of certainty required by the law" (11:1386).

Daubert and the Forensic Disciplines

In response to the aforementioned rulings, many forensic disciplines have opted to critically re-evaluate some of the techniques and methods used in their examinations to ensure the scientific method, where applicable, was followed. According to Sanders (13), the threshold of admissibility may not be equal for some areas of knowledge due to more sophisticated or sensitive measuring equipment, more developed methods, or the ability to control for a greater variety of confounding variables. Disciplines that have been issued non-favorable *Daubert*-based rulings may determine that it is worthwhile to launch research programs to meet and mitigate the specific criticisms of the courts. Conceptualizing methods within the rubric of evidentiary examination in disciplines like forensic anthropology is complicated, in that they employ a combination of traditional scientific methodologies and less rigorous observational methodologies.

According to a recent publication by Grivas and Komar, "many anthropological techniques already meet the criteria for admissibility under Kumho, potentially making many revisions (of current methods) unnecessary" (4:3). There are two issues with this statement. First, it seems to imply that the judge makes a decision to use either the *Daubert* or *Kumho* standard, yet *Kumho* is not a stand-alone standard. The *Kumho* decision allows for flexibility to be considered by the judge in that not all, but some, of the *Daubert* criteria may be applicable to non-scientific expert testimony. The authors do note, however, that the *Kumho* decision allows anthropologists latitude provided that the analysis is both scientific and rigorous (4). Second, it is unknown if current anthropological techniques (with the exception of frontal sinuses) successfully meet any admissibility standards until they are challenged and accepted by the courts. The field can strive for compliance with the *Daubert* criteria, but only the courts can determine if expert testimony has successfully met admissibility standards. It is possible that through the *Kumho* ruling experienced-based expert testimony from forensic anthropologists may be recognized; however, it could also be determined that the observational methods are too subjective.

In evaluating the current research response within the field of anthropology to address issues of admissibility, Grivas and Komar suggest that in attempting to meet the admissibility standards, anthropologists may be "trying needlessly to force powerful qualitative techniques into quantifiable categories" (4:4). In the absence of empirical hypothesis testing, how do we assess the "power" of our methods? Can anthropologists meaningfully label a method powerful in the absence of a scientifically sound methodological or theoretical basis? As professionals and good scientists we should ensure that our methods are valid, transparent, and reliable. While experience is vital in *interpreting* analytical results, experience should not dictate the power of scientific methods within the field of forensic anthropology. Experience is not a replacement for a rational process in reaching an expert decision and the acquisition of experience-based knowledge does not prevent an expert from providing a rational process foundation for that knowledge (13). Even more experience-based anthropological analyses, such as taphonomy and trauma, have a growing theoretical background using empirically based hypothesis testing. Fracture patterns, for example, are the result of principles of bone biomechanics and physics, and predictable patterns can be (and have been) documented. Many recent actualistic studies, including those carried out at the University of Tennessee's Anthropological Research Facility, provide ample empirical framework from which we can draw in evaluating taphonomic change.

The importance of experience level should not be disregarded; however, a practitioner may have many years of experience in which the quality of work has never been evaluated. It is possible that a practitioner has been repeating the same errors in analyses. For example, practitioners misinterpreted the direction of butterfly fractures on skeletal elements for many years before empirical data demonstrated how bone reacts under mechanical loading. Grivas and Komar postulate that the "empirical focus may misrepresent the actual state of the field (of forensic anthropology) as much anthropological testimony appears to be based on more qualitative methodologies" (4:3). This implies that an empirical focus only allows for quantitative and not qualitative methodologies. The empirical approach is the cornerstone for knowledge in both natural and social sciences. Furthermore, quantitative data is based upon

qualitative judgments, and all qualitative data can be described and manipulated numerically.

Despite the uncertainty that disciplines may feel regarding the interpretation of forensic evidence under *Daubert*, recent endeavors within the various fields (including anthropology) have, for the most part, addressed the *Daubert* factors of scientific testing, peer review, and general acceptance (see 1,2,17–21 for examples). Issues of quality assurance including validation, error rates, and professional standards, however, have been less often addressed. Here, we discuss these issues and suggest ideas for consideration and further evaluation.

Quality Assurance

Many forensic disciplines, including identification sciences like anthropology, involve some degree of subjectivity. It is therefore imperative to minimize the risk of error through quality assurance (i.e., proper training, method validation, accreditation, and certification). Quality assurance will help ensure the high quality of anthropological research, assist with establishing method transparency, and provide a secure foundation for forensic anthropologists in the courtroom. Methods with vague descriptions of samples, procedures, variables, or accuracy and bias values (when applicable) should not be considered for use in the forensic setting. Preferred methods, as determined by the analyst, should be selected for scientific reasons and not personal preference or familiarity. Research intended to conduct validation studies of existing techniques must be performed exactly as the method describes or it must be demonstrated that methodological changes do not significantly alter results. Researchers must clearly state method accuracy and precision in a manner that is not only statistically significant, but forensically meaningful. Quality assurance in forensic anthropology can be established through validation studies of analytical methods in order to determine method reliability (precision and accuracy) and through the development of professional standards in the form of best practice protocols.

Validation Studies and Error Rates

Forensic anthropologists must set standards for a theoretical and empirical validation process to guide researchers and practitioners, as well as assist the courts. It is also important to understand that the point of developing methods under the rubric of evidentiary examination is not to completely quantify the field, and that subjectivity does not necessarily equal unreliability. When dealing with the applied sciences, including those involved in forensic anthropological examinations, it is necessary to evaluate the reliability of procedures in its appropriate context. Many principles and methods developed in physical anthropology may be reliable for questions regarding populations or groups, but may require adjustments when applied to problems related to individuals and personal identification.

While anthropologists have taken it upon themselves to validate and improve methods within the field, validation studies are often problematic due to the tendency of researchers to modify or adapt the techniques rather than test the methods as originally presented. In other forensic disciplines including firearms and tool marks, DNA, questioned documents, and chemistry, practitioners often rely on a collection of published validation studies for the technique(s) used. The discipline of forensic anthropology would likely benefit from more specifically aimed validation studies bearing in mind specifics such as sample size, appropriate statistics, intra- and inter-observer error. Recent progress has been made in the area of identification (17–19) and sex determination (18,21), but techniques applied to age estimation (22) and ancestry determination have been addressed less frequently.

Forensic anthropology is fairly unique among the identification sciences in that the techniques used are extremely variable and highly dependent on the evidence available (i.e., the completeness and condition of the skeletal remains), as well as the types of examinations requested. Subjectivity envelops many anthropological methods owing to the variable nature of analyses and more importantly the degree of human variation, but as previously stated, subjectivity does not necessarily equal unreliability. As a consequence to the aforementioned variation, the experimental and statistical approaches used for evaluating the reliability of anthropological techniques are also necessarily very different. Validation studies in anthropology will not result in a binary response, but instead provide an accuracy rate and precision level to be used in determining the reliability of the method.

Method reliability, or method performance, is a multi-level analysis. Concepts of bias, precision, and accuracy form the basis to determine method reliability. These concepts can be evaluated through a multitude of statistical techniques (some more appropriate than others), yet the terms are often conflated in the anthropological literature. Measurement and estimation bias typically leads to a systematic under- or overestimation of the true value (23). Precision refers to the absence of random error and its magnitude is independent of the true value. Precision measures the spread of the data or the statistical variance (24). The levels of bias and precision affect method accuracy, which is the distance between estimated and observed values to the true value. A less accurate technique (e.g., age estimates based on cranial suture closure) is still valuable in a forensic context, especially if it is the most accurate technique at our disposal based on available skeletal material. Anthropologists may be concerned that revealing method error might render the scientific evidence inadmissible in the courts; however, the courts have not determined how much error is too much. Slobogin (16) suggests that under one interpretation of Daubert, high error rates of a method should not be grounds for determining scientific testimony inadmissible because the values can be communicated to a jury for their consideration (16). Currently, it is not necessary to demonstrate that every technique we use is highly reliable, but rather it is important to show that we are (scientifically, statistically) sure of *how* reliable a technique is. The challenge is to employ research designs that adequately test the variable(s) of interest and give us proper measures of reliability.

Anthropological studies have exhibited varied success at properly evaluating the reliability of certain traits used in anthropological analyses. Contemporary anthropologists are using more sophisticated measurement techniques and statistical analyses to evaluate human skeletal remains, and are also increasingly finding ways of quantifying traits that have historically remained fairly subjective and thought to be unquantifiable (frontal sinuses, for example). In many cases, determining the best or most appropriate way to measure and statistically analyze a feature, technique, or approach can be challenging. For the most part, contemporary research within the field presents error values, but the term is often not defined and the potential effect on evidentiary examination is not addressed.

Admittedly, addressing the issue of error rates of forensic anthropological techniques is a rather daunting task. The known or potential error rate can refer to a number of things including the confidence interval, the statistical significance of the result, and the likelihood of being wrong. For the most part it appears that *Daubert* is concerned with methodological error; however, one of the criteria considers the existence and maintenance of standards controlling the technique's operation. This could be interpreted to

include practitioner error. Practitioner error speaks to the individual's qualifications and other quality assurance issues that an agency or laboratory may employ, which are commonly evaluated through proficiency exams. The known or potential error rate is usually addressed through validation studies, while proficiency testing assesses analytical capabilities. In the absence of proper method validation, the courts may attempt to derive method error from proficiency testing results. As professionals, we are concerned with both types of error and at times it may be difficult to evaluate one without assessing the other. Practitioner mistakes, especially those that result in a misidentification, challenge the view of method reliability regardless of the validity of the method (25).

Some practitioners (such as hair examiners and fingerprint examiners) contend that the error for their technique or method, when properly applied, is zero (26). For example, the following was testimony provided by fingerprint examiner William Leo explaining the reasoning behind the zero error rate:

> And we profess as fingerprint examiners that the rate of error is zero. And the reason we make that bold statement is because we know based on 100 years of research that everybody's fingerprint (sic) are unique, and in nature it is never going to repeat itself again (27).

The fallacy in the expert's reasoning is overt, yet this logic is still prevalent in the forensic disciplines. The expert fails to realize that despite the strength of the basis for fingerprint identification (fingerprints are unique), both method and practitioner error still exist. The question of error does not focus only on the uniqueness of fingerprints, but on how reliable the methods of fingerprint examination are in determining a positive match or exclusion.

Other practitioners contend that providing error rates is impossible or unnecessary since they cannot possibly be known or determined for a particular forensic case. Most forensic scientists recognize that in many cases the probability of misidentification or misassociation *for that particular case* would be difficult to determine, but this is precisely why we estimate this error using experimental studies. Based on proper hypothesis testing and statistical analysis of collected data, we can place a probability or confidence interval on our likelihood of correct assessment. Determining the forensic or case-specific error rate, on the other hand, would require an independent examination using another, more discriminating forensic technique (see [28] for an example of hair examinations using both microscopy and mtDNA). Independent examinations may be possible to establish, for example, dental, frontal sinus, or cranial suture identification error rates, but since such examinations are considerably rarer than other forms of forensic identification, sample size may be a problem. Another issue is establishing error rates for the "unique" identifiers that anthropologists currently use for identification, such as antemortem fractures, anatomical anomalies, and presence of surgical implants. Grivas and Komar (4) question the extent to which these types of data are testable and subject to error rates; however, probabilities can be provided for these "unique" identifiers by determining how unique these features are in a forensic population. Practitioners should ask themselves: What is the population frequency of this skeletal defect, anomaly, or fracture? In absence of frequency data, should these identifiers be called unique? In the interim they are still useful for establishing tentative identifications, but multiple lines of evidence should be employed to produce a comprehensive identification. One must also not make the mistake in logic as seen with the fingerprint examiner; no matter how unique an identifier is we must still evaluate error associated with the analytical method.

Professional Standards

Standardization of research results, including appropriate statistical models and levels of precision, in the form of best practice protocols will help ensure a higher quality of anthropological research, and provide a secure foundation for forensic anthropologists in the courtroom. Numerous laboratories including the CIL and the FBI Laboratory are accredited by ASCLD/LAB, a voluntary program for crime laboratories to demonstrate that its management, personnel, operation, technical procedures, equipment, and physical facilities meet established standards. It offers accreditation in specific forensic disciplines such as firearms and trace evidence, but ASCLD/LAB does not offer accreditation in forensic anthropology *per se* (although ASCLD/LAB has been approached about the possibility of adding forensic anthropology to its accreditable disciplines). Several medical examiner systems have had their agencies accredited by the National Association of Medical Examiners (NAME), but (like others) this accreditation program does not include forensic anthropology. Individually, anthropologists have been addressing the issues put forth by the courts, but a more comprehensive approach has been initiated by the CIL and the FBI Laboratory in the creation of a Scientific Working Group. A Scientific Working Group is comprised of members representing a forensic science discipline that meet periodically to discuss and address "best practices" applied in their respective fields in order to ensure quality and consistency in both methods and research. Traditionally, these groups have been sponsored by the FBI Laboratory, which began sponsoring SWGs in the early 1990s in order to improve practices and build consensus with federal, state, and local forensic community partners (29). The ultimate objective of SWGs is to publish and disseminate these guidelines so that they are available to the relevant community and the courts. Since their inception, guidelines and standards set by SWGs have been increasingly recognized and considered by courts, and though voluntary in most cases, adhering to SWG guidelines is widely regarded as good practice.

The most recent addition to the SWG groups is the Scientific Working Group for Forensic Anthropology, or SWGANTH, which is co-sponsored by the CIL and the FBI. The first meeting was held in January 2008, and meetings will be held annually at a minimum. The purposes of SWGANTH, similar to those of other SWGs, are to develop consensus guidelines for the forensic anthropology discipline, and to disseminate these guidelines as well as other relevant studies and findings that may be of benefit to the forensic community. The guidelines are intended to be specific enough to ensure quality and consistency of practice, yet broad enough to be applicable across various jurisdictional types and laboratory settings, as well as be internationally relevant.

SWGANTH members have identified various topic areas in forensic anthropology that they felt warrant review and could benefit from documented, standardized guidelines including: qualifications and training, ethics, statistical methods, aspects of the biological profile, documentation and reporting, and various others. For more information on the Scientific Working Group for Forensic Anthropology including general information, bylaws, a list of current Board Members, and Committee topics, visit http://www.swganth.org. Once drafted, all guidelines will be made available for public review and feedback on this site prior to final publication and dissemination.

Conclusions

It is clear that science and law continue to interact and interrelate, and anthropologists are becoming increasingly aware of the

need to seek out more empirically grounded methods. While the re-evaluation of various anthropological techniques is much needed and has become a recent focus within the professional literature, less attention has been paid to the quality assurance issues of validation, error rates and standardization for methods in forensic anthropology. We believe it is possible for the field to seek and espouse research and methodological techniques that meet higher standards in terms of validation and error estimation. Regarding professional standards, a Scientific Working Group for Forensic Anthropology was recently formed, and should significantly improve the discipline by identifying "best practices" and fostering communication and consensus among practitioners, thereby ensuring quality and consistency in our field.

Acknowledgments

We would like to thank Dan Olson and Stephen Bunch of the FBI Laboratory for their input and discussions on Daubert issues. We also thank Dr. Tracy Rogers of the University of Toronto at Mississauga for her work on the AAFS 2008 abstract and comments regarding this manuscript.

Disclaimer

The research presented in this manuscript was not conducted under the auspices of the Federal Bureau of Investigation (FBI) or the Office of Chief Medical Examiner—New York City (OCME-NYC). The opinions expressed herein are those of the authors and do not reflect the opinions of the US Department of Justice, the FBI, or the OCME-NYC.

References

1. Crowder CM, Rogers TL. Forensic anthropology and meeting evidentiary/legal demands (The fourth era of forensic anthropology: examining the future of the discipline symposium). Proceedings of the 59th Annual Meeting of the American Academy of Forensic Sciences; 2007 Feb 19-24; San Antonio, TX. Colorado Springs, CO: American Academy of Forensic Sciences, 2007.
2. Murray BA, Anderson BE. Forensic anthropology in the courtroom: trends in testimony. Proceedings of the 59th Annual Meeting of the American Academy of Forensic Sciences; 2007 Feb 19-24; San Antonio, TX. Colorado Springs, CO: American Academy of Forensic Sciences, 2007.
3. Grivas CR, Komar D. Daubert and Kumho: implications for anthropologists in the courtroom. Proceedings of the 59th Annual Meeting of the American Academy of Forensic Sciences; 2007 Feb 19-24; San Antonio, TX. Colorado Springs, CO: American Academy of Forensic Sciences, 2007.
4. Grivas CR, Komar D. *Kumho, Daubert,* and the nature of scientific inquiry: implications for forensic anthropology. J Forensic Sci 2008;53(4):771–6.
5. Christensen AM. The impact of *Daubert*: implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). J Forensic Sci 2004;49(3):427–30.
6. *Frye v. United States*, 293 F. 1013 (C.A.D.C 1923).
7. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
8. Godden DM, Walton D. Argument from expert evidence: critical questions and admissibility criteria of expert testimony in the American legal system. Ratio Juris 2006;3(19):261–86.
9. Faigman DL, Kaye DH, Saks MJ, Sanders J. Modern scientific evidence: the law and science of expert testimony. Vols. 1 and 2. St Paul, MN: West, 1997.
10. *General Electric Co. v. Joiner*, 522 US 136 (1997).
11. Kassirer J, Cecil J. Inconsistency in evidentiary standards for medical testimony. JAMA 2002;288(11):1382–7.
12. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).
13. Sanders J. *Kumho* and how we know. Law Contemp Probl 2001;64 (2–3):373–415.
14. Cecil JS. Ten years of judicial gatekeeping under *Daubert*. Am J Public Health 2005;95(S1):S74–80.
15. Moreno JA. Einstein on the bench?: exposing what judges do not know about science and using child abuse cases to improve how courts evaluate scientific evidence. Ohio State Law J 2003;64:531–44.
16. Slobogin C. The structure of expertise in criminal cases. Seton Hall Law Rev 2003;34:105–21.
17. Komar D, Lathrop S. Frequencies of morphological characteristics in two contemporary forensic collections: implications for identification. J Forensic Sci 2006;51(5):974–8.
18. Rogers T, Allard TT. Expert testimony and positive identification through cranial suture patterns. J Forensic Sci 2004;49(2):203–7.
19. Christensen AM. Testing the reliability of frontal sinus outlines in personal identification. J Forensic Sci 2005;50(1):18–22.
20. Christensen AM. Assessing the uniqueness of individual frontal sinus outlines. Am J Phys Anthropol 2005;127(3):291–5.
21. Williams BA, Rogers TL. Evaluating the accuracy and precision of cranial morphological traits for sex determination. J Forensic Sci 2006;51(4):729–35.
22. Crowder CM. Evaluating the use of quantitative bone histology to estimate adult age at death [doctoral dissertation]. Ontario: University of Toronto, 2005.
23. Bruno AW, Moore JL. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography 2005;28:815–29.
24. West MJ. Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias. Trends Neurosci 1999;22:51–61.
25. Simpson EK, James RA, Eitzen DA, Byard BW. Role of orthopedic implants and bone morphology in the identification of human remains. J Forensic Sci 2007;52(2):442–8.
26. Cole S. Is fingerprint identification valid? Rhetorics of reliability in fingerprint proponents' discourse. Law Policy 2006;28(1):109–35.
27. *People v Gomez*, Trial Transcript, No 99CF 0391 (Cal Super Ct Orange Cty. 2002).
28. Houck MM, Budowle B. Correlation of microscopic and mitochondrial DNA analysis of hairs. J Forensic Sci 2002;47(5):964–7.
29. Adams DE, Lothridge KL. Scientific working groups. Forensic Sci Commun 2000;2(3). Available from: http://www.fbi.gov/hq/lab/fsc/backissu/july2000/swgroups.htm.

Additional information and reprint requests:
Angi M. Christensen, Ph.D.
FBI Laboratory, Trace Evidence Unit
2501 Investigation Parkway
Quantico, VA 22135
E-mail: angi.christensen@ic.fbi.gov