The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Assessing Cognitive Bias, Method Validation, and Equipment Performance for the Forensic Anthropology Laboratory**
Author(s):	**Michal Pierce, M.S., D-ABC, Deborrah Pinto, Ph.D., Michal Pierce, M.S., D-ABC**
Document Number:	**254550**
Date Received:	**February 2020**
Award Number:	**2016-DN-BX-K003**

FINAL SUMMARY OVERVIEW

National Institute of Justice Award Number:	2016-DN-BX-K003
Project Title:	Assessing Cognitive Bias, Method Validation, and Equipment Performance for the Forensic Anthropology Laboratory
Principle Investigator:	Michal Pierce, M.S., D-ABC, Quality Director Michal.pierce@ifs.hctx.net
Co-Principal Investigator:	Deborrah Pinto, Ph.D., Forensic Anthropologist Deborrah.pinto@ifs.hctx.net
Submitted by:	Michal Pierce, M.S., D-ABC, Quality Director Michal.pierce@ifs.hctx.net
Submission Date:	December 26, 2018
DUNS and EIN Numbers:	
Recipient Organization:	Harris County Institute of Forensic Sciences, Houston, Texas
Project Period:	Start Date – January 1, 2017, End Date – December 31, 2018

Michal Pierce

Project Purpose

The purpose of this project was to improve the examination and interpretation of physical evidence in forensic anthropology laboratories by improving method validation and equipment performance check procedures related to specific analyses. Although the anthropological methods and equipment tested for this project are generally accepted in the scholarly community and the field-at-large, internal validation and verification, as well as the establishment of performance check procedures, are critical for quality assurance purposes. As an accredited laboratory, the Harris County Institute of Forensic Sciences (HCIFS) Forensic Anthropology Division (FAD) recognized the need to demonstrate that methods and equipment being used are working properly in their laboratory, and that personnel are properly trained and qualified to use such methods and equipment. This internal testing and documentation provides accountability and transparency, indicating that methods, procedures, equipment, and staff have been critically assessed.

Additionally, this project sought to identify where and how cognitive bias impacts forensic anthropological analyses, and how to mitigate potential sources of bias. Specifically, is contextual information relayed through the HCIFS biasing the examiners when performing certain analyses? In order to achieve this goal, the flow of information from the death scene to the laboratory, and within the laboratory, was examined. Although studies have been published that address major sources of cognitive bias, these studies may not have captured the nuances of anthropological activities in the medical examiner setting. Thus, performing internal studies, using the FAD's procedures, would assist in answering the above question and improve the accuracy and reliability of forensic analyses in a medical examiner context.

Project Design and Methods

In order to meet the goals and objectives of this project, a grant-funded Postdoctoral Fellow was hired at the HCIFS for a sixteen month period to work on the studies involved. The Fellow, who had a background in forensic anthropology and practical experience in medical examiner operations, became fully immersed in the FAD and liaised between the FAD and Quality Management Division (QMD). This project was divided into four phases, as described below.

Phase 1: Project Assessment and Initial Training: In the first phase of the project, the Fellow completed orientation and initial training in the FAD and QMD. The Fellow was assigned a staff mentor and worked under the close supervision of the three staff Forensic Anthropologists. The Fellow was also assigned a mentor from the QMD who introduced her to the agency's quality assurance program and the specific standards for FAD activities. Within the QMD, the Fellow learned about issues affecting all forensic analyses, such as measurement uncertainty, cognitive bias, method validation, equipment maintenance, and competency and proficiency testing. This training, coupled with the orientation to the FAD, provided the Fellow with the proper foundation to identify and evaluate factors affecting the accuracy and reliability of analyses performed by the FAD staff.

Phase 2: Validation and Performance Measures: The second phase of the project addressed the reliability of data acquired with specified equipment and software utilized by the FAD. Although these data were assumed to be reliable based on general acceptance by members of the forensic anthropological community, there were no internal validations performed when the equipment and software were received at HCIFS.

In addition to standard calipers, the FAD uses a MicroScribe G2LX digitizer, which digitally captures 3-dimensional coordinate data, for measuring craniums. The 3Skull software program converts these 3-dimensional coordinate data into measurements, or interlandmark distances (ILDs), that can be used to metrically estimate the ancestry, sex, and stature of human skeletal remains. 3Skull does not archive data or produce a database, so the Advantage Data Architect software is used to archive and save these metric data. To validate the digitizer and its corresponding software, one non-forensically significant cranium was pre-marked with the 42 landmarks that are recorded during the digitizing process. The MicroScribe G2LX digitizer was used to collect the coordinate data—all HCIFS Anthropologists recorded these landmarks at least twice—which were captured and converted into ILDs by 3Skull. The ILDs were saved in Advantage Data Architect for later export to analytical software such as Fordisc 3.1. For comparison, all Anthropologists measured the same pre-marked cranium using sliding and spreading calipers.

These digitizer and caliper data were statistically analyzed, and the uncertainty of measurement was calculated for each equipment item. In order to provide adequate training on the use of the digitizer and its associated software for the HCIFS Anthropologists, it was determined that a new standard operating procedure (SOP) was required. Based on the validation process, an SOP addressing the MicroScribe digitizer and the collection of craniometric data was written by the Forensic Anthropology Director and the Fellow.

The primary software used by the FAD to assist with the estimation of ancestry, sex, and stature for unknown decedents is Fordisc 3.1. This software uses multivariate statistical classification methods, specifically discriminant function analysis, to analyze osteometric data, and it was designated as one of the main programs requiring more extensive internal validation.

To perform the validation of the Fordisc software, metric data collected from eight donated, known individuals were evaluated. Cranial and postcranial measurement datasets and Fordisc 3.1 outputs were provided by Dr. Kate Spradley (in consultation with the Fellow who visited Dr. Spradley at her laboratory to train and develop this study) at Texas State University San Marcos' Forensic Anthropology Center. Six of the individuals were considered "type specimens," meaning their osteometric dimensions were typical for their known/reported ancestry and sex, as determined by the composition of the Forensic Data Bank (FDB), the database the Fordisc software uses to classify individuals. The remaining two individuals were considered "atypical," meaning their osteometric data were not typical for their known/reported ancestry when compared to data in the FDB.

To test the software in-house, the metric data for each individual provided by Dr. Spradley was loaded into Fordisc 3.1 at HCIFS. The sex of each individual was provided and entered because a sex-only function (i.e., to estimate the sex of an individual without an associated ancestral group) in the software is not available. However, to assess the functionality of the software to estimate sex—although not independent of ancestry—one dataset was run as a complete unknown (i.e., unknown sex and ancestry). Following the Ousley and Jantz (2012) flowchart, ancestral and sex groups were eliminated until the statistical probabilities (i.e., posterior and typicality) suggested a strong group classification. An additional evaluation for stature was conducted using the provided postcranial measurements.

The HCIFS Fordisc 3.1 data outputs for ancestry/sex and stature estimation were saved for each individual as a Word document. These documents were then compared to the outputs provided by Dr. Spradley. Similar to the MicroScribe validation discussed above, it was determined that an SOP was needed to provide adequate instruction and guidance on the use of Fordisc 3.1 for the HCIFS Forensic Anthropologists. Thus, an SOP addressing the statistical underpinnings and functions of the Fordisc 3.1 software was written.

Phase 3: Cognitive Bias Assessment: In the third phase of the project, the Fellow conducted internal studies designed to identify and mitigate potential sources of cognitive bias affecting the HCIFS Forensic Anthropologists. Throughout this project, the Fellow participated in the following activities to assess potential sources of bias:

1. Attended the morning meetings in which daily cases are discussed prior to examination by the pathologists. These meetings involve viewing scene photographs and listening to decedent history.
2. Observed autopsy consultations (i.e., conversations between the Forensic Pathologists and the Forensic Anthropologist at the autopsy table).
3. Attended scenes with the FAD staff to recover human remains.
4. Conducted supervised anthropological casework, which includes skeletal and cartilaginous analyses requested by the Forensic Pathologists.
5. Observed and conducted supervised casework regarding identification of unknown decedents (i.e., skeletal and dental radiography comparisons).
6. Attended scenes with the Forensic Investigations Division staff to observe the interactions between law enforcement and investigators.

A comparison study was then developed by the Fellow using five selected cases following a retrospective case review. Photographs of the original skeletal, cartilaginous, or casting media from the cases were anonymized to remove any case identifiers and then assigned to two staff Anthropologists for re-analysis. This aspect of the project builds upon the studies designed by Nakhaeizadeh et al. (2013, 2014), in the sense that some Anthropologists received contextual information for a case while some did not; however, this study differed in that no false information was deliberately given to participants. Given that some cases are accessioned into the medical examiner's office with copious amounts of background information while others have little or no information, this approach more closely resembles day-to-day operations at the HCIFS. By presenting varying amounts of information on the same cases to different Anthropologists, the Fellow could ascertain the degree of influence certain information had on an examiner.

The cases were chosen based on the following inclusion criteria: 1) the availability of photographic or radiographic documentation of the original evidence, 2) categories of analysis that have been deemed to be challenging or more prone to the influences of bias, and 3) if potentially biasing information was present in the case management system software. Four analytical categories were chosen for this review: sharp force trauma, blunt force trauma, estimation of the biological profile, and identification based on the comparison of antemortem and postmortem radiographs. Within each category, specific questions were posed to the examiners. For the two sharp force trauma cases, the Anthropologists were asked to determine the number of injuries and the blade class characteristics (i.e., serrated or non-serrated blade). The blunt force trauma case involved the Anthropologist assessing the direction of impact based on a fracture pattern. The radiographic identification was based on the comparison of dental radiographs, and the Anthropologist was asked whether there was enough information to make an identification, and if so, were the radiographs consistent. Finally, the biological profile case asked the Anthropologist to estimate the sex of the individual from photographs of the skull and pelvis.

Phase 3 concluded with Dr. Itiel Dror, a cognitive neuroscience, visiting the HCIFS and providing a two-day workshop on cognitive bias in forensic decision making. The goals for Dr. Dror's consultations were to provide the Anthropologists, Pathologists, and Investigators with fundamental knowledge about human cognition and bias, and to foster discussion about ways to reduce the influence of bias on forensic casework. In addition to the group workshops, Dr. Dror held independent sessions with the Forensic Anthropologists and the Forensic Pathologists to discuss the specific issues and challenges faced by these different groups of practitioners with respect to acquiring case information.

Phase 4: Information Consolidation, Manuscript Development, and Implementation: The final phase of this project involved the consolidation of information and submission of oral presentation abstracts to the American Academy of Forensic Sciences Annual Scientific Meeting. Following the 2019 scientific meeting, a manuscript will be finalized for submission to the *Journal of Forensic Sciences*. The cognitive bias study data was evaluated, and general recommendations for mitigating bias in the FAD were developed. Proposed revisions for operational procedures will be implemented, once approved by management. This work will conclusively help maintain the FAD's impartiality in casework, as well as offer recommendations to non-HCIFS practitioners.

Data Analysis

Quantitative data for the MicroScribe digitizer validation were analyzed as follows: caliper and digitizer data were evaluated by calculating Technical Error of Measurement (TEM). TEM was calculated for each ILD (inter-observer error), and overall TEM was calculated for both the caliper and digitizer data sets (inter- and intra-observer error). A paired sample t-test with equal variance was run to compare the overall TEM for the caliper and digitizer data sets. The digitizer and caliper overall TEM values (i.e., inter-observer error) per ILD were compared to determine whether there was a difference of more than +/-2 mm.

To assess whether data were correctly transferred between the MicroScribe, 3Skull, and Advantage Data Architect, two tests were performed: 1) 3Skull calculates ILDs and will display an error message if an ILD is too large or too small; if this occurred, the Anthropologist recaptured the two landmarks using the digitizer or manually checked the distance with calipers to assure that the ILD was correct; and 2) the ILDs in Advantage Data Architect were compared to the caliper measurements taken by each Anthropologist; while the measurements captured by the digitizer and the calipers may be slightly different (± 1–2 mm), as long as the measurements were similar, it was considered an accurate transfer of these data from the MicroScribe, through 3Skull, and into the Advantage software.

The Fordisc 3.1 validation study necessitated a qualitative analysis. The Fordisc 3.1 data outputs generated at HCIFS were compared to the outputs provided by Dr. Spradley at Texas State University. (Dr. Spradley and the FAD runs the cranial data file version 1.21 and postcranial data file version 1.17 of the Fordisc 3.1software.) The software was deemed valid if the FAD's sex, ancestry, and stature outputs were reasonably similar to those generated by Dr. Spradley. For example, if Dr. Spradley's Fordisc output stated that the individual was a Hispanic female between 60 and 62 inches tall, and the FAD's output stated the individual was a Hispanic female between 61 and 64 inches tall, then the Fordisc 3.1 software was considered to be functioning correctly.

The cognitive bias component of this project was also analyzed qualitatively. The Fellow was able to identify four types of anthropological cases/analyses as having processes that are more susceptible to cognitive bias: 1) assessment of blade class characteristics (i.e., serrated or non-serrated blade) in sharp force trauma analyses; 2) evaluation of the direction of force/impact in motor vehicle vs. pedestrian injuries; 3) development of a biological profile, specifically the estimation of sex, when contextual scene information is available (i.e., clothing on the decedent); and 4) identifications based on comparative medical or dental radiography. For the retrospective case study, the conclusions drawn by each Anthropologist after re-analysis were compared to the original case conclusions. Any discrepancies between the original and the re-analyzed conclusions were further evaluated by the Fellow to determine whether the potentially biasing information provided during the study affected the Anthropologist's decision-making.

Project Findings

Once the Anthropologists at HCIFS recorded 27 ILDs using the calipers, they were able to correctly collect the 42 cranial landmarks using the MicroScribe digitizer. All Anthropologists were able to use the 3Skull software and open the resulting data in Advantage Data Architect without any errors. When data from the Anthropologists were evaluated using TEM, the standard calipers had an overall error of 0.72 mm, and the digitizer had an overall error of 0.69 mm. A paired sample t-test with equal variance indicated there were no statistically significant differences between the measurements generated by the calipers and the digitizer.

The TEM values vary by interlandmark distance. The majority of measurements (70%) had less error when using the digitizer versus the calipers. Three ILDs showed no difference between the digitizer and calipers, and five ILDs showed more error when using the digitizer. Only one ILD (MAL or palate length) had an error higher than 2 mm (using the digitizer). Three ILDs had errors higher than 1 mm when using the calipers and/or the digitizer. Ultimately, there were no ILDs with a difference between the calipers and digitizer exceeding +/- 2 mm. These data indicate that the digitizer can adequately collect craniometric measurements with comparable, or often less, error than the standard calipers. Being a sensitive piece of equipment, the MicroScribe is more precise compared to standard calipers. In order to effectively capture cranial landmarks and reduce measurement errors when digitizing, thorough training to the SOP, practice, and proficiency testing are required.

This validation study also evaluated the functionality of the Fordisc 3.1 software program in the laboratory. All ancestry/sex and stature conclusions were the same between Dr. Spradley and the FAD, therefore meeting the acceptance criteria for this validation study. There were minor value differences between the outputs of the two institutions; however, this was the result of user manipulation of these data and incorporation of different advanced software features, not a reflection of the software's functionality. To conclude, the Fordisc 3.1 software in use at HCIFS is valid and should be incorporated into forensic casework as appropriate.

After one year of observations, shadowing, consultations, and hands-on analyses, the following sources of bias at HCIFS have been identified by the Fellow:

1. Morning meeting: contextual, and, in some cases, biasing information is presented. For example, when a decedent has evident sharp force trauma, it may be stated that the weapon used likely had a serrated blade. While the classification of the blade (serrated or non-serrated, beveled or non-beveled, etc.) is relevant information, presenting it prior to an anthropological analysis of skeletal or cartilaginous tissue can result in a biased interpretation of the injury and tool.
2. Office culture: during an autopsy consultation or when discussing a case with FAD staff, the pathologist or investigator may inadvertently provide too much background information about the decedent or case. Additionally, the general use of personal pronouns for a decedent prior to the Anthropologist's assessment of the remains can bias the assessment of the decedent's sex.
3. Environmental exposure: FAD staff may be exposed to contextual information about a case prior to the decedent's arrival at HCIFS by listening to, reading, or watching local news coverage.
4. Radiographic identifications: simultaneous comparison of antemortem and postmortem radiography can result in a biased interpretation, or potentially an erroneous identification, particularly if the Anthropologist has knowledge of the decedent's tentative identification and circumstances of death. The current peer review process, especially for radiographic identifications, is more subject to confirmation bias than other types of anthropological analyses.
5. Extraneous external information: the availability of contextual information about the case and/or decedent through the agency's case information database can result in bias, if it is reviewed prior to an anthropological analysis. Additionally, phone calls from family or friends requesting the release of the decedent, as well as other external pressures to expedite case completion, have the potential to introduce bias.

The results of the case comparison study were relatively consistent. In each of the five cases, the two Forensic Anthropologists reached the same (or similar) conclusions as the original analysists without any contextual information. However, two cases provided results that require further consideration. First is the radiographic identification case. Based on the original report, it is possible that the original analyst knew some contextual information about the case. For example the original analyst stated that the dentition of the unknown decedent was "consistent with an adolescent," despite the fact that an assessment of age was not the requested analysis. At the time of the original analysis, the following information was available in the case management software: "15 [year] old missing since 5/27/16 found by law enforcement on 5/30/16." The original analyst concluded that the antemortem and postmortem dental radiographs were consistent, and the individual was positively identified. The analyst who re-examined this case was more cautious and was not definitive as to whether there were enough consistencies to make an identification. This suggests two possibilities: 1) the original analyst was biased by contextual information about the case and therefore felt more confident about making the identification, or 2) the second analyst was being overly cautious because they were being observed and "tested." Given that the research for this grant has identified radiographic identifications as being potentially susceptible to cognitive bias, this is something to continue to monitor.

The second case for further consideration is one of the sharp force trauma cases. The original analyst concluded that the sharp force defects were the result of a chopping motion with a non-serrated blade. The analyst who re-examined the case concluded that the blade was serrated or partially serrated. At the time of the original analysis, the following information was available in the case management software: "There is a machete lying on a towel on the floor of the living room…there is blood noted on the blade of the machete," and "Machete used as the weapon." A machete is a non-serrated blade, and is typically used in a chopping motion. As in the previous scenario, it is possible that the original analyst knew this contextual case information prior to the scientific analysis. However, to improve this small experiment, the original analyst was sent the original evidence photos from this case two years later, and with zero contextual information, and was asked to assess the blade class characteristics. Again, the original analyst concluded that the evidence suggested a non-serrated blade. These results indicate that the original analyst was likely not biased by the contextual background information about the case at the time of analysis. This also suggests that classifying blades is challenging and there are no definitive characteristics that are always accurate. The research for this grant has also identified the interpretation of blade class characteristics as an area that requires monitoring and has the potential to be influenced by extraneous contextual information.

Implications for Criminal Justice Policy

This project impacts the forensic and criminal justice communities, particularly Forensic Anthropology, by providing best practices to mitigate potential issues regarding method error and bias. Across the country, few medical examiner offices employ full-time Forensic Anthropologists or have a dedicated forensic anthropology laboratory. Due to the paucity of anthropology laboratories based within a medical examiner system, the sources of error and biases that can affect case analyses have not been examined at length. By conducting this study with Forensic Anthropologists in a medical examiner's office, we have identified areas susceptible to error and bias within this relatively unique setting, and can provide suggestions for reducing or preventing any adverse effects on casework.

In the past, certain forensic disciplines, including forensic anthropology, were less scrutinized than those harder sciences within the crime laboratories (e.g., DNA analysis and toxicology) regarding quality assurance; however, in light of high profile legal cases and congressional pressure, it is apparent that more standardized and quality-oriented practices must be implemented in all forensic disciplines. At both the laboratory and judicial levels, the implications of having an effective quality assurance program for Forensic Anthropology are profound – one that employs best practices to monitor and assess personnel, training, equipment, and the procedures derived from methods/theories. This oversight and documentation limits errors in the laboratory, establishes the surety of the final conclusions and case reports, and ultimately provides transparency within the criminal justice system.

Although these concepts may seem daunting to those practitioners in universities and offices with little institutional support, this project has led to the development of valuable tools that will be shared with the community. As a result of this project, the MicroScribe digitizer and 3Skull and Fordisc 3.1 software programs have been validated within the HCIFS's Forensic Anthropology Laboratory. Two new SOPs have been written to guide FAD staff, as well as other Forensic Anthropologists, in the collection of craniometric data, the use of the MicroScribe digitizer, and the 3Skull and Fordisc 3.1 software programs. Potential sources of cognitive bias have been identified with respect to Forensic Anthropologists in a medical examiner's setting in four specific analytical circumstances. The case comparison study, although small, suggests that contextually irrelevant information has the potential to bias anthropological analyses. Conversely, this study also suggests that concordant conclusions can be made by Forensic Anthropologists based on the same scientific evidence, without the presence of contextual information. Therefore, the community should be advised to avoid (when feasible) as much contextual information as possible prior to analyzing the scientific evidence.

Sharing the results of this project, through national presentations, peer-reviewed publications, and the Organization of Scientific Area Committees (OSAC), will directly impact the forensic sciences and criminal justice system by demonstrating the importance of validating methods and equipment performance and assessing cognitive bias. Additionally, this project re-emphasizes the profound effect of mentoring within Forensic Anthropology. By hiring an emerging Forensic Anthropologist and mentoring her within the context of a medical examiner setting, the Postdoctoral Fellow has become a well-rounded Forensic Anthropologist who can implement research projects in addition to conducting anthropological analyses. At the time this project was proposed, it was stated that upon completion of the FAD mentorship program "the Fellow will be well prepared to seek employment in any office and work comfortably within a team or as a sole practitioner." This project and mentorship program was so successful that the HCIFS hired the Fellow as a new staff Forensic Anthropologist.

References

Nakhaeizadeh S, Dror IE, Morgan RM. 2013. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. Science and Justice.

Nakhaeizadeh S, Hanson I, Dozzi N. 2014. The power of contextual effects in forensic anthropology: A study of biasability in the visual interpretations of trauma analysis on skeletal remains. Journal of Forensic Sciences 59: 1177–1183.

Ousley SD, and RL Jantz. 2012. Fordisc 3 and Statistical Methods for Estimating Sex and Ancestry. In: Dirkmaat DC, editor, *A Companion to Forensic Anthropology*. London: Blackwell, p. 311-329.