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

Document Title:	Forensic Testimony Archaeology: Analysis of Exoneration Cases and its Implications for Forensic Science Testimony and Communications
Author(s):	John Morgan
Document Number:	306259
Date Received:	March 2023
Award Number:	GS10F0114L/OJP2002BF

Forensic Testimony Archaeology: Analysis of Exoneration Cases and its Implications for Forensic Science Testimony and Communications

Final Report

John Morgan, Research Consultant, National Institute of Justice

February 20, 2023

Abstract

Detailed analysis of errors associated with forensic evidence can support the development of targeted and system-based reforms by forensic practitioners and criminal justice system actors and mitigate the risk of future wrongful convictions. The results are reported of a study to examine case factors associated with 732 wrongful convictions classified by the National Registry of Exonerations as being associated with “False or Misleading Forensic Evidence.” A forensic error typology has been developed to provide a structure for the categorization and coding of factors relating to misstatements in forensic science reports; errors of individualization or classification; testimony errors; issues relating to trials and officers of the court; and evidence handling and reporting issues. The typology is designed to distinguish between errors committed by forensic practitioners as part of the examination process and system errors associated with the communication and use of forensic findings. The case review framework and coding typology developed for this study may also be a useful tool to inform root cause analyses or sentinel event reviews that support the comprehensive analysis of wrongful conviction cases involving forensic evidence. The study relied on publicly-available primary and secondary source documentation related to the forensic evidence, an approach designed to support future studies by other researchers.

The study demonstrates that most errors related to forensic evidence are not identification or classification errors by forensic scientists. When such errors are made, they are frequently associated with incompetent or fraudulent examiners, disciplines with an inadequate scientific foundation, or organizational deficiencies in training, management, or resources. More often, forensic reports or testimony miscommunicate results, do not conform to established standards, or fail to provide appropriate limiting information. Just as importantly, actors within the broader criminal justice system—but not under the purview of any forensic science organization—may contribute to errors that are deemed related to forensic science. System issues include reliance on presumptive tests without confirmation by a forensic laboratory, use of independent experts outside the administrative control of public laboratories, inadequate defense, and suppression or misrepresentation of forensic evidence by investigators or prosecutors.

In approximately half of wrongful convictions analyzed, improved technology, testimony standards, or practice standards may have prevented a wrongful conviction at the time of trial. Nonetheless, some forensic analysts produced reports and testimony that did not conform to standards at the time of trial or today. At the same time, unproven methods have been brought into court without scientific validation.

Table of Contents

Overview of the study

A comprehensive examination of the nature and extent of wrongful convictions that may be associated with forensic science errors is presented. Relevant issues include the extent of errors related to forensic testing and methodology; the use of unreliable or unproven methods; the communication of probative value; and the interface of forensic science with investigative and legal practitioners. Public, open-source data has been collected for each case in the study set. The complete dataset has been analyzed to produce a descriptive understanding of wrongful convictions associated with forensic science. Data subsets have been used to evaluate claims regarding forensic errors and elucidate issues in specific disciplines. The analysis is designed to address study objectives:

1. Estimate the number and types of cases in which the misapplication of forensic science has contributed to detected wrongful convictions.
2. Identify whether the cases were categorized as having forensic information that was (1) caused by errors in forensic testing (and identify forensic methods used), (2) based on unreliable or unproven methods, (3) expressed with exaggerated and misleading confidence, and/or (4) fraudulent.
3. Identify if and how the forensic information was used in the balance of evidence to convict.
4. Analyze the sources of errors, including issues relating to specific disciplines, policies, investigation, and prosecution.
5. Identify cases and circumstances where previously-applied forensic techniques and standards were replaced by approaches based on improved technology or scientific understanding.
6. Identify lessons-learned that will help current forensic science and legal practitioners develop systems-based approaches to mitigate factors that could contribute to a wrongful conviction.

The study addresses claims concerning the prevalence and contributing factors associated with forensic-science-associated errors with respect to possible sources in all phases of the collection, analysis, and use of forensic evidence by the criminal justice system.

General statement of results and overarching themes

The development, promulgation, and enforcement of science-based standards can mitigate the risk of case errors associated with forensic science. Many wrongful convictions are associated with examiners whose examinations or testimony do not conform to science-based standards, either because those standards did not exist at the time of trial or were not adopted by their organization, or the examiner was independent of a forensic science organization.

Wrongful convictions provide a basis for consideration of some proposed changes in forensic science policy and practice. Improvements in science, technology, and standards appear to have contributed to a reduction in the contribution of forensic science to wrongful convictions. New methods permit forensic analysts to produce conclusions with greater probative weight. For example, while serology and hair comparison were associated with wrongful convictions prior to 2000, DNA has largely replaced these methods. The application of mass spectrometry, improved microscopy, and forensic databases have also improved the ability of forensic science to identify contributors to evidence and mitigate the risk of wrongful convictions. The adoption of standards with improved scientific foundations has played a similar role. In the pre-2000 period, improvements in fire debris investigation and firearms identification addressed practice deficiencies raised by wrongful conviction cases.

On the other hand, examiners may not use validated methods even after the adoption of consensus standards or may use unproven innovations that introduce unreliable forensic results into criminal trials. The continued development and promulgation of new methods and improved scientific standards should be an important priority to reduce the risk of future forensic errors, but wrongful convictions demonstrate that the establishment of standards is not sufficient. Reforms must include training of both new and existing analysts and the enforcement of reporting and testimony standards at the laboratory and jurisdictional levels.

Forensic science organizations require mechanisms to identify and describe errors, find root causes, and develop mitigation responses in line with the precepts of high-reliability organizations. (Weick & Sutcliffe, 2001) The development of governance structures should take into account the fact that many examiners do not practice within public or accredited forensic science organizations. Many wrongful convictions are associated with examiners who are untrained, do not follow best practices, or practice disciplines that are weakly governed. There is minimal evidence within wrongful convictions that educational level by itself may mitigate the risk of forensic errors. In fact, most forensic medical professionals, bitemark examiners, and forensic pathologists that are associated with forensic examinations in wrongful convictions have medical doctorates or the equivalent.

Some disciplines with inherently subjective interpretation frameworks and direct exposure to task-irrelevant context are vulnerable to forensic examiner variability, cognitive bias, and associated forensic errors. These disciplines include fire debris investigation, blood spatter analysis, forensic medicine, and forensic pathology. There is minimal evidence within wrongful convictions that laboratory-based disciplines that make categorical conclusions—including toxicology and pattern-evidence examiners—produce association errors because of contextual effects. Other criminal justice system practitioners may exhibit cognitive bias and discount forensic results of any type, even to the point of ignoring clearly exculpatory results. Clear and open communication is necessary to prevent forensic errors that may result in wrongful convictions. The design of contextual bias interventions should recognize the need for reliable communication among police, prosecutors, and forensic science organizations. Forensic science reports, testimony, and other communications should reflect a complete understanding of relevant case details, the probative strength of evidence, and technical limitations.

Many wrongful convictions include valid and reliable forensic analysis that was misused by officers of the court in the adjudication of a case. Defendants face a substantial adversarial deficit related to the use of forensic evidence. In many cases, a prosecution is based on forensic results and interpretations, such as a conclusion that a fire had an incendiary origin or seized evidence included a controlled substance. A wrongful conviction may occur when the limitations of forensic conclusions are poorly understood or alternative interpretations are not presented. Officers of the court may lack sufficient understanding or resources to present appropriate objections or independent forensic reviews of evidence. The increasing technical complexity of forensic evidence may exacerbate this issue.

Unlike prior research, this study seeks to establish a framework that can be used outside the wrongful conviction context to analyze errors associated with forensic evidence, establish root causes, and design relevant improvements. A forensic error typology has been developed to provide a structure for the categorization and coding of errors relating to misstatements in forensic science reports; errors of individualization or classification; testimony errors; issues relating to trials and officers of the court; and evidence handling and reporting issues. The typology is designed to distinguish among errors committed by forensic practitioners as part of the examination process, errors associated with the communication of forensic conclusions, and system errors by criminal justice practitioners who use forensic science information. Associated contributing or correlative factors have also been examined, such as the education level of examiners and their organizational context.

Overview of the need for the study

The National Registry of Exonerations (NRE) has recorded over 3,000 cases of wrongful convictions in the United States. (University of California Irvine Newkirk Center for Science & Society, University of Michigan Law School, and Michigan State University College of Law, 2020) As of July 2021, the NRE associated 732 wrongful convictions with “False or Misleading Forensic Evidence.” Despite extensive research and scholarship related to wrongful convictions and forensic science, there is limited data concerning the etiology of forensic science errors in this context. The present study seeks to address this issue through an analysis of the forensic science testimony and case information in known wrongful conviction cases.

The misapplication of forensic science has been associated with a significant fraction of wrongful convictions in the United States, according to the Innocence Project and its analysis of data from the NRE. (Innocence Project, 2020) (The National Registry of Exonerations, 2020) Wrongful conviction researchers are limited by the lack of empirical data on which to base any solid conclusions (Cole S. , 2011) and have come to rely on out-of-date perceptions of forensic practice, while overlooking the increased sensitivity and specificity of newer technologies and methods currently implemented by forensic laboratories. For example, bullet lead comparison and serology have been associated with wrongful convictions, but bullet lead comparison was abandoned 15 years ago, and blood-typing serology was replaced by advancements in DNA technology in the 1990s.

In general, forensic science practitioners and their critics among exoneration advocates have had difficulties in finding common ground for policy and practice reforms. The forensic science community has been limited in its ability to communicate the realities and value of current forensic science practices. The criminal justice system has relied on outdated or incomplete information on wrongful convictions and may have missed opportunities to identify and adopt reforms that are responsive to past problems.

Some attempts have been made to address the gap in research. The NRE, housed at the University of Michigan Law School, publishes an annual review of new exonerations in its database, and these reports indicate that forensic science is playing a limited role in wrongful convictions in recent years. The NRE defines exonerations based on false or misleading forensic evidence based at least in part on forensic information that was (1) caused by errors in forensic testing, (2) based on unreliable or unproven methods, (3) expressed with exaggerated and misleading confidence, or (4) fraudulent. (National Registry of Exonerations, 2021) The NRE claim that 24% of registry cases have involved false or misleading forensic evidence has not been validated or detailed in a manner of utility to the forensic community. (Garrett B. , Wrongful Convictions, 2020) There have been extensive examinations of exonerations related to a subset of cases with faulty forensic evidence (LaPorte, 2017) and reforms of forensic science (Norris, Bonventre, Redlich, Acker, & Lowe, 2017). Nonetheless, the field lacks a comprehensive examination of exoneration data and the implications for current forensic science policy and practice. Without a comprehensive analysis, generalized claims regarding the role of forensic evidence in wrongful conviction cases can confound or conflate the underlying issues, precluding focused improvements for forensic science related policies and practices.

There is an extensive literature of wrongful conviction descriptive studies, analysis, and proposed reforms. (Garrett B. L., 2020) Most of the extant literature has been produced by legal scholars, who rely on isolated case narratives to illustrate points of scholarship. (Leo & Gould, 2009) Social scientists have conducted useful studies of wrongful convictions, although forensic science issues have seldom been the focus of systematic research. The University of Michigan has done substantial work to examine capital cases as an empirical framework for wrongful conviction research (Gross & O'Brien, 2008) and established a database of exonerations—the NRE—that has been useful in elucidating the incidence of major factors in wrongful convictions. (Gross & Shaffer, 2012) Other studies have also examined the capital case framework as a method to study wrongful convictions, demonstrating a variety of causative factors, such as inadequate defense. (Liebman, Fagan, West, & Lloyd, 2000)

Although the NRE contains over 700 records tagged as “False/Misleading Forensic Evidence,” it does not elucidate the basis for this categorization in individual cases or provide insight that links forensic errors to contributing factors or policy issues. In their overarching analysis of the data, Gross and Shaffer provide summary information about forensic errors:

> “The problems with forensic evidence range from simple mistakes to invalid techniques to outright fraud. We see clear examples of all of these, although in some cases it’s impossible to distinguish one type of forensic error from another.” (Gross & Shaffer, 2012)

Gross and Shaffer also found many “no crime” cases in which an individual was convicted of a crime that did not occur. Although many defendants were framed deliberately by their accusers, others were implicated by forensic science errors, especially in cases of arson and “shaken baby syndrome.”

In building the NRE, Gross and Shaffer were hampered by the lack of data in cases that did not involve an innocence organization. Many exonerations were established by “defense lawyers, prosecutors and police officers whose main jobs are arresting, prosecuting and defending the guilty. They often do nothing to call attention to their innocence cases and sometimes actively work to keep them from public view." (Gross & Shaffer, 2012) Often, there is limited documentation to establish the prevalence and nature of errors associated with forensic evidence in these cases, which is also a limitation of the current study.

The interpretation of wrongful convictions can be especially difficult. The literature contains many instances in which study authors express their understandable indignation at the injustice that has been perpetrated on innocent suspects. This tendency can lead to "moral panic" in wrongful conviction analysis, a phenomenon that has also been associated with the work of forensic examiners in wrongful convictions. (Grometstein, 2008) Wrongful conviction analysts may become advocates and may accept the claims of other advocates without doing the required critical analysis. Just as all forensic scientists are subject to confirmation bias, all wrongful conviction researchers are subject to the same limitations. (Nickerson, 1998) Even more so, wrongful conviction researchers are subject to hindsight bias, which is a common problem in retrospective research studies. (Busby, 1999)

Case history analysis in the wrongful conviction literature is particularly vulnerable to two types of error. First, the analyses often emphasize the presentation of the forensic evidence, not the forensic analysis itself. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) Testimony and report errors are a concern, but they can be confused with identification errors, system errors, and other issues. As an extreme example, the FBI review of its pre-2000 hair microscopy reports found that over 90% were in error, although almost all of the errors were based on the review's interpretation of the language used to characterize the strength of the evidence, not source attribution errors. (ABS Group, 2018) Nonetheless, many observers have exaggerated the review's findings to associate hair comparison with "junk science." (Servick, 2016) Latent print examination produces far fewer false positives than hair comparison (Busey, Heise, Hicklin, Ulery, & Buscaglia, 2021) but is vulnerable to the same critique. Even today, there is very little basis to report the statistical strength of a latent print identification. (Neumann & Saunders, 2019) As a result, any latent print testimony could be misleadingly classified as an error, because the examiner makes reports with a limited empirical or statistical basis. In the past, many latent print examiners have testified under a zero-error-rate formulation: "If the scientific method is followed, the error in the analysis and comparative process will be zero." (Cole S. A., The Myth of Fingerprints: A forensic science stands trial, 2000) This formulation has no empirical or statistical basis. Therefore, almost all latent print testimony before the mid-2000's could be characterized as a forensic error on the basis of the zero-error-rate claim. This approach may have merit from the perspective of wrongful conviction advocates, but it does not provide much insight into the problem of wrongful convictions from a research or policy perspective. To address this difficulty, reviews must clarify the methodology used to reach judgments about errors so that the classifications are reproducible by independent researchers. This requirement mirrors the scientific literature, particularly the practice of systematic research reviews. (See, for example, https://www.cochranelibrary.com/about/about-cochrane-reviews.) Researchers should also consider the standards of forensic practice at the time of the trial and judge forensic science testimony against generally accepted standards, not their own subjective views.

Secondly, wrongful conviction researchers may discount information that contradicts an exoneration. One may assume every criminal conviction is associated with some type of incriminating evidence or circumstances. It is also possible that a forensic result is properly performed and reported, while still providing a result that inculpates an innocent suspect. For example, serology may be used to associate many individuals with type A blood antigens, even if a large proportion of the population could be implicated by the result. In part, this issue has arisen because there is a perception that forensic evidence must provide definitive identifications of suspects. Historically, the forensic science community established this framework and may have set expectations that could never be fulfilled. (Bell, et al., 2018) Nuclear DNA analysis has the best foundation to provide absolute identifications, but even DNA does not provide that level of confidence in every case. Even a definitive result may be misinterpreted because an individual's DNA may be present for reasons not related to the crime. There should never be an expectation that forensic results provide some kind of epistemological knowledge. (Cole S. , 2013) Any scientific method produces results with random and systematic errors, so any forensic science method will also be associated with random and systematic errors in addition to inevitable human errors in the method as applied.

More broadly, no conviction or exoneration is based on such an expectation. In this way, science and justice share some philosophical underpinning. Just as a criminal trial in the United States is based on guilt beyond a reasonable doubt, scientific knowledge is limited by uncertainties and observational limitations. These limits extend to the fundamental structure of physics and applied science. The limits of science can be observed in broken bridges and medical errors. The limits of the criminal justice system can be seen in wrongful convictions. Rigorous evaluation of wrongful convictions should reflect a methodology concordant with accident reconstruction or sentinel event analysis. (Hollway, 2021) Just as a pilot may make many good and bad decisions that relate to a plane crash, criminal justice practitioners may make many good and bad judgments along the way to wrongful conviction and exoneration. The exoneration may be associated with a forensic science error, but it may also be associated with forensic results that are valid and contradict the exoneration.

To illustrate, one may consider an example from one case related to latent print examination. Cooley and Oberfeld wrote an influential paper in the Tulsa Law Review in 2007 titled, *Increasing Forensic Evidence's Reliability and Minimizing Wrongful Convictions: Applying Daubert Isn't the Only Problem*. (Cooley & Oberfield, 2007) They analyzed over 50 cases, including the case of Brian Keith Rose, who was on trial for the 2006 murder of Warren Fleming during a carjacking. Cooley and Oberfeld cited a newspaper source that indicated that "prosecutors tied Rose to the shooting 'through partial fingerprints lifted from the victim's Mercedes and a stolen Dodge Intrepid that they said the shooter used to drive away from the parking lot.'" The presiding judge, Baltimore County Circuit Judge Susan M. Souder, excluded the fingerprint evidence as a "subjective, untested, unverifiable identification procedure that purports to be infallible," again as quoted by Cooley and Oberfeld from a press account, although the original court decision was available. (State of Maryland v. Bryan Rose, 2007) They then state, "Following Judge Souder's decision, the State's case fell apart." Further, they quote the defense attorney's statement from the newspaper as follows:

> "Ultimately, the issue of fingerprint admissibility will not go away," said Patrick Kent, who represented Rose in state court and who heads the Maryland state public defender's forensic unit. "The fact that the state has fled to the federal court does not change the fact that fingerprints simply are not admissible, have never been validated and have no place in a courtroom, be it at state or federal level."

Cooley and Oberfeld did not report on any review of the trial transcript, latent print identification, or police investigation, leaving the impression that Rose was wrongly accused, and that latent print identification is unreliable. In fact, Rose had been identified initially because the carjacking and murder had occurred during a police pursuit. In addition to the latent print, Rose's DNA was found on the vehicle. During the subsequent federal trial proceedings, U.S. District Judge Catherine C. Blake admitted the fingerprint testimony and issued an opinion on the scientific reliability of the Analysis-Comparison-Evaluation-Verification (ACE-V) methodology for latent print examination:

> "Accordingly, for the reasons stated above, I have concluded that fingerprint identification evidence based on the ACE-V methodology is generally accepted in the relevant scientific community, has a very low incidence of erroneous misidentifications, and is sufficiently reliable to be admissible under Fed. R. Ev. 702 generally and specifically in this case." (United States of America v. Brian Keith Rose, 2009)

Rose was convicted after he pled guilty to the crime in federal court. (Department of Justice, 2010) The federal Rose conviction was sustained. The defendant was not exonerated. The Rose case demonstrates that the importance of complete documentation and objective analysis to elucidate contributing factors in wrongful conviction cases.

Descriptive analyses

There is an extensive literature that uses the descriptive approach to examine wrongful convictions. (Garrett B. , Wrongful Convictions, 2020) These efforts date back to Borchard's 1932 book, *Convicting the Innocent*, (Borchard, 1932) and the Franks' 1957 book, *Not Guilty*. (Frank & Frank, 1957) The modern era of wrongful conviction literature began with the National Institute of Justice report on DNA exonerations (Connors, Lundregan, Miller, & McEwen, 1996) and the popular press book, *Actual Innocence*, by Scheck et al. (Scheck, Neufeld, & Dwyer, Actual Innocence, 2000) Following his research work on wrongful convictions, Garrett contributed *Convicting the Innocent*, which updated many of the themes in *Actual Innocence*. (Garrett B. , Convicting the Innocent: Where Criminal Prosecutions Go Wrong, 2011) Wrongful convictions have been recognized as a problem in other countries, and the international literature is vast and growing. Most notably, Huff and Killias edited a volume, *Wrongful Conviction: International Perspectives on Miscarriages of Justice*, that included descriptive case studies, historical context, and research summaries from a wide range of countries. (Huff & Killias, 2008) The volume describes the experience of wrongful conviction in different judicial systems, including the adversarial context of the United States, inquisitorial systems, and even the Swiss "objective truth" system. At minimum, these studies indicate that wrongful convictions cannot be remedied through simple solutions, even at the system level.

While the wrongful conviction literature encompasses thousands of publications, the subset related to forensic science issues is of limited empirical value. Leo and Gould's critique of the case history approach of legal scholars is especially relevant to the examination of forensic science errors. (Leo & Gould, 2009) As they state, the literature relies on questionable media accounts and subjective interpretations that provide limited insight into causation. Descriptive case analyses have often ignored the complexities of the organizational and technical context. As a result, it is common for observers to make claims of forensic error and recommend policy and practice changes that have little or no empirical basis in their analysis.

There have been exaggerated claims concerning the prevalence of forensic science errors associated with wrongful convictions. Citing Innocence Project (IP) data, Saks and Koehler estimated that 63% of all wrongful convictions were associated with forensic errors in their 2005 paper in *Science*. (Saks & Koehler, The Coming Paradigm Shift in Forensic Identification, 2005) Even at the time, the 63% figure was misleading, because it was limited to DNA exonerations. Early DNA exonerations were closely associated with forensic serology in the original trial, meaning a possible forensic error could be claimed in almost every case. Further, Saks and Koehler's estimate relied on IP's subjective judgments of the errors in question, which have been demonstrated to be deeply flawed. (LaPorte, 2017)

Further, the wrongful conviction literature suffers from a "circular reasoning" problem because research flaws have been promulgated without adequate review and independent criticism. For example, one of the more sophisticated critiques of forensic science errors arose from Jane Campbell Moriarty's paper on "misconvictions" relating to forensic science. (Moriarty, 2007) Moriarty relies on descriptive case analyses by Saks, Cooley and others to conclude that "misconvictions occur because of bad science, poor laboratory work, and outright fraud." She states that forensic science errors are the second most common contributing factor to wrongful convictions based on the misleading Saks and Koehler paper, as discussed above. (Saks & Koehler, The Coming Paradigm Shift in Forensic Identification, 2005) In discussing the need for a "scientific culture" in forensic laboratories, she again references the Saks and Koehler paper, which maintained that 96% of forensic scientists lack college degrees. In fact, Saks and Koehler misstated the results of an analysis by Furton et al of survey data about job requirements for forensic positions. (Furton, Hsu, & Cole, 1999) It is true that 96% of forensic science positions in 1999 did not require college degrees, but the figure could not take into account those forensic scientists who *exceeded* the minimum requirements for their position. Further, the 1999 paper is now decades out of date, and the vast majority of forensic scientist positions now require a college degree. (National Institute of Justice, 2019) Moriarty devotes a large subsection of her paper to the need for additional accreditation and proficiency testing but does not refer to any scientific publications on the topic. Instead, Moriarty cites Koppl's 2005 paper advocating for the privatization of forensic laboratories. (Koppl, 2005) Koppl's only comment on accreditation related to the concern that it was overseen by the American Society of Crime Laboratory Directors (ASCLD), a claim which itself was incorrect and a common confusion among casual observers. The actual accrediting agency was the ASCLD Laboratory Accreditation Board, which was started by ASCLD but became independent after launch (and is now managed by the fully-independent ANSI National Accreditation Board, https://anab.ansi.org/en/forensic-accreditation). At the time of the Koppl paper, accreditation was also available from Forensic Quality Services, the American Board of Forensic Toxicology, and the National Association of Medical Examiners, among others. (Consortium of Forensic Science Organizations, 2013) Koppl and Moriarty were correct in one aspect: there was a widely recognized need to improve accreditation and quality management in the forensic sciences in the 2000's. Although these scholars have made valuable contributions to the field, their work has been limited by the relative lack of comprehensive research related to forensic science errors and wrongful convictions.

Forensic scientists differ from these observers in their analysis of the origin of errors. They place great emphasis on issues related to sample quality, contamination, and quality assurance. (Organization of Scientific Area Committees for Forensic Science, 2020) Several reviews have failed to appreciate this distinction. For example, the National Academy of Sciences 2009 report (Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, 2009) discussed validation studies, measurement errors, and cognitive bias at length, but it did not address contamination, sample quality or environmental influences. In its summary chapter, the report does mention that the "true value of [fingerprint] evidence is determined by the quality of the latent fingerprint image." The report discussed the pedagogy of the scientific method at length and emphasized the roles of measurement error and human bias. The challenges associated with real-world samples and the inherently retrospective nature of forensic examinations were not addressed. Elsewhere, the report mentioned the need to present “realistic” samples to examiners as part of proficiency testing, but the report did not attempt to define or describe what this means. Overall, the NAS report raised legitimate concerns about the state of forensic science but presented an idealized view of the way that forensic science could be performed. The committee made the case that forensic science should mirror the carefully controlled conditions of scientific inquiry. Certainly, scientific research related to forensic methods should follow such principles and forensic practice must adhere to methods that reflect the scientific consensus. That said, forensic science errors may be related more closely to the difficulties of applied science in a real-world setting than to statistical characterization within idealized frameworks. The report recommends the use of “validation studies” to account for those problems.

The wrongful conviction literature’s treatment of forensic science errors has focused on the identification disciplines, especially feature-comparison methods. The President’s Council on Applied Science and Technology (PCAST) expressed these viewpoints most thoroughly. (PCAST Working Group, 2016) The chapter, Scientific Criteria for Validity and Reliability of Forensic Feature-Comparison Methods, begins by making a distinction between objective and subjective methods:

> “By objective feature-comparison methods, we mean methods consisting of procedures that are each defined with enough standardized and quantifiable detail that they can be performed by either an automated system or human examiners exercising little or no judgment. By subjective methods, we mean methods including key procedures that involve significant human judgment—for example, about which features to select or how to determine whether the features are sufficiently similar to be called a proposed identification.” (PCAST Working Group, 2016)

In other words, PCAST was concerned with “human error, bias, and performance variability across examiners.” The report did not recognize that all forensic techniques involve subjective interpretation issues or that other approaches to fact-finding in courts involve subjective interpretation that may be less reliable than forensic science approaches. Like NAS, it sought to establish an idealized vision of forensic science that eliminates the considerations of crime scenes, contamination, and human judgment. Neither NAS nor PCAST linked their proposed reforms to specific correlates of forensic science errors and wrongful convictions.

Empirical studies

Prior studies have attempted to determine the number of wrongful convictions involving forensic evidence. Few studies have analyzed and described the specific types of errors that are associated with forensic evidence. Multiple attempts have been made to examine wrongful conviction cases in an empirical framework to establish causation. Notably, Gould et al compared erroneous convictions with “near misses” in which a defendant was cleared prior to trial. (Gould, Carrano, Leo, & Young, 2013) Their study found 10 contributing factors, including forensic evidence error, which included:

> “…neglecting to provide the jury with key information such as the victim’s blood type when it would mask the perpetrator’s; overstating the inculpatory nature of the evidence by providing inaccurate or non-existent statistics; and misstating the certainty of the

> results when the forensic technique, such as bite mark, scent, or fiber analysis, does not allow for it." (Gould, Carrano, Leo, & Young, 2013)

Most forensic errors were found at the testimony stage and were not necessarily errors in the testing itself. Hence, the report recommends, "As a result, previous policy recommendations that have focused on improving the quality of forensic laboratory procedures should be revisited to emphasize quality control at the interpretation and testimony stages." Although the study examined these forensic issues, it coded only for forensic discipline and whether an error was present, so it did not examine the nature and incidence of specific error types within the forensic science context.

Two notable studies have examined forensic errors in wrongful convictions more closely. Cooley and Oberfeld examined over 50 case studies in which "unreliable forensic evidence" contributed to wrongful convictions. (Cooley & Oberfield, 2007) Although the authors did not seek to provide systematic analysis, they did establish a baseline of claims concerning factors that cause forensic errors. The paper recommended:

1. Improved judicial oversight of unreliable forensic techniques
2. External and independent crime laboratory oversight
3. Accreditation under international standards
4. Professional certification

The paper did not attempt to provide a direct link between wrongful convictions and these recommendations.

Garrett and Neufeld (GN) produced the most comprehensive examination of forensic errors to date in their 2009 paper. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) They limited their data set to DNA exonerees with available trial transcripts, which at that time included 137 cases, 85 of which included "invalid forensic science testimony" or exculpatory forensic evidence that was withheld at the time of the original trial. Thus, 63% of cases in their study set included forensic errors, as expected given that they limited their analysis to DNA exonerations. Cases included a wide range of pattern evidence, physical evidence, and biological evidence. Notably, they state, "almost half of the valid forensic testimony was not inculpatory and likely did not significantly support the conviction." In other words, erroneous forensic conclusions or testimony were not part of the balance to convict in almost half of these DNA exoneration cases.

GN established a database of exonerations based on DNA testing (https://www.convictingtheinnocent.com/) and published a book on the topic. (Garrett B. , Convicting the Innocent: Where Criminal Prosecutions Go Wrong, 2011) The database relies on information from the Innocence Project and The Innocence Record, which provides more detailed trial transcripts and other information about the cases, including information about forensic testimony. (Garrett B. , Characteristics of Forensic Testimony at DNA Exonerees' Trials, 2011) The 2009 paper acknowledges the limitation facing any review of wrongful convictions, stating, "[O]ne cannot determine whether invalid forensic science testimony was common in the past two decades or today." It is impossible to review the forensic analysis in every past criminal case, and exoneration data is heavily skewed toward cases in which DNA or other evidence was available to justify an overturned conviction.

Garrett and Neufeld divided their study cases into six overarching categories:

1. Non-Probative Evidence Presented as Probative, 48 cases
2. Exculpatory Evidence Discounted, 23 cases
3. Inaccurate Frequency or Statistic Presented, 13 cases
4. Statistic Provided Without Empirical Support, 5 cases
5. Non-numerical Statements Provided Without Empirical Support, 19 cases
6. Conclusion that Evidence Originated from Defendant, 6 cases

The classic case of an identification error was present in only six cases. In most of their data set, wrongful convictions are associated with misinterpretation of the evidence or its miscommunication during testimony, not the erroneous association of evidence with a source. GN did not clarify the basis for their forensic error categorization in many cases or address the resolution of ambiguous issues, if any existed.

GN relied on the case analyses of its coauthors, neither of whom had education or experience in a scientific or forensic discipline. Garrett is a legal scholar (Garrett B. L., 2018) and Neufeld is a defense attorney. (Neufeld) Neufeld had been personally involved in many cases covered by the GN study.

Methodology

Demographic and case variables

The study dataset includes every wrongful conviction in the NRE associated with False or Misleading Evidence as of July 5, 2021, a total of 732 cases. The dataset includes cases from before 1989, the year of the first exoneration based on post-conviction DNA analysis. In recent years, DNA exonerations have become less common. The study set includes 183 DNA-exonerated defendants but only 38 DNA exonerations from 2010 to the present. The NRE cases do not include exonerations in other countries or some exonerations related to major forensic laboratory scandals, such as the Houston crime laboratory investigation (Bromwich M. R., 2007) and the Annie Dookhan and Sonja Farak drug chemistry misconduct cases in Boston. (McDonald, 2019)

All cases have been documented using publicly available sources. All conclusions concerning forensic science issues reference specific documentation in the available case record so that independent researchers have the opportunity to review the study's conclusions. The documentation and findings have been recorded in Case Review Forms for each case that detail relevant documentation and case coding. Coding includes demographic information, including ethnicity, sex, birthdate, date of alleged crime, date of wrongful conviction, and date of exoneration. Crime details include charges, jurisdiction, other evidence, and sentence. Some cases are linked by multiple co-defendants in the same trial or for the same crime. Cases may also be linked because a forensic examiner was involved in multiple wrongful conviction cases, such as an examiner who performed fraudulent or questionable work over many years. Finally, cases may be linked because of fraudulent or negligent work in a forensic science organization that led to multiple wrongful convictions. Exonerations are coded by date, exoneration advocates, DNA exoneration, basis for exoneration, and case outcomes (such as compensation and status of other convictions). The NRE coding for demographic and crime variables was used, except in rare cases in which court and case records disagreed with NRE data.

Each type of known forensic evidence in each case is coded separately. When multiple examiners were involved, each examiner's work was coded separately if possible. In some cases, the delineation of each individual's contribution was not possible, so a unified record was coded. In all, 1,391 forensic examinations were analyzed in the 732 cases. Each defendant is considered a separate case, even if multiple defendants were convicted in connection with the same alleged crime. The number of examinations in individual cases ranged from one to eight. The study uses the term “case error” to refer to any error associated with forensic evidence, whether it applies to an individual examination or overall case. The term “forensic error” is used to describe instances when a forensic examiner committed the error in question. The term “system error” is used to described instances when an error occurred outside the control of an individual forensic examiner. A case error may be a forensic error or system error. An examination or case may include multiple case errors.

For each examiner, the individual’s education level, certification status, and context were coded, if available. Context included whether the individual operated within or outside a forensic science organization or other professional affiliation (e.g., a physician in a hospital). The individual’s organization was also coded, including the nature and independence of their laboratory, level of government, affiliation with a coroner or medical examiner’s office, affiliation with a fire/arson investigation unit, and accreditation status. For many individuals and organizations, certification and accreditation status were not available. Finally, the forensic evidence was coded based on whether it was required in the balance of evidence to convict the defendant. A justification for this coding was recorded in each instance. In general, evidence was not considered “required” if it was non-probative, exculpatory, or weakly probative in a case that included other, strongly inculpatory evidence. For example, many serology cases were sexual assaults in which the majority of the population could have contributed the biological evidence and victim testimony identified the defendant. In those cases, the serological evidence was not required to convict the defendant. Errors could be associated with the forensic evidence, even if it was not considered required to convict.

Study limitations

Overall, detected wrongful convictions now comprise a large-enough sampling to support rigorous analysis of factors that may contribute to wrongful convictions. Nonetheless, the nature of wrongful conviction research prevents the determination of causation. Any findings should be viewed as useful insights that may inform policy and practice, not dispositive conclusions.

Wrongful conviction research is inherently limited by its retrospective analysis and lack of empirical controls. This study seeks to address that issue by considering all detected wrongful convictions that have been associated with false or misleading forensic evidence and using a predefined coding approach that was applied uniformly to all cases. The case documentation, individual coding, and justification for each case coding will be publicly archived for review and use by independent researchers.

In many instances, error coding could be ambiguous and may have been influenced by the limitations or biases of the study author. Ideally, each examination would be reviewed separately by a group of independent experts with relevant backgrounds in scientific research, forensic practice, investigation, and criminal jurisprudence. Resources were not available to support that approach, although experts did review the overall study methodology and a subset of cases. The current study may have benefitted from a unified and consistent approach by the study author. That said, future research may use different review frameworks to produce useful independent assessments.

The study relied solely on publicly available documentation. This limitation was necessary to ensure transparency and the possibility of independent review. In many cases, the available documentation was not sufficient to elucidate the issues in forensic evidence. Trial transcripts and forensic reports were often not available. Complete root-cause analysis would include an assessment of details of the case, forensic examinations, and organizational issues that are not known in most instances. In particular, forensic science organizations are well positioned to ascertain these details, relate contributory factors to root causes, and develop appropriate reforms.

Error typology

The main purpose of this research study is the identification and characterization of forensic-science-associated errors in wrongful conviction cases. All forensic evidence examinations were described and coded using a Forensic Error Typology developed for the study. The typology was supplemented by other coding related to past research reviews, an analysis of claims concerning the root causes of forensic errors, and a complementary coding system based on an approach presented as part of a 2015 conference on forensic science errors at the National Institute of Standards and Technology. (National Institute of Standards and Technology, 2015) A coding guide detailing the error typing system and basis for each type of error classification was developed prior to the coding of the cases.

The study also included coding of forensic examinations with respect to secondary claims regarding wrongful convictions. Coding elements included:

1. No error by this forensic examiner (in order to distinguish system errors from individual errors)
2. Changes related to new technology, improved scientific foundations, or improved standards. One coding element related specifically to cases in which new DNA technology would have provided a more probative result, including in cases using less probative, early DNA methods.
3. Use of a presumptive test to adjudicate a case without confirmatory testing.
4. Exculpatory evidence
5. Errors related to defense experts
6. Possible cognitive or confirmation bias
7. Honest mistake
8. Battle of experts
9. Inadequate documentation

Several models for error classification were considered in the development of the study's error typology. As described above, the GN approach provided a useful foundation for testimony review. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) Forensic scientists apply scientific measurement systems to understand category errors and uncertainties. (Joint Committee for Guides in Metrology, 2008) The NRE classification provides a general framework for non-forensic errors and was incorporated into the descriptive and demographic case analysis. (University of California Irvine Newkirk Center for Science & Society, University of Michigan Law School, and Michigan State University College of Law, 2020) Roberts has proposed a 20-item typology of forensic science errors. (Roberts, 2015) The Roberts approach includes many root causes that cannot easily be discerned in a retrospective analysis, such as institutional distortions and jury issues. Similarly, within the Canadian context, the Federal/Provincial/Territorial Heads of Prosecutions Committee established a Working Group on the Prevention of Miscarriages of Justice in response to wrongful convictions. (FPT Heads of Prosecutions Committee Working Group, 2011 (update)) Although the inquiry did not state a specific error typology, their report did discuss several categories associated with faulty expert testimony:

1. Prosecutorial bias or misleadingly presented evidence to support one theory alone,
2. Evidence presented with exaggerated probative value,
3. Poorly communicated evidence with excessive jargon and terminology,
4. Testimony on contaminated or tainted evidence, and
5. Testimony on evidence reliant on scientifically out-of-date methodologies or evidence reliant on subjective judgments

The NIST error typology provides the most useful delineation of errors related to forensic laboratory practice, including Analyst/Expert Error, Fraud, Methods/Protocol Error, and Instrumentation/Technology Limitations. The NIST approach includes 15 subtypes of Analyst/Expert Error and five subtypes each of Methods/protocol and Instrumentation/Technology Limitations. These 25 subtypes were included in the coding of each forensic examination in the dataset. The FBI's review of hair comparison provided three categories of testimony error. (ABS Group, 2018)

> Error Type 1. The examiner state or implied that the evidentiary hair could be associated with a specific individual to the exclusion of all others.
>
> Error Type 2. The examiner assigned to the positive association a statistical weight or probability or provided a likelihood that the questioned hair originated from a particular source, or an opinion as to the likelihood or rareness of the positive association that could lead the jury to believe that valid statistical weight can be assigned to a microscopic hair association.
>
> Error Type 3. The examiner cites the number of cases or hair analyses worked in the laboratory and the number of samples from different individuals that could not be distinguished from one another as a predictive value to bolster the conclusion that a hair belongs to a specific individual.

Error types 1 and 2 were generalized in this study's Error Typology framework, while type 3 errors were a subset of errors related to mischaracterizations of the probative value of evidence.

Based on analysis of the above frameworks, a five-part Forensic Error Typology was developed for the current study and applied to the analysis of each case and examination in the study dataset. The typology distinguishes among errors related to reporting of results, errors of association, testimony errors, errors by officers of the court, and system errors outside the control of an individual forensic scientist.

Error Type 1

A forensic science report contains a misstatement of the scientific basis of a forensic science examination. All type 1 errors relate to analyses as they were conducted and reported by a forensic examiner *prior* to a trial. Coding elements include:

a. Error originated in a report prepared by a forensic examiner prior to the trial.
b. Report contained a statement of individualization that was not supported by science at the time of trial.

c. Report contained a statement of statistical weight or probability that was not supported by science at the time of trial.
d. Report excluded relevant information that may have supported or refuted its conclusions.
e. Report failed to include relevant reference or background data.
f. Report failed to document its data and methods appropriately and completely.
g. Report excluded exculpatory results.
h. Relevant data was not collected because of resource constraints in the laboratory.
i. Report included statements that impeded the reader's ability to interpret the report.

Error Type 2

A forensic science examination contains an incorrect individualization or association of a piece of evidence with a source or class of sources or the incorrect interpretation of a forensic result that implies an incorrect individualization or association. An error may be intended or unintended. An individualization is a determination that two samples derive from the same source; practically, a determination that two samples derive from sources that cannot be distinguished within the sensitivity of the comparison process. The source may be a person, place, thing, or event. (Organization of Scientific Area Committees for Forensic Science) An error of individualization was coded when it had been determined that the two samples did not derive from the same source. A classification is a determination that two samples derive from the same population sources. The type of source may be a person, place, thing, or event. The population may be distinguished by shared traits or characteristics that may be referred to as class or subclass characteristics. (Organization of Scientific Area Committees for Forensic Science) An error of association was coded when it had been determined that the two samples did not derive from the same population of sources. Coding elements include:

a. Forensic examination contained an incorrect individualization of a piece of evidence with a source.
b. Forensic examination contained an incorrect classification of a piece of evidence with a population of sources.
c. An incorrect individualization or classification was the result of incorrect interpretation of a forensic result.
d. A forensic error resulted from a fraudulent or intended association of evidence with a source.

Error Type 3

Testimony at trial reported forensic science results in an erroneous manner. An error may be intended or unintended. Coding elements include:

a. An error originated in testimony provided during the trial that resulted in wrongful conviction of the defendant.
b. The testimony excluded relevant information that may have supported or refuted the conclusions presented.
c. The testimony included an incorrect individualization or classification.
d. The testimony contained a statement of individualization or classification that was not supported by science at the time of trial.
e. The testimony contained a statement of statistical weight or probability that was not supported by science at the time of trial.
f. The testimony mischaracterized the probative value of the forensic results under standards in place at the time of trial.

g. The testimony conformed to standards in place at the time of trial.
h. The testimony contained a statement of individualization or classification that was not supported by science in 2020.
i. The testimony contained a statement of statistical weight or probability that was not supported by science in 2020.
j. The testimony mischaracterized the probative value of the forensic results under standards in place in 2020.
k. The testimony conformed to standards in place in 2020.

Error Type 4

An officer of the court committed an error related to forensic evidence. Coding elements include:

a. Potentially exculpatory evidence was excluded from consideration during the trial by an officer of the court.
b. Defense associated with the presentation or review of forensic evidence or testimony was adequate.
c. Inadequate pretrial information discovery by defense.
d. Inadequate cross-examination concerning forensic evidence.
e. Independent examination or review not conducted.
f. Relevant objections to forensic testimony not raised.
g. Appeal did not raise relevant forensic issues.
h. Exculpatory forensic results not recognized/presented.
i. Other inadequate defense related to forensic issues.
j. Prosecutor characterized forensic evidence appropriately.
k. Prosecutor mischaracterized scientific validity.
l. Prosecutor mischaracterized statistical interpretation.
m. Prosecutor mischaracterized laboratory methods.
n. Prosecutor mischaracterized source attribution.
o. Prosecution consultant produced erroneous testimony.
p. Other prosecution error related to forensic evidence.
q. There was no error associated with judicial conduct (related to forensic evidence).
r. Novel forensic method not reviewed.
s. Faulty testimony accepted over objection.
t. Faulty jury instructions concerning forensic evidence.
u. Other judicial error related to forensic evidence.

Error Type 5

Potentially probative forensic evidence was not collected, examined, or reported during a police investigation or reported at trial.

a. Crime scene evidence was examined and reported appropriately.
b. Evidence was not collected.
c. Evidence not stored properly.
d. Crime scene analysis not conducted according to standards.
e. Evidence was collected but not sent to laboratory.
f. Evidence was lost or destroyed (pre-conviction).
g. Evidence was not available for postconviction testing.
h. Evidence was processed and reported appropriately in the laboratory.
i. Evidence was not processed in the laboratory.

j. Probative result was not documented.
k. Forensic result was not sent to investigators.
l. Forensic result was not sent to the prosecutor.
m. There was incomplete or unclear communication concerning the results.
n. Chain of custody was not maintained.
o. Investigators used and reported forensic evidence appropriately.
p. Investigators ignored forensic evidence.
q. Investigators suppressed forensic evidence.
r. Other police misconduct concerning forensic evidence.
s. Forensic evidence was reported at trial and shared appropriately with defense counsel.
t. Probative evidence was not reported to defense.
u. Misleading or erroneous information was given to defense about the forensic results.
v. "No source" results were not shared with defense when probative.
w. Other evidence suppression issue.

Inadequate defense was coded when defense counsel failed to ask for relevant information concerning forensic examinations prior to or during a trial; when defense counsel failed to obtain independent examination or review of forensic evidence, if the uncertainties in the analysis or interpretation of the evidence may have been relevant to probative value; if the defense failed to perform appropriate cross-examination of forensic testimony, raise relevant objections to forensic testimony; or a direct appeal did not address issues related to forensic testimony that could impact a case outcome. Judicial conduct error was not coded when forensic evidence was accepted in accordance with precedents related to the methods used in the case. Judicial error was coded when a novel method was employed without an appropriate review within Daubert or another standard that should have applied in the jurisdiction. Judicial error was coded when faulty forensic testimony was accepted (under the standards in place at the time of trial) over the objection of legal counsel.

Evidence collection and analysis decisions may have excluded probative evidence for legitimate reasons relating to resource constraints or imperfect knowledge of the crime at the time of the investigation. Error Type 5 was coded when probative evidence was excluded from consideration or analysis through negligence or malfeasance. Error Type 5 also covered issues related to the management of a forensic science organization, communication of forensic results, police conduct, and evidence suppression.

Testimony and practice standards

Forensic testimony and practice were assessed for conformance with standards at the time of trial and in 2020 at the time of the start of this study. The Department of Justice's Uniform Language for Testimony and Reporting (ULTR) was applied as the primary basis for 2020 testimony standards. (US Department of Justice, 2019) When appropriate, disciplines not specifically covered by the ULTR standards were assessed for conformance with ULTR principles, such as the limitation on citation of examiner experience as a basis for the statistical validity of a conclusion. In many cases, the primary issues involved practice standards and quality assurance, which may or may not have been reflected in the trial testimony. Thus, the selected standards were chosen to encompass practice issues when those issues could be discerned from the public case record.

In some instances, forensic evidence or communications were assessed to have an inadequate scientific basis under the NIST error framework. (National Institute of Standards and Technology, 2015) This assessment is inherently subjective, because the level of scientific foundation considered “adequate” may vary among observers. (Butler, et al., 2020) Thus, reliance on inadequate science was coded only when there was a basis in the consensus views of researchers and forensic standards bodies (for example, the use of canine detection to produce an individualization).

Table 1 provides a summary overview of the primary standards used for the most common forensic disciplines in the study.

Table 1. Forensic discipline standards used in the assessment of forensic evidence in wrongful convictions.

Discipline	**Standards**
Serology	Culliford, B. J., *The Examination and Typing of Bloodstains in the Crime Laboratory.* Washington, DC: National Institute of Law Enforcement and Criminal Justice. (Culliford, 1971) Gaensslen, R. E., *Sourcebook in forensic serology, immunology, and biochemistry.* Washington, DC: National Criminal Justice Reference Service. (Gaensslen, 1983) Federal Bureau of Investigation, *Handbook of Forensic Science.* Washington, DC: National Institute of Justice, US Department of Justice. (Federal Bureau of Investigation, 1982)
Hair comparison	The Laboratory Division, Federal Bureau of Investigation, Proceedings of the International Symposium on Forensic Hair Comparisons. Washington, DC. (The Laboratory Division, Federal Bureau of Investigation, 1986) Houck, M. M., & Budowle, B. Correlation of microscopic and mitochondrial DNA hair comparisons. (Houck & Budowle, 2002) Cases after the publication of the Houck/Budowle paper were coded for error if no attempt was made to confirm an association using DNA testing.
Fire debris investigation	National Fire Protection Association, Guide for Fire and Explosion Investigations (NFPA 921-92). (National Fire Protection Association, 1992) National Fire Protection Association, Guide for Fire and Explosion Investigations (NFPA 921-2017). (National Fire Protection Association, 2017) When appropriate, other versions of this standard were applied. For example, the 2008 version of the standard clearly prohibited the use of negative corpus to conclude an incendiary origin. For cases before 2008, the use of negative corpus would not have been in error at the time of trial if it

	were applied with a full consideration of alternative origins and causes. Prior to 1992, the 1992 version of NFPA 921 was applied as the standard, because a trained investigator would have used the research later cited in NFPA 921 as the basis for conclusions.
DNA	Scientific Working Group on DNA Analysis Methods (SWGDAM), including Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. (Scientific Working Group on DNA Analysis Methods, 2019) ANSI/ASB Standard for Forensic DNA Interpretation and Comparison Protocols. (AAFS Standards Board, 2019) National Academy of Science 1996 report, The Evaluation of Forensic DNA Evidence. (National Research Council, 1996) This report contains information concerning the accepted calculation of random match probabilities. It also includes a discussion of VNTR DNA analysis.
Bitemark comparison	American Board of Forensic Odontology (ABFO) Bitemark Methodology Guidelines. (American Board of Forensic Odontology, 1995) ABFO Standards and Guidelines for Evaluating Bitemarks. (American Board of Forensic Odontology, 2018)
Seized drugs	SWGDRUG guidelines (http://swgdrug.org/archived.htm). Organization of Scientific Area Committees (OSAC) Seized Drug Subcommittee (https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/seized-drugs-subcommittee)
Toxicology	AAFS Standards Board: the ANSI/ASB Best Practice Recommendation 037, First Edition, 2019, Guidelines for Opinions and Testimony in Forensic Toxicology. (AAFS Standards Board, 2019) OSAC Toxicology Subcommittee (https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/toxicology-subcommittee)
Firearms and toolmarks	OSAC Firearms and Toolmarks Subcommittee (https://www.nist.gov/topics/organization-scientific-area-

	committees-forensic-science/firearms-toolmarks-subcommittee) Relevant documents include the draft standard, Range of Source Conclusions and Criteria in Toolmark Examinations. (Organization of Scientific Area Committees, 2021)
Friction ridge	OSAC Friction Ridge Subcommittee, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/friction-ridge-subcommittee. The OSAC web page contains an exhaustive list of historical standards that are relevant to specific cases.
Forensic medicine (pediatric abuse)	American Academy of Pediatrics (AAP) standards, including: Consensus statement on abusive head trauma in infants and young children. (Choudhary, et al., 2018) Guidelines for the Evaluation of Sexual Abuse of Children: Subject Review. (Committee on Child Abuse and Neglect, 2013) (Committee on Child Abuse and Neglect, 1999) (Committee on Child Abuse and Neglect, 1991)
Forensic pathology	American Board of Medicolegal Death Investigators (www.abmdi.org). Death Investigation: A Guide for the Scene Investigator (National Medicolegal Review Panel, 1999) (National Medicolegal Review Panel, 2011)

The study recognizes that the issues in pediatric abuse cases continue to be debated in the legal and scientific communities. (Papetti, Kaneb, & Herf, 2019) As in other disciplines, the error coding reflects that case uncertainties may support multiple, valid interpretations. In those cases, errors may arise if a forensic expert fails to address the limitations associated with conclusions or if a valid interpretation consistent with the defense theory of the case is not presented during the case proceedings. This approach was also followed in the assessment of forensic pathology and fire investigation examinations.

In some cases, the primary issue involved the use of automated databases to conduct latent print searches that may have produced a cold hit to a source. In the determination of an error, the study relies on the date of deployment of a system that could have been used to conduct latent fingerprint searches in the jurisdiction. The study does not code as an error any failure to conduct a latent palmprint database search, which is not generally available at this time.

Wrongful convictions and serology

In their seminal 2009 paper, Garrett and Neufeld consulted with external experts in forensic science, including serology expert Edward Blake. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) Although Blake has been a significant contributor to forensic science research, he has had major disagreements with the practitioner community concerning the interpretation of forensic evidence errors. In 1981, his AAFS presentation asserted that FBI serological analyses were deficient, (Barnett, Blake, & Ogle Jr., 1981) an assertion that was directly countered by a substantive FBI response. (Critique: The Role of the Independent Expert: Several Case Examples, 1981) Blake was personally involved in many of the cases covered by the GN study, which maintained that the serological evidence in many cases included 100% of the population, but the analysts testified that a smaller percentage could have contributed to the biological sample. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) The Garrett-Neufeld interpretation is based on the application of the Blake approach to serological interpretation. As Blake outlines in his 2001 review report (Blake E. , January 9, 2001):

> "…only the detected genetic traits that are foreign to the female victim provide information about potential semen sources. … The only exception to this restriction in interpreting results from commingled body fluids such as vaginal fluid and semen occurs when there is a quantitative showing of the presence of semen at a concentration (ca 1:100 or greater) from which one would expect to be able to detect genetic traits from a secretor semen source. Absent such a showing, no genetic information in regard to the semen source can be inferred from data that is entirely compatible with the female."

Blake conducted research and development of methods for forensic serology from at least 1976. (Sensabaugh & Blake, 1976) Blake and his Ph.D. thesis advisor, George Sensabaugh, were influential in the development of methods to improve the forensic analysis of sexual assault samples. In particular, Sensabaugh and Blake examined the problem of the reliable determination that a sexual assault has occurred, even when sperm is not observed in a sample. (Sensabaugh G. A., 1977) As they state in their final report on the issue, "The finding of sperm unequivocally indicates the presence of semen. However, a significant portion of the male population has been vasectomized and have no sperm in their semen. Moreover, the identification of sperm in vaginal washings and in semen stains is often difficult. For this reason, many crime laboratories use the acid phosphatase test as a presumptive test for semen." Along with other researchers (Hara, 1969), Sensabaugh and Blake identified a semen specific protein, "protein-30," (P30) which was considered unique to seminal fluid and therefore a more definitive test than a finding of acid phosphatase (ACP). There were significant limitations to the use of P30, as it came to be known in forensic laboratories. The precise etiology of P30 was not yet established. Also, as Sensabaugh clarified, "The survival of the protein-30 marker post coitus and in stains needs further characterization. Finally, the sensitivity of the immunological test needs to be improved." (Sensabaugh G. A., 1977)

By the mid-1980's, these issues had been largely addressed. In 1985, Graves et al published a seminal paper in the New England Journal of Medicine on the use of P30 for the identification of semen in vaginal fluid. (Graves, Sensabaugh, & Blake, 1985) Using an ELISA assay, the researchers were able to detect as little as 3 ng/mL of p30 antigen in bodily fluids and detect the marker as long as 27 hours post-coitus. In March of that year, Sensabaugh et al described their recommendations for the analysis and interpretation of serological profiles in sexual assault cases. (Sensabaugh, Bashinski, & Blake, 1985) Their method recommended that two of the three swabs typically collected from a victim be expended to perform a multi-step process based on the quantitative ACP assay.

Sensabaugh makes several points that are salient to the assessment of serological analyses conducted at this time. First, he states, "Seminal fluid is usually detected by finding sperm in the supporting evidence. … Sperm density should then be assessed semiquantitatively from 0 for no visible sperm to 4+ for many in each field." This recommendation was standard practice in forensic laboratories throughout the pre-2000 period, although it was seldom detailed in testimony. Further, the paper recommends, "If no sperm are present, the laboratory may test for other seminal constituents such as ACP, choline, spermine, and semen protein p30." The paper provides a detailed basis for the use of ACP as a definitive indicator of the presence of semen if a threshold level is present. Finally, the paper recognized that "p30 testing is relatively new" and recommended both quantitative and qualitative test methods to detect P30. Finally, the paper states, "PGM and PepA activities decrease quite rapidly in the vagina after coitus; meaningful PGM results are rarely obtained from vaginal swabs collected more than six hours after intercourse and PepA does not appear to survive longer than three hours."

Recognizing the rapid development of serological testing methods and the uncertainty faced by forensic laboratories, the FBI convened a symposium on forensic immunology in 1986. (The Laboratory Division, Federal Bureau of Investigation, 1986) There were 175 attendees, including most of the leading researchers in the field and many of the examiners who were later associated with forensic examinations in wrongful conviction cases. The proceedings updated many of the concepts detailed in the landmark work of Gaennslen on forensic serology. (Gaensslen, 1983) Sensabaugh presented a paper on the future of forensic immunology that detailed the promise and shortfalls of quantitative assays in serology. (Sensabaugh G. , 1986) Sensabaugh stated, "[Q]ualitative tests are but quantitative tests with built in significance thresholds. … [Q]uantitative assessment of semen levels (by acid phosphatase and/or P30 assay) provides a background for the interpretation of genetic marker findings." He outlined significant research hurdles, including the development of quantitative standards, reagents, and an improved understanding of "the inner workings of immunologic test systems." In other talks at the conference, Sensabaugh and Blake presented research concerning the use of quantitation to interpret stains when non-foreign antigens are present in sexual assault samples, such as when a female victim's ABO markers are observed in evidence samples. However, it is clear that the consensus at the time was summarized by representatives of the FBI Laboratory Division: "Many laboratories rely on tests for P30 to identify semen stains <u>when microscopic searches for spermatozoa have failed</u>." (emphasis added) (The Laboratory Division, Federal Bureau of Investigation, 1986)

Interestingly, the proceedings also include a summary of a panel discussion concerning serology reporting procedures and courtroom presentation. The panel recommended that laboratory reports include information about body fluid types and origins examined, the consistency of conclusive genetic marker analyses with the victim(s) or suspect(s), and statistical characterization of the population that "possesses the same combination of genetic markers found both in a stain and in the victim and/or suspect[.]" Further, they were concerned that reports, testimony, and other communications "may still be misinterpreted by the recipients in support of their own theories." Therefore they state, "Conclusionary statements may admit to more than one interpretation and show a bias unless the report also includes a very thorough, detailed description of the examiner's assumptions, prejudices and less probable interpretations of the test results." They discuss the problem that rape shield laws may prevent the identification of a victim's consensual partners, thus complicating the interpretation of serological profiles of evidence. Finally, they presented strategies to prevent the "manipulation of the expert witness" by attorneys, although it was clear that this issue was unavoidable to some extent.

As late as 2000, Gaennslen would confirm the general view that ACP testing was a useful tool to screen and locate semen in stains but those results should be confirmed by microscopic observation of spermatozoa or a P30 test. (Gaennslen, 2000) He stated, "The presence of semen in a stain can thus be identified by finding either sperm cells or a unique component of the seminal plasma." Gaennslen clearly outlined the considerations for confirmation of the contribution of semen to a biological evidence stain. He minimized the importance of a quantitative ACP or other test because "it is not possible to relate a quantity of [ACP] found in some part of a [dried] semen stain back to the liquid specimen data in order to interpret the results of the stain test. The problem is not limited to semen stains or to [ACP]; it is a general problem with the interpretation of quantitative test results from dried stains."

First and foremost, the presence of spermatozoa on properly prepared and examined microscope slides is a "certain method for the forensic identification of semen in case specimens." Gaennslen states that P30 was developed to deal with the "unsolved problem" of the "identification of semen in stains and swabs in the absence of sperm[.]" It took many years for the scientific community to conclude that the various groups detecting P30-like compounds were seeing the same phenomenon and to associate it with the now well-known Prostate-Specific Antigen. (Deguchi, et al., 1991) By the year 2000 when Gaennslen contributed his chapter to the Wecht textbook, P30 testing had become "routine in most forensic laboratories." Today, many laboratories use a qualitative or semi-quantitative P30 test as part of routine serological characterization of sexual assault evidence.

When Garrett and Neufeld published their 2009 study (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009), this background had become almost ancient history in the forensic science community. In relying on the Blake approach to serology, Garrett and Neufeld routinely discounted testimony that indicated the forensic examiner quantified the presence of spermatozoa microscopically, an approach in direct conflict with the consensus in the scientific or practitioner communities during the pre-2000 period. This perspective is important, because 57 of the 92 cases of invalid testimony in their study involved serology. Garrett and Neufeld state that quantification techniques were developed in the mid-1980s to address this issue "but were not widely adopted by crime laboratories until later. None of the cases in the study set with invalid testimony involved the use of techniques for the quantification of semen." In fact, this latter statement is incorrect. There were many wrongful conviction cases in which quantification was performed using microscopic confirmation of spermatozoa, although none referenced the specific protocol advocated by Blake. Further, the full set of procedures used in the cases cannot be inferred from only testimony review, which seldom included extensive details.

In their analysis of wrongful conviction cases, GN did not recognize the value of the microscopic confirmation of spermatozoa, even though it is the most reliable method to confirm the presence of semen. (Blake, Sensabaugh, & Bashinski, A systematic approach to the analysis of semen evidence, 1980) Also, the proper procedure for the preservation of crime scene stains includes the drying of the evidence to prevent bacterial degradation. Unfortunately, the quantification of dried stains includes a great deal of uncertainty. (Gaennslen, 2000) Currently, labs rely on P30 for semiquantitative determination of semen only. (Laux, Tambasco, & Benzinger, 2006) Further, Blake quotes a concentration (1:100) that was not an accepted threshold in the forensic community. Typically, concentrations between 1:200 and 1:500 were accepted, based on guidance from the FBI or the Serological Research Institute (SERI). (Serological Research Institute, 1988)

Quantitation thresholds were at issue in the case of Jules Letemps, in which examiner Rathman found only H antigens consistent with the victim in a stain from a robe worn by the victim after a sexual assault. (Letemps v. Sec'y, Fla. Dep't of Corr., 2015) Letemps was a B secretor, but no B antigens were detected in the stain. Rathman performed a microscopic analysis of the semen stain but found no spermatozoa, leading to speculation that Letemps was sterile or had a low sperm count. Rathman attempted a P30 quantification, stating that “you would need approximately 1 in 200 dilution to be assurance that you would be picking up the blood type of the semen donor in a stain,” a standard she attributed to SERI. The sample’s semen dilution was calculated to be 1 in 322, meaning that Letemps could not be ruled out as a contributor of the semen. In her notes, she recorded that the P30 concentration was 15.49 units. She based her dilution calculation on the assumption that a “neat semen stain” would be measured to be 5000 units. During postconviction proceedings, Gary Harmor, a senior forensic serologist at SERI, stated that SERI had established that neat semen stains should be assumed to be 3000 units to account for the fact that semen stains "will also experience some loss due to drying and degradation." The application of a 3000-unit baseline would yield a dilution ratio of 1:193, just within the semen detection threshold used by the Florida Department of Law Enforcement at the time. Rathman had relied on a SERI training manual from August 1988 for the 5000-unit baseline but SERI had updated its manual in September 1989 to reflect a 3000-unit baseline. The NRE classifies the case as False/Misleading Forensic Evidence, maintaining that she “applied an incorrect standard to determine the stain was too diluted.”

It is noteworthy that the Blake threshold would fail to exclude Letemps in this case, because he uses a 1:100 dilution threshold. More importantly, it is unclear whether the Florida court understood the uncertainties involved in the analysis, including those relating to the quantification of dried stains, variability among males in P30 concentrations in semen, variations in dilution threshold standards, and variations in standards for neat semen stains. On the basis of the reinterpretation of the serology, the court stated, “With this new evidence, Letemps has presented a credible claim of actual innocence. … Harmor's expert testimony effectively rules out both Williams and Letemps as the donor, meaning that the rape was committed by someone other than Letemps.” (Letemps v. Sec'y, Fla. Dep't of Corr., 2015) In this case, the court misinterpreted the limitations of quantitative frameworks to confirm the presence of the male fraction in a biological stain.

Finally, GN reference the Bromwich report, which reviewed the Houston crime lab in 2006. (Bromwich M. R., 2007) The Bromwich report was broadly critical of the practices of the Houston lab, which was associated with eight wrongful convictions in the dataset, excluding the cases associated with field drug testing. The Bromwich report found that the most common mistake was the failure to perform any ABO testing whatsoever in 274 cases, even when sufficient biological evidence was available. They also found:

1. Failure to report a potentially probative finding;

2. Incorrect interpretation or reporting of serology testing results;

3. Reports not supported by documentation;

4. Failure to perform a critical examination;

5. Other errors, including alterations of reports and possible drylabbing.

The Bromwich report held that ACP and P30 tests were generally presumptive in nature and that microscopic sperm identification was the primary method to confirm the presence of a male fraction in evidence. In fact, the report further states, “It is incorrect for the Crime Lab serologists to have concluded that no sperm cells were detected when sperm heads in fact were observed. The consequence is that, in such cases, the Crime Lab erroneously reported that evidence was negative for semen when, in fact, potentially probative genetic marker analysis could have been attempted on the evidence." (Bromwich M. R., 2007)

Basis for interpretation of serological analysis

In serological interpretation of sexual assault samples, the contribution of a male assailant to the serological profile of a biological sample may be masked by the blood group substances from a victim's serological profile. (Culliford, 1971) In addition, serology permits the association of an individual with a class of individuals (or population) that is consistent with the serological profile of the biological sample. Serology does not permit individualization. In any serological or DNA analysis, conclusions should clearly delineate among activity, source, and sub-source hypotheses. (Champod, Biedermann, Vuille, Willis, & De Kinder, 2016) Reports and testimony should clarify the difference between the serological profile of a source, the serological profile of a biological sample, and the serological profile of a population of sources that may have contributed to a biological sample. The current study applies the following standards to the interpretation of serological analyses:

1. Examiner testimony should reflect that serology could not be used to identify an individual or conclude that they were the contributor to a particular piece of evidence.
2. Examiner testimony should reflect uncertainties with respect to masking, contamination, degradation, testing limitations, or other factors relevant to the case in question.
3. Microscopic confirmation of spermatozoa is a sufficient basis to assume that the serological profile of a questioned sample should reflect a male contributor. Regardless of the method used to confirm a male fraction, victim or other contributors to a biological sample should be addressed in the interpretation of the serological profile of a biological sample.
4. The examiner may account for masking when spermatozoa are not present in a biological sample as follows:
    a. Prior to 1986, a semi-quantitative acid phosphatase assay may be used if the limitations of acid phosphatase are clarified; *or*
    b. A semi-quantitative P30 assay may be used if the limitations of P30 are clarified; *or*
    c. The examiner may adhere to a written laboratory policy that would be compliant with the general standards reflected in scientific and practice standards at the time of trial, including those reflected in Gaennslen (Gaensslen, 1983) (Gaennslen, 2000), FBI documents (The Laboratory Division, Federal Bureau of Investigation, 1986) (Federal Bureau of Investigation, 1982), or similar references.
5. Examiner testimony should reflect general procedures and interpretation issues but is not required to include a complete account of testing protocols. Subsidiary documents are used when available to determine if the examiner's lab work, reports, communications, and procedures conformed to generally accepted procedures, legal requirements, and testimony standards.
6. The examiner should not confuse crime, activity, source, and sub-source hypotheses. (Champod, Biedermann, Vuille, Willis, & De Kinder, 2016) In particular, the examiner should not state or imply that the statistical characterization of the serological profile of a possible source is the same as the statistical characterization of the serological profile of questioned evidence.

Case documentation

Case information was collected from publicly available sources, including state and local court systems. Criminal case appeal decisions and related documents were obtained from online repositories, including LexisNexis. When available, civil case decisions and related documents were also obtained. The NRE, convictingtheinnocent.com, and other research archives and publications were used to obtain case information and trial documents. Media accounts were collected to supplement materials to elucidate issues, especially in cases in which other materials were not readily available. In some cases, documentation was supplemented with material from official reviews. For example, the Texas Forensic Science Commission (https://www.txcourts.gov/fsc/) and North Carolina Innocence Inquiry Commission (https://innocencecommission-nc.gov/) maintain documentation of their reviews on public websites. All case documentation was summarized, after which the coding scheme was applied using a Google Forms tool. The information was captured in a Case Review Form for each case that captured all demographic, case, forensic science, and summary information. In all, case documentation exceeded 6000 files. Unlike prior studies, this study did not rely primarily on testimony review. Documents and decisions related to criminal and civil appeals provided information and insights into forensic issues. In general, appeal decisions were considered authoritative and objective determinants of case errors unless a court clearly erred based on other documentation or published scientific research. This approach permitted more systematic and reliable assessment than would otherwise have been possible and mitigated the limitations associated with prior research on this topic. Further, error coding was justified for individual cases and examinations to permit independent assessment of research findings. Finally, the study relied on the predefined Forensic Error Typology and supporting framework to support consistency, objectivity, and reliability of findings.

Overview

The dataset includes 732 cases from 44 states and the District of Columbia. Within the definitions of the study, 622 cases included a case error associated with forensic evidence, and 110 cases did not include an error related to forensic evidence. In 137 cases, a "harmless error" occurred in which the forensic evidence was not required in the balance of evidence to convict. Thus, there were 485 cases in which a forensic error contributed to a wrongful conviction.

Seized drug cases can be considered separately because their distinct etiology. The dataset includes 130 cases in which seized-drug analysis was the only forensic evidence. A forensic laboratory error occurred in one case (Jermaine Dollard), while the remaining 129 were associated with reliance on presumptive field test kits to elicit guilty pleas from defendants. In each of the 129 cases, laboratory analysis did not confirm the presumptive finding, and the conviction was vacated. It should be noted that many drug-related wrongful convictions are not represented in the NRE and therefore not included in this study's dataset. For example, the Dollard case was representative of a much-broader laboratory problem in the Office of the Chief Medical Examiner for the State of Delaware. (Andrews International, 2014) The dataset includes 602 "non-drug" cases that did not solely involve seized drug analysis.

Errors of association may be associated with wrongful convictions. Of the 602 non-drug cases, there were 245 cases in which at least identification or classification error occurred. The errors contributed to the balance of evidence to convict in 209 cases.

732 cases — • Complete dataset

622 cases — • The case included some type of forensic science case error.

485 cases — • A forensic science case error contributed to the wrongful conviction.

355 cases — • Non-drug cases with forensic science case errors that contributed to the wrongful conviction.

245 cases — • Non-drug cases with identification or classification errors that contributed to the wrongful conviction.

A forensic examiner made an error of individualization or classification that contributed to conviction in 209 non-drug cases. More broadly, the dataset includes 305 non-drug examinations that resulted in individualization or classification errors. Individualization errors included 27 bitemark examinations, 20 hair comparisons, 14 latent print comparisons, 12 firearms identifications, and 8 DNA analyses. Classification errors included 51 serology analyses, 34 pediatric abuse examinations, 18 forensic pathology examinations, 17 fire debris investigations, and 9 blood spatter analyses.

In many cases, there were multiple association errors. For example, in the 1987 Steven Chaney case, bitemark examiners produced incorrect identifications associating Chaney with the victim. (Texas Forensic Science Commission, 2016) The latent print examiner made a valid identification of Chaney's thumbprint at the crime scene but produced an invalid association concerning the age of the latent print. The forensic pathologist changed his assessment of the age of a presumed bitemark and did not document the reason for the change. It is possible that the incorrect associations were the result of a bias cascade in which investigation and forensic errors reinforced each other. (Dror, Morgan, Rando, & Nakhaeizadeh, 2017) Notably, the Steven Chaney case was thoroughly reviewed by the Texas Forensic Science Commission (TFSC), so the nature and extent of errors in the case are better understood than in many other cases. In all, nine cases were associated with three or more association errors and 41 with two association errors.

Testimony was provided in 588 cases, and testimony errors occurred in 376 of these cases. In 87 cases, two or more forensic examinations were associated with testimony errors. In 284 cases, the testimony error contributed to the wrongful conviction. In "harmless" testimony errors, hair comparison and serology testimony were the most common disciplines. These cases often included victim eyewitness testimony and other evidence and the hair and serology evidence was presented as minimally probative or not inculpatory at all. Of the 1220 examinations in which testimony occurred, 483 examinations were associated with testimony errors, of which 347 contributed to the wrongful conviction. In 69 instances, the testimony conformed to the standards at the time of trial but would not conform to current standards. In particular, testimony standards for hair comparison, pediatric abuse, and fire debris investigation have changed substantially.

In 516 cases, officers of the court produced errors related to forensic evidence. In 409 of these cases, the forensic evidence contributed to the wrongful conviction. In these cases, there were a total of 625 examinations mishandled by officers of the court. Officers of the court produced errors when they misrepresented forensic evidence, failed to use probative forensic evidence, or failed to counter misrepresentations of forensic evidence. In many cases, these errors were the only issue affecting forensic evidence. In other words, there were no errors in a case related to the collection, analysis, and communication of forensic evidence but an officer of the court made an error in the use of forensic evidence. In 46 cases and 111 individual forensic analyses, an error by an officer of the court was the only error associated with forensic evidence.

Most commonly, defendants lacked adequate defense, including 210 cases in which an independent review or examination could have been relevant to the wrongful conviction and 82 cases in which the defense did not present exculpatory evidence. The study did not examine questions related to the reasons for the adversarial deficit in wrongful convictions, such as a lack of resources. (West, 2010) In many instances, attorneys indicated a lack of understanding of the forensic evidence or demonstrated that lack of understanding in their handling of the case. In 91 instances, a prosecutor mischaracterized forensic evidence, including 23 hair comparison cases. In 37 cases, a judge failed to conduct a review of a novel technique and allowed faulty testimony in 26 cases. These figures may underestimate the incidence of judicial and prosecutor errors related to forensic evidence due to the lack of complete and publicly available transcripts in most cases. The coding is largely based on cases in which an appeals court made findings concerning deficiencies in trial proceedings.

The dataset includes 1,391 forensic examinations, of which 891 had a case error and 500 were valid and not associated with a known case error. Generalizations from the dataset require great care. Many cases were associated with multiple types of forensic examinations and multiple errors. Many examinations were conducted appropriately but system errors undermined the use of the forensic evidence. The Forensic Error Typology includes system issues outside the control of forensic scientists and forensic science organizations. For example, 209 of the examinations were associated with Type 4 or Type 5 errors outside the control of the forensic analyst in the case. The typology also includes issues such as evidence handling and communication that are distinct from considerations in the analysis and interpretation of evidence. Finally, the implications of the experience from wrongful convictions vary among disciplines and jurisdictions depending on the scientific and organizational contexts.

Demographics and crime types

The dataset includes cases from 44 states and the District of Columbia, but roughly half of the cases arise from five states: Texas (177, including 118 drug-only cases), New York (59 cases), Illinois (50 cases), California (45 cases), and Michigan (34 cases). The dataset is heavily weighted toward violent felonies, such as murder/manslaughter (348 cases), sexual assault (173 cases), child abuse (91 cases), robbery (90 cases), burglary (58 cases), and arson (39 cases). Males (642 defendants) outnumbered females (90 defendants). African American defendants (338) constituted almost half of the dataset, outnumbering White defendants (321), Latino/Hispanic defendants (53), and Asian American defendants (9). In general, forensic errors did not correlate with demographic factors. For example, roughly 2/3 of both African American defendant cases and White defendant cases included at least one inadequate defense issue. There were significant differences related to type of alleged crime. African American defendants were more likely to be accused of sexual assault (100 cases), robbery (63 cases) or a drug offense (88 cases), while White defendants were more likely to be accused of murder (172 cases) or child abuse (46 cases). As a result of these alleged-crime factors, African American defendants were associated with a disproportionate share of serology errors (71 cases) and hair comparison errors (45 cases) compared to White defendants (55 cases and 36 cases, respectively). Similar disparities were not observed for DNA analysis, crime scene factors, or police errors associated with forensic evidence.

Defendant age at the time of the alleged crime varied from 11 to 66 years old, with an average age of 29 years. The time between the alleged crime and the wrongful conviction varied from 0 to 28 years, with an average of 1.8 years. The time between the alleged crime and the final exoneration varied from 0 to 58 years, with an average of 14.3 years. For this study, exoneration is coded as the date at which charges are dropped or an official act of government ended the original case (such as a pardon). The time between the wrongful conviction and the final exoneration varied from 0 to 58 years, with an average of 12.4 years. As seen in the Figure 1, the time from crime to exoneration does not provide clear guidance for forensic evidence storage policies, because significant numbers of cases are associated with extended timelines. For example, there are 16 cases in which an exoneration occurred 27 years after the original crime.

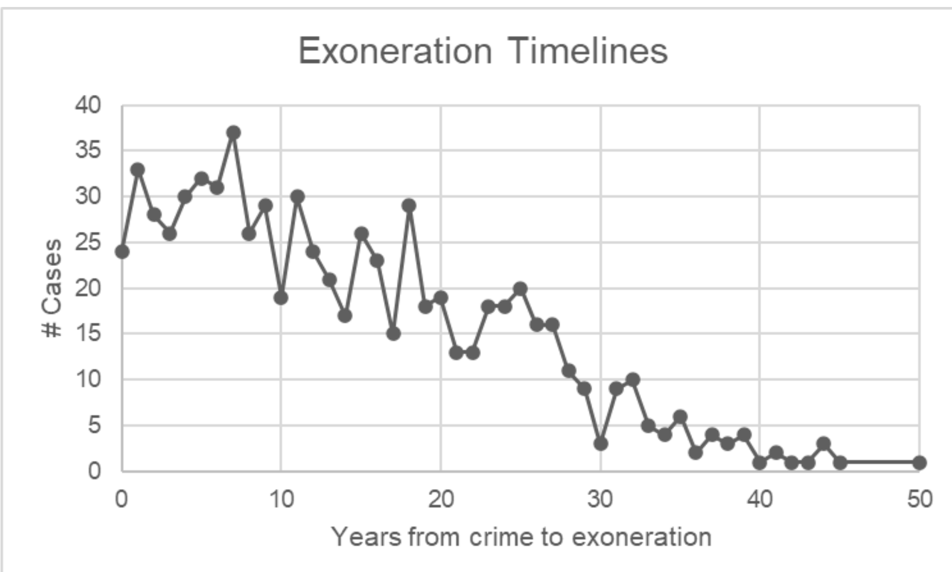

Figure 1. Incidence of number of years that elapse from alleged crime to exoneration in dataset cases.

The dataset's 732 exonerations provide a picture of the changing role of forensic evidence as a contributor to wrongful convictions, as seen in Figures 2 and 3. The figures exclude drug-only cases, which provide significant spikes in incidence in the 2009 to 2015 period. Recent convictions make a minor contribution to known exonerations, although it should be noted that it often takes over a decade for a wrongful conviction to be overturned. That said, Figure 3 shows a clear and steady decline of wrongful convictions associated with False/Misleading Forensic Evidence by the NRE since 1990.

There has been a steady rise in exonerations associated with False/Misleading Forensic Evidence by the NRE, but this may reflect the overall trend in the number of discovered wrongful convictions over the last decade. Increased attention to the problem of wrongful convictions has contributed to increased resources for innocence organizations and conviction integrity units. (Findley & Golden, 2013) Nonetheless, forensic-evidence-associated cases show diverging trends when viewed by year of conviction or year of exoneration. This may indicate that improvements in forensic science technology, policy, and practice have been successful in reducing the risk of wrongful convictions. If so, as the criminal justice system resolves older cases, the number of exonerations associated with forensic evidence issues may decline significantly. Alternatively, many forensic-science-associated errors are related to system issues outside the governance structures of public crime laboratories. If unaddressed, this issue may continue to contribute to detected wrongful convictions in the future.

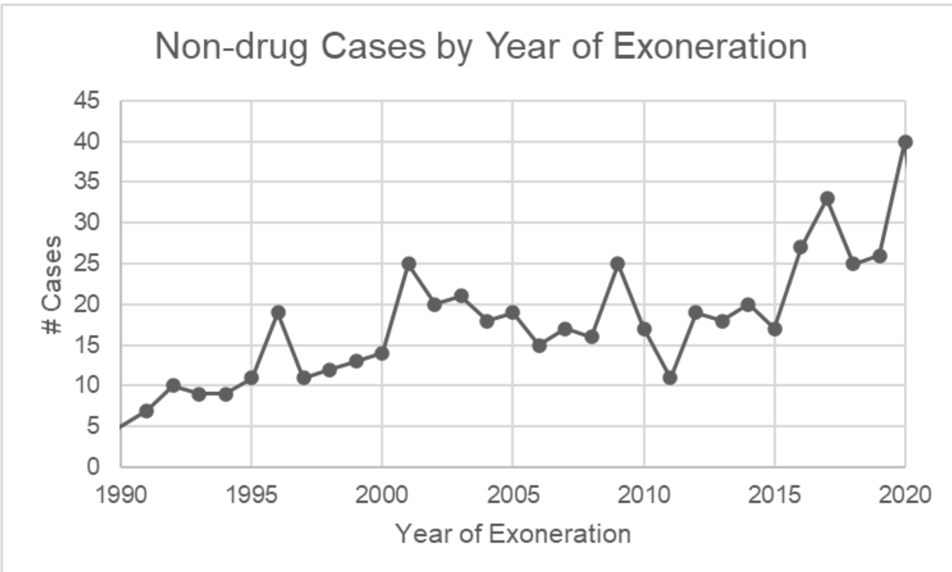

Figure 2. Incidence of exoneration year for non-drug cases in the dataset.

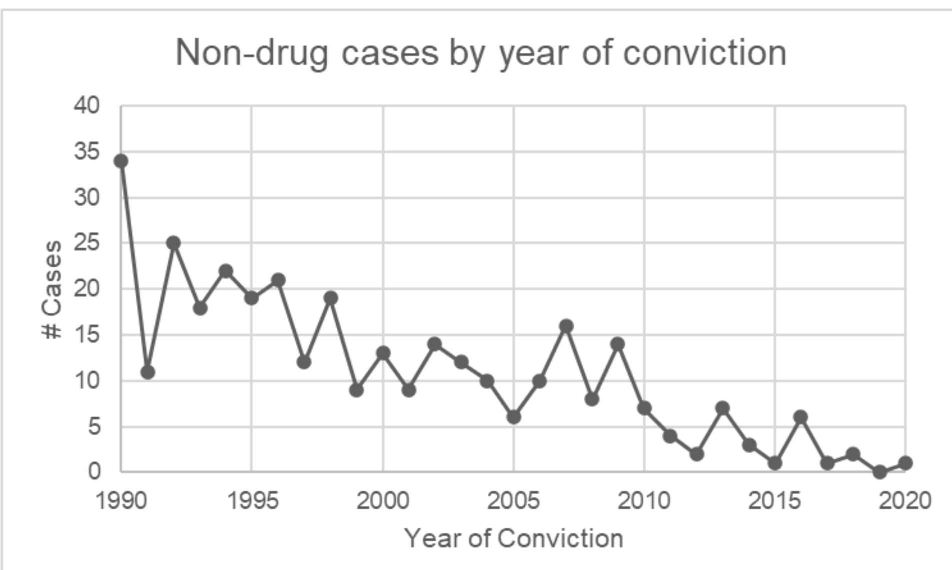

Figure 3. Incidence of conviction years for non-drug cases in the dataset.

During or after 2010, this wrongful conviction dataset included only 34 wrongful convictions that weren't solely the result of field-drug-test errors. These 34 cases involved 85 forensic examinations, of which 52 were associated with a case error. Forensic medicine (18 examinations) and forensic pathology (10 examinations) were the most common disciplines associated with recent case errors. Other disciplines included gunshot residue (4 examinations), crime scene investigation (3 cases), latent prints (3 examinations), DNA (2 examinations), and digital evidence (4 examinations). This breakdown differs from the experience of older cases in which outdated techniques played a role in many wrongful convictions.

Factors associated with exonerations are complex. In many cases, convictions are overturned on issues unrelated to the innocence of the defendant. The dataset includes 205 cases in which inadequate legal defense was associated with the wrongful conviction by the NRE, and many more by this study (485 cases). In many cases, the forensic evidence was presented without appropriate defense review or cross-examination, and appeals courts held that this shortfall prevented a fair trial. In 250 cases, a forensic science issue was a primary basis for vacating a conviction. In 206 of those cases, the wrongful conviction was associated with inadequate defense related to forensic evidence.

DNA exonerations include a substantial portion of wrongful convictions in the dataset (183 cases), but DNA was not dispositive in 57 of these cases. DNA exonerations have become less likely, and only 38 exonerations since 2010 were F/MFE cases in which DNA established a defendant's innocence.

Defendants require external advocates to overturn their wrongful conviction. Innocence organizations were involved in 218 exonerations in the dataset. Conviction Integrity Units contributed to 143 exonerations, including 140 in the post-2010 period. Exonerations were obtained most commonly through successful state-level habeas corpus proceedings (505 cases), though Gubernatorial pardons (62 cases), direct appeals (59 cases), and federal habeas proceedings (41 cases) were also significant contributors. Government compensation was provided to exonerees in 176 cases, and lawsuits were settled or successful in 137 cases. In all, exonerees received $802,602,106 in compensation arising from cases in this dataset. The state of Texas also provides monthly annuities totaling $153,184 to exonerees.

Forensic disciplines

The dataset includes 34 types of forensic evidence categories, including one category for the 12 cases which included "No forensic evidence." Some disciplines were present in a limited number of cases. In addition, some examinations involved highly specific interpretation issues within disciplines. When possible, specific research papers or guidance documents were identified to establish the standards for assessing the forensic evidence work in these situations. Table 2 delineates the number of examinations associated with each type of evidence, the number of case errors associated with that evidence, and the number of examinations presented without case errors. The study defines case errors to include any error related to forensic evidence during the adjudication of a wrongful conviction case. This aligns with the NRE category, False/Misleading Forensic Evidence, which does not distinguish among system errors related to forensic examiners or other criminal justice practitioners.

Table 2. Summary of examinations and errors by discipline in wrongful convictions in the dataset.

Discipline	Number of examina-tions	Case Error	Any Type 1 Error	Any Type 2 Error	Any Type 3 Error	Any Type 4 Error	Any Type 5 Error
Serology	204	139	55	54	90	42	16
Hair comparison	143	84	13	29	72	34	8
Forensic pathology (cause and manner)	136	63	21	18	32	20	2
Seized drug analysis	130	130	0	130	0	1	1
Latent fingerprint	87	40	17	16	21	27	12
Firearms identification	66	26	16	17	16	13	6
Forensic medicine (pediatric sexual abuse)	64	46	21	22	36	8	2
DNA	64	41	14	9	14	8	8
Forensic medicine (pediatric physical abuse)	60	50	20	13	28	11	2
Fire debris investigation (not chemical analysis)	45	35	22	17	30	5	3
Bitemark	44	34	23	32	31	19	4
Fiber/trace evidence	35	16	3	5	7	11	2
Blood spatter (crime scene)	33	19	9	9	13	9	3
Shoe/foot impression	32	21	12	13	14	2	0
Gunshot residue	29	19	10	9	8	9	3

Discipline	Number of examina-tions	Case Error	Any Type 1 Error	Any Type 2 Error	Any Type 3 Error	Any Type 4 Error	Any Type 5 Error
Fire debris chemical analysis	29	17	3	4	6	5	7
Crime scene investigation	26	12	4	0	4	9	2
Dog scent	17	15	11	14	15	2	3
Digital evidence	16	12	4	3	6	4	2
Forensic pathology (time of death)	16	8	1	0	3	1	0
Latent palmprint	15	8	2	1	2	5	1
Forensic pathology (wound assessment, toolmark)	12	6	5	5	5	3	1
Forensic medicine (non-pediatric)	12	3	3	3	3	2	0
Toxicology	12	7	4	4	3	2	2
No forensic evidence	12	0	0	0	0	0	0
Handwriting	10	7	2	2	6	3	0
Forensic pathology (firearm trajectory, blast, distance, handedness)	8	7	2	0	4	2	1
Ballistic trajectory (crime scene)	7	5	1	0	3	2	**1**
Biological evidence	6	6	0	0	0	3	0
Friction ridge (not finger or palm)	5	5	5	5	5	0	0

Discipline	Number of examina-tions	Case Error	Any Type 1 Error	Any Type 2 Error	Any Type 3 Error	Any Type 4 Error	Any Type 5 Error
Tire/tread impression	5	1	0	0	0	1	0
Voiceprint	3	2	2	2	2	2	0
Accident reconstruction	3	3	2	1	2	0	0
Toolmark (not firearms or autopsy)	2	2	2	0	0	0	0
Other	1	1	1	1	1	1	0
Fingernail comparison	1	1	1	1	1	0	0
Soil comparisons	1	0	0	0	0	0	0
Document examination (not handwriting)	0	0	0	0	0	0	0
	1391	**891**	**311**	**439**	**483**	**266**	**92**

Altogether, the study includes 732 cases and 1,391 examinations, or an average of 1.88 examinations per case. The study does not include some types of expert testimony that appear to be classified as forensic evidence in the NRE, including Child Sex Abuse Accommodation Syndrome testimony, gang violence expert testimony, and other psychological assessments. This difference largely accounts for the "No forensic evidence" cases in the dataset. Biological evidence was coded separately in a small number of cases in which it was collected but not subjected to known serological or DNA analyses. Crime scene investigation was coded only in cases in which it was relevant to reports or testimony independent of other forensic evidence. Thus, crime scene reconstruction was coded under crime scene investigation, unless the work was limited to blood spatter or ballistic trajectory analyses. In most cases, crime scene issues were coded under Error Type 5 of the Forensic Error Typology for the related forensic examination. The use of canine detection was coded as a forensic technique only in cases in which it was used to make an identification or classification conclusion that was presented in legal proceedings. The only "Other" examination involved the analysis of agricultural crop yields in a federal case (defendant James Catton) in which the examiner did not collect the reference data needed to support his testimony. The five friction ridge examinations included a lip print examination in the Lavelle Davis case and four ear print examinations in the David Kunze case. All five examinations were associated with an inadequate scientific foundation, possible cognitive bias, and judicial errors related to the failure to review a novel method or the acceptance of faulty testimony over defense objections.

Examiners and forensic science organizations

Very little is known from previous research concerning the characteristics of forensic examiners or forensic science organizations associated with wrongful conviction cases. Some have speculated that forensic examiners lack necessary education or training or appreciation for the scientific process. (Koehler, et al., 2011) Others have speculated that the cases are largely associated with “bad apples” who fall outside of the norms of the broader forensic science profession, although that theory has received criticism. (Thompson W. C., 2007) The characteristics of forensic examiners and forensic science organizations were coded when available, including educational level, organizational context, and certification/accreditation. In many cases, these variables were not available, so findings were only possible in limited circumstances.

There is no basis in wrongful convictions to conclude that the risk of forensic errors is mitigated by the educational level of the examiner. In 324 examinations, the examiner had a doctoral-level degree, such as a PhD or medical degree. Doctoral-level examiners were associated with 231 case errors, including 165 forensic conclusions that contributed to a wrongful conviction. They committed 104 association errors. Almost all case errors related to bitemark comparison, forensic medicine, or forensic pathology were associated with doctoral-level examiners. In 53 instances, doctoral-level examiners produced invalid conclusions that were based on methods with an inadequate scientific foundation. For example, in the cases of Dale Johnston, Roland Cruz, and Alejandro Hernandez, forensic anthropologist Dr. Louise Robbins made erroneous identification decisions based on shoe wear patterns. In the Johnston case, a defense expert testified that Robbins’ research “in the individuality of human footprints had not been sufficiently documented in the scientific literature.” Nonetheless, the trial court and the appeals court allowed the invalid testimony. (State v. Johnston, 1986) Like Robbins, Dr. Bruce Woodling produced unvalidated testimony based on his own research, but in an entirely different field. (People of the State of California vs. Scott Lee Kniffen, 1995), the assessment of pediatric sexual abuse. (Woodling & Heger, The use of the colposcope in the diagnosis of sexual abuse in the pediatric age group, 1986) (Woodling & Kossoris, Sexual Misuse: Rape, Molestation, and Incest, 1981) By his own account, he made substantial income from training medical and legal professionals on his theories and methods. In the Kern County, California pediatric sexual abuse prosecutions, Woodling testified that his “wink response” test was an “accurate and infallible” method to determine if a patient had been previously sodomized. Woodling’s research was not scientifically valid or reliable. The trial court failed to recognize the novelty of the Woodling method and did not perform a Kelly-Frye review. The convictions of Scott and Brenda Kniffen and Alvin and Debbie McCuan largely resulted from Woodling’s invalid testimony.

Many examiners associated with repeated forensic case errors share characteristics with Robbins and Woodling. Fred Zain (seven cases), Arnold Melnikoff (3 cases), and Michael West (6 cases) made several statements in the records of wrongful conviction cases concerning their research work. These individuals maintained that they were on the cutting edge of forensic practice and felt that they were punished for advocating innovations in their fields. For example, Fred Zain was involved in six serology cases in West Virginia that included serious statistical interpretation errors. Nonetheless, Zain had constructed datasets for West Virginia populations across several types of serological markers and published the information in a widely-accepted, peer-reviewed journal (Zain, Fred S., T. A. Smith, and H. B. Myers. "Population Data of Casework in West Virginia on Six Genetic Marker Systems." *Journal of Forensic Science* 34.4

(1989): 1007-1010.). At trial, he referred to these databases and related his statistical characterization of serological profiles directly to his published research work. Nonetheless, Zain had a "long history of falsifying evidence in criminal prosecutions … [and his] pattern and practice of misconduct completely undermined the validity of any forensic work he performed or reported … " (In re: W. Va. State Police Crime Lab., 1993)

Many individuals who testify as forensic science experts operate outside of the context of a forensic science organization. These individuals may not conform to the same standards of accreditation, certification and conduct that are common in public forensic laboratories. The 17 "canine detection" examinations (associated with 15 case errors) and 44 bitemark comparisons (associated with 34 case errors) are clear examples, although most bitemark examiners in wrongful conviction cases were certified by the American Board of Forensic Odontology. Two other cases provide interesting perspective on this issue. In the Hamid Hayat terrorism trial, senior imagery analyst Eric Benn with the Department of Defense testified concerning satellite images of infrastructure around camps in Pakistan. Benn stated that the image details showed the camps were for terrorist training, with odds of 50%, which he characterized as "a good solid possible." (United States v. Hayat, 2017) In the context of intelligence community analysis, Benn provided a typical probability estimate, which was supported by other intelligence information. In a forensic science context, he relied on "domain-irrelevant" information to produce an invalid and biased statistical estimate. (Dror I. E., Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias, 2020) In another case, Cynthia Sommer was convicted of poisoning her husband using arsenic after testing found extraordinarily large amounts of arsenic in his liver and kidneys. (Sommer v. United States, 2013) Key testing was performed by the Environmental Division of the Armed Forces Institute of Pathology (AFIP), a unit that did not normally provide forensic testing. Postconviction review established that there were 16 breaks in the chain of custody of the tissue evidence and that laboratory contamination was likely responsible for the presence of arsenic in the evidence samples. Subsequent testing at NMS Labs confirmed the arsenic levels in the tainted samples, but later testing of other samples by the Quebec Toxicology Center found no arsenic at all. The AFIP is a respected laboratory and Dr. Jose Centeno, the AFIP analyst, had published 19 scientific papers on chemical analysis of arsenic. Still, the chain-of-custody and quality assurance controls associated with forensic samples were unfamiliar to the AFIP Environmental Division, which was not a forensic laboratory.

Case errors may be associated with a wide range of governance structures. Laboratory independence may be justified but wrongful conviction data suggests it is not a panacea that would prevent errors related to forensic evidence. In fact, many examiners associated with errors were completely independent of law enforcement. Of the 234 examinations known to be conducted fully independent of law enforcement, there were 164 case errors, including 119 that contributed to a wrongful conviction. Possible cognitive bias was observed in 83 of these examinations, and there was a failure to follow best practices in 44 instances. These numbers are closely aligned with the overall dataset. Public forensic laboratories were responsible for 839 forensic examinations associated with 578 case errors and 375 case errors that contributed to a wrongful conviction. Possible cognitive bias was observed in 128 of these examinations and failure to follow best practices in 223 examinations. Similar metrics were observed for laboratories at the state and federal levels.

In general, there was little or no discernable relationship between forensic science errors and the educational level of the examiners. In fact, most bitemark and forensic pathology errors were associated with examiners with a Doctoral-level degree. Hair and serology cases were notable exceptions to this trend. The educational level of the examiner is not known in most hair and serology cases. There were 58 cases in which a Bachelor-degree-level examiner was associated with a hair comparison or serology case error, and 20 cases in which a Doctoral-level examiner was associated with a hair comparison or serology case error. That said, none of the 20 Doctoral-level examiners were associated with errors that contributed to a wrongful conviction, while 26 Bachelors-level examiners were associated with errors that contributed to a wrongful conviction. Education level was correlated with improved testimony outcomes and scientific characterizations in hair and serology cases. Overall in hair and serology cases, Bachelors-level examiners were almost twice as likely to commit a testimony error as a doctoral-level examiner. While it is possible that the better-educated forensic analysts had a better understanding of scientific and testimony standards, the gap may also have been due to training, laboratory management deficiencies, or limitations of the available data.

Some forensic practitioners have disregarded scientific studies relevant to their discipline, even many years after practice standards were revised based on new science or technology. This phenomenon was observed in fire debris investigation and medical assessments of pediatric abuse, as described in those sections. Statements by these examiners imply that they continued to rely on training they received prior to the adoption of new standards or that the standards were adopted prior to full consensus and support among forensic practitioners. Thus, any new standards will be effective only if current practitioners are retrained on the changes, including the basis for the changes in scientific research and the reliability of forensic analysis.

Scientific and technological improvements

The study period covers over a century of wrongful conviction cases, although most dataset convictions post-date 1980. Many wrongful convictions are associated with methods and standards that have been replaced by techniques with improved probative value based on scientific and technological advances. For example, while serology and hair comparison were associated with wrongful convictions prior to 2000, DNA has largely replaced these methods. Also in the pre-2000 period, improvements in fire debris investigation and firearms identification addressed practice deficiencies raised by wrongful conviction cases. On the other hand, unproven innovations have introduced unreliable forensic results into criminal trials.

Scientific and technological improvements may have improved the probative value of 607 examinations in the dataset, including 484 that were associated with a case error. In 186 cases and 335 examinations, a case error that contributed to a wrongful conviction may have been prevented by the application of improved science or technology. Clearly, the ability of the forensic science community to develop new, validated methods can contribute to the reduction of forensic science errors. Relevant scientific and technological improvements are not limited to DNA analysis but include a wide range of analytical methods with improved sensitivity and selectivity, database improvements, and standards that rely on improved scientific foundations.

DNA now provides more probative information in all 347 hair and serology examinations in the dataset, as well as many of the 44 bitemark examinations. When applied correctly, hair comparison and serology are primarily used to eliminate suspects or associate a defendant with a class of sources in a population. In contrast to DNA, hair comparison and serology could not be used reliably to identify an individual. As a result, case errors associated with hair comparison or serological analyses are rarely observed in wrongful convictions when DNA was available after 2000. In the 2009 trial of William Campbell, examiner Michael Trimpe from the Hamilton County coroner's office implied a hair individualization when he stated, "In all of my years looking at hairs, I've had one case where I couldn't tell the difference between two people, and they were an Asian brother and sister." (State v. Campbell, 2019) There is no public record that a DNA test was attempted on the hair evidence. In the 2008 Knolly Brown case, serology was used to support the prosecution theory that the victim's blood was on the defendant's jacket. (State of North Carolina vs. Knolly Brown, Jr., 2015) There was also a hair fragment that "was found to be microscopically similar to the pubic hair standard from Knolly Brown, Jr." In that case, the hair was sent for mitochondrial DNA testing. The exculpatory DNA report was sent to the Rocky Mount Police Department and the Rocky Mount District Attorney by next-day UPS air but were lost. Postconviction, the North Carolina Innocence Inquiry Commission exonerated Brown on the basis of Y-STR testing of the sexual assault kit. They also found the exculpatory DNA report on the hair evidence and other lost evidence in a Rocky Mount police officer's personal storage unit. The case is a good example of the limits of forensic technology when poor communication and evidence handling practices undermine the reliability of an investigation. Other post-2000 cases used serology to identify the presence of blood or other biological fluids or determine the time of deposition of a stain, issues which are not generally amenable to DNA analysis and remain subject to significant scientific uncertainty.

The 130 drug-only wrongful convictions were based on colorimetric field test kits, but improved technologies now exist to provide more reliable presumptive testing. (Fedchak, 2014) Improved analytical methods are also relevant to laboratory-based confirmation testing. Prior to 2000, many laboratories applied gas chromatography (GC) using flame ionization detection (FID) or similar methods with limited ability to identify specific chemical constituents. FID has been replaced with mass spectrometry in seized drug analysis, toxicology, and fire debris chemical analysis. In 11 cases, the use of mass spectrometry may have prevented a wrongful conviction.

Interjurisdictional latent print searches using Automated fingerprint identification systems (AFIS) can produce cold hits that were not technologically feasible until recent years. (Stokes, 2019) Improved AFIS technology could have impacted 13 wrongful conviction cases in the dataset. In some cases, prints were found at the scene that were believed to be probative but the defendant was excluded as a source. The 1975 conviction of Edward Carter was overturned after an AFIS hit to an alternate suspect. (Carter v. City of Detroit, 2016) In the 1980 Elmer Daniels case, the victim's notebook was found in Daniels' home but never associated with the correct source—his brother—until a postconviction AFIS search in 2018. Daniels' brother was an alternate suspect in the case, which had relied on an unreliable hair comparison from FBI examiner Michael Malone. (Otterbourg, 2020) In the LaMonte Armstrong case, a latent palmprint was never identified until a postconviction AFIS cold hit in 2012. (Armstrong v. City of Greensboro, 2016) When used in a valid manner during a thorough investigation, AFIS technology may prevent similar wrongful convictions.

The limitations of GC-FID were central to the 1991 conviction of Patricia Stallings, whose child died of a rare genetic disorder, Methylamalonic Acidemia (MMA), which produced high levels of methylmalonic acid in the child's bloodstream. (Tipton, 1991)Two laboratories confirmed that the child's blood contained lethal levels of ethylene glycol using GC-FID. The elution times of ethylene glycol and methylmalonic acid are similar, but the two chemicals are easily distinguished by mass spectrometry. The mistake was discovered postconviction by Dr. William Sly of St. Louis University, whose laboratory was responsible for one of the incorrect ethylene glycol conclusions. (Shoemaker, Lynch, Hoffmann, & Sly, 1992) Sly spiked a blood sample from the child with ethylene glycol and generated a double-headed GC peak. In addition to demonstrating the value of mass spectrometry, the case also reflects cognitive bias concerns. The possibility of MMA was known at the time of trial, but medical and police investigators discounted it after the initial misdiagnosis by the child's pediatrician and the faulty lab results. One test was not consistent with the prosecution theory, because it showed high levels of ethylene glycol days after Stallings had been allowed to visit her child. In fact, the presumed blood levels of ethylene glycol were not metabolically possible. Stallings' defense attorney did not attempt to obtain an independent review of the lab tests because he claimed was "no evidence" to support an MMA diagnosis.

The limitations of GC-FID impacted wrongful convictions even in cases in which the technique was applied in a valid and reliable manner. In the 1988 Davey James Reedy case, Virginia Department of Forensic Sciences examiner Amy Lawrence "concluded that Reedy's T-shirt and underpants showed the presence of gasoline." A postconviction expert, Dr. David Stafford, criticized Lawrence's use of a liquid gasoline standard instead of an evaporated gasoline standard, but he concluded that the chromatograms were inconclusive, not exculpatory. (Reedy v. Wright, 2002) Governor Terry McAuliffe granted a pardon to Reedy, contending that the original analysis was "discredited [and] inaccurate" but that contention was largely based on the improved probative value of mass spectrometry, not on the validity of Lawrence's chemical analysis. (Bondurant, 2015)

Testimony standards have also improved as scientific research and technological developments improve the foundational basis of forensic disciplines. Improvements have occurred in response to wrongful convictions. Firearms identification provides a "classic" example. In eight cases in the early 20th century, firearms identification was associated with wrongful convictions. The key case involved Nelson Green and Charles Stielow, who were convicted of a double homicide on the basis of bullet comparison testimony by Alfred Hamilton, an untrained examiner. (People v. Stielow, 1916) Firearms experts Charles Waite of the New York Office of the Attorney General and optician Max Poser determined postconviction that Hamilton's analysis was deeply flawed, leading to Stielow's exoneration. (Borchard, 1932) Waite worked with physicist John Fisher, Major Calvin Goddard, and chemist Philip Gravelle to establish the Bureau of Forensic Ballistics in New York City. Gravelle developed the comparison microscope, which is still used in ballistic examinations today. The researchers jointly developed methods for firearms identification that provided an objective, research-based foundation for their discipline. (Goddard, 1926)

More recently, similar improvements have improved the basis for fire debris investigation, pediatric abuse assessment, and DNA analysis. Improved standards were relevant to 20 fire debris cases. In particular, the National Fire Protection Association (NFPA) 921 standard and associated standards have addressed well-established shortcomings in the scientific basis for fire debris reports and testimony, including those presented in wrongful convictions. (Lentini, The Evolution of Fire Investigation and Its Impact on Arson Cases, 2012) In pediatric abuse cases, the American Academy of Pediatrics (AAP) has developed several consensus documents concerning diagnosis and reporting, (Narang, Fingarson, & Lukefahr, 2019) (Committee on Child Abuse and Neglect, American Academy of Pediatrics, 1999) although the guidelines remain controversial. (Papetti, Kaneb, & Herf, 2019) Nonetheless, the updated AAP guidelines would have affected 31 case errors in pediatric abuse cases in the dataset, including cases related to infant abusive head trauma and pediatric sexual abuse. In general, the availability of standards does not guarantee that forensic evidence will be used in a valid and reliable way. In 29 pediatric abuse examinations, a case error would have occurred regardless of the availability of updated AAP standards because the examiner did not follow the best practices associated with the field. In 21 instances, examiners made an error that associated a defendant with the pediatric abuse and did not conform to any standard, even those in place at the time of trial. In 62 cases, the defendants lacked adequate defense, most commonly not having an independent review or examination of the evidence related to a conclusion of abuse. In cases involving subjective analysis, fact-finders require access to valid interpretations that are consistent with both the prosecution or defense theories of the case. In this sense, a valid interpretation is one that is consistent with the accepted standards of the discipline. For these cases, the adversarial deficit contributed to the wrongful conviction despite the possibility that valid interpretations under AAP standards may have been consistent with the defense theory of the case.

A significant subset of cases is associated with unvalidated methods or methods with an inadequate scientific foundation. These cases include instances when an examiner exceeded the standards of the field at the time of trial. The cases also include instances when the field had not developed science-based standards. For example, unvalidated methods were applied in bitemark comparisons and canine detection. An inadequate scientific foundation was used in a very wide range of disciplines when an examiner exceeded the limits of science in making conclusions, reporting results, or in testimony. In many cases, the examiner had been trained on older methods of interpretation and didn't update their approach as standards evolved. Fire debris investigation and forensic medicine practitioners exhibited this "illusion of validity" problem, as detailed in those sections of this paper. (Kahneman & Tversky, 1996)

Best practices and examiner variability

The NIST error typology was presented as part of a 2015 conference on forensic science errors (National Institute of Standards and Technology, 2015) and is included in the coding of cases within this study. The NIST typology is primarily concerned with errors that occur in laboratory processes, interpretations, and reports. Aspects of the NIST typology complement this study's Forensic Error Typology.

Many case errors relate to a failure to follow best practices, processes, or methods. For the purposes of this study, "best practices" include the consensus, documented standards and practices of trained forensic examiners at the time of trial. As described above, many practitioners fall outside traditional forensic contexts or within poorly governed contexts that fail to enforce appropriate standards. The development and enforcement of standards may be the most important way to mitigate the likelihood of examiner errors that contribute to wrongful convictions. In 334 cases, 389 examiners did not follow best practices. Best practice failures were slightly more associated with independent consultants (37%) than public laboratory examiners (33%). Best practice failures were less likely to be associated with examiners from fully-independent laboratories (22%) and examiners with doctoral-level degrees (21%) and more likely to be associated with examiners from organizations that report to law enforcement (45%). It should be noted that these data have significant limitations, because the affiliation and education-level is unknown for many examiners. Also, many public-laboratory-associated best practice failures were associated with seized-drug cases that relied on presumptive field drug test kits. In seized-drug cases, the forensic science organization produced valid and reliable results, but convictions were obtained based on guilty pleas relying on the presumptive tests only. When considering non-drug cases, examiners in public laboratories were associated with best practice failures only 21% of the time. Examiners from organizations that report to law enforcement were associated with best practice failures in 27% of non-drug cases. Also, best practice failures in public forensic science organizations demonstrate a steady decline in this dataset over the last 40 years, as seen in Figure 4. The significant failure of field drug test kits is excluded from this graph, although those cases were classified as best practice failures. The decline may be due to improvements in quality assurance, technology, or accreditation.

Figure 4. Incidence of best practice failures in non-drug examinations in wrongful convictions by year of conviction.

Forensic examiner variability is associated with disciplines in which subjective interpretations cause examiners to reach different conclusions when reviewing the same empirical data. For example, one expert making a post-mortem interval finding may place greater weight on body temperature and rigor mortis while another expert may emphasize the presence or absence of blowfly eggs. In such a case, there is no established standard or scientific research to reconcile the difference of opinion.

Many wrongful convictions associated with forensic examiner variability are also associated with inadequate defense. It is often necessary for a defense attorney to understand the implications when an alternative interpretation of evidence may be exculpatory. This understanding should lead to appropriate cross-examination of a prosecution witness, independent review or examination of the evidence, or other actions that support the defense theory of the case.

Forensic examiner variability and case errors were observed in 82 cases and 105 examinations. Variability was most commonly associated with forensic medicine (26 examinations with case errors), bitemark comparison (21 instances) and forensic pathology (11 instances). Variability was often associated with possible cognitive bias (77 examinations), a battle of experts (8 cases), inadequate scientific foundation (31 examinations), examiner training deficiencies or incompetence (27 examinations), consultants (30 examinations), and highly-educated examiners (68 examinations by examiners with doctoral-level degrees). Examiner variability were not associated with identification or classification errors in 48 cases. In these cases, examiner variability occurred within the range of valid interpretations within practice standards.

Bitemark comparison demonstrates most of these associations. In the 15 cases and 21 examinations associated with variability and case errors among bitemark examiners, 18 of the examiners were consultants and 20 had a doctoral degree (generally associated with dentistry). In the Cristini/Moldowan case, five examiners played some role over three trials and two decades. (Zalman & Windell, 2019) Prosecution examiner Alan Warnick testified that Moldowan made a bite on the victim's neck and Cristini made one on the victim's arm. After hearing that the Warnick conclusions were reviewed by Norman Sperber, examiner Pamela Hammel agreed with the Warnick conclusions. Hammel recanted postconviction after learning that Sperber had not done a review. Warnick denied the Hammel claim concerning Sperber, and the odonatological basis for Hammel's variable conclusions remains unclear. Examiners James Woodward and Raymond Rawson concluded that no bitemark could be associated with the defendants. The defendants were acquitted in their retrials.

Analyst incompetence was associated with 78 examinations, 65 of which were part of the evidence to convict a defendant. In some cases, it is difficult to assess whether the analyst was incompetent, fraudulent, or both. Also, an incompetent examiner may have contributed to many undiscovered errors prior to the discovery of a wrongful conviction. For example, in the 2002 court martial of Roger House, analyst Phillip Mills with the US Army Criminal Investigation Laboratory (USACIL) produced faulty DNA analysis linking the defendant to used condoms associated with the alleged crime. (House v. United States, 2011) Retesting established that House was excluded as a source of any biological material and led to a three-year, $1.4 million investigation that demonstrated that Mills had produced errors in 55% of his casework. (Taylor & Doyle, 2011) In the 1982 Ronald Carden case, Arkansas Medical Examiner, Dr. Fahmy Malak, incorrectly identified the presumed victim on the basis of birthmarks and did not conduct a fingerprint or dentition check. (Masterson, 2015) The remains were then cremated. Later, it was discovered that the remains were misidentified by Malak. The remains were then correctly associated with murder victim Mildred Honeycutt. Honeycutt was murdered by William Perry, who also murdered his wife the month after Carden was convicted of murder. The case also included a faulty hair examination using evidence hair that was not suitable for comparison. Malak was involved in many controversial rulings in Arkansas before he was removed in 1992. (Dake, 2017)

Training deficiencies were associated with 53 examinations, including eight shoe/foot impression comparisons, six firearms identifications, six canine detections, five fire debris investigations, and a wide variety of other disciplines. Untrained individuals may provide forensic analysis outside of their expertise or the controls established in an accredited forensic laboratory setting. For example, the six firearm identification cases arose from untrained examiners in cases in the early 20th Century and a university-based examiner who applied compositional bullet lead analysis in the Philip Cannon case.

Incompetent forensic examiners may engage in fraudulent activity to cover their mistakes. For example, Joyce Gilchrist appears to have altered or destroyed evidence in cases to prevent independent examinations or review. (Federal Bureau of Investigation, 2001) She was also reprimanded by two forensic associations in 1987 and 2000 for unethical conduct and giving unreliable testimony.

Forensic examiners may alter findings in light of biasing information from investigators. It should be noted that these changes may often be the result of *conscious* decisions by incompetent examiners, not unconscious bias. This phenomenon was demonstrated during the New York State Police fingerprint scandals of the 1980's and 1990's. (Roth, 1997) In those cases, untrained and uncertified examiners planted evidence to implicate defendants that they were convinced were guilty. Examiners David Harding and Robert Lishansky argued about the fabricated prints in the Shirely Kinge case because each man was concerned about the "glory" associated with solving the case. One planted "evidence" print was credited to Harding, while the other was credited to Lishansky. The fraudulent work was discovered when Harding applied for a federal job and boasted about the fabrication of evidence while with the state police. That said, there were elements that were discernable at the time, including poor documentation, poor evidence management, and destructive alteration of evidence. In addition, these shortfalls were noted by investigator David McElligott, who questioned the authenticity of the prints *before the Kinge trial*. Also, certified examiners Linus Rautenstrauch and Martin Hughes noted the crime scene and evidence irregularities, but McElligott did not sufficiently follow up on their concerns.

Forensic errors are often associated with examiners at the extreme ends of professional expertise. As expected and observed many times in the data set, untrained individuals produce unreliable results. Also, highly trained experts may advocate for findings that are "on the cutting edge" but not reliable. Bitemark examiners, early DNA scientists, and forensic pathologists are all doctoral-level experts that have provided erroneous forensic testimony. Typical, well-trained forensic scientists may be more reliable because they employ established methods and standards. The "cutting-edge" effect may be observed in cases in which "moral panic" or similar considerations skew expert judgments. (Grometstein, 2008) In the 1997 death of a three-month-old child, Iowa State Medical Examiner Thomas Bennett concluded that the child's parents had shaken the baby on the basis of brain and eye hemorrhages. (Siegel, Judging Parents as Murderers on 4 Speck of Blood, 1999) Bennet said the shaking would have lasted no more than a second or two. Postconviction review by forensic pathologist Jerry Jones found that there was minimal or no blood in evidence photographs and no basis to conclude that the child had been abused in any way. Bennett said later, "Not all shaken babies are going to have visible bleeding behind the eyes. You need 18 hours after the injury to see a subdural hemorrhage develop. But the kids aren't living that long. That's why you don't see those hemorrhages.....History moves slowly and is very cruel to those who try to move fast." Reportedly, after the case was dismissed, he told the county attorney, "My job is to stick my neck out for you, and here you are stabbing me in the back."

Cognitive bias

Extensive research and commentary have supported the claim that cognitive bias plays a major role in forensic science errors related to wrongful convictions. (Bonventre, 2021) Most research has focused on the influence of contextual information on the reliability and biasability of pattern evidence examiners. (Cooper & Meterko, 2019) Most notably, the erroneous latent print identification of Brandon Mayfield raised concerns about confirmation bias and related cognitive bias effects. (Stacey, 2004) Mayfield was never brought to trial or wrongfully convicted. Observers warn that the forensic science community has ingroup biases that prevent an accurate understanding of the relative role of forensic science errors in wrongful convictions. (Scherr & Dror, 2021) Although hundreds of research studies and academic reviews have been devoted to this topic, few studies have attempted to connect theories concerning cognitive bias to data from wrongful convictions. (Lentini, Anatomy of a Wrongful Arson Conviction: Sentinel Event Analysis in Fire Investigation, 2014)

There are difficulties in the assessment of cognitive bias in wrongful convictions because there are significant uncertainties in the retrospective documentation of the basis for an individual's conclusions. In the current study, coding was based on the assessment of "possible cognitive bias," meaning that some documentary evidence supports the possibility that cognitive bias influenced the examiner's conclusion, reporting, or testimony. The consideration of cognitive bias was not limited to contextual bias. An examiner may have exhibited one of many types of cognitive bias, or "predictable deviations from rationality," as suggested by Croskerry and others in studies of diagnostic failures in clinical medicine. (Croskerry, Singhal, & Mamde, 2013) There was no attempt to code subcategories of cognitive bias. Thus, the cases associated with "possible cognitive bias" represent a population of cases that can be studied because bias effects could be documented in the case history, not necessarily the entire population of cases in which bias may have been present or a systematized coding of bias types.

All types of bias were coded as “possible cognitive bias,” if that could be documented and regardless of the bias type or origin. Possible cognitive bias may have included any type of bias, but bias was not coded when there was no objective basis in the documentation to support the claim. In some cases, examiners themselves made statements that support the possibility of bias. In other cases, contextual information or collegial influence can be documented as an influence on the examiner. The coding was necessarily subjective, and other observers might reasonably differ in their analysis of individual cases.

In disciplines that make source and sub-source conclusions, examples of cognitive bias include target bias (such as “teasing the points” in latent print examination) or authority bias (in which a non-blind verification inevitably confirms the original analysis). For example, these issues were documented by the Department of Justice review in the Mayfield case, which included target bias effects attempting to fit the Mayfield print to the evidence print and authority bias related to the failure to perform blind verifications. (110) Similar examples of these issues in wrongful convictions are described in the section on latent print comparison .

In disciplines that make activity-level conclusions, the effects may be more complex and context-dependent, but the context may relate to the case, the opinions of other experts, or the Semmelweis reflex (holding to established norms from training or colleagues). For example, the problem of established norms contributed to many wrongful convictions associated with fire debris investigation. As described in that section, many investigators relied on their training prior to the adoption of NFPA 921 well after the adoption of that standard in 1992. Errors may also arise due to deference to the opinions of colleagues, as seen among the forensic pathologists in the Souter case described in this section below. Base rate bias can also be observed in some cases, such as the Monroe case described in the crime scene investigation section.

Cognitive bias factors were most prominent in disciplines that require consideration of contextual information and use subjective interpretation frameworks at the conclusion level—such as forensic pathology and fire debris investigation. The clear and open communication of forensic results is necessary to prevent errors by police investigators and other criminal justice system actors, who may discount or ignore exculpatory forensic results due to their own biases. The effective use of reliable forensic evidence may play a critical role to reduce the risk of future wrongful convictions.

Possible cognitive bias was observed in 348 examinations, including 275 examinations that contributed to a wrongful conviction. There was significant variability among disciplines. Possible cognitive bias was associated with no seized-drug analyses, latent palm print comparisons, and toxicology analyses and very few fire debris chemical analyses or DNA analyses. Other disciplines were much more closely associated with possible cognitive bias, including canine detection (15 out of 17 examinations), bitemark comparison (30 out of 44 examinations), fire debris investigation (27 out of 45 examinations), forensic medicine (55 out of 124 examinations), and forensic pathology cause and manner determinations (36 out of 136 examinations). In some respects, these results are not surprising, because forensic disciplines with the greatest exposure to contextual information tend to be most closely associated with possible cognitive bias, while laboratory-based analyses tend to less closely associated with possible cognitive bias. This general trend is seen in the respective figures for other laboratory-based methods: latent fingerprint analysis (14 out of 87 examinations), serology (32 out of 204 examinations), and hair comparison (27 out of 143 examinations).

Disciplines that rely on subjective interpretation frameworks are vulnerable to cognitive bias, as seen in the 110 forensic pathology or forensic medicine examinations associated with both case errors and possible cognitive bias. The associated forensic errors may align with cognitive bias frameworks other than confirmation bias. In many cases, these forensic experts exhibited congruence bias, in which alternative hypotheses are discounted or ignored even when contradictory information is present. These cases may also be accompanied by deference bias to a colleague. When one expert makes an initial determination, their colleagues may be reluctant to disagree and produce congruent interpretations that may not be supported by the evidence.

In the Larry Pat Souter case, neuropathologist Dr. Steven Bauserman initially made a valid interpretation that the victim's injuries could have been caused by a broken whiskey bottle found on the road near the body or by being hit by a car. (Larry Pat Souter, Petitioner-Appellant, v. Kurt Jones, Warden, Respondent-Appellee, 2005) The Michigan state police crime laboratory determined that glass particles from the victim's body were inconsistent with automobile headlight glass and were not brown in color like the bottle, producing an ambiguous set of results for the investigation. The Newaygo County medical examiner, Dr. Ronald Graeser, issued a report which stated the injuries "may well have been inflicted" by the whiskey bottle. After the case then lay dormant for eight years, Graeser produced a more definitive report, saying the injuries were caused by the bottle and that it was "virtually impossible" that it could have been caused by a car. Graeser had been trained by Bauserman and another colleague, Dr. Stephen Cohle, who both now testified that the victim's injuries were "consistent with" being struck by the bottle. In this case—unusually—there was a defense expert, Dr. Lawrence Simpson, who had been consulted by the police during the initial investigation and testified that the bottle could not have caused the injuries. Postconviction, Bauserman and Cohle recanted their trial testimony. Cohle was the only certified forensic pathologist among the three prosecution experts. He later stated that he was "strongly influenced" to support Graeser's interpretation. Among other issues, Graeser had told Cohle incorrectly that the bottle had a sharp edge at the time of the death but it had lost the edge over the years. Bauserman stated that his opinion that the bottle caused the wounds was speculation and then defered to Cohle as to the true cause. Bauserman and Cohle also alleged "deficiencies in Dr. Graeser's education and training." Souter's conviction was overturned, and other information pointed to the likelihood that the death was caused by a hit-and-run driver. At various points in the proceedings, Graeser, Bauserman, and Cohle all changed their view of the injuries on the basis of biases that had little or nothing to do with the medical or physical evidence in the case. The Souter case is an extreme example, but it is indicative of a broader set of issues in subjective interpretation disciplines.

It is notable that the eight fingerprint case errors were associated the presence of exculpatory latent print evidence that was discounted or ignored by police. Police investigators, prosecutors, and defense attorneys exhibit choice-supportive bias when discounting exculpatory forensic results, a phenomenon seen in more than 60 wrongful conviction cases overall. In other words, investigators may prefer to maintain their tunnel vision about their theory of a case and discount contrary information as unimportant or flawed.

In many wrongful conviction cases, forensic science failed to prevent miscarriages of justice by correcting false theories developed by police investigators. (Cole S. A., Forensic Science and Wrongful Convictions: From Exposer to Contributor to Corrector, 2012) This failure was often due to miscommunication or misunderstanding of forensic science results. The clearest example may be the DNA evidence in the Maurice Patterson case. (Bluhm Legal Clinic) (Mills, 2013) The victim in the case, Robert Head, was stabbed to death, and two knives were seized at the crime scene. Patterson and another man, James Starkey, were identified as possible suspects. The commercial DNA testing lab found that one of the knives had blood from both victim Head and suspect Starkey, but the report from the Illinois State Police crime laboratory only mentioned Starkey as a source. Because it was known that Starkey had been injured during the incident, the detective and prosecutor assumed that the knife was the weapon that injured Starkey and was not the weapon that killed Head. There was no claim that the miscommunication was anything but an honest mistake. It appears that the laboratory did not appreciate the context of the case and therefore only reported the presence of blood from any suspects, not the victim. Meanwhile, the investigators did not clarify the issues with the crime laboratory to fully understand the result that found Starkey's blood on the knife.

The Patterson case—and many others like it—demonstrate that forensic evidence can only prevent wrongful convictions if contextual information is well-understood by the forensic professionals in a case. This does not obviate the need for fingerprint examiners (and other forensic analysts) to be concerned with confirmation bias, target bias, and task-irrelevant information. Instead, it supports the need for forensic professionals who play a primary role in the communication of information within the laboratory and with investigators and fact-finders outside the laboratory. Worries about contextual information exposure in forensic analysis should not prevent the appropriate prioritization of evidence, communication of results, and coordination with investigators to ensure the full implications of exculpatory or unexpected results are fully understood.

<u>Forensic disciplines</u>

Serology

204 serology examinations
- 190 cases in which serology was used
- 139 examinations with errors of any type ("case errors"
- 42 case errors contributing to conviction

System errors
- 95 instances: inadequate defense
- 28 cases: exculpatory evidence ignored or discounted
- 17 cases: police or prosecutor misconduct

Forensic errors
- 54 identification or classification errors
- 90 testimony errors
- 25 evidence collection, contamination or chain of custody errors
- 51 failures to maintain standards/best practices

Serology includes a wide range of methods, including blood typing and bodily fluid analysis. (Gaensslen, 1983) Before the routine use of STR-based DNA profiling in the late 1990's (CODIS and NDIS Fact Sheet — FBI), biological evidence was routinely analyzed using serological blood typing and related methods. The scientific foundations of forensic serology were established in the early part of the 20th Century. By the 1970's, the field had developed standard analytical and interpretation frameworks, including a clear and robust approach to statistical characterization. (Gaensslen, 1983) Serology remains in use in forensic laboratories as a screening tool, especially with respect to the identification of the type and amount of bodily fluids present in a sample.

There were 139 case errors related to serological examinations, but only 42 cases in which that evidence was required for conviction. Many serology examinations were weakly probative in sexual assault cases, and the case outcomes depended primarily on victim testimony. Thus, although there were numerous serology errors, only 30% of these errors were relevant to the balance of evidence to convict the defendant. Many serology examinations were associated with errors that arose prior to courtroom testimony. In 51 cases, examiners did not follow standards or best practices. In many cases, the error may have preceded the laboratory, such as the 23 cases with 25 examinations in which evidence was contaminated or mishandled. In another 24 instances, examiners did not demonstrate competence in their analysis. For example, Oklahoma examiner Joyce Gilchrist caused errors in six connected cases ("The Beatrice Six") after she failed to accurately type a reference sample from the actual perpetrator, who was then released.

In general, because serology produces results of limited probative value, it was used primarily to eliminate suspects. For example, in the case of Troy Webb, forensic analyst David Pomposini found foreign A blood group substances in rape kit swabs that could not have been contributed by the defendant, who was a nonsecretor. (Pomposini forensic testimony, 1989) Instead of excluding Webb, Pomposini instead made an unfounded speculation that there were "two or more seminal fluids present in that mixture." Further, he continued to evade the issue on cross-examination, which attempted to establish the exculpatory nature of the serology. There was no basis in the serology or case evidence for multiple assailants or a consensual male contributor to the sample.

Serology was used to imply an individual source in only three cases, including two cases associated with Fred Zain, whose work was discredited by an ASCLD-LAB investigation in 1993. (McNamara & Linhart, 1993) In other cases, the testimony narrowed the class of possible sources to a significant degree that was could be considered highly probative but did not rise to the level of an individualization. For example, in the 1990 Mark Bravo case, examiner Richard Catalani narrowed the possible sources to 3% of the population because Bravo and the actual perpetrator shared an A blood type and an unusual PGM marker. Although Catalani's laboratory analysis produced valid results, his testimony divided the 3% figure in half to eliminate female contributors. That invalid interpretation was elicited by the prosecutor's direct examination, but the responsibility for the error rests on the forensic examiner, who failed to represent the science correctly in response to the prosecutor's questioning.

The examiner did not make an error in many serology cases in which another individual made a case error. These examinations overlapped to a great extent with errors from officers of the court. For example, there were 28 defendants who should have been exculpated by serology results but were nonetheless wrongfully convicted. In part, this issue was closely related to nine cases in which probative serology results were not reported to the defense and 24 cases in which misleading serology information was communicated to the defense. In some cases, the defense failed to recognize the value of exculpatory or potentially exculpatory evidence when it was clearly probative in the circumstances of the case. To some extent, these issues were inherent to serology, because foreign blood markers could arise from a variety of factors, including contamination, consensual partners, or test variability. Ideally, examiners would be consistent in their interpretations and would provide consistent explanations of these limitations without regard to whether the serological results were inculpatory, exculpatory, or non-probative.

In 18 cases, a prosecutor clearly misrepresented or suppressed serology results, but the extent of that problem may be underestimated in the current study. Clearly, the prosecution has a responsibility to disclose information about forensic results, but there is less clarity concerning the responsibility of the prosecution to disclose interpretations of the evidence that may support the defense theory of a case. The coding assumes that forensic scientists have a greater obligation in this regard. Hence, there were 23 instances in which serology results were poorly communicated in a forensic report and 33 instances in which communication to the defense was lacking. Poor communication included nine cases in which probative serology results were not reported to the defense and 24 cases in which misleading serology information was communicated to the defense.

Qualitatively, it appears that prosecutors demonstrated better understanding of serology and other forensic evidence than defense lawyers. They were more likely to present the scientific limitations of serology when the evidence was inconsistent with the prosecution theory of the case. For example, there were many cases in which microbial contamination may have produced spurious results in ABO immunological analyses. Some prosecutors demonstrated their understanding with clear questions designed to elucidate the issue and undermine any interpretation of the results that might have been exculpatory. Similar performance by defense lawyers seems to have been much less common. It is possible this difference was due to the closer relationship between forensic scientists and prosecutors, but that is impossible to discern from the data in the current study.

Serology had a well-established scientific foundation throughout the study period (1956 to the present for serology cases). Nonetheless, many prosecutors, defense attorneys, and judges exhibited a poor understanding of the implications of serological evidence in particular cases, even making statements to that effect during trials. Although scientific and technological improvements have improved the reliability and probative value of forensic science, the impact includes additional complexity that may not be well-understood by non-scientists. Wrongful convictions associated with serology demonstrate that the reliability of forensic science may depend on the education of legal professionals about the limitations and interpretation of forensic evidence.

Of 32 serological analyses associated with possible cognitive bias, 29 were also associated with inadequate defense. This linkage was often closely related to gaps in scientific understanding. In essence, a serologist would present a questionable or invalid interpretation consistent with the prosecution theory of the case, but a defense attorney would not provide adequate cross-examination to challenge the testimony. In the 1986 Ernest Sonnier case, the defendant was a B secretor, but no B blood group substances were found in the seminal fluid from the sexual assault kit and the victim's jeans. Examiner David Coffman speculated that the victim "flushed out" the markers during menstruation. (Possley, Ernest Sonnier, 2018) Although the defense established that both attackers had ejaculated, they were unable to challenge Coffman on the research establishing that spermatozoa can be found even when coitus has occurred before menstruation. (Morrison, 1972) Sonnier was exonerated postconviction by DNA and latent print matches to an alternate suspect.

Evidence handling may compromise biological evidence on account of contamination or other consequences, as seen in 25 cases involving serological evidence. In the Charles Fain case, autopsy washings were not stored in a refrigerator or freezer, causing bacterial contamination that compromised the serological analysis. The case was also compromised by the failure to deliver probative autopsy swabs to the crime laboratory for testing. (State v. Fain, 1989) The Gregory Taylor case involved uncertainties about a link between the victim's vehicle and Taylor's vehicle. (State v. Gregory Flynt Taylor, 2009) These uncertainties led one investigator to return to the crime scene at night to spray extensive amounts of luminol in the hopes of reconstructing the events the led up to the murder. Later, phenolphthalein test results were highly variable, in part because luminol may compromise confirmatory such tests in the laboratory. (Luedeke, Miller, & Sprague, 2016) It is unclear if the scene investigator communicated the extent of luminol usage at the scene to the laboratory. In the Paul Kordonowy case, the presence of A antigens could have been exculpatory because the defendant was an O secretor. Examiner Julie Long noted that the victim's consensual partner was an O secretor and discussed possible bacterial contamination because the samples had been stored for two years in a moist environment. (State v. Kordonowy, 1991) Oddly, a postconviction analysis by Edward Blake held that Long's testimony was in error, stating, "There is no evidence whatsoever that bacteria produce water soluble ABO antigens of any sort much less ABO antigens of type "A"." (Blake E. , Montana v. Paul Kordonowy Report 1 File 03-128, 2003) This was misleading, because it was well-established that bacterial contamination could produce false positives in ABO immunoassays. (Culliford, 1971) Bacteria produce proteins that cause false positives on ABO immunoassays, a phenomenon that is quite distinct from any contention that bacteria produce ABO antigens.

In the case of Jerry Watkins, forensic analyst Carol Kohlman relied on autopsy findings that established the likelihood of bacterial contamination based on ethanol levels in the toxicology report and the fact that the body had been found five days after the probable day of the murder. (Forensic Testimony Transcript, 1986) Forensic pathologist John Pless testified that the body had a level of alcohol "consistent with what we see in bodies that have been dead in excess of 48 hours" due to bacterial fermentation. Further, Kohlman reported that the P30 test failed so that she could not confirm the presence of a male fraction in the sample. No competent forensic serologist would have attempted to include or exclude any suspect, although other research has argued incorrectly that Kohlman's contamination interpretation was without basis. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009)

In 30 cases, laboratory analysis errors led to errors in forensic reports or testimony. The Gary Dotson case was the first DNA exoneration in the United States. The Dotson case is unusual because the original serology was closely reviewed and repeated by both prosecutor and defense experts, as detailed in the Blake report on the case. (Blake E. , FSA File No. 85-035, 1985) Illinois State Police serologist Mark Stolorow reexamined the serology in 1985, as did Edward Blake later that year. The original examiner, Timothy Dixon, did not examine or report sperm concentrations or any other method to account for masking. Postconviction, it was established that the evidence samples were entirely from the alleged victim and her consensual partner, not Dixon. The partner was an O secretor, but his H antigens were not observed by Dixon. Stolorow confirmed the male fraction in the sample using microscopic confirmation of spermatozoa and detected B and H antigens consistent with the alleged victim and her consensual partner. As Blake correctly noted and Stolorow implied, Dixon had used low-quality H-lectin that had not been subject to appropriate checks and control testing. (Blake E. , FSA File No. 85-035, 1985) Blake and Stolorow agreed that the Dixon error primarily related to poor quality assurance and poor documentation.

In 39 cases, the laboratory error involved inadequate reference testing to support the interpretation. For example, the Bromwich report found that the Houston Police Department

Crime Laboratory often failed to obtain reference samples from victims, a practice which required them to assume the blood type or secretor status of the victim based on the evidence sample. (Bromwich M. R., 2007) In Georgia, the 1983 Calvin Johnson conviction relied on serology and other evidence. The serology accounted for masking using microscopic confirmation of spermatozoa with appropriate interpretation limitations clearly delineated on cross-examination but did not include necessary reference testing of the victim's consensual partner. (Tillman, 1983)

Even when laboratory work was valid and reliable, a misinterpretation could lead to a forensic error. In the William Barnhouse sexual assault case, the serologist detected spermatozoa but concluded the serological testing was inconclusive. (Barnhouse v. City of Muncie, 2020) The serologist failed to account for Barnhouse's congenital Klinefelter's Syndrome, which prevented him from producing spermatozoa. Therefore, he should have been excluded as a contributor to the biological evidence. The case also included a hair identification error.

In the 1990 Marvin Mitchell case, examiner David Brody found only blood group substances from an O secretor on a mixed blood and semen stain. (Mitchell v. City of Boston, 2001) Mitchell was an A secretor and the victim was an O secretor, and Brody did not exclude Mitchell. Brody did not provide any discussion of bodily fluid testing, microscopic confirmation of spermatozoa, or quantification with ACP or P30. Had he done so, Mitchell could have been excluded. On cross-examination, the defense did not challenge Brody's laboratory testing or interpretation but did require Brody to repeat that Mitchell's A blood group substances were not present. (Commonweath v. Marvin Mitchell, 1990) Brody also testified in the Neil Miller case. In that case, he clearly delineated the blood group substances found in the various biological evidence samples and referenced microscopic confirmation of spermatozoa on smear swabs. Miller found foreign B antigens but did not speculate on that topic, though the prosecutor made an unsubstantiated claim that it came from a consensual partner.

It should be noted that the Garrett-Neufeld study (GN) found Brody's analysis faulty in both the Mitchell and Miller cases on the basis of "masking." (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) This study's findings diverge from GN, which did not find that microscopic confirmation of spermatozoa was a sufficient basis to determine a male fraction was present in a sample and account for masking by the victim's blood group substances. As discussed above, microscopic confirmation of spermatozoa is the optimum method for confirmation of the male fraction in a biological evidence sample. Other methods, including the use of P30, are useful when spermatozoa are not present. In 52 cases, GN attributed a masking or serological analysis error when the forensic examiner had used appropriate methods to account for masking and interpret the serological results. Cases not covered elsewhere in this paper are detailed in Table 3. Although these cases generally reflect appropriate consideration of masking, they do implicate other forensic and system errors.

Table 3. Analysis of masking and other serology errors in wrongful conviction cases.

Case and reference	Masking interpretation	Other serology error
Herman Atkins **(People of the State of California vs Herman Atkins, 1988)**	Microscopic confirmation, ACP; addressed masking in testimony	Failed to account for ABO nonsecretors with PGM 2+1+, 2+, or 1+. Cross-examination did not address serology.

Chester Bauer **(State of Montana vs. Chester Bauer, 1984)**	Microscopic confirmation, ACP, P30, and choline; addressed masking in testimony.	False probability when dividing by two on basis of gender fraction.
Orlando Boquette **(McNamara forensic testimony, 1983)**	Victim was nonsecretor, so masking only relevant in stains containing blood. Seminal fluid stains showed no BGS (omitted in GN paper).	Poor testimony stated Boquette excluded as contributor of A BGS. Implied that failure to confirm semen in two samples meant nonsecretor was included.
Mark Bravo **(Catalani forensic testimony, 1990)**	Masking not addressed by examiner (GN omission).	False probability when dividing by two on basis of gender fraction.
Dennis Brown **(Brown v. Miller, 2008)**	Forensic testimony not available for review. Appeals court ambiguous on masking issue.	Examiner told police investigator that Brown was "positively identified" by ABO testing.
Victor Burnette (Garrett B.)	Forensic testimony not available for review. Analyst clarified that Burnette was nonsecretor and could not be excluded.	No error can be documented.
Kevin Byrd **(Bolding forensic testimony, 1985)**	Microscopic confirmation, ACP; limited testimony to nonsecretors.	Failed to collect victim reference sample to determine secretor status.
Roy Criner **(Criner v. State, 1992)**	Clarified all masking considerations; did not detail quantification in testimony.	Did not account for nonsecretors. Blood on Criner screwdriver not analyzed.
Alan Crotzer **(Wilbarger forensic testimony, 1982)**	Microscopic confirmation of spermatozoa	False probability when dividing by two on basis of gender fraction. Failed to account for PGM 1.
Charles Dabbs **(Dabbs. v. Vergari, 1990)**	Microscopic confirmation, with summary quantification stated in testimony; ACP. Victim was nonsecretor; no masking issue.	Failed to collect necessary elimination samples.
Dwayne Dail	Analyst testified Dail not inculpated because of uncertainties in quantification.	No error

(Milks forensic testimony, 1989)		
Willie Davidson (Testimony of Mary Jane Burton, 1981)	Microscopic confirmation	Judge misinterpreted testimony.
Gerald Davis and Dewey Davis (McNamara & Linhart, 1993)	P30 confirmation	Lab results falsified.
Frederick Daye (Testimony of Annette Peer, 1984)	Microscopic confirmation, ACP; clarified masking issues on cross-examination.	No error
Luis Diaz (Testimony of George Borghi, 1980)	Microscopic confirmation, ACP; clarified masking issues.	No error
Alejandro Dominguez (Testimony of William Wilson, 1990)	Microscopic confirmation, P30; clarified masking issues.	Poor communication during testimony
Lonnie Erby (State v. Moore, 2013)	No transcript available to substantiate claim of error.	No error
Dennis Fritz, Ronald Williamson (Williamson v. State of Oklahoma, 1991)	Microscopic confirmation but failed lab tests.	Judge failed to provide funds for independent review. Examiner may have made testing errors.
Larry Fuller (Testimony of Bruce Carr, 1981)	Victim was nonsecretor, so masking not relevant.	Poor communication and interpretation of serology; incorrect statistic.
Anthony Green (Testimony of Joe Serowick, 1988)	ACP, P30.	Failed to determine victim secretor status or possibility of victim contribution to sample.
Clarence Harrison	Microscopic confirmation. Clarified masking issues.	No error
Kenneth Ireland (Bloss & McElligott, 2014)	No transcript available. Microscopic confirmation used but no ABO BGS found, so masking not relevant.	No error
Calvin Johnson	Microscopic confirmation. Masking issues clarified.	Failed to conduct elimination testing

Ronald Jones	Microscopic confirmation. Masking issues clarified.	No error
Paul Kordonowy	Microscopic confirmation. Masking issues clarified.	Evidence storage inadequate
Carlos Lavernia	Microscopic confirmation. Masking issues clarified.	No error
Leonard McSherry **(People v. McSherry, 1992)**	No transcript available. Masking not a relevant issue due to fecal contamination in underwear.	Contamination correctly used to discount ABO response; GN omitted this information.
Brandon Moon **(Stolorow & Maddox, 2011)**	Microscopic confirmation, P30; masking addressed in interpretation.	Postconviction review established that laboratory testing was invalid.
Donald Reynolds, Billy Wardell	Microscopic confirmation.	Incorrect interpretation that assailant was O secretor; misleading testimony
LaFonso Rollins	Stipulation only. Masking not relevant because serology control sample failed.	Poor communication with defense concerning serology
Peter Rose	P30, ACP; masking issues clarified. Failed control presented appropriately.	No error
James Tillman	Microscopic confirmation, ACP; masking and other issues clearly presented in complex case.	Poor evidence handling may have contaminated evidence.
Patrick Waller **(National Registry of Exonerations, 2019)**	No transcript available.	Mathematical error. GN and NRE do not clarify that PGM marker may arise regardless of ABO secretor status.
Earl Washington	Microscopic confirmation. Detailed forensic report but no test method details.	Inadequate defense; Exculpatory serological results not recognized.
Joseph White		
Willie Williams	Microscopic confirmation. Clarified masking issues, presented valid statistics.	No error

Anthony Woods	Microscopic confirmation. Clarified masking issues.	No error

In 18 cases, the GN masking interpretation is supported by the case history. In the case of William Harris case, GN correctly concluded, “The testimony ignored masking and falsely divided the Invalid statistic in half.” The victim, defendant, and evidence all showed the same serological profile: type O (or H antigens) and PGM 1+. (Zain forensic testimony, 1987) Forensic analyst Fred Zain concluded that only O secretors with PGM type 1+ could be included among possible male contributors on the unsupported assumption that all of the samples were pure seminal fluid. In addition, he divided the population of possible contributors in half “being the combination if they were males” and further intimated that the population was even smaller because only 3% of West Virginians were African American. He had no basis in the serology to make that distinction. The cross-examination failed to clarify these issues.

In 27 cases, GN made an incorrect determination of a masking error. In some cases, the victim was a nonsecretor, so the masking issue did not apply. In the 1984 Carlos Lavernia case, analyst Patrician Hulin with the Texas Department of Public Safety testified that sperm was observed on a rectal swab and a stain on shorts. (Lavernia v. Lynaugh, 1988) The victim was an O secretor and Lavernia was an O nonsecretor. She found H blood group substances on the shorts and concluded that he could not be eliminated as the contributor of the sperm. She further stated that any nonsecretor or O secretor could be included, which included 58% of the population. (Hulin forensic testimony, 1984) Hulin applied the appropriate techniques and interpretation available to her in 1984 in making this conclusion. Her testimony clearly delineated the full range of males who were consistent with the stain’s serological profile.

In the 1985 Ronald Jones case, GN claimed that masking issues meant that “no donors could be excluded” from vaginal swab evidence that exhibited A antigens. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) (Garrett B. , Convicting the Innocent: Where Criminal Prosecutions Go Wrong, 2011) The victim was a Type A secretor and Jones was a nonsecretor. Garrett states, “A competent analyst would have explained that any man could have been the rapist.” In her testimony, analyst Mary Ann Furlong testified that she obtained a positive acid phosphatase indication on vaginal, oral, and rectal swab but was able to confirm spermatozoa microscopically only on the vaginal swab. (Furlong forensic testimony, 1989) She clearly delineated the masking issue and further clarified, “That would also be saying that there was enough semen there to detect a mixture which, you know, when semen is deposited it is going to be diluted somewhat by the vaginal fluids also.” This statement and her statistical analysis were clearly based on appropriate and reliable testing and interpretation frameworks.

Many serological errors were identified with a small number of examiners and organizations. In the Nebraska Beatrice Six case, the actual perpetrator had been identified as a suspect in Nebraska but was arrested in Oklahoma. (Joseph E. White v. Richard Smith et al, 2011) Oklahoma examiner Joyce Gilchrist failed to accurately type the suspect’s ABO serology, and the individual was then released. The Nebraska police then pursued an alternate theory that resulted in six wrongful convictions. Gilchrist made errors in four other cases in the dataset, but the serology was either exculpatory or minimally probative. In the 1983 David Bryson case, Gilchrist’s hair and serology testimony inculpated the defendant, but the postconviction reviews clearly established that her work was fraudulent in the case. (Bryson v. Macy, 2009)

In the Jeffrey Todd Pierce case, Gilchrist stated her opinion that the semen donor was an O-secretor, like [Pierce].” (Pierce v. Gilchrist, 2007) In fact, contrary to Gilchrist’s testimony, no spermatozoa was found when the samples were reexamined in 2001. Also, no confirmatory P30 testing was performed. Gilchrist’s work was thoroughly reviewed by SERI. (Wraxall & Fedor, 2001) They found that Gilchrist mistyped Pierce as a PGM 1, when he was actually a PGM 2-1. Gilchrist failed to confirm the male fraction in the samples with *either* microscopic spermatozoa or P30.

The Chicago Police Department Crime Laboratory was associated with 17 serology case errors, including six associated with examiner Pam Fish and five with examiner Michael Podlecki. In the 1986 Ford Heights Four case, all four defendants (Marcellius Bradford, Calvin Ollins, Larry Ollins, and Omar Saunders) and the victim were nonsecretors. (Ollins v. O'Brien, 2005) Separately, examiner Fish associated an evidence hair with Saunders. Bradford pled guilty, the Ollins brothers were convicted in a joint trial, and Saunders was convicted in a separate trial. Fish’s testimony included false statements that were evident in the 1988 trials:

1. She testified that she had done “hundreds of thousands of genetic marker determinations in her career.

2. She assumed that there were multiple attackers and used the multiple court venues to imply that markers foreign to the defendants were associated with the other alleged assailants who were not present. As non-secretors, none of the defendants could have contributed the ABO blood group substances.

3. She stated that 50% of African Americans are secretors. No figure below 70% is supported by the research literature.

4. She refused to admit on cross-examination that Calvin Ollins was an O nonsecretor on the basis that his blood group substances were not found in his saliva. Instead, she stated that she could not determine his secretor status from his blood, a true but evasive answer.

5. She testified that the semen on the vaginal swab “had to be deposited by Larry Ollins plus at least one other person”, implying that she had made a conclusion that Larry Ollins was a contributor to the semen on the vaginal swab.

As part of a postconviction review, defense expert Edward Blake presented a detailed analysis of this case. (Blake E. , January 9, 2001) In addition to the testimony, Blake had access to the Fish lab reports of 10/30/1986 and 10/20/1987 that detailed her laboratory analysis. Blake labelled Fish’s work in this and other wrongful conviction cases “scientific fraud.” That said, Blake made two errors in his analysis of Fish’s work. First, he maintained that the serology implied that there were multiple assailants because “A PGM test result that reveals three PGM alleles which could not originate from a female victim demonstrates the presence of at least two semen sources.” This interpretation failed to note that two of the PGM markers in the vaginal swab (2- and 1+) were associated with the victim, so only one PGM marker from the vaginal swab (the 1-) could be associated with the assailant because of masking. In fact, there were indeed two assailants, but their serological profiles are unknown. (Ollins v. O'Brien, 2005) Further, Blake concluded that “even Saunders cannot be eliminated as a potential contributor to the vaginal swab specimen.” That is incorrect, because Saunder’s PGM profile, 2+2+, was not observed in the vaginal swab. As a result, Saunders was the only defendant that could be conclusively excluded from the vaginal swab. It is also the case that the prosecution theory of the case was not consistent with the H antigens in the vaginal swab, because none of the Ford Heights Four were O secretors and the victim was a nonsecretor. There had to be an O secretor source among the semen contributors, and none of the defendants fell into that classification.

In short, Fish and Blake committed serious errors in their analyses that were likely heavily influenced by their preexisting biases about the case. Fish was aware of the implications of the H antigens in the vaginal swab, the nonsecretor status of the defendants, and her conflicting misrepresentations across multiple trials. The case was based on a fraudulent set of reports and testimony. It demonstrates the need for quality assurance mechanisms that detect and prevent false reporting by incompetent or fraudulent examiners.

In 64 cases, valid serological findings were presented, including 11 cases in which the serology was part of the balance of evidence to convict the defendant. With rare exceptions, all serological profiles can be associated with a class of sources, so it is possible for serology to implicate an innocent suspect who shares the same profile as the actual source. In the 1991 Ben Salazar case, examiner Devane Clark determined that the perpetrator must have contributed B blood group substances and a PGM 2- marker to the evidence. (Clark forensic testimony, 1992) Three other PGM markers (1-, 1+, and 2+) could have been contributed by the victim or her husband. The victim had testified that the assailant was a Hispanic male (and had also identified Salazar). Clark relied on the Serological Research Institute's Blood Stain Analysis Manual that established that the profile could be associated with only 2% of the Hispanic population. Salazar was exonerated by postconviction dqAlpha and polymarker DNA testing, implying that he was extremely unfortunate to have shared an unusual serological profile with the unknown perpetrator. (National Registry of Exonerations, 2019) In their 2009 review of DNA exonerations, GN classified the testimony in this case as a "probability error" because Clark did not include consideration of AB secretors explicitly in his testimony. The GN classification is in error, because Clark clearly described the considerations in his analysis, provided a complete analysis in his forensic report, and referenced B secretors, which includes AB secretors. More importantly, AB secretors are such a small percentage of the population that they represent less than the rounding error associated with uncertainties in the subset of the Hispanic population associated with the serological profile. The profile would still be associated with 2% of the Hispanic population, whether Clark had ignored AB secretors or not.

In 42 cases, an error related to serological findings contributed to the balance of evidence to convict. The serology testimony was valid in five of these cases. In the Paul House case, FBI examiner Paul Bigbee presented valid serological findings, but the evidence was compromised by poor storage conditions. (House v. Bell, 2006) It is possible that Bigbee ignored or failed to document spillage, but that has not been conclusively established. The Patrick Willis case provides an example of serology error in a more recent case. (People v. Willis, 2019) The crime occurred in 1992, but the trial did not take place until 2013. At that time, DNA testing demonstrated that Willis was the source of sperm from an oral swab of the victim, who was a sex worker. The serologist testified that the sperm may have been deposited six to 12 hours before death. The serologist provided appropriate limitations to the analysis of sperm deposition time and clarified that the uncertainties were substantial. (Warren, 2013) The prosecutor's closing argument did not reflect those uncertainties and was based on a six-hour time window prior to death. In 2019, the Willis conviction was vacated, in part because there was evidence that the victim was alive at a later time, meaning the murder occurred at a time other than the time at which Willis deposited the sperm. In 2021, a hung jury voted 11 to 1 for acquittal, and the charges were subsequently dismissed. In this case, the prosecutor's misrepresentation contributed to the error, but there is no basis to conclude that the forensic science work was in error.

Hair Comparison

143 hair comparisons

- 139 cases in which hair comparison was used
- 84 hair comparisons with errors of any type ("case errors")
- 59 errors contributing to conviction

System errors

- 19 cases: police or prosecutor misconduct
- 41 cases: inadequate defense
- 6 cases: hair comparison was exculpatory

Forensic errors

- 29 identification or classification errors
- 51 testimony errors based on standards at time of trial
- 77 testimony errors based on current standards
- 27 examinations with possible cognitive bias

Hair comparison testimony was closely related to serology testimony in the era before the availability of DNA analysis, but hair comparison errors followed a different overall pattern. Hair comparison errors were much more likely to contribute to the balance to convict (60% of the time). The dataset includes 143 hair comparisons in which there were 84 case errors. In 20 cases, hair examiners produced a report or testimony that stated or implied an individualization, and the large majority of these hair examinations contributed to the balance to convict.

Forensic hair microscopy provides special challenges for the analysis of post-conviction exonerations. Some observers believe hair microscopy has no forensic value, while the National Academy Sciences "found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA." (Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, 2009) Among many others, Paul Gianelli has written extensively on the weaknesses of microscopic hair comparison. (Giannelli, Microscopic Hair Comparisons: A Cautionary Tale, 2010) Between 1974 and 1986, Barry Gaudette, a forensic scientist with the Royal Canadian Mounted Police (RCMP), attempted to provide an empirical framework for hair microscopy examinations. (Gaudette & Keeping, 1974) Gaudette found that highly-trained examiners could perform reliable and accurate hair comparisons under idealized conditions. The work was criticized for its methodology and statistical analysis. Gaudette's participants compared single hairs to single hairs, but casework examiners typically use many reference hairs due to variations in an individual's hair morphology, even when the samples are limited to particular areas of the body. Therefore, the study examined the ability to distinguish sub-sources (individual hairs), not sources (individual persons with a range of hair characteristics). When citing the Gaudette study, examiner testimony did not generally reflect this subtle but important distinction. The Gaudette study participants coded for 23 characteristics and 96 total possible variations, but the coding system did not always consider whether the variations were from corresponding regions of a hair sample. Examiners did not apply the Gaudette methodology in casework and often used only a subset of the 23 characteristics employed in the Gaudette study. Testimony review appears to indicate that these distinctions were not appreciated by officers of the court. Gaudette attempted to answer criticisms of his research, and independent researchers tried to duplicate his work with mixed results. Hepworth and Wickenheiser summed up the research thusly:

> "[W]ith the application of rigid selection criteria, the frequency of coincidental matches in forensic science hair comparisons is low … [but] routine hair classification is not feasible, because of inconsistency in examiner discrimination." (Hepworth & Wickenheiser, 1990)

During the pre-2000 period, the FBI and many hair comparison examiners recognized these difficulties. The FBI convened a conference in 1985 to establish standards for the conduct and reporting of hair comparisons. (The Laboratory Division, Federal Bureau of Investigation, 1985) The symposium included a recommendation that the Gaudette study should be cited only on cross-examination in response to a question about research supporting the application of hair comparison as a forensic tool. Most importantly, the 1985 symposium recommended six forms of conclusion language t to report hair comparison examinations, including three possible statements that could be used to imply that the source of an evidence hair could be a particular individual. The three inclusion statements were:

1. The questioned hair is consistent with having come from John Doe.
2. The questioned hair could have come from John Doe.
3. John Doe qualifies as being the donor of the questioned hair.

These statements did not imply a "match" or statistical characterization. It is unclear whether investigators, officers of the court, or juries interpreted examiner decisions differently based on the three different language choices, nor is it clear that there was a valid scientific basis for the three distinctions. The practitioners at the FBI's 1985 symposium may have understood the differences but generally did not communicate this understanding in testimony. As established by the FBI review of hair comparisons (ABS Group, 2018), practitioners abandoned the distinctions soon after the 1985 symposium. Within a few years, FBI examiners—and presumably most other examiners—would typically report comparisons using only the language of conclusion #1, the "consistent with" formulation. Their motivations are difficult to discern, but it is very likely that they implicitly realized that the differences in language among the conclusions were not meaningful and therefore chose the clearest and strongest language.

The FBI did not formally adopt report or testimony guidance or other standards on the basis of the symposium. (ABS Group, 2018) In 1995, the laboratory did begin to monitor hair comparison testimony by its examiners, but the program was limited to aspects related to accreditation (such as the demeanor of examiners in the courtroom), not the actual content of testimony. The monitoring program was extended to "accuracy" of testimony in 1998, but it is unclear how this was implemented. By 2000, the FBI Laboratory had instituted routine mitochondrial DNA testing of hair evidence in conjunction with hair microscopy, effectively relegating the practice to a screening role similar to that of serology. In 2002, Houck and Budowle conducted a retrospective study of hair microscopy examinations in the FBI Laboratory between 1996 and 2000 and established the relative value of mt-DNA for hair analysis, which has been called "a landmark in forensic science because it was the first study to systematically and comprehensively analyze a large collection of previous casework to measure the frequency of false-positive associations." (PCAST Working Group, 2016) Houck and Budowle found association errors in 9 out 80 (11%) cases in which a positive association had been reported. Therefore, it is appropriate to assume that there were association errors in many pre-DNA hair comparison cases. That does not mean that each case resulted in a false conviction, because other probative evidence may have been sufficient to establish guilt beyond a reasonable doubt.

In a different study, Dimick et al examined 691 casework hairs. (Dimick, Melton, Higgins, Lindstrom, & Nelson, 2005) They were able to obtain a mt-DNA profile in 92% of cases, although contamination and DNA mixtures complicated their analysis in 9% of cases. Other research studies have indicated that hair comparison may have some value, especially if used as a screening method or in combination with DNA analysis. The controversy of hair morphological comparison continues to the present day, and hair comparison continues to be used, though primarily as a screening and presumptive analysis tool. (Koch, 2017)

The FBI conducted a thorough review of its hair microscopy comparisons in collaboration with the Innocence Project and the National Association of Criminal Defense Lawyers (NACDL). They reported in 2015 that:

> "[T]he examiners' testimony in at least 90 percent of trial transcripts the Bureau analyzed as part of its Microscopic Hair Comparison Analysis Review contained erroneous statements. Twenty-six of 28 FBI agent/analysts provided either testimony with erroneous statements or submitted laboratory reports with erroneous statements."

The most detailed examination of these cases has been presented in the subsequent 2018 "root cause" analysis by an outside contractor to the FBI. (ABS Group, 2018) That report discussed the hair examiner errors in much more detail. Although the report did not specifically address post-conviction exonerations, it was undoubtedly in response to concerns related to the large proportion of exonerations related to hair microscopy and the subsequent criticisms of the field by the NAS, among others. The FBI and ABS Group did not examine whether the associations were correct or whether the FBI's testimony in these cases resulted in a wrongful conviction. The analyses were limited to the language of reports and testimony produced by examiners. In total, the FBI reviewed 3,499 cases in which hair comparison by 35 different examiners contributed to a conviction prior to 2000. They found that at least 90% of the reports and transcripts contained errors, but the vast majority of these errors were related to the use of language that conformed to the testimony recommendations of the 1985 FBI hair symposium.

> "Almost all report errors (over 98%) described the questioned hair using the phrases consistent with having originated from [individual's name] or consistent with having come from [individual's name)." (emphasis and brackets in the original) (ABS Group, 2018)

In other words, the review did not find that hair examiners had routinely made association errors. Instead, the review found that the examiners adopted language that was deemed after 2010 to be an overstatement of the statistical and scientific basis for microscopic hair comparison analysis. As part of the collaboration among the FBI, IP, and NACDL, the groups arrived at the following conclusion on this topic:

> It's important to note that microscopic hair comparison analysis is a valid scientific technique still conducted by the FBI Laboratory. The science of microscopic hair comparisons is not the subject of the review. In 1996, the FBI Laboratory developed and implemented mitochondrial DNA (mtDNA) analysis in conjunction with probative hair analysis because it is the most effective protocol for the forensic examination of hair, and it provides a more meaningful association than either technique used alone. (Federal Bureau of Investigation, 2015)

Nonetheless, many observers have used the FBI's conclusion to promote the perception that hair comparison—and by extension all pattern evidence examination—is based on "junk science." (Scheck, Barry Scheck Lectures on Wrongful Convictions, 2005)

This background has implications for the development of a full understanding of the contribution of hair comparison errors to false convictions. First, it is necessary to make a clear distinction among types of errors. Although report and testimony errors are important, they are fundamentally different from errors in the association of evidence with an individual source or class of sources. In addition, the appropriate scientific foundation or statistical characterization of a result may vary over time or based on the views of the observer. The FBI's mid-2010's analysis disagreed with hair comparison interpretations that arose from its own 1985 symposium. The subsequent root-cause analysis faulted a general lack of management direction concerning hair comparison testimony standards. Therefore, it may be appropriate to conclude that the primary issue in pre-2000 hair comparison was the lack of enforced, consensus standards. In the attribution of forensic errors related to wrongful convictions, examiners cannot be expected to have followed ex-post-facto standards that were developed years or decades after the original trial. They certainly can't be expected to satisfy the expectations of future critics with sententious views about the nature of the scientific foundations of their discipline. Rather, forensic scientists should conform to the testimony standards in general acceptance at the time they were working. For the current study, the current DOJ Uniform Language for Testimony and Reports (US Department of Justice, 2019) was used as the standard for testimony in the modern context, while the 1985 FBI Symposium (The Laboratory Division, Federal Bureau of Investigation, 1985) was used as the standard for testimony in the pre-2000 time frame.

This study found 15 case errors in wrongful convictions associated with FBI Laboratory hair examiners. In six cases, victim testimony or other evidence constituted the primary evidence against the defendant. In the 1983 Anthony Caravella case, FBI examiner Andrew Podalak produced an exculpatory report that linked an evidence hair to an alternate suspect. (Podolak forensic testimony, 1984) The link was discounted by the prosecution at trial on the basis of secondary transfer, and police detectives later indicated that they may have contaminated the evidence through poor evidence handling. A postconviction DNA cold hit was not associated with the alternate suspect identified by the hair comparison. (Spadaro v. City of Miramar, 2013) Nonetheless, the case demonstrates that investigators often discount exculpatory forensic evidence. The evidence handling issue may have reflected poor investigative practices more broadly, because the alternate suspects were known but never thoroughly investigated prior to Caravella's trial and conviction.

In the Gussie Vann case in Tennessee, FBI examiner Chester Blythe limited his testimony to microscopic similarity and made a valid conclusion that the evidence hair samples were "probably" from an adult. (State v. Vann, 1998) The case included erroneous testimony from a state medical examiner, Ronald Toolsie, who was discovered to have been abusing drugs for many years, including during the Vann trial. (State v. Vann, 2010) Several other forensic professionals gave valid testimony in the case, including Blythe and trace and serology examiners from the Tennessee Bureau of Investigation laboratory. As in the Caravella case, valid forensic work was undermined by other case errors.

Some hair comparison errors related to changes in testimony standards since the time of trial. In many cases, examiners used language that stated an evidence hair was "consistent with" an exemplar or made similar statements that are not permitted under the Uniform Language for Testimony and Reports. (US Department of Justice, 2019) In 51 cases, the examiner made a testimony error that would not conform to standards at any time. For example, several examiners showed a poor understanding of the Gaudette hair comparison studies or otherwise mischaracterized the foundational science of the discipline. (Gaudette & Keeping, 1974) Some examiners testified that the Gaudette studies could be used to establish a random match probability (RMP). Also, there was often a poor understanding of the methodology used by Gaudette. Some examiners cited the Gaudette study but relied on a much more limited number of comparison features to associate a suspect with the evidence than the Gaudette participants used in their comparisons. Examiners did not recognize that the Gaudette studies were based primarily on Caucasians, making them of limited utility when examining hairs from individuals with African ancestry. In many cases, the examiner made invalid statistical statements that undermined otherwise valid conclusions and testimony. In the Timothy Bridges, case, examiner Elinos Whitlock cited the Gaudette study but clarified that there was no basis for hair random match probability estimates. (State v. Bridges, 1992) He then testified that, based on "his personal experience … the likelihood of two Caucasian individuals having indistinguishable head hair, it is very low. A conservative estimate for that probability would be . . . approximately one in a thousand." This statistical estimate was invalid. The trial court and appeals court accepted the testimony at the time, but the district attorney advocated for Bridges' exoneration based on exculpatory DNA evidence and the Whitlock testimony error. Postconviction trace evidence examiner Skip Palenik concluded that the comparison itself was valid but criticized the interpretation as "beyond the limits of science." (Bantz, 2017)

In the Gary Dotson trial, examiner Timothy Dixon presented hair testimony in addition to faulty serological analysis. The hair comparison testimony was valid based on the standards in place at time of trial because it was limited to the "consistency" of microscopic characteristics. GN judged the hair testimony to be "valid" and "vague" although the Dixon testimony would not conform to the current ULTR hair comparison standard. The prosecutor inflated the probative value of the hair comparison in his summation: "… the pubic hair found among her pubic hair happens to match the defendant's." (Warden) Further, the hair comparison may have relied on a hair that was unsuitable for forensic comparison. (Blake E. , FSA File No. 85-035, 1985) It is now established that hair comparison errors are very likely in cases involving light-colored hair as in the Dotson case. (Houck & Budowle, 2002)

In many cases, hair comparison testimony would have conformed to the standards of the 1985 symposium but not the ULTR. In the connected cases of Santae Tribble and Cleveland Wright, two different FBI examiners testified at the two different trials. (Tribble v. District of Columbia, 2016) Both examiners limited their testimony to "microscopic characteristics," but they also cited their experience from prior casework to maintain that it was "very rare" to observe "hairs of two individuals that show the same characteristics." Such statements do not conform to the requirements of the ULTR. The prosecutor then mischaracterized the probative value of the hair comparison, (Kaye, 2017) saying:

> But he couldn't reject Santae Tribble's because it was exactly the same. And the only reason he said could be is because there is one chance, perhaps for all we know, in ten million that it could [be] someone else's hair. But what kind of coincidence is that? ... The hair is a great deal more than "could be." And if you listened to Agent Hilverda and what he really said, the hair exactly matched Santae Tribble's hair, found in the stocking."

Regardless of the forensic examiners' testimony, the prosecutor's statement was highly misleading and scientifically invalid. Postconviction DNA testing excluded Tribble and Wright as sources of the evidence hairs. That said, it is possible that Tribble and the actual source shared sufficient characteristics to justify the original testimony, if it had been appropriately limited and not mischaracterized by the prosecutor.

Several overturned convictions were associated with the FBI's notification to local jurisdictions arising from their evidence reviews. (Comey, 2016) In the Dennis Butler case, FBI examiner

Myron Scholberg had examined 16 characteristics on the questioned hairs and the defendant's head hairs and stated that three questioned hairs were "microscopically the same of alike" the reference hair. Although he repeatedly clarified that hair comparison could not be used to positively identify a source, he also cited past case work to support his ability to distinguish hairs from different sources to with an implied RMP of four in 10000. This latter aspect of his testimony was invalid. As in the Tribble/Wright case, the prosecutor exaggerated the probative value of the hair comparison evidence. In 2020, the appeals court majority stated the evidence to convict included "false testimony that Butler's hair microscopically matched hair found on the victim." As in many cases, the court failed to distinguish between exaggerated testimony and other issues. In fact, the Butler's hair did microscopically match the evidence hair, so the Scholberg testimony was not "false" in that respect. As the dissent in that case stated, Scholberg repeatedly clarified the limitations of hair comparison on cross-examination and redirect. (United States v. Butler, 446 U.S. App. D.C. 247, 955 F.3d 1052 (2020)) Butler's conviction was overturned on the basis that the hair testimony denied him a fair trial and the charges against him were dismissed.

FBI examiner Michael Malone was associated with three hair comparison errors and two trace evidence examination errors in wrongful convictions. Malone's testimony did not conform to the 1985 or 2020 standards. In the Elmer Daniels case, Malone associated both head and pubic hair with the defendant, saying, "When you get a double, what we call a double match like this, it would increase the probability tremendously." (Otterbourg, 2020) A 2018 AFIS search associated latent prints from the crime scene with Daniels' brother, who was an alternate suspect at the time of the original trial in 1980. It is not clear why the print was not compared to the alternate suspect at that time. Malone provided faulty hair and trace testimony in the case of Juan Matta-Ballesteros, who had other convictions related to drug trafficking. (Matta-Ballesteros v. United States, 2017) For both types of evidence, Malone exaggerated the probative value of his comparisons and implied that he could individualize the source of hairs and fibers. The Matta-Ballesteros case was one of 162 cases in which the DOJ Office of Inspector General found that Malone had produced faulty hair and fiber testimony. (Bromwich M. , 1997) Notably, the internal review found that the actual examinations were probably correct in the Matta-Ballesteros case, but Malone's exaggerations undermined what might otherwise have been considered valid forensic evidence and a sustained conviction.

Fraudulent forensic work was substantiated in only two cases. In other words, most case errors arose from honest mistakes, misinterpreted evidence, exaggerated testimony, or failures by the prosecutors or defense attorneys. In three Montana cases, Arnold Melnikoff produced serious errors that misrepresented the standards of the discipline and the Gaudette research. It has not been established that Melnikoff's work was based on deliberate fraud. Melnikoff sued the Washington State Police after his dismissal. In the civil case, the court asked Gaudette to review Melnikoff's testimony. (Arnold Melnikoff v. Washington State Patrol, 2005) Gaudette concluded that Melnikoff's work "did not meet the standards of practice expected of a fully qualified and competent hair examiner." Gaudette noted that Melnikoff (like Malone) multiplied the probabilities associated with head and pubic hair comparisons, which was invalid because the characteristics are not independent variables.

Implications for Statistical Interpretation Frameworks

In wrongful convictions, serological testimony errors included mistaken statistical conclusions. (Garrett & Neufeld, Invalid Forensic Science Testimony and Wrongful Convictions, 2009) These errors usually resulted from problems in evidence handling, laboratory analysis, or misinterpretation, not calculation errors. Most commonly, errors related to the invalid application of population groups in analysis. For example, a suspect may have been a possible contributor to a crime scene stain, but the fraction of the population that shared the defendant's serological profile is not the same as the fraction of the population that may have shared with crime scene stain's serological profile. At minimum, a crime scene stain's ABO profile could have included contributions from any nonsecretors, who comprise approximately 20% of the overall population. In many cases, examiners did not clarify this point of interpretation. Further, when a defendant was a nonsecretor and "could not be excluded" as a contributor to a stain, some examiners failed to clarify that it was not possible to exclude nonsecretors from any stain based on an ABO profile. Serological interpretation provides a cautionary example of the drawbacks of statistical or quantitative models for forensic interpretation. In wrongful convictions, serological statistical models based on well-established science were compromised by cognitive bias, poor evidence handling, laboratory testing errors, and interpretation errors. One cannot conclude that quantitative interpretation frameworks will prevent similar forensic science errors in other disciplines.

Hair comparisons and serology seldom contribute to any conviction presently. In this study, all hair and serology examinations were associated with the possibility that DNA technology would provide more reliable and probative results. That said, the issues regarding scientific research, random match probabilities, and statistical characterization remain relevant. ENFSI and SWGDAM have each expanded the use of statistical characterization of forensic evidence in recent years, but issues related to the validity of the underlying population data are strikingly similar to those seen in wrongful conviction cases. (Champod, Biedermann, Vuille, Willis, & De Kinder, 2016) The ENFSI guidelines require the use of a likelihood ratio approach, even when dealing with types of evidence in which the likelihood or rareness of a positive association is not established empirically. Hair comparisons do not have an empirical basis for statistical characterization, hence the application of the ENFSI methodology would constitute an error under the FBI's 2012 review of hair comparison cases. (ABS Group, 2018) ENFSI cites wear marks on shoes for the purpose of individualization as a good example in which a local database of seized shoes could be used to make a population estimate. In the current study set, there were 11 cases in which shoe/foot impression evidence errors contributed to a wrongful conviction, five of which included an incorrect individualization. Dr. Louise Robbins was the leading advocate for the use of wear impressions to individualize shoeprints and was involved in three of these cases. Her approach would have conformed to the ENFSI guidance but produced invalid and incorrect results that led to the conviction of innocent defendants. ENFSI has emphasized the importance of transparency to mitigate the possibility that invalid population data would cause case errors. However, the history of wrongful convictions demonstrates that transparency is inadequate in this regard. First, officers of the court often fail to understand the statistics related to forensic analyses, even to the point of failing to recognize clearly exculpatory results. Also, local population databases have contributed to wrongful convictions because they contained skewed populations or were interpreted wrongly. For example, Fred Zain was involved in six serology cases in West Virginia that included serious statistical interpretation errors. Nonetheless, Zain had constructed datasets for West Virginia populations across several types of serological markers and published the information in a widely-accepted, peer-reviewed journal (Zain, Smith, & Myers, 1989). At trial, he referred to these databases that were very similar to the types envisioned by ENFSI, but his misinterpretations were not discovered for many years. In short, the history of wrongful convictions argues against the use of statistical frameworks without a valid, empirical foundation.

As observed in the history of the Gaudette work and its impact on wrongful convictions, forensic scientists may take great care to understand the limits of research studies that report statistical measures but cannot be used to support RMP. Further, forensic scientists must consider the extent to which a particular study or set of studies exhibit construct validity. (O'Leary-Kelly & Vokurka, 1998) Like the Gaudette work, many research studies use idealized frameworks to control empirical design and may not reflect actual forensic practice. Even if Gaudette had established an RMP framework, forensic examiners could not have relied on the paper to report an RMP because of the differences between study methodology and practice methodology. Finally, the application of statistical methodology requires that forensic examiners fully understand statistics, including concepts related to population constructs. As discussed, hair comparison examiners failed to recognize the difference between source and sub-source propositions, an error that resulted in mistaken beliefs about the relevance of various population estimates in the Gaudette studies. Also, serologists also failed to make similar distinctions, despite extensive training and clear scientific guidance. Valid statistical frameworks may lead to errors in the absence of training and management review, including testimony review.

Fire debris investigation

45 fire debris investigations
- •43 cases in which fire debris analyses was used
- •35 fire debris investigation errors of any type
- •33 fire debris investigation errors contributing to conviction

29 fire debris chemical analyses
- •Assessed separately from fire debris investigation
- •17 fire debris chemical analysis errors of any type
- •9 fire debris chemical analysis errors contributing to conviction

System errors
- •16 cases: police or prosecutor misonduct
- •30 cases: inadequate defense

Forensic errors
- •21 classification errors (4 involving chemical analysis)
- •36 testimony errors (6 involving chemical analysis)
  30 examinations: possible cognitive bias
  22 failures to conform to best practices or standards
  26: reference data or documentation errors

In general, the field of fire debris investigation has been deeply transformed by the experience of wrongful convictions. More than half of fire-debris-related wrongful convictions occurred in 1990 or before. The advent of NFPA standards has played a major role in the mitigation of the risk of wrongful convictions since 1990, especially NFPA 921. (National Fire Protection Association, 1992) The adoption of NFPA standards in practice was variable throughout the 1990’s and 2000’s, especially with regard to changes in NFPA 921 after its initial adoption in 1992. In many cases, investigators trained prior to the issuance of NFPA 921 continued to apply methods they had learned in their early careers. For example, these investigators would make conclusions that did not account for flashover, accidental/natural causes, and high-temperature burning.

In seven convictions occurring as late as 2010, investigators used a “negative corpus” to conclude arson by eliminating other causes. NFPA 921 itself has been ambiguous on the negative corpus issue, with some versions discouraging its use only. The 2008 version clarified that findings must be based on evidence observations, not “speculative information” such as negative corpus. Nonetheless, in several cases, investigators used negative corpus to conclude arson without a full investigation of other possible causes, an approach which was never in keeping with any version of NFPA 921. The 2007 Joseph Awe case is instructive. Three prosecution experts testified that the fire was arson, relying on a consultant's view that an electrical circuit was not responsible. (State v. Awe, 2010) It is unclear if the experts were aware that the consultant had omitted evidence of electrical arcing from his final report. There was no evidence of accelerant, and investigators relied on the point of lowest and deepest char to determine the origin of the fire. As examiner James Sielehr stated,"[O]nly when you can eliminate each and every [accidental possibility] can you refer to this as an incendiary fire. That's how we do it." The only defense expert was an electrical engineer with no training or experience in fire debris analysis. Postconviction, defense experts John Lentini and Mark Svare maintained that the fire originated in an electrical service panel, but the conviction was overturned largely on the basis of the changed standards concerning negative corpus. (Hall D. , 2013)

Like forensic pathology, fire debris investigation relies on subjective analyses, contextual information, and the possibility of multiple, valid interpretations. Trained investigators may differ concerning their findings or the confidence in their findings, even when analyzing the same set of physical evidence. As a result, it is critical for defense attorneys to access independent experts in cases in which arson is not definitively established. In several wrongful conviction cases, defense attorneys did not choose to consult an independent expert or call one as a defense witness. They may have had a poor understanding of the uncertainties in fire debris investigation, especially as NFPA 921 evolved over the last 30 years. Also, they may have based their defense theory on an alternate suspect who may have committed arson, so they did not see value in questioning the arson theory itself. As a result of such issues, inadequate defense appears to be implicated in most wrongful convictions involving fire debris analysis. The courts do not always recognize the importance of independent review by the defense in arson cases. The 2007 Daniel Carnevale conviction—also a negative corpus case—was upheld on appeal in 2012 because the appeals court held that a defense expert would not have had access to the physical evidence. (Commonwealth v. Carnevale, 2012) The evidence had been destroyed "per ATF policy and procedures" in 1998 after no perpetrator had been identified in the five years after the fire. In fact, the ATF examiner at trial had also testified based on memory, not physical evidence or reports.

In several cases, the fire debris investigator did not use chemical analysis to confirm the presence of accelerants at a scene. In some of these cases, the investigator ignored or discounted chemical analysis that did not find accelerants. One subset of interest is the use of accelerant-detecting canines, which were relied upon to determine the presence of accelerants independent of laboratory analysis. In the most extreme cases, canine "hits" were deemed more reliable than chemical analysis because dogs were presumed to be more sensitive. In the 2006 conviction of Robert Yell, accelerant detecting canine "PJ" alerted on several spots in the trailer crime scene, but chemical testing did not confirm accelerants. (Yell v. Commonwealth, 2007)The trial court declined to do a formal Daubert analysis, attributing canine scent tracking as "non-scientific, experience-based knowledge." The Kentucky Supreme Court reviewed the issue, with the prosecution arguing that "the dog's ability to accurately detect accelerants through its sense of smell is not dependent on scientific explanation." The court held that the jury could weigh the canine detection evidence adequately within the overall context of the case, including the negative chemical findings. The dissent noted that there was no documentation on the dog's accuracy rate, that most courts have considered canine accelerant alerts to be novel scientific evidence, and cited DeHaan's analysis that canine alerts are not reliable in the absence of laboratory confirmation. (DeHaan & Icove, 2011) The dissent noted the illusion of validity that many people place on canine detection, which may produce a prejudicial effect in a case. (Einhorn & Hogarth, 1978) Yell's conviction was overturned in 2013 on basis of inadequate defense and a reassessment of the scientific basis for the fire investigation conclusions. (State of Wisconsin v. Joseph Awe, 2013)

Some observers have claimed broadly that many inculpatory fire debris findings have been based on "junk science." (Giannelli, Junk Science and the Execution of an Innocent Man, 2013) The use of the term "junk science" may be misleading in cases in which the examiner attempted to apply valid fire science but failed to follow necessary documentation or practice standards, as seen in 18 cases covering 22 fire debris examinations in the dataset. In the 2008 Victor Caminata case, examiner Michael Jenkinson completed a poor documentation of the fire scene and a flawed and incomplete reconstruction. (Plaintiff's Reply Brief, 2016) That said, an independent review by the federal Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATFE) found that Jenkinson "clearly demonstrate[d] adherence" to the scientific method. (Caminata v. Cnty. of Wexford, 2016) The distinction matters current science-based standards are a valid and reliable basis for findings of incendiary cause and should not be conflated with "junk science" in other disciplines. Also, scientific research will not translate to practice in the absence of standards, effective training, and quality assurance mechanisms.

Many cases demonstrate the organizational and professional challenges faced by the field of fire debris investigation. In one case, the commander of a newly formed arson task force provided poorly documented findings when he attended the fire while conducting a training exercise. (Transcript of Evidentiary Hearing, 2015) In other cases, prosecutors misrepresented evidence or—in the 1989 Latta case—made the initial arson finding themselves after attending the fire scene. (Latta v. Chapala, 2005) As outlined in prior research, wrongful conviction cases provide clear examples of confirmation bias in the discipline. (Lentini, Anatomy of a Wrongful Arson Conviction: Sentinel Event Analysis in Fire Investigation, 2014) Organizational and professional shortfalls are closely related to bias issues; 77% of examinations associated with best practice errors were also associated with possible cognitive bias.

In six arson wrongful convictions, no forensic error was found. In the Connecticut cases of Martina Jackson and Speciale Rose Morris, the Connecticut State Police found petroleum distillates presumed to be from a gasoline can found at the scene. The Connecticut State Police Fire and Explosion Investigation Unit concluded an incendiary cause. Jackson and Morris pled guilty but were never sentenced. Alternate suspects were identified and charged. (Shugarts, 2018) In the meantime, the original chemist had retired and his replacement would not testify on the basis of the original finding. A retest failed to find the accelerant, and all charges were dismissed, including those against Jackson and Morris. Volatile organic compounds are known to evaporate after prolonged shelf-storage, so there is no basis to conclude that the original chemical analysis or fire debris investigation were flawed. (Hsieh, Horng, & Liao, 2003) In the case of Walter Forbes, there is no basis to conclude that the fire investigation was flawed, although the 38-year-old conviction was overturned on the basis of recanted eyewitness testimony. (Associated Press, 2020)

Canine detection

17 canine detection examinations

- 17 cases in which canine detection was used
- 15 canine detection examinations with errors of any type
- 15 canine detection examination errors contributing to conviciton

14 system errors

- 8 cases: police or prosecutor misconduct
- 8 cases: objections not raised by defense
- 13 cases: judicial error (including 9 cases in which canine detection was not recognized as a novel forensic method)

14 forensic errors

- 14 identification or classification errors
- 15 testimony errors
- 15 cases: possible cognitive bias
- 2 cases: fraudulent examination

Canine detection is inherently oriented around crime scene work and police investigation. Some canine detection instances were not coded as forensic evidence because the use of dogs was limited to investigative work and was not the subject of known courtroom testimony. In 17 cases, canine detection was used to make forensic conclusions about the presence of an individual at a crime scene or accelerants at a fire scene, including seven false individualizations. Testimony often exceeded the limits of science, although some cases preceded the development of clear standards in the field. In nine cases, such as those involving the use of canine detection lineups, judges failed to recognize the novelty of the method and did not conduct a Daubert or similar review. In four cases, the defense objected to elements of faulty testimony, but the judge did not sustain the objection. In these cases, the failure was clearly identified in appeals decisions. Almost all canine detection cases were associated with cognitive bias, because the handler did not take appropriate steps to limit cueing signals to the dog. One handler, John Preston, was involved in three wrongful convictions in Brevard County, Florida and produced fraudulent reports and testimony in two of those cases. Although Preston is an extreme example, dog handlers generally did not recognize the limitations of canine detection. A typical situation was detailed by Judge Nancy Gertner in the James Hebshie case appeal:

> “It is not an understatement to say that Lynch, the dog handler, was permitted to testify to an almost mystical account of Billy's powers and her unique olfactory capabilities. He presented unsubstantiated claims about the dog's accuracy. He was allowed to go on at great length about his emotional relationship with the dog and his entirely subjective ability to interpret her face, what she thought, intended, and the "strength" of the alert she gave in this case. Finally, Lynch was permitted to testify that the dog did not alert to anything else on the premises, as if the dog had been allowed to range widely on the fire scene (she was not), and as if the dog's failure to alert had evidential value (it does not).” (United States v. Hebshie, 2010)

Judge Gertner supported Hebshie's inadequate defense claim on the basis of a failure by defense to recognize and object to this inappropriate testimony.

Bitemark Comparison

44 bitemark examinations

- 28 cases in which bitemark comparison was used
- 34 bitemark comparisons with errors of any type
- 33 errors that contributed to conviction

System errors

- 14 cases: inadequate defense
  6 cases: evidence handling or chain of custody error
- 5 cases: police or prosecutor misconduct
  5 cases: misleading information about forensic results given to defense

Forensic errors

- 32 individualization or classification errors
- 31 testimony errors
- 30 examinations associated with possible cognitive bias
- 24 "honest mistakes"
- 14 cases of postmortem artifact misinterpretation

The dataset includes 44 instances of bitemark comparisons, including cases in which more than one examiner made a comparison and testified. In all, there were 34 case errors, 32 errors of identification or classification, and 31 instances in which there was faulty testimony. The large majority of bitemark case errors affected the balance to convict in the cases. Bitemark comparison is the only discipline in which identification and classification errors by forensic examiners were present in almost all examinations with a case error.

In 30 examinations, bitemark examiners demonstrated that cognitive biases may have affected their work. There were 21 examinations associated with forensic examiner variability. For example, in the Roy Brown case, Edward Mofson, Lowell Levine, and Homer Campbell all provided bitemark comparisons. (Santos, 2006) Mofson actually testified five different times over the years, four times for the prosecution and once for the defense. Levine did an analysis for the prosecution that concluded the bitemark was an exclusion. There was nonetheless a case error related to his work, because the Levine analysis was suppressed by the prosecutor and not disclosed to the defense. Campbell testified for the defense that it was an exclusion, and there were no case errors associated with his examination.

Bitemark comparison demonstrates the limits of some approaches to forensic improvement. In the 33 examinations in which the educational level of the examiner was known, all examiners had doctoral-level degrees, usually in a dentistry-related field. In 31 instances, the examination was performed by an independent consultant. This is unsurprising, given that the discipline has been largely the province of part-time odontologists. The education level and independence of bitemark examiners did not prevent errors. For example, examiners did not follow best practices of the American Board of Forensic Odontology (ABFO) in 20 instances and misrepresented the scientific basis for their conclusions in 24 instances. In general, bitemark examiners were not subject to the governance or quality control mechanisms of public crime laboratories. Bitemark comparisons were performed within a public crime laboratory in only two cases. Like other disciplines in a similar position—such as canine detection and digital evidence—bitemark examiners may have benefited from more stringent enforcement of standards and a closer relationship to the governance structures of the rest of the forensic science community.

Some bitemark comparison errors were associated with “bad apple” examiners. Michael West contributed erroneous testimony in six wrongful convictions, including four cases in which he made dubious claims about postmortem artifacts that may not have been bitemarks. The ABFO disciplined West and removed his certification in 1993, although he continued to testify in cases. (Brewer v. Steven Timothy Hayne & Michael H. W., 2015) He provided invalid individualization testimony in the 2001 Stubbs/Vance case. He also provided invalid video enhancement testimony in that case, though the defense failed to object to that testimony. (Stubbs v. State, 2003) It would be a mistake to assume that the bad-apple problem was the primary issue in bitemark comparison wrongful convictions. ABFO Diplomates testified 22 times in wrongful convictions and were associated with 20 errors. Respected ABFO leaders, such as Robert Barsley and Richard Souviron, produced misidentifications. Souviron’s work in the Robert DuBoise case also involved a misinterpretation of postmortem artifacts and testimony that did not conform to ABFO standards. (Office of the State Attorney 13th Judicial Circuit, 2020)

Public documents do not include documentation of the feature details that examiners used in wrongful conviction cases, so there is limited basis to determine the level of difficulty of the comparisons. Qualitatively, many comparisons appeared to be clear errors contradicted by postconviction DNA analysis or postconviction reanalysis of the bitemarks themselves. In some cases, postconviction experts stated that the bitemark was a clear exclusion or was not suitable for forensic comparison. In other disciplines, wrongful convictions tend to be more closely aligned with difficult comparisons or untrained examiners, and errors by recognized leaders in the disciplines are almost unheard of. This phenomenon—expert variability that extends even to highly-experienced and respected examiners—is widely observed in only two disciplines, bitemark comparison and forensic pathology. In the case of bitemark comparison, it may reflect inherent limitations in the field. In other words, wrongful convictions demonstrate the uncertainties in the use of human skin as a registration medium for bitemark impressions and may suggest that the technique is untenable for any level of individualization or classification.

It should be noted that the ABFO has changed its standards considerably over the study period. Current standards permit only three conclusions: excluded, not excluded, and inconclusive. (American Board of Forensic Odontology, 2018) None of the testimony in wrongful conviction cases conformed to the current guidelines, and the most recent conviction (Crystal Dawn Weimer in 2006) predates the 2018 ABFO revision by over a decade. Dr. Homer Campbell used a similar formulation in his testimony, which included three conclusions: exclusion, consistent with, and “reasonable degree of dental certainty that those teeth did in fact make that mark.” In a 1987 case involving two codefendants (Calvin Washington and Joe Sidney Williams), Campbell said that Williams’ dentition was “consistent with the injury found on the decedent.” (Hall M. , 2015)That formulation has a similar semantic meaning to the ABFO’s “not excluded” conclusion. Arguably, Campbell’s testimony was acceptably limited in scope, but his conclusion was incorrect. Williams was exonerated by postconviction DNA testing. Campbell provided an accurate exclusion in the Roy Brown case. (Santos, 2006) He used the “reasonable degree” identification formulation in the Steven Chaney case, which was also a DNA exoneration and became a landmark case in Texas’ decision to disallow bitemark testimony. (Texas Forensic Science Commission, 2016)

In the Robert Stinson case, three odontologists (including the defense examiner) agreed that Johnson made the eight bitemarks observed on the victim at autopsy. (Robert Lee Stinson v. James Gauger, Lowell Johnson, and Raymond Rawson, 2015) The bitemarks were considered distinctive because it appeared that the biter was missing an upper lateral tooth. Stinson was missing his upper right incisor, which is not a lateral tooth, and had numerous irregularities in his dentition. Dr. Lowell Johnson, who was certified by the American Board of Forensic Odontology (ABFO), initially associated the impressions with a missing lateral tooth, then changed his conclusion to a missing incisor. During a one-to-three hour examination, ABFO-certified examiner Raymond Rawson reviewed Johnson's findings and agreed with them

Seized drugs

The study set includes 130 cases in which seized drug analysis was used to convict an individual. In all 130 cases, the seized drug analysis was in error and was required for the conviction. Notably, the dataset does not include cases from large-scale scandals in this field, such as the Annie Dookhan and Sonja Farak drug chemistry misconduct cases in Boston. (McDonald, 2019) The cases include 13 jurisdictions and demonstrate the limitations of colorimetric drug test kits, as demonstrated in the research literature. (Philip & Fu, 2018) Better technologies, such as Raman-based instruments, now exist but colorimetric kits are still in general use. (Fedchak, 2014)

In at least 25 cases, less than 1 gram of material was seized. Only one case went to a criminal trial (Dollard), while the remainder involved guilty pleas. In Texas, where 118 of the cases were adjudicated, the state has revised its policies on drug test kits and no longer allows drug convictions in the absence of confirmation tests, even in the case of a guilty plea. (Texas Forensic Science Commission, 2018) The Texas Forensic Science Commission report on this subject reflects many concerns about field drug testing, including officer safety, false positives, poor performance for impure samples, poor performance with novel psychoactive substances, and the subjectivity of color perception. (Texas Forensic Science Commission, 2018)

Two cases involved synthetic cannabinoids. The Roumaldo Lerma case included a positive field test for marijuana, but the substance was actually AB-CHMINACA. (Possley, Roumaldo Lerma, 2015) (Banuelos & Miller, 2015) The Clayton Martel case also included a positive field test for marijuana, but the substance was actually AB-FUBINACA. (Possley, Clayton Martel, 2015) (Salazar & Noyola, 2013) Both compounds are commonly referred to as "Spice/K2" and were subsequently listed as Schedule 1 controlled substances by the Drug Enforcement Administration. (Drug Enforcement Administration, 2017)

Almost all seized-drug cases used field drug test kits and did not involve an error in a forensic laboratory. These errors are generally attributed to forensic practice, but forensic science organizations may not play a primary role in the management of field drug testing. As in many wrongful convictions in other disciplines, field drug test errors often occur outside the purview of forensic science organizations and the governance and quality assurance mechanisms associated with public crime laboratories.

Toxicology

In 11 cases involving 12 toxicological analyses, there were seven case errors, including four classification errors and three testimony errors. In four cases, there were issues involving evidence collection or handling. Quality assurance and reporting issues played a significant role in several cases. In the Cruz-Romero case, the police used an Intoxilyzer that had failed calibration checks, although those failures did not coincide with the time of the instrument's use in the specific case. The trial court declined to admit evidence of the Intoxilyzer malfunctions because the machine had passed reliability checks near the time of the defendant's test. The appeals court disagreed, saying "it is improper for a court to base a relevancy determination on its own findings of fact." (State of Idaho v. Carlos Adrian Cruz-Romero, 2016)

The Sommer and Stallings cases, already discussed above, related to quality assurance issues in laboratory analysis. In the Eric Smith case, a military court convicted Smith on the basis of contaminated urine drug samples. Postconviction DNA testing demonstrated that the sample had foreign DNA of an "unknown and unrelated individual." (Ashton & Bernton, 2016) It is not clear if the contamination occurred at the sampling site at Joint Base Lewis-McChord (JBLM) or the testing facility at Tripler Army Medical Center. Other quality assurance issues have arisen at Tripler. (Maria Delacruz vs. Tripler Army Medical, 2007)

In the Virginia LeFever case, the interpretation of the toxicology results was complicated by uncertainties in the way that poison had been administered. (State v. LeFever, 1991) Police had confiscated garbage bags that included food materials laced with arsenic and strychnine. The digestive tract of the decedent was free of lesions, but he did have high levels of both toxins in his blood. Objects containing strychnine, arsenic and amitriptyline were found in the lower sigmoid colon, suggesting that the items had been inserted rectally. The interpretation of the toxicology and cause and manner of death were valid, but toxicologist James Ferguson had misrepresented his credentials. (LeFever v. Ferguson, 2013) That led to a decision for a new trial, and the prosecution dismissed the now 22-year-old charges.

The case of Hannah Overton involved the interpretation of hypernatremia in a child who may have self-administered a lethal amount of salt or a salt-containing spice. (Ex parte Overton, 2014) The defense team failed to recognize the importance of its own expert's deposition on the origin of the child's hypernatremia. That report was never entered into evidence, so the valid and exculpatory interpretation was not properly considered by the court. Overton has received a certificate of actual innocence.

Forensic pathology

172 forensic pathology examinations

- 148 cases in which forensic pathology was used
- 84 forensic pathology errors of any type
- 60 errors contributing to conviction

System errors

- 9 cases: evidence handling or chain of custody errors
- 56 instances: inadequate defense
  9 cases: misleading information provided to defense
- 6 cases: police or prosecutor misconduct

Forensic errors

- 23 identification or classification errors
- 44 testimony errors
- 52 examinations: possible cognitive bias
  23 examinations: misinterpretation of postmortem artifacts

Forensic pathology was among the most common types of forensic evidence in the study set, with 172 examinations, 84 case errors and 60 case errors that contributed to the balance of evidence to convict the defendant. All examiners had a doctoral degree, when the examiner's degree was known. Most examiners resided in a public laboratory, including 92 within a medical examiner office and 14 within a coroner's office. Among the eight case errors associated with coroner's offices, five of the examiners had doctoral degrees and three had unknown levels of education. There were also ten consultants, nine of whom had known doctoral degrees and most of whom were affiliated with a public forensic science organization not directly involved in the case (i.e., they were operating independently of their primary position).

In general, forensic pathologists committed honest mistakes. No fraudulent forensic pathology examinations were found in any wrongful conviction case, although some examiners produced findings that raised ethical questions (e.g., Stephen Hayne in Mississippi, associated with four wrongful convictions in this dataset). Rather, wrongful convictions data demonstrate two difficult challenges for the forensic pathology discipline. First, there may be multiple valid interpretations that can be derived from the same fact pattern, as evidenced by forensic examiner variability (coded in 18 instances). Second, forensic pathologists often consider case elements that do not directly relate to medical findings. It is unclear from the data in wrongful convictions if there is a clear basis to determine what evidence is task-relevant to a forensic pathology analysis, an issue under some debate among researchers and practitioners. (Dror, et al., 2021) To some extent, these issues may derive from gaps in interpretation standards, but it is unclear if current initiatives directly address interpretative variability, bias, and task relevance in forensic pathology. In any case, forensic pathologists are associated with a disproportionate share of examinations associated with cognitive bias, as discussed earlier in this paper.

Interpretative variability may be underestimated in the current study because many cases include only one trial expert in forensic pathology. This expert's opinion usually supports the prosecution theory of the case. If a death certificate does not list a homicide cause, it is unlikely for a suspect to come to trial, so this fact alone usually implies that the forensic pathologist's conclusions will support the prosecution and conviction. For example, in the 1981 murder case involving Randolph Arledge, forensic pathologist Dr. Nina Hollander testified "that the knife in question was consistent with being able to inflict the wounds that the deceased Carol Armstrong received." (Arledge v. State, 1985) DNA testing later exculpated Arledge and identified an alternate suspect. Nonetheless, Hollander's testimony was limited and valid, even if the knife was not the actual murder weapon.

There are uncertainties in forensic pathology findings that may or may not be reflected on a death certificate or in trial testimony. Effective defense may require the use of an independent expert to review the forensic pathology. Independent review may offer an alternative interpretation or raise issues regarding the limitations or errors in the prosecution expert's findings. In 52 instances, the defense did not obtain an independent review of forensic pathology findings that may have impacted the probative value of the findings. In 28 of these instances, there was no clear testimony error by the prosecution expert. In some cases, an appeals court made a clear decision that the defense was inadequate in its failure to present its alternative interpretation.

Postconviction evidentiary hearings often include consideration of forensic pathology findings. In recent years, innocence organizations have used evidentiary hearings to provide multiple experts who present interpretations that are compatible with the exoneration of the defendant. In many cases, the original trial did not include a defense expert or any discussion of the alternative interpretation. Whereas the original trial was heavily weighted in favor of prosecution experts, the evidentiary hearing may be heavily weighted in favor of defense experts. It is unclear if there is a basis to determine the "most valid" interpretation in these circumstances, although some experts do suggest that an inconclusive finding may be the best alternative for many probative issues. Most pertinently, there is little or no mechanism for the substantive resolution of competing interpretations *among the experts*, as opposed to by courts.

Forensic examiner variability may be related in some cases to uncertainties about the definition and use of task-irrelevant information by forensic pathologists. In many cases, a death certificate finding may be based on case information that is not directly relevant to medical information. In extreme cases, forensic pathologists will change findings based on police investigator views or input, even if that input contains no information relevant to the medical assessment.

The interaction between other forensic disciplines and forensic pathology demonstrates the difficulty and variability of pathological interpretation. Blood spatter, gunshot residue, and ballistic trajectory findings may all affect forensic pathology assessments, even when medical data contradicts the findings of other forensic disciplines. In some ways, this puts the forensic pathologist in the position to decide the ultimate guilt of a suspect, because prosecutors, judges, and juries often defer to the judgment of the forensic pathologist without a clear understanding of the limits of the pathologist's expertise. They have little or no basis to object to the use of non-medical information by the forensic pathologist.

An interesting example of forensic examiner variability and the use of task-irrelevant information occurred in the Michael Skakel case, which largely revolved around the victim's time-of-death estimate. The original expert, medical examiner Dr. Elliot Gross, put the time-of-death at between 10 pm and 5 am. Defense expert Dr. Joseph Jachimczyk put time of death at 10 pm but relied on witness information—not medical data—to narrow the possible time-of-death estimate. At the trial (held 27 years after the murder), the new Connecticut State Medical

Examiner, Dr. Harold Carver, said it was not possible to determine the victim's precise time of death and presented appropriate alternative interpretations consistent with a time of death between 930 pm and 10 pm and consistent with a time of death any time before 1 am. Carver acknowledged that the victim's family had begun to search for her by 1 am. Jachimczyk repeated his 10 pm estimate and stated that it was supported by witness statements about barking dogs at that time, a fact which was not necessarily task-relevant to a medical finding. The postconviction court overturned the conviction in agreement with Jachinczyk's view. As the dissenting opinion noted, "Although Jachimczyk was, of course, free to consider nonmedical evidence such as curfews and barking dogs in forming his opinion as to the likely time of death, there was no suggestion that he had any special expertise in the fields of teenage or canine behavior." The coding for this case includes four expert opinions, including the cause and manner determination by Gross and three time-of-death determinations. Only Jachimczyk's work was associated with a case error, which included the mischaracterization of the probative value and scientific foundation of his time-of-death determination.

An extreme example of a "battle of experts" occurred in the 1998 Neal Robbins case involving the death of a small child. The defense theory of the case held that the child's injuries were due to poor CPR efforts by her mother and other responders. Dr. Patricia Moore, assistant medical examiner in Harris County, concluded that death was homicide and "suffocation by compression." Dr. Robert Bux testified that the death should have been ruled "undetermined" because of the uncertainties related to the resuscitation attempts and indications that the child was still alive when admitted to the emergency room. (Robbins v. State, 2002) Three more experts weighed in postconviction to support the Bux conclusion, and Moore changed her view of cause of death to "undetermined." Another expert, Dr. Linda Norton, ruled that the death was "homicide" by suffocation, and the death certificate was changed to reflect that conclusion. (Ex parte Robbins, 2014) As the dissent in Robbins' successful habeas corpus appeal noted, "At worst, the result of a finding that five out of six pathologists cannot determine the cause of Tristen's death is only an admission that science cannot resolve the issue of whether Tristen's death was the result of a homicidal act. The jurors would have to decide that crucial question based upon the rest of the evidence." (Ex parte Robbins, 2014) As in other "battle of experts" cases, the critical issue is whether the various valid interpretations were clearly presented to the jury. It will never be possible to determine if Moore or Bux or the other experts were "right" in this case. It is possible to conclude that the possible interpretations were presented at trial, (Robbins v. State, 2002) so no forensic error should be associated with the forensic pathology in this case. That said, it does raise important concerns regarding the variability of forensic pathology decisions and the possible influence of cognitive bias.

Another example is the Robert Weitzel case. Psychiatrist Weitzel was convicted of 5 counts of manslaughter or misdemeanor homicide of elderly patients who had been prescribed large doses of psychotropic drugs and morphine. (Park, 2002) The case appeared to encompass issues related to end-of-life care, and the findings of prosecution and defense experts largely involved whether Weitzel had provided an appropriate standard of care. Later, it became clear that the issue involved prescription drug fraud for which the Drug Enforcement Administration had been investigating Weitzel during this period. As US District Judge Dee Benson stated, "This is a case of a doctor addicted to narcotics who was defrauding his patients and pharmacies to get it for himself. It's as simple as that." (Deseret News, 2002) Although Weitzel was acquitted on retrial of responsibility for the patient deaths, the medical opinions were appropriate and defensible on both sides. Although the case was a "battle of the experts," the experts agreed that Weitzel provided substandard care to his patients. They were not aware of the context regarding his drug addiction. Although it did not include deliberate context management, the Weitzel case does demonstrate that context management controls in forensic pathology may have unexpected, negative outcomes.

Forensic pathology continues to contribute to wrongful convictions. While other disciplines have benefited from technology improvements (DNA, mass spectrometry) or improved standards (fire debris, pediatric abuse assessment), such improvements do not appear to have affected the incidence of wrongful convictions associated with forensic pathology.

Forensic medicine

136 forensic medicine examinations	•93 cases in which forensic medicine was used •124 examinations involving pediatric abuse •99 forensic medical examinations with errors of any type •83 errors that contributed to conviction
System errors	•77 cases: inadequate defense •2 cases: police or prosecutor misconduct •6 cases: judicial errors
Forensic errors	•38 identification or classification errors •67 testimony errors 44 forensic report errors •58 examinations: possible cognitive bias 28 examinations: failure to follow best practices

Pediatric abuse cases are largely distinct from other types of cases in the dataset. They seldom include other types of forensic evidence or any involvement from "traditional" forensic science organizations. Only 18 of the assessments were connected to a public forensic science organization. Thus, most forensic medicine examinations were done independently of law enforcement by doctoral-level practitioners under standards set by a medical association, the American Academy of Pediatrics (AAP). In theory, this structure should have provided important mechanisms to support the independence of the forensic analysts and mitigate the risk of errors that contribute to wrongful convictions. In practice, the picture is more complex.

First, AAP standards have improved considerably over the last 25 years. In part, this was a response to wrongful convictions in the 1990's, particularly those related to alleged pediatic sexual abuse. The AAP has emphasized improvements to evidence collection, documentation, differential diagnosis, and recognition of limitations of evidence in ambiguous cases. Defense lawyers continue to question the validity of AAP standards, particularly in the area of abusive head trauma findings in cases involving infants and small children. (Tuerkheimer, 2009) (Donohue, 2003) These objections mirror divisions within the relevant medical research community. A total of 26 pediatric abuse examinations involved forensic examiner variability. These instances were primarily associated with the assessment of pediatric head trauma, usually in relation to shaken baby syndrome (SBS) or abusive head trauma (AHT).

Most examiners conformed to AAP standards at the time of trial, but their testimony would not conform to current standards because of the changes in AHT and sexual abuse standards. Some examiners did not conform AAP standards at time of trial. For example, AHT findings have always required a triad of findings related to optic and neural system damage, but some practitioners made findings based on only one element of the triad. Other practitioners failed to recognize underlying health conditions that may have contributed to medical findings. Of particular concern, many AHT and other pediatric physical abuse findings have discounted issues related to medical interventions in hospitals or other medical facilities that may have contributed to a child's injuries. In some cases, medical professionals associated with medical facilities testified to abuse and discounted possible medical errors, but it was unclear if appropriate context and bias concerns were addressed in the investigation or trial. Overall, many examinations were associated with possible cognitive bias, but only three with a fraudulent examination.

Pediatric abuse cases include difficult problems of interpretation for the medical practitioner who provides forensic testimony. In many cases, errors will occur when an examination is presented without an appropriate discussion of the limitations of scientific and medical interpretations. Sometimes, the defense does not (or cannot) provide a valid, alternative interpretation consistent with the defense theory of the case. Finally, appeals courts may overturn convictions on the basis of victim or witness recantations, even when the forensic evidence was valid and reliable. The assessment of the forensic testimony requires consideration of each factor.

A very large fraction of cases involved some issue with defense, including lack of independent examinations that would have been relevant to the probative value of the medical findings. Many defendants may not have access to the resources necessary to advance medical assessments consistent with the defense theory of a case. It is also possible that many wrongful convictions in pediatric abuse cases remain undetected because exonerations in the dataset often required affidavits or testimony from multiple defense experts. Given the evidence for inadequate defense during trials, it is likely that many convicted individuals lack the resources to retain appropriate experts and mount a successful appeal. Interestingly, exonerations in pediatric abuse cases are disproportionately associated with White defendants (91 cases) over African American defendants (23 cases), a phenomenon which may arise from postconviction resource disparities.

In the Brian Franklin case, the 13-year-old victim alleged that the defendant had sexually assaulted her. Examining physician Dr. Jan Lamb testified that "that there was a rupture in a certain area of the hymen indicative of blunt force trauma and that the injury observed on B.R. would be consistent with her bleeding at the time of the offense." (Ex parte Franklin, 2002) A forensic serologist testified for the defense that the color of a blood stain on the victim's clothing could have indicated that it was too recent to be deposited within the time frame of the alleged assault. Postconviction, the victim indicated that she had also been sexually assaulted by her stepfather. At the postconviction evidentiary hearing, Lamb clarified that she would not have been able to distinguish whether the victim had been assaulted by Franklin, her stepfather, or both individuals. (Ex parte Franklin, 2002) In the context of the trial, she had not implicated Franklin specifically and had provided valid and reliable testimony that reflected the physical evidence and its limitations. There was no forensic error in the case.

In the 1991 trial of Andrew Anthony Taylor, the associate medical examiner of Miami-Dade County, Dr. Valerie Rao, testified concerning bruises on the victim and colposcope-based observations of injuries consistent with sexual abuse. The victim recanted as an adult, attributing the false allegation to physical abuse from her mother. (Andrew Anthony Taylor vs.

State of Florida, 2017) As detailed in the decision concerning Taylor's petition for postconviction compensation, Dr. Rao "persuasively and credibly testified" concerning the victim's injuries and the context and limitations of her findings.

There were 63 examinations that were required for conviction and associated with testimony errors. Most often, the examiner mischaracterized the probative value of the evidence. The American Academy of Pediatrics (AAP) has modified its guidelines on specific issues regarding child abuse cases. (Committee on Child Abuse and Neglect, 1999) (Choudhary, et al., 2018) In 13 instances, the testimony conformed to guidelines in place at the time of trial but would not be in conformance with current guidelines. For example, in the 1995 Audrey Edmonds "shaken baby syndrome" case, all medical experts supported a finding of severe and inflicted physical trauma on the basis of the victim's injuries. The main issue concerned the lucid interval following the abuse, an issue which was relevant to the likelihood that Edmonds or another caregiver had been responsible for the abuse. Prosecution and defense experts testified to alternative views of the lucid interval interpretation and timing of the injuries. The research concerning abusive head trauma and lucid interval has advanced considerably in recent years, now supporting lucid intervals up to 72 hours. (De Leeuw, Beuls, Jorens, Parizel, & Jacobs, 2013) (De Leeuw, Beuls, Parizel, Jorens, & Jacobs, 2013) Notably, the AAP and other consensus committees do not yet recognize that lucid interval may be present in fatal abusive head trauma cases. During the Edmonds trial, there was an appropriate "battle of the experts" that provided the valid, alternative interpretations for consideration, although the trial would have benefited from the updated research on the lucid interval question.

Forensic medical interpretation requires extensive documentation, full disclosure of medical history, and science-based standards. These elements were not always present, such as in the 15 cases associated with errors related to missing reference or background data. The 2004 case of Randy Liebich is instructive. Medical testimony and autopsy findings by forensic pathologist Dr. Darinka Mileusnic-Polchan led to the conclusion that the decedent child's swollen brain and other symptoms were indicative of abuse by the defendant and to discount the possibility of a post-trauma lucid interval. (People v. Liebich, 2016) Postconviction, Mileusnic-Polchan revised her conclusions, citing new information about the child's preexisting health conditions that indicated abuse may have occurred from five to ten days before death. In fact, the wrongful conviction was closely related to significant gaps in the communication of salient medical findings to Mileusnic-Polchan and other medical experts. Among other items, Mileusnic-Polchan was not informed concerning a flawed CT scan interpretation and the child's preexisting pancreatitis and myocarditis. (People v. Liebich, 2016)

In the 1985 Harold Snowden case, the key issue was the detection of vaginal Gardnerella vaginitis (GV) in an alleged female victim and gonorrhea in the throat of an alleged male victim. All three alleged victims were under six years old. At the time, GV was considered a sexually transmitted disease, but the AAP established in a 1992 paper that it could be found in the absence of abuse. (Ingram, et al., 1992) Also, the gonorrhea test can have false positives, an issue possibly evidenced in the case when a subsequent test of the alleged male victim was negative. The AAP still considers gonorrhea to be a reliable indicator of a sexually-transmitted disease. The defense presented an expert who raised the issue of the reliability of gonorrhea tests for consideration during the Snowden trial. The NRE apparently classifies this case as associated with F/MFE on the basis of invalid testimony from a child psychologist who testified that "99.5% of children tell the truth" about abuse. (Snowden v. Singletary, 1998)

Pediatric sexual abuse cases raise difficult challenges for the criminal justice system and forensic practitioners. AAP recognizes that pediatric sexual abuse will not produce physical findings in every case. (Committee on Child Abuse and Neglect, 1999) In wrongful conviction cases, some medical practitioners provided misleading testimony that abuse was indicated on the basis of no physical findings while relying on non-medical case findings. This has never conformed to AAP standards. In other cases, medical practitioners testified that abuse was *possible* but not substantiated in the absence of medical findings. Such testimony conforms to AAP standards and was not coded as an error in the present study. Similarly, in more recent cases, medical practitioners have continued to present abusive head trauma findings but with additional context about the inherent uncertainties. Appropriately limited abusive head trauma findings conform to AAP standards and have not been coded as errors in the current study. It should be noted that pediatric abuse cases are complex, so a study finding concerning the reliability of one expert's testimony cannot be used to draw conclusions about the guilt or innocence of a defendant or the appropriateness of an exoneration.

The recent exoneration of Dane Krukowski and Codie Lynn Stevens provides a notable example of these issues. Their three-month-old child exhibited stomach issues and injuries on a hospital visit in 2015. The parents had visited a chiropractor, who had "adjusted" the child a few weeks before with the intent to treat the child's illness. Eight different experts weighed in on the child's condition at trial and in the postconviction hearings. The emergency room physician concluded non-accidental trauma early in the child's treatment, and hospital radiologists supported that conclusion with findings of healing fractures. A neurologist found retinal and subdural hemorrhages indicative of "shaken baby syndrome" or abusive head trauma. Defense experts said the injuries could have been caused by a bathtub fall. In postconviction, a biomechanical engineer concluded in error that a fall could be the *only* possible source of the child's injuries. Only one expert, Allegheny County Chief Medical Examiner Karl Williams, provided a full picture of the possible injury sources, although only during the postconviction review. Williams cited the chiropractic manipulation as a possible source of the baby's injuries, clarified that the medical findings were not consistent with current AAP guidelines for a conclusion of abusive head trauma, and criticized the hospital-based experts for a failure to perform an appropriate differential diagnosis. Nonetheless, Williams left open the possibility that the child could have been abused or that the parents failed to provide appropriate medical care. Williams was the only expert associated with a forensic science organization, and it is possible this background was conducive to a more balanced view of the evidence in the case.

Like forensic pathology, forensic medical assessments associated with pediatric abuse continue to contribute to wrongful convictions at a steady pace. The cases are difficult to overturn, and exonerating evidence such as DNA is generally not available. Hence, there may continue to be a pediatric abuse exonerations, but case uncertainties may continue to limit the overall incidence.

DNA

64 DNA examinations

- 63 cases in which DNA was used
- 41 DNA analyses with errors of any type
- 19 errors contributing to conviction

System errors

- 7 cases: evidence handling or chain of custody errors
- 30 cases: inadequate defense
  18 cases: police or prosecutor misconduct

Forensic errors

- 9 identification or classification errors
- 14 testimony errors
- 7 examinations associated with possible cognitive bias
- 22 examinations that would be impacted by improved DNA methods

DNA analysis is not immune to analyst errors. The nine identification errors included some failures to follow practice standards and some cases in which the biological evidence was limited or a mixture of multiple sources. The 2016 Mayer Herskovic case included both elements. A sample from the complainant's sneaker contained 97.5 picograms of DNA material, less than the amount (125 picograms) required by the kit manufacturer for reliable DNA testing. (People v. Herskovic, 2018) The New York City Office of the Chief Medical Examiner (OCME) used a forensic statistical tool that concluded that the DNA sample was 133 times more likely to have originated from a mixture of the complainant and Herskovic than the complainant and an unknown person. The analysis relied on Caucasian population estimates that did not take into account Herskovic's Hasidic ethnicity. Also, two Herskovic alleles were missing from the composite DNA profile produced by OCME, while only four alleles matched Herskovic's profile after accounting for masking from the complainant. The OCME dropped the unvalidated statistical tool after the Herskovic conviction was overturned. (People v. Herskovic, 2018)

The subjective elements of the DNA interpretation in the Herskovic case may have been influenced by cognitive bias. The issue of cognitive bias in DNA analysis has been highlighted in the research literature (Jeanguenat, Budowle, & Dror, 2017) and a 2002 wrongful conviction in the Kerry Robinson case. (Dror & Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, 2011) The Robinson case was complicated by DNA mixture interpretation issues. Georgia Bureau of Identification (GBI) analyst Brad Pearson obtained 13 alleles from a sexual assault kit sample, but 11 were associated with Tyrone White, who had confessed and implicated Robinson. (Robinson v. State, 2003) The other two alleles matched Robinson and a third suspect, Sedrick Moore. Pearson said Robinson "could not be excluded" as a potential donor, but also stated on redirect that the chance of finding other individuals whose alleles would match those in the sample was "very, very low." Postconviction, Greg Hampikian became an advocate for Robinson's exoneration and produced a research paper in which 16 out of 17 "context-free" examiners disagreed with Pearson's conclusion. (Dror & Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, 2011) The result has been interpreted as a form of target bias in which Pearson searched for the Robinson alleles in the electropherograms, which were reproduced in part in the paper. The paper does not clarify that 11 of the 13 alleles came from White or that Pearson held that either Robinson or Moore could have contributed the other two alleles. Thus, the target bias interpretation may be untenable. Instead, the significant problems may have been poor DNA mixture interpretation methods at the time of trial, Pearson's exaggerated statistical testimony, and the lack of laboratory policy on the interpretation of partial profiles and mixtures. Hampikian's research may have been influenced by his own cognitive biases as an advocate for Robinson.

The development and enforcement of DNA analysis standards has played a critical role in the prevention of forensic errors. Seven DNA wrongful convictions are associated with failures to follow best practices or standards, including the Herskovic and Robinson cases. The Houston Police Department (HPD) laboratory provides an useful case study in the impact of gaps in the management of DNA laboratories. The HPD's wrongful conviction of Josiah Sutton has been detailed by independent reviews by William Thompson (Thompson W. , 2003) and Michael Bromwich (Bromwich M. R., 2007). According to Bromwich's report, the DNA analysis was compromised by failed differential extractions, failed controls, poor temperature control of the dqAlpha assay, lack of technical review, poor documentation, and statistical misinterpretation. The defense attorney took funds to do independent DNA testing but did not actually obtain the testing. The HPD experience is an extreme example, but the importance of rigorous and enforced standards and quality assurance persists as a common theme across wrongful conviction case histories.

Police or prosecutor misconduct was disproportionately represented within DNA cases. In nine of the 18 cases associated with police or prosecutor misconduct, exculpatory evidence was suppressed or ignored. In the 1998 Nathaniel Hatchett case, DNA testing excluded Hatchett as a contributor to semen found in the sexual assault kit. The DNA evidence was discounted because of Hatchett's false confession, and it was theorized that the victim's husband was the source. It was not disclosed to the defense that the husband also did not match the semen DNA. (Hatchett v. City of Detroit et al, 2010) Similarly, in the 1997 Keith Cooper case, a partial DNA profile excluded Cooper as a possible contributor, but that fact was not shared with the defense. Postconviction DNA testing was able to produce a complete profile and a cold hit to the source. In the Buncombe Five case, the district attorney suppressed exculpatory DNA evidence after police had elicited several false confessions. (State v. Kaginyera/Wilcoxson, 2011) One defendant, Kenneth Kagonyera filed for DNA testing in August of 2001, unaware that the results had exculpated him and his four codefendants in March. He provided a false confession in November. The North Carolina Innocence Inquiry Commission formally exonerated the Buncombe 5 in 2011. (State v. Kaginyera/Wilcoxson, 2011) In the 1990's, it was not uncommon for exculpatory DNA to be discounted when it did not identify an alternate suspect. Because dq-Alpha, polymarker, and RFLP methods weren't well-suited to cold hits, many wrongfully-convicted defendants were exonerated only after the advent of standardized STR markers in the late 1990's. Further, there was reasonable skepticism about DNA reliability in the scientific (National Research Council (US) Committee on DNA Technology in Forensic Science, 1992) and legal (Baird, Neufeld, & Scheck, 1990) communities. These perceptions may have contributed to the discounting of DNA results by law enforcement into the post-2000 period.

Latent fingerprint and palmprint comparison

102 latent print examinations

- 95 cases in which latent fingerprint or latent palmprint evidence were used
- 48 latent print comparisons with errors of any type
- 18 errors contributing to convicition

System errors

- 5 cases: evidence handling or chain of custody errors
- 21 cases: exculpatory latent print evidence present
  18 cases: police or prosecutor misconduct

Forensic errors

- 17 identification or classification errors
- 23 testimony errors
- 8 fraudulent examinations
- 10 examinations: best practices and standards not followed

The dataset includes 102 latent print comparisons, with 48 case errors, 18 of which contributed to a wrongful conviction. Most examinations were valid and many did not inculpate or exculpate the defendant. Exculpatory fingerprints or palmprints were present in 21 cases, meaning print evidence was clearly probative and reasonably associated with an assailant other than the defendant.

The case errors included 17 incorrect individualizations, eight of which were fraudulent. For example, three fraudulent identifications were associated with a New York State Police scandal of the late 1980's involving planted fingerprint evidence. (Roth, 1997) In the Jonathan Irons case, the exculpatory latent print was correctly analyzed but suppressed by the examiner, who was also the primary police investigator. (State ex rel. Schmitt v. Green, 2020) In the 1968 DePalma case, examiner James Bakken falsified the prints using a photocopy machine. The 2001 Stephan Cowans case was likely at first an "honest mistake" of misidentification by an untrained examiner who did not follow the standards of the discipline. (Smith R. , 2004) The examiner realized his mistake before the trial but continued to mislead fact-finders. The Cowans case is the clearest example of possible cognitive bias in a wrongful conviction influencing a latent print identification, because the examiner appears to have been influenced by the knowledge that the murder victim was a Boston police officer.

Many observers have assumed that contextual bias is a major cause of pattern evidence errors that lead to wrongful convictions. (Dror I. , Biases in forensic experts, 2018) There is limited basis in wrongful convictions to substantiate that claim. In the Aquirre-Jarquin case, examiner Donna Birks testified that Aguirre-Jarquin was the source of a latent palmprint that was not suitable for comparison and relied on only seven points of comparison. (Clemente Javier Aquirre Jarquin v. State of Florida, 2008) A postconviction expert from the Florida Department of Law Enforcement, Christina Barber, confirmed that Birks used an unsuitable print in this case and five others, including one incorrect identification. In the Aguirre-Jarquin case, Barber found that four of the seven points of comparison used by Birks did not align with features in the defendant's print. The error was not related to known cognitive bias, though target bias (looking for features of the suspected source in the questioned evidence) may have played a role. (Dror I. E., Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias, 2020) In the Lana Canen case, an untrained, uncertified examiner made a similar misidentification based on too few points of comparison and subject to confirmation bias. The examiner later recanted his trial testimony. He stated that he "was trying to help out Elkhart City [police]." (Lana Canen v. Dennis Chapman, 2017) In the Warney case, examiner Robert Garland associated a latent print with the defendant on the basis of only three features of comparison, a conclusion that was outside the standards of the discipline and exceeded the limits of science. (Warney v. Monroe County, 2009)

In two unusual cases, defendants were able to retain defense experts who provided exculpatory testimony about print identifications. In the Beniah Dandridge case, the latent print attributed to Dandridge actually belonged to his son, and the comparison was viewed as difficult. (Double Loop Podcast, 2021) A defense expert testified at the trial that the print was an exclusion, but the prosecutor made inflammatory statements about the expert, who was a retired FBI examiner. (Possley, Beniah Alton Dandridge, 2015) Although the prosecution examiner did not follow best practices, she was experienced and well-trained. As in the Aquirre-Jarquin case, the only possible cognitive bias may have been target bias.

The Richard Jackson case included three prosecution examiners who produced erroneous identifications and two defense examiners who testified correctly that the prints were exclusions. The original error was not subject to blind verification, and it is likely that the two concurring examiners were influenced by confirmation bias. Although only one prosecution examiner was certified, there is no other basis to determine that the examiners were not trained, and an appeals court found that the errors were honest mistakes. (Jackson v. Paparo, 2002)

In assessing the fact patterns in these cases, the roles of poor training and failures to follow the standards of the discipline should not be underestimated. Even in the New York State Police scandal cases, other fingerprint examiners had pointed out the repeated documentation and practice failures of the examiners who later were found to be planting evidence. (Roth, 1997) In any case, procedures to limit task-irrelevant contextual information will not be successful if a forensic science organization is not capable of maintaining conformance with practice standards. Further, the case errors were not subtle or difficult to determine from technical review. Several cases involved the use of prints unsuitable for comparison or very small numbers of points of comparison. Other cases involved experienced defense experts who directly contradicted the identifications, a clear and unusual "red flag," especially in latent print cases.

There were 23 testimony errors, including 15 in which the probative value of the evidence was mischaracterized. For example, in the Steven Chaney case, the examiner correctly identified Chaney's print but made speculative statements about the age of the print. (Ex parte Steven Mark Chaney, Applicant, 2018) The same examiner made similar statements about shoeprints in the Chaney case. In the Michael Seri case, the examiner provided misleading testimony that implied a possible identification. He stated that he "couldn't rule out the possibility Seri's prints were on the two library books because police didn't take enough of his handprint for comparison." (Martineau, 2003) Eventually, an alternate suspect was matched to the prints.

In the Glenn Ford case, the examiner made a highly unusual class-level association of Ford with a latent print on the basis of a whorl pattern in both the latent and Ford's reference prints. The examiner said 35% of the population would be included as possible contributors to the print. This case was coded as using an unvalidated method outside the best practices of the discipline, exhibiting cognitive bias, reporting a statistical characterization without a scientific basis, an incorrect classification based on an incorrect interpretation, invalid testimony concerning the statistical and probative value of the evidence, and inadequate defense.

As seen above, most misidentifications and testimony errors were associated with nonconformance with the standards of the discipline. In short, wrongful convictions suggest that certified latent print examiners applying the standards of the discipline are extremely reliable. There is no evidence in wrongful conviction cases to conclude that latent print examiners make errors of identification except in exceedingly rare instances. This contention is now supported by extensive research. (Ulery, Hicklin, Buscaglia, & Roberts, 2011) It has been argued that latent print errors may be common but aren't detected because they are not associated with DNA evidence that could exculpate a suspect. (Cole S. A., More Than Zero: Accounting for Error in Latent Print Identification, 2005) Most latent-print-associated exonerations were resolved without the use of DNA evidence, and most exonerations overall do not rely on DNA. Further, the NRE has now documented about 3,000 exonerations, including the 732 associated with possible forensic errors. Based on this accumulated data, it is reasonable to conclude that latent print errors are not a major contributor to wrongful convictions, although it is impossible to use wrongful conviction data to determine the rate of latent print errors in the field overall.

The data does support one clear gap: latent prints are underutilized. Valid and probative exclusions were discounted in many cases. In several cases (Miles, Philen, Nash), technological improvements in AFIS systems would have identified an alternative suspect who was the actual perpetrator. Although the Aguirre-Jarquin case involved a print unsuitable for comparison, many other defendants claim that suitable and probative prints were not used when clearly probative. Overall, as many as 21 defendants might have been exonerated by latent print evidence that was fully utilized at the time of trial. The underutilization of exculpatory latent prints should be taken into account in the design of forensic science reforms and case processing strategies. (Rairden, Garrett, Kelley, Murrie, & Castillo, 2018)

Ballistic evidence

102 ballistic evidence examinations	•85 cases in which ballistic evidence was used •62 firearms identification cases with 66 examinations, 26 errors of any type, and 14 errors contributing to conviction •28 gunshot residue cases with 29 examinations, 19 errors of any type, and 11 errors contributing to conviction •7 ballistic trajectory cases with 7 examinations, 5 errors of any type, and none contributing to conviction
System errors	•22 cases: inadequate defense 9 cases: police or prosecutor misconduct 9 gunshot residue evidence collection, contamination, or crime scene errors
Forensic errors	•9 firearms identification errors 8 gunshot residue classification errors •24 testimony errors 17 examinations: possible cognitive bias

As used in this section, ballistic evidence is considered to broadly include firearms identification, gunshot residue, and ballistic trajectory analysis at the crime scene. Compositional bullet lead analysis (CBLA, 3 examinations, 3 case errors) was included under firearms identifications.

Mistaken firearms identifications rarely lead to wrongful convictions in the modern era. Of the nine identification errors associated with firearms identification, five arose prior to 1930 and the development of the comparison microscope and associated practice standards. Two modern cases, Desmond Ricks and Darrel Siggers, were associated with poor evidence handling and adherence to standards in the Detroit police crime laboratory, which was closed in 2008 for issues related to ballistics comparison. (Ricks v. Pauch, 2020) (Anderson, 2018)

In the 1986 Anthony Hinton case, autopsy bullets were matched to a gun recovered from Hinton's mother's mattress and established a link among three separate shootings. Two prosecution examiners testified that the six recovered bullets "had all been fired from the same gun: the revolver found at Hinton's house." (Hinton v. Alabama, 2014) The judge granted $1000 to the defense to hire an independent expert, but Hinton's attorney was not able to find a competent examiner for that amount. The defense expert, Andrew Payne, concluded that corrosion in the revolver made it impossible to conclude that any particular bullet had been fired from that gun and stated that the six evidence bullets did not match one another. On cross, the prosecutor discredited the defense witness, pointing out that Payne had only one eye (making it impossible for him to use a comparison microscope as intended). Payne's only relevant training was a degree in civil engineering from 1933, more than 50 years before the trial. It was later determined that the state's examiners had not made a valid match because they didn't record any land or groove information for any of the six evidence bullets. Thus, all trial experts in the Hinton case exhibited gross incompetence similar to that seen in the Detroit laboratory.

Finally, the Curtis Flowers case involved the identification testimony with an unsupported probabilistic statement and relied on slugs that were of no value for forensic comparison. (Flowers v. State, 2003)

Two cases demonstrate some of the unique complexities of firearms identification as a discipline and are considered identification errors by some observers. (Siegel, A System on Trial : Sentencing the Wrong Man to Die, 1987) In the 1973 Joseph Brown case, the victim was shot with a .38 special round, but Brown's handgun was a .38 regular. The .38 regular handgun can only accommodate shorter rounds than typical .38 special ammunition. FBI examiner Courtland Cunningham made an inconclusive determination, stating that the Brown handgun could not be eliminated because many .38 regular handguns were modified to permit the loading of .38 special ammunition. (Joseph Green Brown, Petitioner appellant, v. Louie L. Wainwright, Secretary Florida Department of Corrections, 1986) Such modifications were well-known at the time. Cunningham had testified to the same issue in the Kennedy assassination case, in which Lee Harvey Oswald used a .38-200 British "regular" handgun that had been modified to use .38 special ammunition. (Testimony of Courtland Cunningham and Joseph D. Nicol, 1964) Because Cunningham was not able to inspect the Brown handgun, he was unwilling to eliminate it as a weapon that might be associated with the autopsy bullets.

The Patrick Pursley case involved two bullets—one recovered at autopsy and one from a car dashboard—and two shell casings from the crime scene. Pursley's 9mm Taurus pistol was identified by examiner Daniel Gunnell as the murder weapon "to the exclusion of all other firearms." (People v. Pursley, 2018) Defense expert David Boese disagreed, saying the murder weapon was a Taurus 9mm, but not necessarily Pursley's. Postconviction in 2011, Illinois State Police (ISP) examiner Russell McLain performed new test fires to enter the case into the IBIS database. His reexamination generally agreed with Pursley's conclusions. ISP conducted yet another examination of the evidence in 2012. ISP examiner Beth Patty agreed with Pursley's conclusions about the cartridge casings but reached an inconclusive determination on the recovered bullets. Pursley reexamined the evidence and agreed with Patty, attributing the inconclusive finding to handling or degradation of the evidence materials. Finally, another examiner, John Murdock, weighed in with his own examination using higher-power microscopes than the ones available to Gunnell, Boese, McLain, or Patty. Murdock excluded the Pursley weapon as the source of the evidence bullets and casings, though he clarified that his exclusion was only possible because of the higher magnification microscopy that he was able to perform. A fifth examiner, Chris Coleman, performed a verification of Murdock's conclusions. Coleman knew that Murdock had performed the comparisons but was not aware of Murdock's results. Coleman agreed with Murdock's findings. Afterward, he performed an open, technical review of Murdock's notes and comparisons. At an evidentiary hearing, Gunnell stated that Murdock's observations supported an inconclusive finding, not an elimination. The conviction was overturned on the basis that Murdock's reanalysis constituted newly discovered evidence. (People v. Pursley, 2018)

The Pursley case is unlike any other firearms identification case in the dataset. All other misidentifications can be associated with training and practice deficiencies. The six Pursley examiners largely conformed to consensus standards, though Gunnell overstated the statistical weight of his conclusion at trial. The case is a good example of the value of improved technology, which was an important element of Murdock's "newly discovered evidence." The case also demonstrates that firearms examination is vulnerable to expert variability. In some respects, this is unsurprising. Bullet comparison, in particular, must consider shot-to-shot variability, which in some firearm-ammunition combinations may be comparable to variations from one gun to another. As a result, firearm identification, elimination, and inconclusive findings are inherently subjective. That said, the process may be more predictable because it is inherently a physical phenomenon, so it is essentially free of the biological variability of disciplines like friction ridge comparison. Researchers are making progress on the development of mathematical frameworks for ballistics toolmark comparisons (Zheng, n.d.). In the meantime, it should be noted that the forensic examiner variability seen in the Pursley case was observed in only two other cases in the dataset, the 1924 Hoffman case and the Hinton case described above, both of which involved issues in training, documentation, and practice standards.

Some outdated methods were associated with wrongful convictions, though these methods may have been applied in alignment with contemporaneous testimony standards. CBLA was used in three cases, including the Philip Cannon conviction in 2000. (Cannon v. Polk County/Polk County Sheriff, 2014) By that time, many laboratories had abandoned CBLA as unreliable, and the Oregon state laboratory declined to perform the testing. Prosecutors used an Oregon State University researcher, Michael Conrady, to conduct and present the CBLA. The judge did not sustain a defense motion to exclude CBLA as unreliable. Conrady concluded that the evidence and reference bullets were “analytically indistinguishable.” This statement was technically true. Within a scientific context, the language would have conveyed that the compositions of the two sources were the same within Conrady’s definition of error limitations. In court, the language conveyed an identification conclusion. No known wrongful convictions are associated with CBLA after the National Research Council report on the subject. (National Research Council, 2004)

GSR is implicated in a significant number of wrongful convictions. GSR sampling may be done poorly or at a time when possible secondary transfer undermines the probative value of a result. Confirmatory testing may be done poorly or not at all. In some cases, a conclusion of GSR was associated with an insufficient number of particles or a finding of only lead in a particle, thus limiting the reliability of the analysis. The limitations of GSR may not be communicated effectively to investigators, prosecutors, or the court. The 1998 Tyrone Jones case included possible issues on these elements. Maryland State Police examiner Joseph Harant found two particles of GSR, one with two elements and one with three elements. (Hanes, 2005) The association of a particle with GSR based on two elements is consistent with current standards but is not accepted by all laboratories. (Gunshot Residue Subcomittee Chemistry Scientific Area Committee Organization of Scientific Area Committees for Forensic Science, 2020) The case included possible secondary transfer. Harant did not provide the testimony, which was given by a colleague, Daniel Van Gelder, and the defense claimed that Van Gelder mischaracterized the probative value of the GSR. (Jones v. State, 2000) As in the Jones case, court decisions related to GSR reflect uncertainties about GSR interpretation issues that are difficult to resolve in a legal context.

An outdated method associated with wrongful convictions is atomic absorption spectroscopy (AAS) for gunshot residue (GSR) testing. Like the chemical methods used in CBLA, AAS is an accepted scientific method for chemical analysis. The application of AAS in a forensic context requires an empirical foundation on compositional variations among sources, sampling issues, and the limitations of AAS in particular, among other considerations. AAS is a reliable method to determine the overall composition of samples that may be associated with GSR, but it lacks the ability to use morphological analysis or determine the chemical composition of specific particles with a morphology consistent with GSR. (Dalby, Butler, & Birkett, 2010) Further, AAS is an inherently destructive test and has some limitations for the testing of barium and antimony. Forensic laboratories have transitioned to the use of scanning electron microscopy (SEM) and the complementary technique, energy dispersive x-ray spectroscopy (EDS) to address the shortcomings of AAS. In general, the wrongful convictions associated with AAS were not overturned on the basis of the limitations of AAS as a method. Like other GSR-associated cases, the primary issues involved sample collection and contamination concerns. In the 1996 Glover/Johnson/Wheatt (GJW) murder case, the AAS testing and testimony were valid and reliable, but GSR sampling was not conducted at the crime scene. (State v. Glover, 2016) One defendant, Derrick Wheatt, was sampled eight hours after the shooting and had a high level of GSR, while another defendant, Laurese Glover, was negative for any GSR. The possibility of GSR contamination from the police vehicle or other sources undermined the otherwise reliable forensic analysis. (Berk, Rochowicz, Wong, & Kopina, 2007)

The issue of cognitive bias arose in the postconviction proceedings of the GJW case. (Transcript of Evidentiary Hearing, 2008) The original examiner, Dr. Sharon Rosenberg with the Cuyahoga County's Coroner's Office, had testified, "I have no knowledge of the situation. I normally do not make any attempt to be involved with the police examinations or the reports until after all facts are in. This is not something I care to have any influence from outside sources on." Postconviction defense expert John Kilty, formerly with the FBI laboratory, criticized Rosenberg's view, saying that the contextual information concerning the sampling delay should have been used to conclude that any GSR testing would be "inherently flawed." He stated, "We had the biggest phone bill in the FBI laboratory because we'd call the contributors … to get verifiable information about the case … because it's our obligation to interpret the results after we do the exam." In more formal parlance, Kilty argued that contextual information would be required to triage forensic analysis and to resolve activity-level propositions. In the GJW case, GSR results could have been related to at least three types of activities: the shooting; being a bystander to the shooting; or exposure to contamination after arrest. Even if a chemist performing AAS or SEM/EDS is unaware of contextual information, another individual must use this information to provide a reliable interpretation of the results. In essence, Kilty argued that the GJW wrongful conviction was based on the failure to conduct a criminalistic analysis to support the GSR chemical analysis.

Crime scene investigation

Twenty-six crime scene investigations (CSI's) were coded as distinct forensic science examinations. This represents a small subset of the number of CSI's performed in wrongful convictions, but sufficiently detailed data about the crime scene process could only be obtained in a small number of cases. Crime-scene-related errors were identified in 89 other examinations connected to other forensic disciplines. For example, serology examinations were compromised by crime scene errors in 16 cases, and fire debris investigation in eight cases. In many cases, it is difficult to determine if evidence handling issues arose at the crime scene or later in the process. There were 115 examinations compromised by some type of evidence handling or chain of custody issue. This subset included failures to collect or store evidence properly at the scene as well as failures to send the evidence for analysis or when the evidence was lost.

Crime scene investigation may be compromised by early assumptions about the nature of the incident under investigation. This phenomenon is most readily observed in cases in which a forensic pathologist must make a determination of homicide or suicide. Because suicide is much more common, investigators may not secure a location or collect evidence as they would at a crime scene. (van den Eeden, de Poot, & van Koppen, 2019) In the Beverly Monroe murder conviction, the case was initially treated as a suicide and very little evidence was collected from the scene—a clear example of base rate bias. There were failures to collect cigarette butts, search for a suicide note or other documents, preserve the decedent's clothing, dust for fingerprints, or collect trace evidence. (Monroe v. Angelone, 2003) When investigators then changed their view, the case relied on forensic findings that had been changed or compromised by the poor initial investigation. Similar issues compromised the three accident reconstruction cases in the dataset (Judith Fritz, Dana Payne, and William Campbell).

In the John Tomaino case, a cartridge casing was found in the fold of a blanket under the decedent's leg. (People v. Tomaino, 1998) The casing and other evidence were considered dispositive that the case was a homicide. In Tomaino's retrial, it became clear that poor crime scene work had compromised the entire investigation. (Herbeck & Prohaska, 1991) The bed moved during crime scene processing. The decedent had been moved by the crime scene photographer to allow him to see the entry wound. The sheets and bedding were not collected until days later and after cleaners had moved them to a garage.

Chain of custody issues occurred during the transfer of evidence between laboratories. In the Vidale McDowell case, the forensic pathologist collected fingernail clippings and claimed they were taken to the crime laboratory by homicide detectives, but the evidence was lost at an unknown point. (Mullen, 2004)

Blood spatter examinations are also closely related to crime scene investigations. In six cases, blood spatter evidence was not collected or documented properly at the crime scene. In the Brad Jennings case, crime scene investigators prematurely discounted the possibility of homicide, leading to inadequate collection of evidence related to blood spatter and potential gunshot residue (GSR). (Jennings v. Nash, 2020) The victim's robe was seized months after her death, thus limiting the probative value of a positive test for blood on the garment and a negative GSR test. In the Birch/Henning case, Dr. Henry Lee produced a speculative conclusion that attempted to account for the absence of blood spatter on the defendants' clothing. (Henning v. Commissioner of Correction, 2019) Lee relied on photographs of blood spatter from the crime scene and seized evidence that had not been subject to conclusive testing for the presence of blood. Although Lee's conclusion may have been one valid interpretation of the blood spatter, he overstated the probative value of his findings and demonstrated clear prosecution bias on the basis of unrelated information (such as the discovery of the defendant's credit card in the victim's toilet). Similar contextual bias challenges exist for other disciplines that cannot avoid contextual information. In the case of blood spatter, wrongful conviction data supports the findings of research with regard to contextual effects and general reliability. (Taylor, Laber, Kish, Owens, & Osborne, 2016)

Cases with no forensic error

Across all 732 cases in the dataset, there were 500 forensic examinations that were not associated with any errors, although many cases included other forensic evidence associated with errors. In 97 cases, there were no errors of any type associated with any forensic evidence. In these cases, the forensic evidence was collected, analyzed, and reported correctly and used appropriately in trial proceedings. The conclusions of this study diverge from conclusions coded as "False/Misleading Forensic Evidence" by the NRE. The basis for the NRE coding is unknown in many cases. Further, the documentation of forensic science issues may be variable for the NRE and the current study, leading to uncertainties in case assessment.

There are 34 "no error" cases that include hair comparison or serology. The NRE appears to rely on the GN interpretation for serological interpretation, an approach which does not reflect the validity of microscopic confirmation of spermatozoa to account for masking effects. The NRE also attributes errors to many inculpatory hair comparisons. As seen in the 1991 Timothy Bridges case, it is possible to conduct a valid hair comparison that encompasses a class of possible sources that include both an innocent defendant and the actual contributor. As observed in that case, hair comparison testimony may be valid if it is presented within appropriate standards. (US Department of Justice, 2019)

In some instances, a wrongfully convicted person may be exonerated when new technology provides exculpatory evidence. DNA exonerations are one example. Some cases in the 1990's used early DNA methods that required large amounts of sample or produced only weakly probative results. The original DNA work was not in error in these cases, because the technology to perform more sensitive and selective analyses was not yet developed and available to forensic science organizations. Similarly, advances in chemical analysis and AFIS systems now enable more powerful forensic analysis. In the 1988 LaMonte Armstrong case, a latent palmprint was used to make a cold hit when it was entered into an AFIS n 2012. (Armstrong v. City of Greensboro, 2016) There was no forensic error, because AFIS systems were not sufficiently developed to make the cold hit at the time of trial in 1995. Similar AFIS cold hits exonerated several wrongfully convicted defendants.

In many disciplines, standards have changed considerably, particularly in fire investigation and pediatric abuse assessment. There are 19 wrongful conviction cases that involved valid fire investigations and pediatric abuse assessments. The original analysis and testimony was coded as valid if the examiner recognized and conveyed the limitations and uncertainties in the discipline.

The assessment of some cases may be considered ambiguous or debatable. In the 1993 Sean Ellis murder case, simultaneous latent fingerprint impressions were used to inculpate his alleged coconspirator, Terry Patterson. The Patterson case was noteworthy as a court test for the acceptance of simultaneous impressions, which involve the use of points of comparison using multiple partial prints deposited at the same time. The Massachusetts court rejected the simultaneous prints on the basis that the technique had not achieved general acceptance (Commonwealth v. Patterson, 2005). The Scientific Working Group on Friction Ridge Analysis, Study and Technology established a standard for simultaneous impression examination after the Patterson decision in 2008. (Scientific Working Group on Friction Ridge Analysis, Study and Technology, 2008) Ellis was convicted separately from Patterson. Ellis was not inculpated by any latent print evidence, though the Patterson association was introduced at his trial. Ellis was associated with the murder weapon by firearms identification and witness testimony, and the firearms examination was valid and reliable. His conviction was overturned based on corruption issues in the Boston police department related to the murder victim. In any case, the simultaneous print was not a misidentification of Patterson, and no forensic error was associated with Ellis' conviction.

In 12 cases, no forensic evidence was considered, according to public documentation of the cases. In the 2006 Emmanuel Mervilus and Ayodeji Oladapo cases, police investigator John Kaminskas supported polygraph testing using unreliable statistics, but that is not classified as a forensic technique under the current study. (Kaminskas v. Office of the Attorney General, 2019) The Oladapo case also involved Child Sexual Assault Accommodation Syndrome testimony, but no physical evidence or other forensic examination. (State v. A.O., 2009) In the John Quattrocchi case, psychologists testified concerning the reliability of flashback recollections of alleged sexual abuse by the defendant. (State v. Quattrocchi, 1996) Psychological assessments are not included as forensic examinations in this study. In the 2008 Ignacio Ixta case, detective Alex Arnett of the Oxnard Police Department testified as a gang expert, but that testimony does not constitute a forensic issue. (People v. Ixta, 2012) There may have been forensic examinations related to a bullet that went through the victim and landed in a wall at the crime scene, but there is no basis to support any forensic error if such examinations were conducted.

The federal Ali Sadr case was connected to electronic documents and Internet Service Provider records supporting Sadr's alleged support for a conspiracy to violate sanctions against Iran and related financial crimes. Case proceedings were affected by disclosure violations by prosecutors. (United States v. Nejad, 2020) No substantive information concerning digital evidence examinations was documented in the public case records, and it appears that the substance of any digital evidence work was unrelated to the prosecutor violations in any case. Muneer Deeb was convicted in a 1985 murder-for-hire trial. In the related Calvin Washington and Joe Sidney Williams cases, the defendants were wrongfully convicted on the basis of flawed bitemark comparisons. Nonetheless, the bitemarks did not inculpate Deeb and were not relevant to his conviction, which was largely based on his relationship with murder suspects David Wayne Spence, Gilbert Melendez, and Anthony Melendez, all of whom were convicted and not exonerated.

Six of the 12 cases involved the 1973 Rochester mafia investigation, which was compromised by police misconduct and falsification of evidence. Several defendants were later convicted on racketeering and conspiracy charges, and the overturned convictions are dubious exonerations at best. (United States v. Russotti, 1983) The fabricated evidence included surveillance logs, but there was no known document examination supporting the convictions. The chief of the Rochester Fire Department falsified arson findings in conspiracy with the defendants, but that misconduct was separate from the evidence in the trials covered by the six cases in the NRE. (Arson Laid to Ex-Rochester Fire Chief, 1975)

Summary

Wrongful conviction research provides unique insights into the nature of errors related to forensic evidence. As the current research shows, the reliability of forensic science depends on a wide range of system factors, including the development and enforcement of effective quality assurance and governance mechanisms across the forensic sciences. In addition, it is critical that forensic conclusions are communicated to and understood by other criminal justice system actors in a manner that permits accurate and reliable decision-making by fact-finders. Most importantly, forensic science organizations should treat wrongful convictions as sentinel events that elucidate system deficiencies within specific laboratories and laboratory systems. Like other high-reliability organizations, forensic science organizations will benefit from a preoccupation with errors and the development of root-cause-analysis capabilities to investigate errors whether they result in a wrongful conviction, near-miss, or other consequence. Although other criminal justice system actors could also benefit from this approach, forensic science organizations possess technical expertise and quality assurance mechanisms that make them well-positioned to take the lead.

The variability of governance mechanisms across forensic disciplines may contribute to the risk of wrongful convictions. In particular, many experts provide testimony in criminal trials without the review and oversight that are practiced in public forensic science organizations. Although some disciplines have standards-setting bodies, lessons from the history of bitemark comparison demonstrate that these bodies may be slow to adopt reforms and may not have sufficient enforcement mechanisms in any case. Interestingly, the state of Texas has seen the highest incidence of wrongful convictions associated with false or misleading forensic evidence and is also the jurisdiction with the most comprehensive approach to the governance of forensic experts. It should be a goal of future research to study wrongful convictions, near misses, and other data to evaluate the effectiveness of governance mechanisms in forensic science.

Courts are poorly equipped to handle this task. Although judges may often be excellent auto-didacts, they nonetheless have made significant errors in the handling of forensic evidence in wrongful convictions. They accepted unvalidated forensic evidence or failed to accept valid and reliable testimony that could have been exculpatory. They frequently failed to account for the adversarial deficit faced by defendants who lack the resources to retain forensic experts to support their cases. Officers of the court demonstrated poor understanding of scientific evidence. Often, this criticism has been raised solely with respect to the acceptance of unvalidated techniques. In wrongful convictions, the reliable use of validated forensic evidence is observed at least as often. Certainly, improved training and resources for indigent defense may alleviate this issue. Courts should also recognize the importance of governance mechanisms for forensic science outside the courtroom and place greater emphasis on the importance of scientific consensus review in their deliberations.

Wrongful convictions demonstrate that forensic science improvement can improve criminal justice outcomes in a lasting and impactful way. New technologies have improved the probative value of forensic evidence and could have prevented many wrongful convictions had they been available. Further, the development, promulgation, and enforcement of science-based standards have also improved the reliability and probative value of forensic evidence. In particular forensic disciplines—such as firearms identification, DNA analysis, and fire debris investigation—the development of standards was needed to prevent continued miscarriages of justice. The standards development process is ongoing and should be considered a necessary component of systems to reduce the risk of future wrongful convictions. Standards-setting bodies should consider the role of emerging and unvalidated techniques in many wrongful convictions. There is a limited basis for police investigators and courts to use to accept or reject new methods, and it may be useful to develop guidelines for the appropriate validation and use of new technologies or methods that are not covered by current standards.

This study provides an introduction to the analysis of forensic evidence in wrongful conviction cases. It was limited only to cases that the NRE has classified as related to F/MFE and therefore did not examine the role of forensic evidence in the large majority of detected wrongful convictions. Further, it did not examine near-miss cases or many overturned convictions that are not included in the NRE. The study used a methodology that relied on retrospective case review, public source documentation, and the author's subjective judgment on error coding. Further research is warranted to examine specific issues and disciplines, expand the reach of documentation and cases, and apply more rigorous methodologies. This study's documentation and analysis may provide a useful starting point for future studies.

Note: Additional information concerning Case Review Form format, individual Case Review Forms, study data, study metadata, and the calculation of results has been provided in prior reports, including the Final Report for Phase 2 of this study.

Recommended further research

The current study provides the first, significant attempt to characterize claims of false or misleading forensic evidence in wrongful convictions in 14 years. First and foremost, such research should be conducted on a continuing basis to elucidate wrongful-conviction factors that may be addressed in policy and practice.

Further recommended research includes:

1. Models for root-cause-analysis and organizational approaches in forensic science organizations. Forensic science organizations require reliable guidance concerning cost-effective improvements that improve the reliability of forensic science, including the examination and description of high-reliability organization frameworks.
2. Approaches to the adoption of novel forensic technologies and the development of science-based standards. The forensic science community has limited mechanisms to identify novel techniques and develop scientifically valid standards prior to their premature introduction into criminal trials.
3. Subjective interpretation and cognitive bias. Many disciplines are based on subjective interpretations at the conclusion level, including forensic pathology and forensic medicine. The understanding, structure, and etiology of bias in these disciplines is poorly understood, despite the many parallels to clinical medicine, in which bias issues have been deeply researched. Additional research to inform policy and practice would benefit these forensic disciplines.
4. Use of forensic science in the criminal courts. Wrongful convictions demonstrate the courts may misuse forensic evidence that is otherwise reliable. In addition, courts may accept invalid evidence that lacks sufficient scientific foundation or practice standards. Finally, defendants face an adversarial deficit in the use forensic evidence, a phenomenon that may be exacerbated by the increasing sophistication of technology and forensic science. These issues have been subject to minimal, rigorous research, which could elucidate causative factors and useful policy and practice improvements.
5. Pattern evidence. Wrongful convictions demonstrate that the use of pattern evidence may be associated with errors based on cognitive bias, poor scientific foundation, poor discriminatory power (such as handwriting or bitemark analysis), poor training, failure to conform to standards, and fraud. Additional research is needed to understand the correlates of these issues in police investigation, evidence collection/prioritization, and the relative roles of target and contextual biases.
6. Fire debris investigation. Wrongful convictions related to fire debris investigation have been documented more thoroughly than in other disciplines, but there has been limited research to examine the factors associated with errors, such as crime scene investigation methods, interpretation approaches, types of physical evidence, and the relationship between fire debris investigation and other evidence in a case investigation. Additional research may elucidate common elements and provide a more thorough guide for interpretation for the fire investigator.
7. Governance. Wrongful convictions demonstrate the current governance frameworks may lack the mechanisms required to ensure that valid and reliable forensic science is conducted in all disciplines and jurisdictional contexts. Further research should elucidate governance mechanisms that have proven effective (or ineffective) in providing adequate governance, including state forensic science commissions, funding, certification, database/technology access, discipline-specific governance, accreditation, quality assurance, and leadership/management training.

Tables and Figures

Bibliography

AAFS Standards Board. (2019). *Guidelines for Opinions and Testimony in Forensic Toxicology.* Colorado Springs, CO: American Academy of Forensic Sciences.

AAFS Standards Board. (2019). *Standard for Forensic DNA Interpretation and Comparison Protocols.* Colorado Springs, CO: ANSI/AAFS Standards Board.

ABS Group. (2018). *Root Cause Analysis of Microscopic Hair Comparison Analysis.* Federal Bureau of Investigation.

American Academy of Pediatrics, Committee on Adolescence. Sexual assault and the adolescent. (1994). *Pediatrics, 94*, 761-765.

American Board of Forensic Odontology. (1995). *ABFO bite mark methodology guidelines.* American Academy of Forensic Sciences.

American Board of Forensic Odontology. (2018). *Standards and Guidelines for Evaluating Bitemarks.* American Board of Forensic Odontology. Retrieved from http://abfo.org/wp-content/uploads/2012/08/ABFO-Standards-Guidelines-for-Evaluating-Bitemarks-Feb-2018.pdf

American Board of Forensic Odontology. (2018). *Standards and Guidelines for Evaluating Bitemarks.* American Academy of Forensic Sciences.

Anderson, E. (2018, August 13). After 34 years behind bars, man gets new trial in Detroit. *Detroit Free Press*.

Andrew Anthony Taylor vs. State of Florida, 17-002295VWI (Division of Administrative Hearings, Florida November 28, 2017).

Andrews International. (2014). *Review of the Office of the Chief Medical Examiner: Report for the Department of Health and Social Services.* State of Delaware.

Arledge v. State, 10-84-079-CR (Court of Appeals of Texas, Tenth District, Waco January 17, 1985).

Armstrong v. City of Greensboro, 1:15CV282 (United States District Court for the Middle District of North Carolina March 31, 2016).

Arnold Melnikoff v. Washington State Patrol, DISM-04-0046 (Personnel Appeals Board, State of Washington 2005).

Arson Laid to Ex-Rochester Fire Chief. (1975, September 13). *New York Times*, p. 27.

Ashton, A., & Bernton, H. (2016, March 7). Panel backs JBLM doctor who was jailed over contaminated drug test. *The News Tribune*.

Associated Press. (2020, December 20). Wrongfully convicted Michigan man freed after 38 years. *The Times Herald (Port Huron, Michigan)*, p. A2.

Baird, M., Neufeld, P., & Scheck, B. (1990). DNA Testing-Is Forensic DNA Testing Reliable. *ABA Journal, 76*, 34.

Bantz, P. (2017, December 11). Charlotte settles wrongful conviction suit for $9.5M. *North Carolina Lawyers Weekly*.

Banuelos, S., & Miller, J. (2015). *Laboratory Report #1, Roumaldo Lerma.* Houston, TX: Houston Forensic Science Center.

Barnett, P., Blake, E., & Ogle Jr., R. (1981). The Role of the Independent Expert. *American Academy of Forensic Sciences*, (pp. 1-14). Los Angeles, CA.

Barnhouse v. City of Muncie, 1:19-cv-00958-TWP-DLP (United States District Court for the Southern District of Indiana, Indianapolis Division November 4, 2020).

Bell, S., Sah, S., Albright, T., Gates, J., Denton, M., & Casadevall, A. (2018, May). A call for more science in forensic science. *Proceedings of the National Academy of Sciences of the United States of America*, 4541-4544. doi:10.1073/pnas.1712161115

Berk, R., Rochowicz, S., Wong, M., & Kopina, M. (2007). Gunshot residue in Chicago police vehicles and facilities: an empirical study. *Journal of forensic sciences, 52*(4), 838-841.

Beyler, C. L. (2009). *Analysis of the Fire Investigation Methods and Procedures Used in the Criminal Arson Cases AGainst Ernest Ray Willis and Cameron Todd Willingham.* Huntsville, TX: Texas Forensic Science Commission.

Blake, E. (1985). *Illinois v. Gary Dotson Post Conviction Evaluation of the Conventional Serology Evidence of Timothy Dixon.*

Blake, E. (2003). *Montana v. Paul Kordonowy Report 1 File 03-128.* Richmond, CA: Forensic Science Associates.

Blake, E. (January 9, 2001). *Review Report, Illinois v. Ollins et al.* Forensic Science Associates.

Blake, E., Sensabaugh, G., & Bashinski, J. (1980). A systematic approach to the analysis of semen evidence. *Fall Semiannual Seminar of the California Association of Criminalists.*

Bloss, W., & McElligott, S. (2014, March 31). In the Matter of the Claim of Kenneth Ireland against the State of Connecticut, Analysis of Damages.

Bluhm Legal Clinic. (n.d.). *Maurice Patterson.* Retrieved from Bluhm Legal Clinic Center on Wrongful Convictions: https://www.law.northwestern.edu/legalclinic/wrongfulconvictions/exonerations/il/maurice-patterson.html

Bondurant, T. (2015, December). *Davey Reedy Granted Full Pardon in 1987 Fatal Arson.* Retrieved from Gentry Locke Attorneys: https://www.gentrylocke.com/news/davey-reedy-granted-full-pardon-in-1987-fatal-arson/

Bonventre, C. (2021). Wrongful convictions and forensic science. *Wiley Interdisciplinary Reviews: Forensic Science*, e1406.

Borchard, E. M. (1932). *Convicting the Innocent: Sixty-Five Actual Errors of Criminal Justice.* Garden City, NJ: Garden City Publishing.

Bowers, C. M. (2002). Identification from Bitemarks: Scientific Issues. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Science in the Law: Forensic Science Issues* (pp. 244-293). St. Paul, MN: West.

Brewer v. Steven Timothy Hayne & Michael H. W., 3:13-cv-898-HTW-LRA (United States District Court for the Southern District of Mississippi, Northern Division April 16, 2015).

Bromwich, M. (1997). *The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases.* Washington, DC: US Department of Justice Office of Inspector General.

Bromwich, M. R. (2007). *Final Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room.* Washington, DC: Fried, Frank, Harris, Shriver & Jacobson LLP . Retrieved from http://www.hpdlabinvestigation.org/

Brown v. Miller, 06-30887 (US Court of Appeals, Fifth Circuit February 27, 2008).

Brundage, J. (1998). The Identification of Consecutively Manufactured Gun Barrels. *AFTE Journal, 30*(1), 438-444.

Bryson v. Macy, CIV-05-1150-F (United States District Court for the Western District of Oklahoma April 30, 2009).

Budowle, B., Lindsey, J. A., DeCout, J. A., Koons, B. W., Giusti, A. M., & Comey, C. T. (1995). Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQalpha using a multiplex amplification and typing procedure. *Journal of Forensic Science, 40*(1), 45-54.

Bunch, S. G., Smith, E. D., Giroux, B. N., & Murphy, D. P. (2009). Is a Match Really a Match? A Primer on the Procedures and Validity of Firearm and Toolmark Identification. *Forensic Science Communications, 11*(3).

Busby, J. (1999). The Effectiveness of Collective Retrospection as a Mechanism of Organizational Learning. *The Journal of Applied Behavioral Science, 35*, 109-129.

Busey, T. A., Heise, N., Hicklin, R. A., Ulery, B. T., & Buscaglia, J. (2021). Characterizing missed identifications and errors in latent fingerprint comparisons using eye-tracking data. *PloS one, 16*(5), e0251674.

Butler, J., Iyer, H., Press, R., Taylor, M., Vallone, P., & Willis, S. (2020). *NIST Scientific Foundation Reviews.* US Department of Commerce. National Institute of Standards and Technology. doi:https://doi.org/10.6028/NIST.IR.8225

Caminata v. Cnty. of Wexford, 16-1451 (United States Court of Appeals for the Sixth Circuit November 16, 2016).

Cannon v. Polk County/Polk County Sheriff, 3:10–cv–00224–HA (United States District Court, D. Oregon. December 18, 2014).

Carter v. City of Detroit, 11-15322 (United States District Court for the Eastern District of Michigan, Southern Division January 27, 2016).

Champod, C., Biedermann, A., Vuille, J., Willis, S., & De Kinder, J. (2016, March 12). ENFSI Guideline for Evaluative Reporting in Forensic Science, A Primer for Legal Practitioners. *Criminal Law & Justice Weekly, 180*, 189-193.

Choudhary, A., Servaes, S., Slovis, T., Palusci, V., Hedlund, G., Narang, S., . . . Offiah, A. (2018). Consensus statement on abusive head trauma in infants. *Pediatric Radiology*, 1048-1065.

Clemente Javier Aquirre Jarquin v. State of Florida, SC06-1550 (Supreme Court of Florida February 7, 2008).

Cole, S. (2011). Forensic Science and Wrongful Convictions: From Exposer to Contributor to Corrector. *New Eng. L. Rev.*, 711-736.

Cole, S. (2013). Forensic culture as epistemic culture: The sociology of forensic science. *tudies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44*(1), 36-46.

Cole, S. A. (2000). The Myth of Fingerprints: A forensic science stands trial. *Lingua Franca, 10*(8), 54-62.

Cole, S. A. (2005). More Than Zero: Accounting for Error in Latent Print Identification. *Journal of Criminal Law and Criminology, 95*, 985-1078.

Cole, S. A. (2012). Forensic Science and Wrongful Convictions: From Exposer to Contributor to Corrector. *New England Law Review, 46*, 711-736.

Comey, J. (2016, 2 26). *Director Comey Letter to State Governors Requesting Additional Case Information from Prosecutors.* Retrieved from Federal Bureau of Investigation: https://www.fbi.gov/file-repository/comey-letter-to-governors.pdf/view

Committee on Child Abuse and Neglect. (1991). Guidelines for the evaluation of sexual abuse of children. *Pediatrics, 87*(2), 254-260.

Committee on Child Abuse and Neglect. (1999). Guidelines for the Evaluation of Sexual Abuse of Children: Subject Review. *Pediatrics, 103*, 186-191. doi:10.1542/peds.103.1.186

Committee on Child Abuse and Neglect. (2013). The Evaluation of Children in the Primary Care Setting When Sexual Abuse Is Suspected. *Pediatrics, 132*(2), e558-e567.

Committee on Child Abuse and Neglect, American Academy of Pediatrics. (1999). Guidelines for the Evaluation of Sexual Abuse of Children: Subject Review. *Pediatrics*, 186-191.

Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council. (2009). *Strengthening Forensic Science in the United States: A Path Forward.* Washington, DC: National Academies Press.

Committee to Assess the Feasibility, A. a. (2008). *Ballistic Imaging.* Washington, DC: National Academies Press.

Commonwealth of Virginia vs. Willie Davidson (Circuit Court of the City of Norfolk May 27, 1981).

Commonwealth v. Carnevale, 200615299 (Common Pleas Court of Allegheny County, Pennsylvania, Criminal Division January 26, 2012).

Commonwealth v. Patterson, SJC-09478 (Supreme Judicial Court of Massachusetts December 27, 2005).

Commonweath v. Marvin Mitchell, 072574 (Superior Court, Commonwealth of Massachusetts January 18, 1990). Retrieved 12 10, 2020, from https://convictingtheinnocent.com/exoneree/marvin-mitchell/

Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial.* Washington, DC: National Institute of Justice.

Consortium of Forensic Science Organizations. (2013, December). *Accreditation of Entities Providing Forensic Science Services.* Retrieved from American Society of Crime Laboratory Directors: https://www.ascld.org/wp-content/uploads/2014/02/CFSO-Accreditation-Paper-December-2013.pdf

Cooley, C. M., & Oberfield, G. S. (2007). Increasing Forensic Evidence's Reliability and Minimizing Wrongful Convictions: Applying Daubert Isn't the Only Problem. *Tulsa Law Review, 43*(2), 285-380.

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: a systematic review. *Forensic science international, 297*, 35-46.

Criner v. State, 09-90-096 CR (Court of Appeals of Texas, Beaumont January 8, 1992).

(1981). *Critique: The Role of the Independent Expert: Several Case Examples.* US House of Representatives.

Croskerry, P., Singhal, G., & Mamde, S. (2013). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Qual Saf, 22(Suppl 2)*, ii58-ii64.

Culliford, B. J. (1971). *The Examination and Typing of Bloodstains in the Crime Laboratory.* Washington, DC: National Institute of Law Enforcement and Criminal Justice.

Dabbs. v. Vergari, 17797/88 (Supreme Court of New York, Westchester County November 21, 1990).

Dake, D. (2017, April 10). Coroner Talk Podcast.

Dalby, O., Butler, D., & Birkett, J. (2010). Analysis of Gunshot Residue and Associated Materials—A Review. *Journal of Forensic Sciences, 55*(4), 924-943.

De Leeuw, M., Beuls, E., Jorens, P., Parizel, P., & Jacobs, W. (2013). History of an abusive head trauma including a lucid interval and a retinal hemorrhage is most likely false. *The American journal of forensic medicine and pathology, 34*(3), 271-276.

De Leeuw, M., Beuls, E., Parizel, P., Jorens, P., & Jacobs, W. (2013). Confessed abusive blunt head trauma. *The American journal of forensic medicine and pathology, 34*(2), 130-132.

Deguchi, T., Kuriyama, M., Shinoda, I., Okano, M., Ban, Y., Matsui, A., . . . Kawada, Y. (1991). Immunological comparison between prostate-specific antigen and γ-seminoprotein. *Urological Research, 19*(1), 25-30.

DeHaan, J., & Icove, D. (2011). *Kirk's fire investigation.* Pearson Higher Ed.

Dennis Patrick BROWN, Plaintiff-Appellee, v. Nace Jerry MILLER, in his individual ando, 06-30887 (United States Court of Appeals, Fifth Circuit February 27, 2008).

Department of Justice. (2010, January 11). *Brian Rose Pleads Guilty To January 2006 Murder Of Carjacking Victim.* Retrieved from Department of Justice: https://www.justice.gov/archive/usao/md/news/archive/BrianRosePleadsGuiltytoJanuary2006MurderOfCarjackingVictim.html

Deseret News. (2002, September 12). Weitzel gets year in jail cell for fraud. *Deseret News*.

Dimick, G., Melton, T., Higgins, B., Lindstrom, L., & Nelson, K. (2005, January). Forensic mitochondrial DNA analysis of 691 casework hairs. *Journal of Forensic Sciences, 50*(1), 73-80.

District of Columbia Department of Forensic Sciences. (2013). *FBS06--P30 Antigen Test for the Presence of Semen.*

Donohue, M. (2003). Evidence-Based Medicine and Shaken Baby Syndrome. *Am. J. Forensic Med. & Pathology, 24*, 239-241.

Double Loop Podcast. (2021, April). Double Loop Podcast Compares Fingerprints.

Dror, I. (2018). Biases in forensic experts. *Science, 360*(6386), 243.

Dror, I. E. (2020). Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. *Analytical chemistry, 92*(12), 7998-8004.

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science and Justice, 51*(4), 204-208.

Dror, I., Melinek, J., Arden, J., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. (2021). Cognitive bias in forensic pathology decisions. *Journal of Forensic Sciences, 66*(5), 1751-1757.

Dror, I., Morgan, R., Rando, C., & Nakhaeizadeh, S. (2017). The Bias Snowball and the Bias Cascade Effects: Two Distinct Biases that May Impact Forensic Decision Making. *Journal of forensic sciences, 62*(3), 832-833.

Drug Enforcement Administration. (2017, 04 10). *Schedules of Controlled Substances: Temporary Placement of Six Synthetic Cannabinoids (5F-ADB, 5F-AMB, 5F-APINACA, ADB-FUBINACA, MDMB-CHMICA and MDMB-FUBINACA) into Schedule I.* Retrieved from Federal Register: https://www.federalregister.gov/documents/2017/04/10/2017-07118/schedules-of-controlled-substances-temporary-placement-of-six-synthetic-cannabinoids-5f-adb-5f-amb

Einhorn, H., & Hogarth, R. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological review, 85*(5), 395.

Eugene Johnson v. Rich Gansheimer, 1: 06-CV-2816 (US District Court Northern District of Ohio Western Division November 11, 2008).

Ex parte Franklin, 72 S.W.3d 671 (Court of Criminal Appeals of Texas April 10, 2002).

Ex parte Overton, WR-75,804-02 (Court of Criminal Appeals of Texas September 17, 2014).

Ex parte Robbins, WR-73, 484-02 (Court of Criminal Appeals of Texas November 26, 2014).

Ex parte Steven Mark Chaney, Applicant, WR-84,091-01 (Court of Criminal Appeals of Texas December 19, 2018).

Fedchak, S. (2014). *Presumptive Field Testing Using Portable Raman Spectroscopy.* January.

Federal Bureau of Investigation. (1982). *Handbook of Forensic Science.* Washington, DC: National Institute of Justice, US Department of Justice.

Federal Bureau of Investigation. (2001). *Summary of case reviews of Forensic Chemist, Joyce Gilchrist, Oklahoma City Police Department Crime Laboratory.* Oklahoma City, OK: US Department of Justice.

Federal Bureau of Investigation. (2015, 4 19). *FBI/DOJ Microscopic Hair Comparison Analysis Review.* Retrieved from Federal Bureau of Investigation:

https://www.fbi.gov/services/laboratory/scientific-analysis/fbidoj-microscopic-hair-comparison-analysis-review

Findley, K., & Golden, L. (2013). The innocence movement, the innocence network, and policy reform. In *Wrongful Conviction and Criminal Justice Reform* (pp. 107-124). Routledge.

Firearms and Toolmarks Subcommittee of the Organization of Scientific Area Committees. (n.d.). *Standard Scale of Source Conclusions and Criteria for Toolmark Examinations.* National Institute of Standards and Technology.

Flanagan, R. J., Cuypers, E., Maurer, H. H., & Whelpton, R. (2020). *Fundamentals of analytical toxicology.* John Wiley & Sons.

Florida v. Luis Diaz (1980).

Flowers v. State, 1999-DP-01369-SCT (Supreme Court of Mississippi February 20, 2003).

FPT Heads of Prosecutions Committee Working Group. (2011 (update)). *Report on the Prevention of Miscarriages of Justice.* Department of Justice, Government of Canada. Retrieved from https://www.justice.gc.ca/eng/rp-pr/cj-jp/ccr-rc/pmj-pej/pmj-pej.pdf

Frank, J., & Frank, B. (1957). *Not Guilty.* Doubleday & Company.

Friction Ridge Subcommittee of the Organization of Scientific Area Committees. (n.d.). • *Guideline for the Articulation of the Decision-Making Process Leading to an Expert Opinion of Source Identification in Friction Ridge Examinations .* National Institute of Standards and Technology.

Friction Ridge Subcommittee of the Organization of Scientific Area Committees. (n.d.). *Standard for Friction Ridge Examination Conclusions .* National Institute of Standards and Technology.

Furton, K., Hsu, Y., & Cole, M. (1999). What Educational Background Do Crime Laboratories Require from Applicants? *Journal of Forensic Sciences, 44*(1), 128-132.

Gabrielson, R., & Sanders, T. (2016, July 7). Busted: Examining Chemical Field Tests. *New York Times*, pp. 1-36.

Gaennslen, R. E. (2000). Forensic Analysis of Biological Evidence. In C. H. Wecht, *Forensic Sciences* (29 ed., Vol. 1). New York: Matthew Bender and Co.

Gaensslen, R. E. (1983). *Sourcebook in forensic serology, immunology, and biochemistry.* Washington, DC: National Criminal Justice Reference Service. Retrieved from https://www.ncjrs.gov/pdffiles1/pr/160880_unit_5.pdf

Garrett, B. (2011). *Characteristics of Forensic Testimony at DNA Exonerees' Trials.* Retrieved from Convicting the Innocent: https://www.convictingtheinnocent.com/wp-content/uploads/2016/10/garrett_forensics_appendix.pdf

Garrett, B. (2011). *Convicting the Innocent: Where Criminal Prosecutions Go Wrong.* Harvard University Press.

Garrett, B. (2020, January). Wrongful Convictions. *Annual Review of Criminology, 3*, 245-259.

Garrett, B. L. (2018, March). *Brandon Garrett Resume.* Retrieved June 22, 2021, from Duke Law School - Duke University: https://law.duke.edu/sites/default/files/cv/garrett_resume_2018-3.pdf

Garrett, B. L. (2020). Wrongful Convictions. *Annual Review of Criminology, 3*(1), 245-259.

Garrett, B. (n.d.). *Victor Burnette.* Retrieved from Convicting the Innocent: https://convictingtheinnocent.projects.law.duke.edu/exoneree/victor-burnette/

Garrett, B., & Neufeld, P. (2009, March). Invalid Forensic Science Testimony and Wrongful Convictions. *Virginia Law Review*, 1-97.

Gaudette, B. D., & Keeping, E. S. (1974). An attempt at determining probabilities in human scalp hair comparisons. *Journal of Forensic Sciences, 19*, 599-606.

Giannelli, P. (2010). *Microscopic Hair Comparisons: A Cautionary Tale.* School of Law, Case Western Reserve University.

Giannelli, P. (2013). Junk Science and the Execution of an Innocent Man. *New York University Journal of Law & Liberty, 7*(2), 221.

Goddard, C. (1926). Scientific Identification of Firearms and Bullets. *Journal of Criminal Law and Criminology, 17*(2, August), 254-263.

Gould, J. B., Carrano, J., Leo, R., & Young, J. (2013). *Predicting Erroneous Convictions: A Social Science Approach to Miscarriages of Justice.* NCJRS. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/241389.pdf

Graves, H. C., Sensabaugh, G. F., & Blake, E. T. (1985). Postcoital Detection of a Male-Specific Semen Protein--Application to the Investigation of Rape. *The New England Journal of Medicine, 312*, 338-343.

Grispino, R. R. (1990). The effect of luminol on the serological analysis of dried human bloodstains. *Crime Lab Digest, 17*(1), 13-23.

Grometstein, R. (2008). Wrongful Conviction and Moral Panic: National and International Perspectives on Organized Child Sexual Abuse. In C. R. Huff, & M. Killias (Eds.), *Wrongful Conviction: International Perspectives on Miscarriages of Justice* (pp. 11-32). Philadelphia, PA: Temple University Press. Retrieved from https://www.ncjrs.gov/App/publications/Abstract.aspx?id=247359

Gross, S. R., & O'Brien, B. (2008, December). Frequency and Predictors of False Conviction: Why We Know So Little, and New Data on Capital Cases. *Journal of Empirical Legal Studies, 5*(4), 927-962.

Gross, S., & Shaffer, M. (2012, June). *Exonerations in the United States, 1989-2012*. Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf

Gunshot Residue Subcomittee Chemistry Scientific Area Committee Organization of Scientific Area Committees for Forensic Science. (2020). *Standard Practice for Expert Opinions on the Interpretation of Primer Gunshot Residue (pGSR) Analysis by Scanning Electron Microscopy/Energy Dispersive X-Ray Spectrometry.* National Instute of Standards and Technology.

Hall, D. (2013, March 26). After 6-year ordeal and nearly 3 years in prison, Joseph Awe is a free man. *Wisconsin State Journal*.

Hall, M. (2015, October 20). Another Texas exoneration calls bite mark evidence into question. *Texas Monthly*.

Hamby, J. E., Brundage, D. J., & Thorpe, J. W. (2009). The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries. *AFTE Journal--Volume 41 Number 2--Spring 2009, 41*(2), 99-110.

Hanes, S. (2005, January 23). Evidence under suspicion. *Baltimore Sun*, pp. A1, A12-13.

Hara, I. (1969). Immunoelectrophoretic Studies of the Protein Components in Human Seminal Plasma (Especially Its Specific Component). *Japan. J. Leg. Med., 23*, 117.

Hatchett v. City of Detroit et al, 08-CV-11864 (US District Court, Eastern District of Michigan, Southern Division 02 10, 2010).

Henning v. Commissioner of Correction, SC 20137 (Supreme Court of Connecticut June 14, 2019).

Hepworth, D. G., & Wickenheiser, R. A. (1990). Further Evaluation of Probabilities in Human Scalp Hair Comparisons. *Journal of Forensic Sciences, 35*(6), 1323-1329.

Herbeck, D., & Prohaska, T. (1991, November 20). Jury in retrial acquits man of shooting wife to death. *Buffalo News*.

Hinton v. Alabama, 13-6440 (Supreme Court of the United States February 24, 2014).

Hollway, J. F. (2021). *Instilling a Culture of Continuous Learning from Criminal Justice Systems Errors: A Multi-Stakeholder Sentinel Event Review Process in Philadelphia.* National Institute of Justice.

Houck, M. M., & Budowle, B. (2002). Correlation of microscopic and mitochondrial DNA hair comparisons. *Journal of Forensic Sciences, 47*(5), 964-967.

House v. Bell, 04-8990 (Supreme Court of the United States June 12, 2006).

House v. United States, 08-758C (United States Court of Federal Claims July 8, 2011).

Hsieh, C., Horng, S., & Liao, P. (2003). Stability of Trace-level VOLatile Organic Compounds Stored in Canisters and Tedlar Bags. *Aerosol Air Qual. Res., 3*, 17-28. doi:https://doi.org/10.4209/aaqr.2003.06.0003

Huff, C. R., & Killias, M. (Eds.). (2008). *Wrongful Conviction: International Perspectives on Miscarriages of Justice.* Philadelphia, PA: Temple University Press.

In re: W. Va. State Police Crime Lab., 190 W. Va. 321 (Supreme Court of Appeals of West Virginia November 10, 1993).

Ingram, D., White, S., Lyna, P., Crews, K., Schmid, J., Everett, V., & Koch, G. (1992). Gardnerella vaginalis infection and sexual contact in female children. *Child Abuse & Neglect, 16*(6), 847-853. doi:https://doi.org/10.1016/0145-2134(92)90086-7

Innocence Project. (2020, January 28). *Overturning Wrongful Convictions Involving Misapplied Forensics*. Retrieved from Innocence Project: https://www.innocenceproject.org/overturning-wrongful-convictions-involving-flawed-forensics/

Jackson v. Paparo, 00-3413 (United States District Court for the Eastern District of Pennsylvania October 25, 2002).

James Kluppelberg vs. Jon Burge et al, 13 CV 3963 (US District Court for the Northern District of Illinois Eastern Division November 5, 2015).

Jeanguenat, A., Budowle, B., & Dror, I. (2017). Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias. *Science and Justice, 57*(6), 415-420.

Jennings v. Nash, 6:18-cv-03261-NKL (United States District Court for the Western District of Missouri, Southern Division January 15, 2020).

Joint Committee for Guides in Metrology. (2008). *JCGM 100: 2008, Evaluation of Measurement Data--Guide to the Expression of Uncertainty in Measurement* (First ed.). Saint-Cloud, France: International Bureau of Weights and Measures.

Jones v. State, No. 1962, September Term, 1999 (Court of Special Appeals of Maryland June 9, 2000).

Joseph E. White v. Richard Smith et al, 4:09CV3145 (US District Court for the District of Nebraska 08 08, 2011).

Joseph Green Brown, Petitioner appellant, v. Louie L. Wainwright, Secretary Florida Department of Corrections, 785 F. 2d 1457 (U.S. Court of Appeals for the Eleventh Circuit March 17, 1986).

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*(3), 582-591.

Kaminskas v. Office of the Attorney General (Supreme Court of New Jersey January 17, 2019).

Kaye, D. (2017, November 17). *Comedic License.* Retrieved from Forensic Science, Statistics & the Law: http://for-sci-law.blogspot.com/2017/11/comedic-license.html

Koch, S. (2017). Microscopy of Hair Part 1: A Practical Guide and Manual for Human Hairs. *Forensic Science Communications, 6*(1), 1-44.

Koehler, J., Mnookin, J., Cole, S., Fisher, B., Dror, I., Houck, M., . . . Siegel, J. (2011). The Need for a Research Culture in the Forensic Sciences. *UCLA L. Rev., 58*, 725-780.

Koppl, R. (2005). How to Improve Forensic Science. *European Journal of Law and Economics, 20*, 255-286.

Lana Canen v. Dennis Chapman, 16-1621 (United States Court of Appeals, Seventh Circuit January 27, 2017).

LaPorte, G. (2017, September). Wrongful Convictions and DNA Exonerations: Understanding the Role of Forensic Science. *NIJ Journal*, 1-10.

Larry Pat Souter, Petitioner-Appellant, v. Kurt Jones, Warden, Respondent-Appellee, 03-1528 (United States Court of Appeals,Sixth Circuit January 18, 2005).

Latta v. Chapala, 2:03-CV-41 (United States District Court, N.D. Indiana, Hammond Division October 25, 2005).

Laux, D., Tambasco, A., & Benzinger, E. (2006). *Forensic Detection of Semen II. Comparison of the Abacus Diagnostics OneStep ABAcard p30 Test and the Seratec PSA Semiquant Kit for the Determination of the Presence of Semen in Forensic Cases.*

Lavernia v. Lynaugh, 87-1918 (United States Court of Appeals, Fifth Circuit May 5, 1988).

LeFever v. Ferguson, 2:11-cv-935 (US District Court for the Southern District of Ohio, Eastern Division July 15, 2013).

Lentini, J. (2012). The Evolution of Fire Investigation and Its Impact on Arson Cases. *Criminal Justice, 27*(1), 12-18.

Lentini, J. (2014). Anatomy of a Wrongful Arson Conviction: Sentinel Event Analysis in Fire Investigation. *International Symposium on Fire Investigation, Science and Technology*, (pp. 1-17). College Park, MD.

Leo, R. A., & Gould, J. B. (2009). Studying Wrongful Convictions: Learning from Social Science. *Ohio St. J. Crim., 7*, 7-30.

Letemps v. Sec'y, Fla. Dep't of Corr., 6: 13-cv-718-Orl-31KRS (United States District Court for the Middle District of Florida, Orlando Division July 20, 2015).

Levine, B., & Kerrigan, S. (Eds.). (2020). *Principles of Forensic Toxicology 5th ed. 2020 Edition.* Springer.

Liebman, J. S., Fagan, J., West, V., & Lloyd, J. (2000). Capital Attrition: Error Rates in Capital Cases, 1973-1995, 78 Tex. L. Rev. 1839 (2000). *Tex. L. Rev., 78*, 1839-1865.

Luedeke, M., Miller, E., & Sprague, J. (2016). The effects of Bluestar® and luminol when used in conjunction with tetramethylbenzidine or phenolphthalein. *Forensic science international, 262*, 156-159.

Manas, B., Nilesh, P., Manika, A., & Betina, C. (2016). Bite Marks -- Revisited. *Journal of Forensic Science & Criminology, 4*(5).

Manas, B., Nilesh, P., Manika, A., & Betina, C. (2016). Bite Marks -- Revisited. *Journal of Forensic Science & Criminology, 4*(5), 503-508.

Maria Delacruz vs. Tripler Army Medical, 05-00571 (US District Court for the District of Hawaii 07 24, 2007).

Martineau, K. (2003, July 22). He Paid for Police Mistake. *Hartford Courant*, p. 75.

Masterson, M. (2015, 04 25). Failing trust and justice. *Arkansas Democrat-Gazette*.

Matta-Ballesteros v. United States, LA CV16-02596 (US District Court for the Central District of California May 22, 2017).

McDonald, D. (2019, September 25). 24,000 charges tossed because they were tainted by former Amherst lab chemist's misconduct. *Boston Globe*. Retrieved 05 06, 2020, from https://www.bostonglobe.com/metro/2019/09/25/charges-tossed-because-they-were-tainted-former-amherst-lab-chemist-misconduct/MUPgdHeLy8bdrzI5KGtvIN/story.html

McNamara, J. J., & Linhart, R. R. (1993). *ASCLD/LAB Investigation Report, West Virginia State Police Crime Laboratory, Serology Division, South Charleston, West Virginia.* ASCLD/LAB.

Mills, S. (2013, September 4). Suspect in assault linked to homicide. *Chicago Tribune*, pp. 1-12.

Mitchell v. City of Boston, 98-CV-11674-PBS (US District Court for the District of Massachusetts January 26, 2001).

Monroe v. Angelone, 02-6548, 02-6625 (United States Court of Appeals,Fourth Circuit March 26, 2003).

Morgan, J. S. (2016). *Forensic Optical Topography: A Landscape Study.* Washington, DC: National Institute of Justice .

Moriarty, J. C. (2007). "Misconvictions," Science, and the Ministers of Justice. *Neb. L. Rev., 86*(1), 1-42.

Morrison, A. (1972). Persistence of spermatozoa in the vagina and cervix. *British Journal of Venereal Diseases, 48*(2), 141.

Mullen, A. (2004, January 21). Confessions and recantations. *Detroit Metro Times*.

Narang, S., Fingarson, A., & Lukefahr, J. (2019). Abusive Head Trauma in Infants and Children. *Pediatrics, 145*(4). doi:145(4):e20200203

National Fire Protection Association. (1992). *National Fire PrGuide for Fire and Explosion Investigations (NFPA 921-92).* Quincy, MA: NFPA.

National Fire Protection Association. (2017). *NFPA 921, Guide for Fire and Explosion Investigations.*

National Institute of Justice. (2011). *The Fingerprint Sourcebook.* Washington, DC: National Criminal Justice Reference Service. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/225320.pdf

National Institute of Justice. (2019). *Needs Assessment of Forensic Laboratories and Medical Examiner/Coroner Offices: A Report to Congress.* Washington, DC: Department of Justice.

National Institute of Standards and Technology. (2015). Proceedings of the 2015 International Symposium on Forensic Science Error Management. In J. M. Butler (Ed.). Retrieved from https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1206.pdf

National Medicolegal Review Panel. (1999). *Death investigation: a guide for the scene investigator.* Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

National Medicolegal Review Panel. (2011). *Death Investigation: A Guide for the Scen Investigator, Technical Update.* Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

National Registry of Exonerations. (2019, 11 8). *Ben Salazar.* Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3605

National Registry of Exonerations. (2019, 11 15). *James Waller.* Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3713

National Registry of Exonerations. (2021). *Annual Report.* Newkirk Center for Science & Society at the University of California Irvine, the University of Michigan Law School & Michigan State University College of Law.

National Research Council (US) Committee on DNA Technology in Forensic Science. (1992). *DNA Technology in Forensic Science.* Washington, DC: National Academies Press.

National Research Council. (1996). *The Evaluation of Forensic DNA Evidence.* Washington, DC: National Academies Press.

National Research Council. (2004). *Forensic analysis: Weighing bullet lead evidence.* National Academies Press.

Neufeld, P. (n.d.). *Peter Neufeld/Partner.* Retrieved June 21, 2021, from NSB: https://www.nsbcivilrights.com/our-people/peter-neufeld/

Neumann, C., & Saunders, C. P. (2019). *Foundational Research Into the Quantification of the Value of Forensic Evidence for Complex Evidential Forms Arising from Impression and Pattern Evidence: Final Summary Overview Report.* National Institute of Justice.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175-220.

Norris, R., Bonventre, C., Redlich, A., Acker, J., & Lowe, C. (2017). Preventing Wrongful Convictions: An Analysis of State Investigation Reforms. *Criminal Justice Policy Review, 30*(4), 597-626.

Office of the State Attorney 13th Judicial Circuit. (2020, September 15). *State Attorney's Office launches review of convictions based on bite mark evidence following discovery of DuBoise's wrongful conviction.* Retrieved from Office of the State Attorney 13th Judicial Circuit Hillsborough: https://www.sao13th.com/2020/09/15/

O'Leary-Kelly, S., & Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of operations management, 16*(4), 387-405.

Ollins v. O'Brien, 03 C 5795, 03 C 7175 (United States District Court for the Northern District of Illinois, Eastern Division March 28, 2005).

Organization of Scientific Area Committees. (2021). *Standard Scale of Source Conclusions and Criteria for Toolmark Examinations.*

Organization of Scientific Area Committees for Forensic Science. (2020, March 23). *OSAC Research and Development Needs.* Retrieved from National Institute of Standards and Technology: https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/osac-research-and-development-needs

Organization of Scientific Area Committees for Forensic Science. (n.d.). *OSAC Lexicon.* Retrieved from National Institute of Standards and Technology: https://lexicon.forensicosac.org/Home/Index

Otterbourg, K. (2020, 12 14). *Elmer Daniels.* Retrieved from National Registry of Exonerations: http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=5470

Papetti, R., Kaneb, P., & Herf, L. (2019). Outside the Echo Chamber: A Response to the "Consensus Statement on Abusive Head Trauma in Infants and Young Children:. *Santa Clara Law Review, 59*, 299-366.

Park, L. (2002, November 23). Jury clears Dr. Weitzel after two hours of deliberation. *Standard-Examiner Davis Bureau*.

PCAST Working Group. (2016). *Report to the President: Forensic Science in Criminal Courts, Ensuring Scientific Validity of Feature Comparison Methods.* Washington, DC: President’s Council of Advisors on Science and Technology.

People of the State of California vs Herman Atkins, CR 28832 (Court of Appeal State of California Fourth Appellate District August 22 to 25, 1988). Retrieved from https://convictingtheinnocent.com/exoneree/herman-atkins/

People of the State of California vs. Frederick Rene Daye, D002073 (Court of Appeal of the State of California Fourth Appellate District May 24, 1984).

People of the State of California vs. Scott Lee Kniffen, F004423 (Supreme Court of the State of California 1995).

People v. Herskovic, 2017-02494 (Supreme Court of New York, Appellate Division, Second Department October 10, 2018).

People v. Ixta, B230515 (Court of Appeal of California, Second Appellate District, Division Six June 28, 2012).

People v. Jones, 156 Ill. 2d 225 (Circuit Court of Cook County 1989). Retrieved from https://convictingtheinnocent.com/exoneree/ronald-jones/

People v. Liebich, 2-13-0894 (Appellate Court of Illinois, Second District March 28, 2016).

People v. McSherry, B048607, B066689 (Court of Appeal of California, Second Appellate District, Division Five December 17, 1992).

People v. Pursley, 2-17-0227 (Appellate Court of Illinois, Second District May 2, 2018).

People v. Stielow, 160 N.Y.S. 555 (Supreme Court of New York, Special Term, Erie County July 30, 1916).

People v. Tomaino, 0143 (Supreme Court of New York, Appellate Division, Fourth Department March 13, 1998).

People v. Willis, A139858, A 154925 (Court of Appeal of California, First Appellate District, Division Four April 29, 2019).

Peterson, P. E., Dreyfus, C. B., Gische, M. R., Hollars, M., Roberts, M. A., Ruth, R. M., . . . Soltis, G. L. (2009, October). Latent Prints: A Perspective on the State of the Science. *Forensic Science Communications, 11*(4).

Philip, M., & Fu, S. (2018). A review of chemical 'spot tests: A presumptive illicit drug identification technique. *Drug Testing and Analysis, 10*(1), 95-108.

Pierce v. Gilchrist, CIV-05-1519-C (United States District Court, W.D. Oklahoma January 16, 2007).

Possley, M. (2015, 10 12). *Beniah Alton Dandridge.* Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4768

Possley, M. (2015, 10 28). *Clayton Martel.* Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4784

Possley, M. (2015, 10 30). *Roumaldo Lerma.* Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4783

Possley, M. (2018, 06 13). *Ernest Sonnier.* Retrieved from National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=5342

Rairden, A., Garrett, B., Kelley, S., Murrie, D., & Castillo, A. (2018). Resolving latent conflict: What happens when latent print examiners enter the cage? *Forensic science international, 289*, 215-222.

Reedy v. Wright, CL00000-23 (Circuit Court of the City of Roanoke, Virginia May 13, 2002).

Reid, M. E., & Lomas-Francis, C. (2004). *The Blood Group Antigens Facts Book, 2nd ed.* New York: Elsevier Academic Press.

Ricks v. Pauch, 17-12784 (United States District Court for the Eastern District of Michigan, Southern Division March 23, 2020).

Robbins v. State, 1939-00 (Court of Criminal Appeals of Texas October 23, 2002).

Robert Lee Stinson v. James Gauger, Lowell Johnson, and Raymond Rawson, 13-3343 (United States Court of Appeals, Seventh Circuit August 25, 2015).

Roberts, P. (2015). Paradigms of forensic science and legal process: a critical diagnosis. *Phil. Trans. R. Soc. B, 370: 20140256*, 1-11.

Robinson v. State, A02A2042 (Court of Appeals of Georgia, Second Division February 11, 2003).

Roth, N. (1997). *The New York State Police Evidence Tampering Investigation Report to the Honorable George Pataki.* Ithaca, New York.

Saks, M. J., & et al. (2016). Forensic bitemark identification: weak foundations, exaggerated claims. *Journal of Law and the Biosciences*, 538-575.

Saks, M. J., & Koehler, J. J. (2005). The Coming Paradigm Shift in Forensic Identification. *Science, 309*, 892-895.

Salazar, M., & Noyola, A. (2013). *Laboratory Report, Clayton Martel.* Houston, TX: Houston Police Department Forensic Services Command, Crime Laboratory Division.

Santos, F. (2006, December 21). In Quest for a Killer, an Inmate Finds Vindication. *New York Times*.

Scheck, B. (2005). Barry Scheck Lectures on Wrongful Convictions. *Drake Law School Series on the Justice System.* Drake University Law School.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual Innocence.* New York: Random House.

Scherr, K., & Dror, I. (2021). Ingroup biases of forensic experts: perceptions of wrongful convictions versus exonerations. *Psychology, Crime & Law, 27*(1), 89-104.

Scientific Working Group on DNA Analysis Methods. (2017). *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.* Washington, DC: Federal Bureau of Investigation.

Scientific Working Group on DNA Analysis Methods. (2019). *Interpretation Guidelines for Mitochondrial DNA Typing by Forensic DNA Testing Laboratories.* Washington, DC: Federal Bureau of Investigation.

Scientific Working Group on Friction Ridge Analysis, Study and Technology. (2008). *Standard for Simultaneous Impression Examination.*

Sensabaugh, G. (1986). Perspectives on the Future of Forensic Immunology. *Proceedings of the International Symposium on Forensic Immunology* (pp. 111-116). Quantico, Virginia: The Laboratory Division, Federal Bureau of Investigation.

Sensabaugh, G. A. (1977). *Identification and Individualization of Semen in the Investigation of Rape.* National Institute of Law Enforcement and Criminal Justice.

Sensabaugh, G. F., & Blake, E. T. (1976). Genetic Markers in Human Semen: A Review. *Journal of Forensic Science, 21*(4), 784-796.

Sensabaugh, G. F., Bashinski, J., & Blake, E. T. (1985, March). The Laboratory's Role in Investigating Rape. *Diagnostic Medicine*, 46-53.

Serological Research Institute. (1988). *Training Manual.* Richmond, CA: Serological Research Institute.

Servick, K. (2016, March 7). Reversing the legacy of junk science in the courtroom. *Science*. doi:doi: 10.1126/science.aaf4158

Shoemaker, J., Lynch, R., Hoffmann, J., & Sly, W. (1992). Misidentification of propionic acid as ethylene glycol in a patient with methylmalonic acidemia. *J. Pediatr., 120*(3), 417-421. doi:10.1016/s0022-3476(05)80909-6

Shugarts, J. (2018, July 20). State police arrest fourth suspect in Oxford arson. *Republican American Archives*.

Siegel, B. (1987, May 10). A System on Trial : Sentencing the Wrong Man to Die. *Los Angeles Times*.

Siegel, B. (1999, July 11). Judging Parents as Murderers on 4 Speck of Blood. *Los Angeles Times*.

Smith, R. (2004). *Request for Latent Print Consultantion Services.* Meridian, Mississippi: Ron Smith & Associates.

Smith, T. P., Smith, A., & Snipes, J. B. (2016). A Validation Study of Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework. *Journal of Forensic Sciences, 61*(4), 939-946.

Snowden v. Singletary, 94-4303 (US Court of Appeals, Eleventh Circuit February 18, 1998).

Sommer v. United States, 09cv2093-CAB (United States District Court for the Southern District of California December 5, 2013).

Spadaro v. City of Miramar, 11-61607-CIV-COHN/SELTZER (United States District Court for the Southern District of Florida February 7, 2013).

Stacey, R. A. (2004). A Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case. *J. Forensic Identification*, 706.

State ex rel. Schmitt v. Green, WD83688 (ourt of Appeals of Missouri, Western District, Writ Division April 28, 2020).

State of Florida v. Anthony Caravella, 84-339CF (Circuit Court of the 17th Judicial Circuit July 31, 1984).

State of Idaho v. Carlos Adrian Cruz-Romero, 42994 (Court of Appeals of the State of Idaho March 31, 2016).

State of Illinois vs. Alejandro Dominguez (1990).

State of Maryland v. Bryan Rose, K06-0545 (Circuit Court for Baltimore County 2007).

State of Montana vs. Chester Bauer, 83-407 (Supreme Court of the State of Montana June 4, 1984).

State of North Carolina vs. Knolly Brown, Jr., 08-CRS-50309 (North Carolina Innocence Inquiry Commission Hearing in the General Court of Justice, Superior Court Division, County of Edgecombe December 9 & 10, 2015).

State of Ohio vs. Anthony Green (1988).

State of Texas v. Ben Salazar, 0921724 (District Court of Travis County, 299th Judicial District 1992). Retrieved from https://convictingtheinnocent.com/exoneree/ben-salazar/

State of Texas v. Larry C. Fuller (August 25, 1981).

State of Wisconsin v. Joseph Awe, 07 CF 54 (State of Wisconsin Circuit Court Marquette County March 21, 2013).

State v. A.O., 198 N.J. 69 (Supreme Court of New Jersey March 4, 2009).

State v. Awe, 2009AP633-CR (Court of Appeals of Wisconsin, District 4 March 4, 2010).

State v. Boquette, 443 So. 2d 295 (1983). Retrieved from https://convictingtheinnocent.com/exoneree/orlando-boquete/

State v. Bridges, 9126SC657 (Court of Appeals of North Carolina October 20, 1992).

State v. Byrd, 424823 (262nd District Court of Harris County, Texas Trial Court 1985). Retrieved from https://convictingtheinnocent.com/exoneree/kevin-byrd/

State v. Campbell, C-170666 (Court of Appeals of Ohio, First Appellate District, Hamilton County May 17, 2019).

State v. Crotzer, 425 So. 2d 159 (Circuit Court for Hillsborough County 1982). Retrieved from https://convictingtheinnocent.projects.law.duke.edu/exoneree/alan-crotzer/

State v. Dail, 402 S.E.2d 421 (1989). Retrieved from https://convictingtheinnocent.projects.law.duke.edu/exoneree/dwayne-dail/

State v. Fain, 15414 (Supreme Court of Idaho April 4, 1989).

State v. Glover, 102828, 102829, and 102831 (Court of Appeals of Ohio, Eighth Appellate District, Cuyahoga County May 5, 2016).

State v. Gregory Flynt Taylor, Wake County 91CRS71728 (North Carolina Innocence Inquiry Commission September 3, 2009). Retrieved from http://innocencecommission-nc.gov/wp-content/uploads/state-v-taylor/state-v-taylor-brief.pdf

State v. Harris (1987). Retrieved from https://convictingtheinnocent.com/exoneree/william-harris/

State v. Jerry Watkins, 30S00-8708-CR-764 (09 19, 1986). Retrieved from https://convictingtheinnocent.com/exoneree/jerry-watkins/

State v. Johnston, 412 (Court of Appeals of Ohio, Fourth Appellate District, Hocking County August 6, 1986).

State v. Kaginyera/Wilcoxson (North Carolina Innocence Inquiry Commission September 22, 2011).

State v. Kordonowy, 90-317 (Supreme Court of Montana December 23, 1991).

State v. Lavernia (1984). Retrieved from https://convictingtheinnocent.com/exoneree/carlos-marcos-lavernia/

State v. LeFever, CA-3535 (Court of Appeals of Ohio, Fifth Appellate District, Licking County November 18, 1991).

State v. Mark Bravo, VA003313 (Superior Court of Los Angeles County 1990). Retrieved from https://convictingtheinnocent.projects.law.duke.edu/exoneree/mark-bravo/

State v. Moore, ED99077 (Court of Appeals of Missouri, Eastern District, Division One October 22, 2013).

State v. Quattrocchi, 95-343-C.A. (Supreme Court of Rhode Island July 31, 1996).

State v. Troy Web, 397 S.E.2d 539 (1989). Retrieved from https://convictingtheinnocent.com/exoneree/troy-webb/

State v. Vann, 03-S01-9706-CR-00068 (Supreme Court of Tennessee, Eastern Section, At Knoxville September 21, 1998).

State v. Vann, E2009-01721-CCA-R9-CD (Court of Criminal Appeals of Tennessee, At Knoxville May 25, 2010).

Stokes, J. (2019). Technical note: Next generation identification — A powerful tool in cold case investigations. *Forensic Science International, 299*(June), 74-79.

Stolorow, M., & Maddox, L. (2011). *Affidavit, Appendix B, Brandon Lee Moon Investigation.* Texas Forensic Science Commission.

Stubbs v. State, 2001-KA-01361-SCT (Supreme Court of Mississippi March 20, 2003).

Taylor, M. C., Laber, T. L., Kish, P. E., Owens, G., & Osborne, N. K. (2016). The reliability of pattern classification in bloodstain pattern analysis—part 2: bloodstain patterns on fabric surfaces. *Journal of forensic sciences*, 1461-1466.

Taylor, M., & Doyle, M. (2011, 03 20). Army slow to act as crime-lab worker falsified, botched tests. *McClatchy*.

Testimony of Courtland Cunningham and Joseph D. Nicol. (1964, April 1). *Warren Commission Report and Hearings*. Washington, DC.

Texas Forensic Science Commission. (2016). *Forensic Bitemark Comparison Complaint Filed by National Innocence Project on Bhealf of Steven Mark Chaney - Final Report.* Austin, TX: Texas Forensic Science Commission.

Texas Forensic Science Commission. (2018). *Report in Compliance with HB-34 (85th Legislature).* Austin, TX.

The Laboratory Division, Federal Bureau of Investigation. (1985). Proceedings of the International Symposium on Forensic Hair Comparisons. Washington, DC. Retrieved March 30, 2020, from https://www.ncjrs.gov/pdffiles1/Digitization/116592NCJRS.pdf

The Laboratory Division, Federal Bureau of Investigation. (1986). Proceedings of the International Symposium on Forensic Immunology. Quantico, Virginia.

Thompson, W. (2003). *Review of DNA Evidence in State of Texas v. Josiah Sutton (District Court of Harris County, Cause No. 800450).* Irvine, CA: KHOU.

Thompson, W. C. (2007, December). Beyond Bad Apples: Analyzing the Role of Forensic Science in Wrongful Convictions. *Southwestern Law Review, 37*, 971-994.

Tillman, L. B. (1983). State v. Johnson. Trial testimony. Retrieved 12 27, 2020, from www.convictingtheinnocent.com

Tipton, V. (1991, September 22). 'Fight's Over' Freed From Prison, Patricia Stallings Tries To Start Life Again. *St Louis Post-Dispatch*, pp. 1, 43.

Tribble v. District of Columbia, 2013 CA 003237 B (Superior Court of the District of Columbia, Civil Division February 26, 2016).

Tuerkheimer, D. (2009). The Next Innocence Project: Shaken Baby Syndrome and the Criminal Courts. *Washington Law Review, 87*(1).

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences, 108*(19), 7733-7738.

United States of America v. Brian Keith Rose, Criminal No. CCB-08-0149 (United States District Court, D. Maryland. December 8, 2009).

United States v. Hayat, 2:05-CR-0240 (United States District Court for the Eastern District of California August 17, 2017).

United States v. Hebshie, 02cr10185-NG (United States District Court for the District of Massachusetts November 15, 2010).

United States v. Nejad, 18-cr-224 (US District Court for the Southern District of New York June 9, 2020).

United States v. Russotti, 83-1054, 83-1055, 83-1056, 83-1069 (US Court of Appeals for the Second Circuit August 29, 1983).

University of California Irvine Newkirk Center for Science & Society, University of Michigan Law School, and Michigan State University College of Law. (2020, January 28). Retrieved from The National Registry of Exonerations: https://www.law.umich.edu/special/exoneration/Pages/about.aspx

US Department of Justice. (2019, March 19). *Uniform Language for Testimony and Reports.* Retrieved from US Department of Justice: https://www.justice.gov/olp/uniform-language-testimony-and-reports

US Department of Justice. (2020). *US Department of Justice.* Retrieved from Uniform Language for Testimony and Reports: https://www.justice.gov/olp/uniform-language-testimony-and-reports

van den Eeden, C. A., de Poot, C. J., & van Koppen, P. J. (2019). The forensic confirmation bias: a comparison between experts and novices. *Journal of forensic sciences, 64*(1), 120-126.

Victor Caminata v. Michael Jenkinson, 16-1451 (US Court of Appeals for the Sixth Circuit August 26, 2016).

Warden, R. (n.d.). *First DNA Exoneration.* Retrieved April 3, 2020, from Bluhm Legal Clinic Northwestern Pritzker School of Law: https://www.law.northwestern.edu/legalclinic/wrongfulconvictions/exonerations/il/gary-dotson.html

Warney v. Monroe County, 08-0947 (US Court of Appeals for the Second Circuit November 13, 2009).

Warren, J. (2013). *Time Since Intercourse: Studies and Estimates.* UNT Health Science Center. Association of Forensic DNA Analysts and Administrators. Retrieved from https://afdaa.org/2013/wp-content/uploads/2013/01/Time-Since-Intercourse-Studies-and-Estimates-JWarren.pdf

Washington State Patrol Crime Laboratory Division. (2019). *Biochemical Analysis Procedures.* Retrieved from https://www.wsp.wa.gov/forensics/docs/crimelab/manuals/technical/dna/Biochem_Procedures_Manual_Revision_20.pdf

Weick, K. E., & Sutcliffe, K. M. (2001). *Managing the Unexpected.* San Francisco: Jossey-Bass.

West, E. (2010). *Court findings of ineffective assistance of counsel claims in post-conviction appeals among the first 255 DNA exoneration cases.* Innocence Project.

Williamson v. State of Oklahoma, F-88-501 (Court of Criminal Appeals of Oklahoma May 15, 1991).

Woodling, B., & Heger, A. (1986). The use of the colposcope in the diagnosis of sexual abuse in the pediatric age group. *Child abuse & neglect, 10*(1), 111-114.

Woodling, B., & Kossoris, P. (1981). Sexual Misuse: Rape, Molestation, and Incest. *Ped. Clinics N. Amer., 28*(2), 481.

Wraxall, B., & Fedor, T. (2001). *Oklahoma City Police Department Crime Laboratory Serology Audit.* Serological Research Institute.

Wright, F. D. (2011). Human Bitemarks, NAS Report and Daubert. *National Science and Technology Council, Committee of Science, Subcommittee on Forensic Science.* Washington, DC.

Yell v. Commonwealth, 2006-SC-000327-MR (Supreme Court of Kentucky December 20, 2007).

Zain, F., Smith, T., & Myers, H. (1989). Population Data of Casework in West Virginia on Six Genetic Marker Systems. *Journal of Forensic Science, 34*(4), 1007-1010.

Zalman, M., & Windell, J. (2019). The Bite Mark Dentists and the Counterattack on Forensic Science Reform. *Albany Law Review, 83*, 749.

Zheng, X. (n.d.). *NIST Ballistics Toolmark Research Database*. Retrieved from National Institute of Standards and Technology: https://tsapps.nist.gov/NRBTD

Appendix 1. Codebook

This codebook was provided with a spreadsheet of deidentified data to the National Archive of Criminal Justice Data, which is expected to make the dataset available for secondary analysis. The codebook is based on the study coding guide but includes additional information to guide secondary researchers or researchers who may conduct future research related to wrongful convictions and forensic science.

Background

The study collected data concerning wrongful convictions that may be associated with forensic science errors. The study population included cases from the National Registry of Exonerations (NRE) (University of California Irvine Newkirk Center for Science & Society, University of Michigan Law School, and Michigan State University College of Law, 2020) that were coded, "False or Misleading Forensic Evidence" as of July 5, 2021. This data set includes 732 cases covering the period from 1900 to 2021.

This guide was developed to assist in the accurate and reliable coding of wrongful conviction cases related to forensic science errors. The guide was used as a reference document for the completion of the Case Review Form, which is attached as Appendix A. Data entry for the Case Review Form was implemented in Google Forms. The guide was designed to provide a basis for decisions in the Case Review Form on issues likely to arise in the cases, including the study's error typology and accepted testimony standards. The reviewer was permitted to deviate from the guidance in this document if the justification was provided in the Case Review Form in the appropriate comment box of the relevant section. The Case Review Form includes all information elements listed below. As long as the form was completed fully, all study information was collected. The form was treated as a confidential work product and was shared with study participants and stored on a protected, Department of Justice file system.

The archived dataset does not include identifiable information about the individuals involved in the cases or qualitative information recorded in the Case Review Forms. The archived dataset includes the quantitative case coding. This coding guide provides the description of the data and calculations that have been used to derive findings about the dataset. Each dataset record includes an Archive Record Number that is unique to each case and Examination Code that is unique to each forensic examination coded during the study. The Examination Code includes the Archive Record Number and a letter identifier for the specific forensic examination in the case.

Note that some included variables may have some value in the identification of the defendant or forensic examiner in a case, but only if the individual has already been identified by other means.

The original dataset included the following variables that have been removed to prevent reidentification:

1. Last name
2. First name
3. Ethnicity
4. Sex
5. Birthday
6. Age at crime

7. NRE link (i.e., web link to page on the National Registry of Exonerations website)
8. Date that crime occurred
9. Crime year
10. State
11. County of record (trial)
12. NRE Coding
13. Child victim/case type
14. Drug case only. Note: This field contained information about the nature and geographic origin of the drug case that has been removed. The field is retained to indicate if the case involved only drug possession charges (coded as "Yes") or involved other charges (coded as "No").
15. Conviction date
16. Conviction year
17. Sentence
18. Linked cases (for cases with multiple defendants or linked by the forensic examiner)
19. Type of link among cases (defining the type of link among cases)
20. Exoneration date
21. Exoneration year
22. Official acts of exoneration (including type of appeal)
23. Exoneration advocates
24. Post-exoneration outcomes (including civil case outcomes)
25. Compensation (amounts of publicly-disclosed government compensation, civil lawsuit awards, or civil lawsuit settlements related to the wrongful conviction)
26. Compensation annuity (annual government compensation related to the wrongful conviction, if any)
27. Name of examiner
28. Forensic science organization

Qualitative fields are not included in the dataset, such as comments, case descriptions, or supporting information concerning coding decisions. NRE coding related to other case elements (e.g., false confessions) are not included in the dataset.

For the purposes of this study, forensic science errors may include errors related to evidence collection, forensic laboratory practice, testimony, or court proceedings. In many cases, the error is related to forensic science but may not have been the responsibility of a forensic examiner.

The study methodology includes several steps designed to mitigate the possibility of errors related to the interpretation of forensic science errors. All cases have been documented using publicly available sources as much as possible. All findings referenced case documentation, evidentiary standards, or other sources to justify coding decisions. All conclusions concerning forensic science errors referenced specific documentation in the available case record. The study used an expert panel that reviewed the findings to produce a consensus final assessment in a subset of cases.

All cases were coded using a predefined Case Review Form. The elements of the Case Review Form are provided below. This information relates to the coding process only. Dataset variable ranges and definitions are provided separately in Section 5.

Section 1. Defendant and Case Information

The following data was recorded from the NRE and verified, if possible, using independent documentation. The information was documented in the Case Review Form in the sections, Demographics, Crime, and Trial.

Demographics

1. Record Number: A separate record number was assigned to each defendant. This variable was randomized in the archive file.
2. NRE: Was this case in the NRE database? (YES/NO)
3. NRE data link
4. Last Name
5. First Name
6. Birthdate
7. Ethnicity
8. Sex

Crime

9. Most serious charge
10. Date that alleged crime occurred
11. State (where crime occurred)
12. County (where crime occurred)
13. Other crimes alleged
14. Arson alleged?
15. Child victim?
16. Homicide?
17. Misdemeanor?
18. Drug case only?
19. Shaken baby case?

Trial

20. Date convicted
21. Sentence
22. Co-Defendant Confessed?
23. Federal case?
24. Jailhouse Informant?
25. Guilty Plea?
26. Mistaken Witness Identification?
27. False Confession?
28. Perjury or False Accusation?
29. Official Misconduct?
30. Inadequate Legal Defense?
31. Linked cases
32. Type of link among cases
33. Comments (for crime and trial; used to clarify any entries)

Exoneration

1. Date exonerated
2. Exoneration advocated by? (Prosecution, Conviction Integrity Unit, Innocence Organization, State Innocence Commission, Other, None)
3. DNA Exoneration?
4. Posthumous exoneration?
5. Actual innocence?

   Select YES only if there has been a finding of actual innocence by an official government act, including retrial, appeal, Innocence Commission finding, or civil suit. Keep in mind that a wrongful conviction may relate to factors other than the guilt or innocence of the defendant.
6. Basis for actual innocence finding (documenting the official action that provides the basis for a finding of actual innocence)
7. Basis for exoneration other than innocence

   This may include a habeas corpus proceeding or retrial that does not lead to conviction. If "other", describe the basis in the Exoneration comments field in this section.
8. Comments
9. Case outcome
10. Trial transcripts
11. Other official case documentation
12. Other case documentation

This section also includes coding for linked cases and comments concerning the coding of the crime and trial information. Cases may be linked by multiple co-defendants in the same trial or for the same crime. Cases may also be linked because a forensic examiner was involved in multiple wrongful conviction cases, such as an examiner who performed fraudulent or questionable work over many years. Finally, cases may be linked because of fraudulent or negligent work in a forensic science organization that led to multiple wrongful convictions. For example, a crime laboratory may have performed many DNA analyses based on the misapplication of statistical methods.

Section 2. Evidence and Testimony

The study collected data to address claims of case errors related to forensic evidence in wrongful convictions. The form was marked as "*undetermined" if data was unavailable or not definitive*. These items were documented in the Case Review Form in sections: Exoneration, Forensic Evidence and Other Reviews. The form permitted the documentation of any number of forensic evidence types. Each separate analysis was treated as an "examination" and coded separately in the Forensic Evidence , Other Reviews, Error Context, and Error Classification sections. If multiple examiners completed analyses for a particular evidence type, each analysis was coded separately if it was possible to distinguish their contribution. Thus, two collaborating examiners may have been coded together as one examination.

Forensic Evidence

1. Probative evidence in the case that is not forensic evidence (text box). This information was not be coded in detail in the study.
2. Was probative value of forensic evidence needed to obtain conviction?

    a. Answer YES ***only*** if other probative evidence did not exist or was completely circumstantial.
    b. Document the justification if YES.
    c. Answer NO if a subjective interpretation of the intent of the jury would be required to make this finding.
3. Comments (at end of section)
    a. Provide additional information about forensic evidence. This field can be used for a longer description of the case. There is a separate comments section in the Error Classification section to describe the forensic evidence testimony and its implications for the overall case.

For each type of forensic evidence, document the following:

4. Type of evidence
5. Name of forensic examiner
6. What was the education level of the forensic examiners in the case? (Level and type of highest degree attained)
7. Extraneous case information: (YES/NO) In some cases, the forensic examiner may have been aware of case information that was not relevant to the forensic examination. For example, some fire debris investigators may be aware of the circumstances of a case that do not relate to the assessment of a fire scene. This knowledge may raise issues related to cognitive bias. The awareness of such information does not constitute a forensic science error but may be associated with errors. Answer YES only if the forensic examiner's knowledge of extraneous case information is documented in a reliable source.
8. Name(s) of forensic science organizations involved in the case (i.e., the organization, if any, that the examiner worked for at the time)
9. Type of organization: The organization may be a public crime laboratory, a private company that performs forensic analysis on a contract basis, or an independent consultant.
10. Accreditation status of forensic science organization at the time of the trial, if available.
11. Independence of the forensic science organization: Three options are provided. The forensic science organization may reside inside a police department, such as some fingerprint or digital evidence units. The forensic science organization may be organized as a separate entity but organized under a law enforcement agency, such as many pubic crime laboratories. The forensic science organization may be fully independent of law enforcement as an organization, such as commercial laboratories or many medicolegal death investigation entities.
12. Comments (concerning the particular type of forensic evidence)

<u>Other reviews</u>

13. Other research reviews of the forensic evidence in this case and findings.
14. Was the case classified as False or Misleading Forensic Evidence in the NRE database? (YES/NO)
15. Garrett/Neufeld: Was this case in the GN data set? (YES/NO)
16. Was the GN analysis correct?
17. Were the Garrett/Neufeld findings correct? (YES/NO; if NO, list findings and document the research error.) For this question, use the basis Garrett/Neufeld used to determine

the validity of testimony. If the testimony was in error on the basis that Garrett/Neufeld applied, mark this "YES". If the testimony was valid on the basis that Garrett/Neufeld applied, mark this "NO" and explain.

18. How did GN classify the errors? (List GN findings.)

    1) Non-Probative Evidence Presented as Probative
    2) Exculpatory Evidence Discounted
    3) Inaccurate Frequency or Statistic Presented
    4) Statistic Provided Without Empirical Support
    5) Non-numerical Statements Provided Without Empirical Support
    6) Conclusion that Evidence Originated from Defendant
    7) Masking and Quantitation Errors in Serology
    8) Withholding Forensic Evidence
    9) Gross Error in Analysis
    10) Failure to Conduct Elimination Testing or Comparison
    11) Invalid Prosecution Use of Forensic Science
    12) Failures of Defense Counsel
    13) Judicial Rulings on Forensic Science

This table clarifies the interpretation of the Garrett-Neufeld data set within the framework of the current study. Additional information about the current study's error typology is provided in the Error Context and Error Classification sections.

Garrett/Neufeld	"Best fit" to current study's error typology	Comments
Non-Probative Evidence Presented as Probative	Error Types 3c or 3d. Incorrect interpretation or characterization of a forensic result that implies an incorrect individualization or association or a different probative value than is acceptable under the standards of the discipline.	Corresponding testimony errors present forensic evidence that was exaggerated or improperly presented as probative. These errors associate evidence with a source based on an unscientific and unjustified interpretation. In the Garrett/Neufeld data set, most of these errors relate to serological evidence with limited probative value.
Exculpatory Evidence Discounted	Error Type 3c. Exclusion of relevant information that may support or refute the	Corresponding type 3c testimony errors exclude or minimize relevant, exculpatory forensic results or present exculpatory evidence as inculpatory. Error Types 1c (when this is done in the report phase) and

	conclusions of the report	5 (evidence not collected, examined or reported) may apply.
Inaccurate Frequency or Statistic Presented	Error Type 1b. Statement of statistical weight or probability that is not supported by science as defined by current standards of practice	Corresponding type 1b errors include cases in which the statistical weight of the evidence has been miscalculated. In these cases, a scientifically valid statistical calculation could have been done but was not reported correctly by the examiner.
Statistic Provided Without Empirical Support	Error Type 1b. Statement of statistical weight or probability that is not supported by science as defined by current standards of practice	Corresponding type 1b errors include cases in which there is no scientific basis for the calculation of a statistical weight of the evidence. This error would occur if a statistical weight is reported without an actual calculation having been completed.
Non-numerical Statements Provided Without Empirical Support	Error Type 1d. Statements that lack clarity or mislead a reader's ability to interpret the report.	Corresponding type 1d errors include cases in which the language used by the forensic scientist does not conform with testimony standards at the time of trial. Nonconformance with current standards should be reported under item 45.
Conclusion that Evidence Originated from Defendant	Error Type 2. Incorrect individualization or association.	All type 2 errors would correspond to this definition.
Masking and Quantification Errors in Serology	Error Type 1c. Exclusion of relevant information that may support or refute the conclusions of the report.	Corresponding type 1c errors include cases in which serological evidence was reported without full consideration of the problem of masking (such as victim serological contributions) or quantification (such as determination of the presence of semen).
Withholding Forensic Evidence	Error Type 5. Probative evidence not collected, examined or reported.	All type 5 errors would correspond to this definition.
Gross Error in Analysis	Error Type 2. Incorrect individualization or association.	Corresponding errors imply significant negligence or incompetence (but not fraud) on the part of the forensic examiner. Specific circumstances vary depending on the case.

Failure to Conduct Elimination Testing or Comparison	Error Type 5b. Forensic evidence not examined or a forensic result is excluded from an examiner's report.	Corresponding type 5b errors occur when an analyst fails to analyze relevant reference samples to eliminate other possible sources that may be associated with the evidence.
Invalid Prosecution Use of Forensic Science	Error Type 4c. Mischaracterization of forensic evidence by legal counsel	Corresponding type 4c errors include cases in which a prosecutor provided a misleading interpretation of accurate and reliable forensic testimony.
Failures of Defense Counsel	Error Type 4b. Inadequate defense associated with the presentation or review of forensic evidence or testimony	All type 4b errors would correspond with this definition.
Judicial Rulings on Forensic Science	Error Type 4d. Errors associated with judicial conduct, including acceptance of forensic evidence not aligned with evidentiary standards or incorrect instructions to jurors.	All type 4d errors would correspond with this definition.

Section 3. Error Context

The main purpose of this research study was the identification and characterization of forensic science errors in wrongful conviction cases. The Error Classification section of the Case Review Form provides a framework for the examination of each type of evidence relevant to a case.

Each forensic analysis was coded separately, when possible.

The Case Review Form provided questions that related to the error typology, claims regarding forensic science errors, and the context of forensic science errors, including:

1. Provide a general description of the substance and relevance of each type of forensic evidence. This comment section was used to provide a qualitative description of forensic science errors or other issues.
2. Was there a forensic science error related to this type of evidence?
   This field is "Yes" in any instance in which an error of any type occurred.
3. Did the testimony include an error corresponding to the NIST error typology? There were three drop-down boxes that covered the NIST typology. There was also comment box to document additional items or further information.

    a. The first drop-down box covered the main categories of the NIST typology, including Analyst/Expert Error, Fraud, Methods/protocol error and Instrumentation/Technology Limitations.
    b. The second drop-down box covered all types of Analyst/Expert Error. Measurement errors were coded under the Methods and Instrumentation drop-down box.
    c. The third drop-down box covered Methods and Instrumentation.
    d. Comments for additional information concerning the NIST typology.
4. Would technological improvements be used today to perform a forensic test with an improved scientific foundation and probative value?
    a. YES/NO
    b. Forensic method that would be employed
    c. Does the forensic method have an improved scientific foundation?
    d. Does the forensic method have an improved probative value?

    Note: Nuclear DNA analysis was assumed to have been possible in all cases involving serology evidence or out-of-date DNA methods (such as RFLP or DQ Alpha). Mitochondrial DNA was assumed to have been possible in cases involving hair microscopy evidence in the time period after 2000. In all cases in which an updated DNA method is now available, the new technique would be considered to have improved probative value and scientific foundation.

5. Was a presumptive test relied upon to adjudicate a case in which a confirmatory test would have produced a more reliable or probative result?
   Presumptive tests were performed in many cases involving drug or serological testing. In some of these cases, a plea or other conviction may have been based on a presumptive test in the absence of a confirmatory test. The field was YES only if the presumptive test was relied upon to obtain conviction in the absence of other forensic testing or probative evidence. Further delineation of the issue was documented in the Error Typology section, such as an incorrect association or inadequate laboratory practice.

Additional considerations

The remaining part of the Error Typology section was used to define any forensic science errors associated with the evidence based on the Error Typology section of this Coding Guide. The questions were organized with respect to the five main categories of the Error Typology. The Case Review Form included consideration of whether the evidence was presented correctly against two standards: the generally accepted standards at the time of the trial and the generally accepted standards in 2020 at the time of this study. This Coding Guide provided appropriate standards for most forensic disciplines that were encountered during the review in the Testimony Standards section.

The Case Review Form included items that ask whether the testimony included an error of individualization or association, as proven by subsequent analyses. The form included a comments section for the reviewer to document the re-analysis or new testing that demonstrated that the original match was an erroneous identification or association.

An individualization is a determination that two samples derive from the same source; practically, a determination that two samples derive from sources that cannot be distinguished within the sensitivity of the comparison process. The source may be a person, place, thing or event. (From Organization of Scientific Area Committees definition of individualization, see http://lexicon.forensicosac.org/Term/Home/index)

A classification is a determination that two samples derive from the same population of sources. The type of source may be a person, place, thing, or event. The population may be distinguished by shared traits or characteristics that may be referred to as class or subclass characteristics. An error of association occurs when it is determined that the two samples do not derive from the same population of sources.

A forensic science error occurs when it is determined that the two samples do not derive from the same source based on subsequent analysis. An exoneration may occur even in cases in which an association was correct, so a forensic science error finding was based on a reanalysis that definitively demonstrated that the original work produced an incorrect result.

Great care is needed when examining issues related to populations and source associations. For example, the source of a feather may be claimed to be a family (birds), genus (raptor), species (golden eagle) or a specific, individual bird. In principle, claims tying the feather to any of these could be made and could be erroneous or not depending upon how “population” is defined. These considerations were reflected in the analysis of possible classification errors.

NIST Error Typology

The complete NIST typology includes (National Institute of Standards and Technology, 2015):

Analyst/Expert Error

1. Errors due to human bias (i.e., cognitive bias, confirmation bias)
2. Forensic examiner variability
3. Errors due to improperly collected or improperly labeled evidence from crime scenes
4. Errors due to break in the chain of custody
5. Errors due to contamination and mislabeling of evidence
6. Errors due to mishandling (i.e., losing samples, sample mix-ups, sample mislabeling and sample contamination)
7. Errors due to misinterpretation of evidence
8. Errors due to misinterpretation of data
9. Errors in poorly following best practices, processes and methods
10. Errors due to poor documentation and transcriptions
11. Errors due to inadequately trained personnel
12. Errors due to analyst incompetence
13. Errors due to failure to review the analysis of the original analyst
14. Errors due to misinterpretation of post-mortem artifacts (i.e., artifacts due to resuscitation, exhumation, decomposition, embalming, rigor mortis, toxicological, environmental)
15. Measurement errors (i.e., systematic and random)

Fraud

1. Errors due to examiner fraud
2. Errors due to falsified reports
3. Errors due to suppression of exculpatory evidence
4. Errors due to exaggeration of test results

5. Errors due to false testimony about test results

Methods/protocol error

1. Errors due to unvalidated methods
2. Errors due to methods without scientific underpinnings
3. Errors due to inaccurate and misleading statistics
4. Error rates in scientific techniques
5. Measurement errors (i.e., systematic and random)

Instrumentation/Technology Limitations

1. Errors in software packages
2. Error rates in technology solutions
3. Laboratory equipment errors (i.e., poor or no calibrations)
4. Errors due to deficiencies in laboratory reference materials
5. Measurement errors (i.e., systematic and random)

Fraud was not coded under the NIST typology because it was captured in other coding elements under the Forensic Error Typology.

Cognitive Bias

The NIST error typology includes a category, "Errors due to human bias (i.e., cognitive bias, confirmation bias)." This category was coded under the heading, "Possible cognitive bias" and included any situation in which all of the following were true:

1. A type of human cognitive bias was present.
2. The bias could be documented in the public case record.
3. The bias affected the forensic examination.

There are difficulties in the assessment of cognitive bias in wrongful convictions because there are significant uncertainties in the retrospective documentation of the basis for an individual's conclusions. In the current study, coding was based on the assessment of "possible cognitive bias," meaning that some documentary evidence supports the possibility that cognitive bias influenced the examiner's conclusion, reporting, or testimony. The consideration of cognitive bias was not limited to contextual bias. An examiner may have exhibited one of many types of cognitive bias, or "predictable deviations from rationality," as suggested by Croskerry and others in studies of diagnostic failures in clinical medicine. (Croskerry, Singhal, & Mamde, 2013) There was no attempt to code subcategories of cognitive bias. Thus, the cases associated with "possible cognitive bias" represent a population of cases that can be studied because bias effects could be documented in the case history, not necessarily the entire population of cases in which bias may have been present or a systematized coding of bias types.

All types of human bias were coded as "possible cognitive bias," if that could be documented and regardless of the bias type or origin. Possible cognitive bias may have included any type of bias, but bias was not coded when there was no objective basis in the documentation to support the claim. In some cases, examiners themselves made statements that support the possibility of bias. In other cases, contextual information or collegial influence can be documented as an influence on the examiner. The coding was necessarily subjective, and other observers might reasonably differ in their analysis of individual cases.

Types of cognitive bias that were included in “possible cognitive bias” and observed in the dataset included:

Contextual bias (case-based): The forensic examiner was influenced by case context and produced an error consistent with the expectations from that case context. Example: The examiner was aware that the suspect had confessed to the crime and produced an erroneous association of evidence with the suspect. Note that coding of “possible cognitive bias” in this instance would require documentation of both elements, i.e., it can be documented that the examiner was aware of the confession and the examiner made an erroneous association consistent with that context.

Contextual bias (forensic information): The forensic examiner was influenced by the results of other forensic evidence and produced an error consistent with the expectation from those results. Example: A serologist was aware of a hair comparison association and produced an interpretation of the serology that erroneously discounts an exculpatory analytical result. This example is specific, but the issue may arise from any type of forensic examination or information.

Contextual bias (moral panic): The forensic examiner was influenced by the nature of the alleged crime or community revulsion concerning the alleged crime, leading to an incorrect or misleading analysis or interpretation of evidence. Example: A fire debris investigator bases an incendiary-cause interpretation on the perceived lack of remorse of an arson suspect concerning the deaths of victims in the fire.

Base rate bias: The forensic examiner was influenced by experience or prior knowledge to have an expectation about the case and produced an error without sufficient foundation in the objective record. Example: A crime scene investigator assumes that a shooting is the result of a suicide because suicide is much more common than homicide. The investigator fails to collect evidence at the scene to substantiate that view or other case information.

Target bias: The forensic examiner was influenced by evidence features to have an expectation about reference features, leading to an incorrect result or incorrect representation of a result. Example: A fingerprint examiner mad an incorrect match based on a close non-match reference print without appropriate consideration or documentation of non-matching features.

Semmelweis reflex: The forensic examiner continued to employ outdated methods based on prior experience or training, leading to an incorrect analysis or interpretation of evidence features. Example: A fire debris investigator did not use the current version of NFPA 921 to conduct an investigation and based a forensic interpretation on unvalidated methods from pre-NFPA-921 training.

Authority bias: The forensic examiner was influenced by a colleague’s analysis and interpretation of evidence, leading to a failure to perform a thorough and reliable forensic analysis and an error in analysis or interpretation. Example 1: A pattern evidence examiner agreed with an incorrect evaluation of pattern evidence based on a non-blind review. Example 2: A forensic pathologist agreed with an unreliable interpretation of evidence based on a colleague’s interpretation and cited the need for professional deference.

Choice-supportive bias: A criminal justice practitioner discounted or ignored an exculpatory result or interpretation of forensic evidence that contradicted the predominant theory of a case. A forensic examiner may exhibit choice-supportive bias in their analysis, interpretation, or communication about a forensic result. Example 1: A forensic serologist testified that a male defendant could not be excluded from evidence although the serological profile of a crime scene stain contained spermatozoa but did not contain blood group substances from the defendant. Example 2: A police investigator claimed that a non-matching, probative latent print did not exculpate a suspect in a case.

Note: Cascade bias is the promulgation of case errors based on prior investigative or forensic errors. It is commonly associated with authority bias or choice-supportive bias.

Calibration bias: A forensic examiner exaggerated the probative value of evidence outside the scientific limits of the forensic method. In this situation, "calibration" refers to the ability of the examiner to interpret or communicate the perceived probative value of evidence in alignment with the actual value of the evidence. Example: A medical practitioner based an interpretation of pediatric sexual assault on victim statements, not medical findings.

Law enforcement bias: A forensic examiner produced an incorrect analysis or interpretation with the intent to align the forensic evidence with a conclusion that the suspect in the case was the source of the evidence. Example: A trace examiner did not document or preserve evidence collection and analysis with the intent to minimize information that could be used by a defendant in legal proceedings.

Prestige bias: A forensic examiner exaggerated or misrepresented evidence based on their desire for prestige within their organization or profession. Example: An anthropologist provided an individualization of the source of an evidence shoeprint based on the anthropologist's unvalidated research findings.

Other types of bias may have been present in the dataset but were not observable in the documented case record. The study did not attempt to produce a detailed coding of bias types and effects.

Section 4. Error Classification

Each error type was noted with a justifying description in the Description and Comments field at the beginning of the section for the type of evidence under review. The form asked multiple questions about each error type to assist the reviewer in classifying errors. A description of the elements of the error typology is provided here.

Error Type 1. A forensic science report that contains a misstatement of the scientific basis of a forensic science examination. The misstatement may include:

a. Statement of individualization that is not supported by science as defined by current standards of practice;
b. Statement of statistical weight or probability that is not supported by science as defined by current standards of practice;
c. Exclusion of relevant information that may support or refute the conclusions of the report;
d. Statements that lack clarity or mislead—whether intended or unintended—a reader's ability to interpret the report.

All type 1 errors relate to analyses as they were conducted and reported by a forensic examiner *prior* to a trial. Testimony may reflect errors in forensic science reports. In that case, the error is recorded as type 1 or type 2. In some cases, testimony will include errors that are not present in forensic science reports. In that case, the error is recorded as type 3. Keep in mind that “current standards of practice” may refer to standards at the time of trial or standards in place in 2020. The Case Review Form clearly delineates these two standards in two different sections.

Error Types 1a and 1b are interpretation errors and are determined based on testimony standards and scientific knowledge at the time of the trial. Statements of individualization include language that associates evidence with a specific person (such as the defendant) or source (such as a particular firearm). Statements of statistical weight or probability will describe the strength of an association based on comparisons with a population of sources. In determining this error, it is not relevant if the association is correct, only if the association has been misrepresented with respect to its scientific foundation or statistical weight.

Example: A serological analysis associates a crime scene stain to the defendant based on blood type. The analyst reports that the defendant is the source of the crime scene stain. In this case, the analyst has made an erroneous statement of individualization, because blood type cannot be used to identify an individual as the sole source.

Example. A serological analysis associates a crime scene stain to the defendant based on blood type. The analyst reports that only 5% of the male population shares the same blood type, but the correct percentage is 20%. The analyst has made an erroneous statement of statistical weight. Note that the serological data may be accurate, even if the interpretation is in error.

Error Type 1c includes situations in which evidence reached the laboratory but relevant data was not collected or reported. The evidence may not have been analyzed because of resource constraints or poor judgment. The evidence may not have been analyzed sufficiently to produce a reliable result (such as when reference or background data is not collected). All relevant evidence may have been analyzed, but results were not documented appropriately or reported completely. The error must be relevant to the probative value of the evidence as understood at the time of the trial. It is not necessary to examine every piece of crime scene evidence if further analysis would not have produced additional probative information.

Error Type 1c should not apply on the sole basis that a scientific or technological improvement might have provided a more reliable or probative result had it been available at the time of the trial. That issue may be documented elsewhere on the coding form.

Error Type 1d applies to a forensic science report that results in a relevant misinterpretation of the results of the analysis. Relevance refers to a change in the probative value of the forensic evidence. The report may be misleading because of the negligence or incompetence of the writer of the report.

If the original report is not available, it may not be possible to determine if an Error Type 1d occurred. Some forensic testimony is provided by police investigators or others who may not have performed the original analysis. The typology is based on a judgment concerning whether an error arose because of the analyst’s original misinterpretation (type 1 error) or during the presentation of testimony (error type 3).

If a report is available and is deemed vague or unclear, then it is a type 1d error. If the testimony is unclear but the quality of the forensic report cannot be definitively determined, then it is a type

3 error and was not coded under type 1d. If the analysis is incorrect but clearly stated, it is not a type 1d error.

The report may be misleading because of fraudulent intention of the writer of the report. Thus, fraudulent conduct may include Error Type 1d (if the report is unclear and misleading), Error Type 2d (if the analysis led to a fraudulent association), and/or Error Type 3 (if the examiner provides perjurious testimony).

Example: In a fire debris case, the fire debris investigator may have concluded that the difference in temperature between the floor and ceiling were due to the presence of accelerant, but the examiner did not report that accelerant residue was not found inside the house. This may be Error Type 1c (excluding relevant data). The examiner also concluded that the fire was "very aggressive" and made other, qualitative, non-scientific statements in the report and testimony. This may be Error Type 1d (misleading language in a forensic report).

Error Type 2. A forensic science report that contains an incorrect individualization or association of a piece of evidence with a source or class of sources or the incorrect interpretation of a forensic result that implies an incorrect individualization or association. An error may be intended or unintended.

e. Incorrect individualization
f. Incorrect classification
g. Incorrect interpretation of a forensic result
h. Fraudulent or intended association of evidence with source

All type 2 errors relate to analyses as they were conducted and reported by a forensic examiner prior to a trial. The association may be an individualization or a classification.

As noted above under Additional Considerations, an individualization is a determination that two samples derive from the same source; practically, a determination that two samples derive from sources that cannot be distinguished within the sensitivity of the comparison process. The source may be a person, place, thing, or event. (From Organization of Scientific Area Committees definition of individualization, see http://lexicon.forensicosac.org/Term/Home/index) An error of identification occurs when it is determined that the two samples do not derive from the same source.

A classification is a determination that two samples derive from the same population sources. The type of source may be a person, place, thing, or event. The population may be distinguished by shared traits or characteristics that may be referred to as class or subclass characteristics. An error of association occurs when it is determined that the two samples do not derive from the same population of sources.

Type 2 errors include *actual errors*, i.e., the class or source was associated with crime scene evidence because the forensic method was applied or interpreted incorrectly. Statements of individualization will include language that associates a specific person (such as the defendant) or source (such as a particular firearm) with crime scene evidence. Statements of classification will include language that associates a class of people (such as all people with blood type A) or sources (such as all shoes of a particular brand and model) with crime scene evidence.

Type 2a and 2b errors occur when methods are misapplied to produce inaccurate data. For example, a latent print examiner may describe inaccurate points of comparison or an instrument may provide faulty results.

Type 2c errors occur when an incorrect data interpretation produces an incorrect individualization or classification. This error type differs from Type 2a and 2b errors, which occur when the method is misapplied to produce erroneous data. Type 2c errors are misinterpretations of accurate data that relied on reliable analytical methods.

Type 2c errors may or may not be associated with type 1 errors. In both cases, the data has been misinterpreted. A type 2c error occurs when the misinterpretation results in an incorrect individualization or classification.

Type 2d errors include deliberate falsification of forensic reports, if the report's conclusions include a false identification or classification.

Example: A drug chemist produces seized drug report without performing the required analysis. This is a type 2d error.

Error Type 3. Testimony at trial that reports forensic science results in an erroneous manner. An error may be intended or unintended. Type 3 errors may include:

l. Misstatements concerning the scientific basis of a forensic science examination;
m. Misstatements concerning the statistical basis of a forensic science result;
n. Incorrect interpretation of a forensic result that implies an incorrect individualization or association;
o. Incorrect characterization of a forensic result that implies a different probative value than is acceptable under the standards of the discipline;
p. Exclusion of relevant information that may support or refute the conclusions of a forensic report.

Error Type 3 covers testimony concerning forensic science results provided as part of a trial that resulted in a wrongful conviction. The testimony may be provided by the forensic scientist who performed the analysis, a representative of the forensic science organization, a police investigator, or other individual. Errors that occurred prior to the trial during the forensic analysis and reporting (type 1 or 2 errors) are distinguished from errors that occurred during the testimony at the trial (type 3 error). When an error is coded in more than one place, the justification is provided as part of the documentation of the error.

The Case Review Form permitted testimony to be assessed based on the standards in place at the time of the trial or the standards of the present day. In some cases, the testimony would be considered appropriate based on standards that have been changed in the intervening years. This study's goals included an analysis of testimony errors with this distinction in mind. Error Type 3 coding is based on an assessment of testimony based on past and present testimony standards separately. This guide provides a basis for these decisions in the Testimony Standards section.

The Garrett-Neufeld review used a unique set of standards for the validity of testimony developed for the purposes of that study only. Therefore, Garrett-Neufeld's findings and this study's review of those findings were documented in the Case Review Form under Other Reviews, not within the Error Typology.

Error Type 3a applies when testimony does not reflect an accurate understanding of the scientific basis for the forensic science examination. The testimony may contain an error relating to the description of the basis of the method or its application in the case. The testimony may misstate the scientific research basis for the validity or reliability of the method.

Error Type 3b applies when testimony mischaracterizes the statistical basis for the result of a forensic science examination. In this case, the testimony should include a statement that relates to a quantitative estimate of the statistical basis. Qualitative statements should only be assessed under this error type when used to justify the probative value of the result within a statistical framework, such as when an examiner states the number of times that the method has been employed by the examiner or within the examiner's laboratory (which is appropriate when discussing an examiner's expertise but not when justifying the probative value of a forensic result). This error applies when a statistic was based on faulty assumptions, even if the calculation based on those assumptions was correct.

Error Type 3c applies when testimony incorrectly associates a piece of evidence with a source or class of sources. This error applies only when the association occurred during testimony and was not based on an error from a forensic report. In other words, an error that occurs in the laboratory would be classified under Error Type 1 or 2. An accurate forensic report that is mischaracterized as an identification or association would be a type 3c error.

Error Type 3d applies when testimony includes an incorrect characterization of a forensic result that implies a different probative value than is acceptable under the standards of the discipline. In this case, testimony reflects an incorrect qualitative interpretation. The probative value of the evidence may be misstated to imply that the result could only apply to an individual or a small set of individuals. In some cases, the result may exonerate the defendant but the testimony is presented in a manner that inculpates the defendant.

Error Type 3e applies when testimony fails to provide information or data that may exculpate the defendant. The information must be relevant to the probative value of a forensic result. The information may support or refute the conclusions of a forensic report. In this case, the information was collected or analyzed but is not provided during testimony. This error does not apply if the evidence wasn't collected or analyzed or shared with defense counsel as a Brady violation (Error Types 1c or 5). Keep in mind that an examiner may not have had the opportunity to provide exculpatory information as a result of a judge's decision during the trial. In this case, Error Type 3e does not apply, but Error Type 4d may apply.

Error Type 4. The mischaracterization of forensic testimony at trial by an officer of the court, including but not limited to:

a. Exclusion of potentially exculpatory forensic evidence;
b. Inadequate defense associated with the presentation or review of forensic evidence or testimony;
c. Mischaracterization of forensic evidence by legal counsel;
d. Errors associated with judicial conduct, including the acceptance of forensic evidence not aligned with evidentiary standards or incorrect instructions to jurors;

Error Type 4 includes circumstances in which officers of the court are responsible for forensic science errors. In these cases, forensic science professionals or police investigators may have been responsible for errors, but those errors are coded elsewhere.

Error Type 4a applies when an officer of the court excluded probative evidence. This error does not relate to Brady violations, which are covered under Error Type 5. This error applies when a prosecutor's direct examination excludes discussion that should have elucidated the probative value of forensic evidence. The error also applies when a prosecutor's direct examination excludes discussion of exculpatory forensic results that were available at the time of the trial. The interpretation of this error should be based on an examination of trial transcripts or appeals that clearly demonstrate that the probative value of evidence would have been misinterpreted because its limitations or implications were mischaracterized during the trial.

Error Type 4b relates to inadequate defense counsel. The error may apply when defense counsel fails to ask for relevant information concerning forensic examinations prior to or during a trial. It may also apply when defense counsel fails to obtain independent examination or review of forensic evidence, if the uncertainties in the execution or interpretation of the evidence may be relevant. The error may apply if defense fails to perform appropriate cross-examination of forensic testimony, raise relevant objections to forensic testimony, or appeal decisions related to forensic testimony that could impact a case outcome.

Error Type 4c applies when forensic evidence is mischaracterized by legal counsel. The mischaracterization must relate to the probative value of the forensic evidence. The mischaracterization may relate to source attribution, scientific validity, statistical interpretation, or laboratory methods. The error may apply when a prosecutor directly employs a consultant to review or perform a forensic examination, if the consultant's work is in error or used improperly by the prosecutor during the trial.

Error Type 4d applies to judicial conduct related to forensic evidence. It is not an error of judicial conduct to accept forensic evidence in accordance with precedents related to the methods used in the case. Judicial error may occur if a novel method is employed without an appropriate review within Daubert or other standard that should apply in the jurisdiction. Judicial error may occur if faulty forensic testimony is accepted (under the standards in place at the time of trial) over the objection of legal counsel. Judicial error may occur if jury instructions mischaracterize the probative value of forensic evidence or the scientific or statistical basis for forensic testimony.

Error Type 5. Errors associated with potentially probative forensic evidence that was not collected, examined or reported during a police investigation or reported at trail.

x. Crime scene evidence not collected, stored, or sent to a laboratory for analysis
y. Forensic evidence not examined or a forensic result is excluded from an examiner's report.
z. Investigator suppresses or ignores forensic evidence or report.
aa. Forensic evidence is not reported at trial or is not shared with defense counsel.

Error Type 5 includes situations in which probative evidence is available for analysis and reporting but is not used in the trial. The error does not apply if the evidence would not have been probative using the methods employed at the time of the trial, such as small amounts of biological evidence that would not have been useful for serological analysis. The error may apply to any part of the investigative or legal process. The reason for the error should be documented, including organizational and practice considerations.

The reviewer should keep in mind that evidence collection and analysis decisions may have excluded probative evidence for legitimate reasons relating to resource constraints or imperfect knowledge of the crime at the time of the investigation. Error Type 5 should be used only when it is clear that probative evidence was excluded from consideration or review through negligence or malfeasance.

Error Type 5a applies when relevant crime scene evidence is not made available for forensic analysis. The error only applies when a trained investigator would have recognized the potentially probative value of the evidence. The error may apply if the evidence was not collected at the crime scene, if the evidence was available and intact for collection. The error may apply if potentially probative evidence was collected but not sent to a laboratory for analysis for any reason, including resource constraints. The error may apply if probative evidence was lost or destroyed. This error applies if probative evidence was not available during the postconviction phase, if the jurisdiction's standards for evidence retention were not followed.

Error Type 5b applies when potentially probative evidence was not examined, or results excluded form an examiner's report. This error does not apply if the forensic examiner failed to complete analysis of reference samples or other analysis needed to understand the probative value of evidence (which is documented under Error Type 1c). Error Type 5b applies when the evidence reaches a laboratory but is not processed. Error Type 5b also applies when a probative result is not sent to investigators or prosecutors.

Error Type 5c applies when a probative forensic result is suppressed or ignored during an investigation. This error applies when a police investigator fails to recognize exculpatory information or share the information with other individuals in the case, including the defendant. The investigator's failure should be documented within the official record of the case.

Error Type 5d applies when Brady violations or other circumstances occur which prevent appropriate defense access to forensic evidence. This error applies when a probative forensic result has been obtained in the laboratory but is not reported to all interested parties. This error also applies when a forensic examiner provides erroneous or misleading information concerning a forensic result. In some cases, forensic results may be negative, i.e., no source can be associated with the evidence. In these cases, Error Type 5d applies when the negative result would have had a substantive impact on the investigation of the crime or the weight of probative evidence at trial.

Testimony Standards

Forensic testimony and practice were assessed for conformance with standards at the time of trial and in 2020 at the time of the start of this study. The Department of Justice's Uniform Language for Testimony and Reporting (ULTR) was applied as the primary basis for 2020 testimony standards. (US Department of Justice, 2019) When appropriate, disciplines not specifically covered by the ULTR standards were assessed for conformance with ULTR principles, such as the limitation on citation of examiner experience as a basis for the statistical validity of a conclusion. In many cases, the primary issues involved practice standards and quality assurance, which may or may not have been reflected in the trial testimony. Thus, the selected standards were chosen to encompass practice issues when those issues could be discerned from the public case record.

In some instances, forensic evidence or communications were assessed to have an inadequate scientific basis under the NIST error framework. (National Institute of Standards and Technology, 2015) This assessment is inherently subjective, because the level of scientific foundation considered "adequate" may vary among observers. (Butler, et al., 2020) Thus, reliance on inadequate science was coded only when there was a basis in the consensus views of researchers and forensic standards bodies (for example, the use of canine detection to produce an individualization).

Table 1 provides a summary overview of the primary standards used for the most common forensic disciplines in the study. Discipline-specific details are provided in subsequent sections.

Table 4. Forensic discipline standards used in the assessment of forensic evidence in wrongful convictions.

Discipline	**Standards**
Serology	Culliford, B. J., *The Examination and Typing of Bloodstains in the Crime Laboratory.* Washington, DC: National Institute of Law Enforcement and Criminal Justice. (Culliford, 1971) Gaensslen, R. E., *Sourcebook in forensic serology, immunology, and biochemistry.* Washington, DC: National Criminal Justice Reference Service. (Gaensslen, 1983) Federal Bureau of Investigation, *Handbook of Forensic Science.* Washington, DC: National Institute of Justice, US Department of Justice. (Federal Bureau of Investigation, 1982)
Hair comparison	The Laboratory Division, Federal Bureau of Investigation, Proceedings of the International Symposium on Forensic Hair Comparisons. Washington, DC. (The Laboratory Division, Federal Bureau of Investigation, 1986) Houck, M. M., & Budowle, B. Correlation of microscopic and mitochondrial DNA hair comparisons. (Houck & Budowle, 2002) Cases after the publication of the Houck/Budowle paper

	were coded for error if no attempt was made to confirm an association using DNA testing.
Fire debris investigation	National Fire Protection Association, Guide for Fire and Explosion Investigations (NFPA 921-92). (National Fire Protection Association, 1992) National Fire Protection Association, Guide for Fire and Explosion Investigations (NFPA 921-2017). (National Fire Protection Association, 2017) When appropriate, other versions of this standard were applied. For example, the 2008 version of the standard clearly prohibited the use of negative corpus to conclude an incendiary origin. For cases before 2008, the use of negative corpus would not have been in error at the time of trial if it were applied with a full consideration of alternative origins and causes. Prior to 1992, the 1992 version of NFPA 921 was applied as the standard, because a trained investigator would have used the research later cited in NFPA 921 as the basis for conclusions.
DNA	Scientific Working Group on DNA Analysis Methods (SWGDAM), including Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. (Scientific Working Group on DNA Analysis Methods, 2019) ANSI/ASB Standard for Forensic DNA Interpretation and Comparison Protocols. (AAFS Standards Board, 2019) National Academy of Science 1996 report, The Evaluation of Forensic DNA Evidence. (National Research Council, 1996) This report contains information concerning the accepted calculation of random match probabilities. It also includes a discussion of VNTR DNA analysis.
Bitemark comparison	American Board of Forensic Odontology (ABFO) Bitemark Methodology Guidelines. (American Board of Forensic Odontology, 1995) ABFO Standards and Guidelines for Evaluating Bitemarks. (American Board of Forensic Odontology, 2018)
Seized drugs	SWGDRUG guidelines (http://swgdrug.org/archived.htm). Organization of Scientific Area Committees (OSAC) Seized Drug Subcommittee (https://www.nist.gov/topics/organization-

	scientific-area-committees-forensic-science/seized-drugs-subcommittee)
Toxicology	AAFS Standards Board: the ANSI/ASB Best Practice Recommendation 037, First Edition, 2019, Guidelines for Opinions and Testimony in Forensic Toxicology. (AAFS Standards Board, 2019) OSAC Toxicology Subcommittee (https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/toxicology-subcommittee)
Firearms and toolmarks	OSAC Firearms and Toolmarks Subcommittee (https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/firearms-toolmarks-subcommittee) Relevant documents include the draft standard, Range of Source Conclusions and Criteria in Toolmark Examinations. (Organization of Scientific Area Committees, 2021)
Friction ridge	OSAC Friction Ridge Subcommittee, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/friction-ridge-subcommittee. The OSAC web page contains an exhaustive list of historical standards that are relevant to specific cases.
Forensic medicine (pediatric abuse)	American Academy of Pediatrics (AAP) standards, including: Consensus statement on abusive head trauma in infants and young children. (Choudhary, et al., 2018) Guidelines for the Evaluation of Sexual Abuse of Children: Subject Review. (Committee on Child Abuse and Neglect, 2013) (Committee on Child Abuse and Neglect, 1999) (Committee on Child Abuse and Neglect, 1991)
Forensic pathology	Various standards and research were utilized, including guidelines from the College of American Pathologists and National Association of Medical Examiners. **Invalid source specified. Invalid source specified.** American Board of Medicolegal Death Investigators (www.abmdi.org).

	Death Investigation: A Guide for the Scene Investigator (National Medicolegal Review Panel, 1999) (National Medicolegal Review Panel, 2011)

The study recognizes that the issues in pediatric abuse cases continue to be debated in the legal and scientific communities. (Papetti, Kaneb, & Herf, 2019) As in other disciplines, the error coding reflects that case uncertainties may support multiple, valid interpretations. In those cases, errors may arise if a forensic expert fails to address the limitations associated with conclusions or if a valid interpretation consistent with the defense theory of the case is not presented during the case proceedings. This approach was also followed in the assessment of forensic pathology and fire investigation examinations.

In some cases, the primary issue involved the use of automated databases to conduct latent print searches that may have produced a cold hit to a source. In the determination of an error, the study relies on the date of deployment of a system that could have been used to conduct latent fingerprint searches in the jurisdiction. The study does not code as an error any failure to conduct a latent palmprint database search, which is not generally available at this time.

Serology

Note: In general, the following represents the understanding of the science of forensic serology as it was practiced before 2000 and a summary guide for the determination of forensic science errors related to serology. The definitive texts include Robert Gaensslen's serology sourcebook (Gaensslen, 1983) and Bryan Culliford's guide to laboratory practice. (Culliford, 1971)

Basis for interpretation of serological analysis

In serological interpretation of sexual assault samples, the contribution of a male assailant to the serological profile of a biological sample may be masked by the blood group substances from a victim's serological profile. (Culliford, 1971) In addition, serology permits the association of an individual with a class of individuals (or population) that is consistent with the serological profile of the biological sample. Serology does not permit individualization. In any serological or DNA analysis, conclusions should clearly delineate among activity, source, and sub-source hypotheses. (Champod, Biedermann, Vuille, Willis, & De Kinder, 2016) Reports and testimony should clarify the difference between the serological profile of a source, the serological profile of a biological sample, and the serological profile of a population of sources that may have contributed to a biological sample. The current study applies the following standards to the interpretation of serological analyses:

7. Examiner testimony should reflect that serology could not be used to identify an individual or conclude that they were the contributor to a particular piece of evidence.
8. Examiner testimony should reflect uncertainties with respect to masking, contamination, degradation, testing limitations, or other factors relevant to the case in question.
9. Microscopic confirmation of spermatozoa is a sufficient basis to assume that the serological profile of a questioned sample should reflect a male contributor. Regardless of the method used to confirm a male fraction, victim or other contributors to a biological sample should be addressed in the interpretation of the serological profile of a biological sample.

10. The examiner may account for masking when spermatozoa are not present in a biological sample as follows:
    a. Prior to 1986, a semi-quantitative acid phosphatase assay may be used if the limitations of acid phosphatase are clarified; *or*
    b. A semi-quantitative P30 assay may be used if the limitations of P30 are clarified; *or*
    c. The examiner may adhere to a written laboratory policy that would be compliant with the general standards reflected in scientific and practice standards at the time of trial, including those reflected in Gaennslen (Gaensslen, 1983) (Gaennslen, 2000), FBI documents (The Laboratory Division, Federal Bureau of Investigation, 1986) (Federal Bureau of Investigation, 1982), or similar references.
11. Examiner testimony should reflect general procedures and interpretation issues but is not required to include a complete account of testing protocols. Subsidiary documents are used when available to determine if the examiner's lab work, reports, communications, and procedures conformed to generally accepted procedures, legal requirements, and testimony standards.

The examiner should not confuse crime, activity, source, and sub-source hypotheses. (Champod, Biedermann, Vuille, Willis, & De Kinder, 2016) In particular, the examiner should not state or imply that the statistical characterization of the serological profile of a possible source is the same as the statistical characterization of the serological profile of questioned evidence.

Forensic serology uses non-DNA biomarkers to identify the source of biological evidence. Most biomarkers are proteins expressed in different types in species, individuals, or bodily fluids. Most commonly, evidence technicians collect samples that appear to be blood to locate a crime scene, associate a weapon with a crime, or eliminate suspects. For the purposes of wrongful conviction analysis, the FBI notes the following two limitations of serology (Federal Bureau of Investigation, 1982):

1. It is not possible to identify human blood as coming from a particular person.
2. The race of the person from whom blood came cannot be conclusively ascertained; nor can the age of a dried stain be determined.

There are more than 15 serological biomarkers that have been used in forensic analysis. Most methods consume sample, so the forensic scientist must select those examinations that are both practical and probative. This document covers only the most common serological analysis types. **A finding of forensic science error was not be based primarily on whether a laboratory exhausted all possible serological methods or failed to use a particular method that—in hindsight—may have been useful in the case**.

Detection of blood

Serological analysis may determine if a sample is blood using presumptive tests. The tests are subject to possible false positive and false negative results. In some cases, methods to detect blood may interfere with methods to establish ABO type or other serological or DNA testing.

Primary methods include:

1. Kastle-Meyer: Hemoglobin in blood reacts with hydrogen peroxide in a solution of colorless phenolphtha*lin* to produce bright pink phenolphtha*lein*. Hemoglobin reduces

hydrogen peroxide to water, losing electrons in the peroxidase process. Phenolphthalin donates electrons to hemoglobin to continue the reaction and turns into phenolphthalein in the process. The hemoglobin is acting as a catalyst of the hydrogen peroxide reduction, so the test can be sensitive to small blood concentrations.

a. Pink indications after 30 seconds should be ignored, because chemical oxidants (bleach, copper salts, etc.) will turn solutions pink even without the addition of hydrogen peroxide. False positives may arise from metallic salts, bacteria, saliva, malt extract, and vegetable extracts. Because of uncertainties related to the test procedure and false positives, the Kastle-Meyer test is presumptive only and should not be used to state that a sample is blood with absolute certainty. Appropriate conclusions would include:
    i. No blood detected.
    ii. Indications of blood detected or presumptive presence of blood detected.
    iii. Staining consistent with blood was detected. (Washington State Patrol Crime Laboratory Division, 2019)
b. Good laboratory practice includes tests with positive blood control to verify working reagents and the use of a negative control.
c. Test was used more routinely after the 1967 Supreme Court case, Miller v. Pate, 386 U.S. 1 (1967). In that case, a visual determination of blood was not confirmed by chemical testing and the defendant was acquitted. The prosecution was aware at the time of trial that the stain was paint, not blood, although it was reported that the Illinois state crime laboratory had identified the stain as serological blood type A.
d. There are other catalytic reactions used to determine the presence of blood that work in a similar manner to the Kastle-Meyer test, including tetramethyl benzidine (producing a blue color) and leucomalachite green (producing a blue-green color).

2. Luminol: Luminol is a chemiluminescent compound. The luminescence is greatly enhanced via a chemical reaction with hemoglobin, thus permitting the detection of small amounts of blood. Other body fluids do not enhance the luminescence, with the exception of fecal matter. Older blood stains are more easily observed, because the breakdown products of hemoglobin provide a stronger reaction with luminol. Luminescence is visible in the dark and for a limited period of time (minutes). (Grispino, 1990)
    a. Luminol should not be used to estimate the age of a stain based on the level of observed luminescence.
    b. Common materials that may produce false positives include bleach, coffee stains, metals, and rust (e.g., carpet tacks). Because of interferents, luminol should be a presumptive test only.
    c. Procedures should include a photograph of the luminescent material to document the presumed presence of blood.
    d. Luminol interferes with other serological tests, with the exception of ABO testing, because the chemical breaks down enzymes in samples. Luminol interferes with DNA testing. In general, luminol use should be confined to stains that will not be used for subsequent analyses.
3. Precipitin: This test, sometimes in the form of Ouchterlony double immunodiffusion, is a confirmatory test for blood that may also be used to determine the species of origin of

blood. Antiserum is produced from animals who are injected with human blood. The antibodies from the antiserum will bind with human blood to form a precipitate in solution. In the Ouchterlony method, the antiserum and suspected blood are diffused in opposite directions across an electrophoretic gel, producing a visible precipitin arc or line in the middle of the gel.

a. In good laboratory practice, a confirmatory preciptin test should be conducted to verify the results of presumptive tests, such as luminol or catalytic tests.
b. Variations on the preciptin method may be used for many other serological analyses, including ABO testing.

ABO System

The ABO blood group system was the primary serological method used in forensic science until the emergence of DNA technology. Humans express many thousands of ABO antigens on the surface of their blood cells. The body makes antibodies to the antigens that may be encountered from incompatible blood cell types or other biological sera that contain these antigens. Most humans secrete the soluble antigens into other bodily fluids, including semen, so antigens may be found in the absence of blood. These individuals are "secretors." Non-secretors do not secrete antigens into other bodily fluids. There are four basic ABO types, including A, B, O, and AB. The blood type is determined from an encoding region on chromosome 7, and it has now been definitively established that hundreds of variant ABO alleles may be expressed. The O antigen is referred to as the H antigen, which also serves as a precursor to A and B antigens. The blood type, AB, included individuals who expressed both the A and B antigen. In general, traditional serology identified two types of A antigen, dubbed A_1 and A_2, which could also be observed in A_1B and A_2B blood types. Many other ABO phenotypes have now been identified, some of which may have been used in cases pre-2000 in particular cases. A child receives one of three alleles that encode the ABO antigen from each parent. ABO-based serology was routinely used in paternity cases for many years, but its application to criminal cases was very different.

The population may be divided as follows:

Blood type	Red blood cell antigens	Antigens in serum for secretors	Antigens in serum for non-secretors	Antibodies in serum for secretors and non-secretors
A	A	A	None	Anti-B
B	B	B	None	Anti-A
AB	A and B	A and B	None	None
O	None	None	None	Anti-A and Anti-B

The frequency of occurrence of ABO groups in U.S. populations has been described by Gaensslen and others. Gaensslen provides population data on subgroups that may not be representative of the entire population, such as tumor patients. The range of representative population estimates is as follows:

Population	O	A	B	AB	Secretor?
Gaennslen estimates					
Caucasian	38 to 47%	37 to 43%	9 to 16%	2 to 6%	72 to 77%
African American	44 to 53%	22 to 27%	17 to 23%	2 to 5%	
Hispanic	55 to 62%	27 to 32%	9 to 12%	1 to 4%	
Asian American	32 to 45%	27 to 37%	21 to 25%	5 to 9%	
Consensus (Reid & Lomas-Francis, 2004)					
Caucasian	44%	43%	9%	4%	
African American	49%	27%	20%	4%	
Asian American	43%	27%	25%	5%	

In some cases, a forensic examiner may choose to use subpopulation data from Gaensslen or one of the scientific publications cited by Gaensslen to support an ABO analysis, although care should be taken to ensure that the research data is reliable and sufficient to draw conclusions.

Common Methods

Landsteiner Agglutination: The Landsteiner test relied on the agglutination of red blood cells when mixed with antisera that contain antibodies to the blood type. The sensitivity of Landsteiner tests is reduced when stains are older and cells have denatured.

Absorption-elution is a variation on the Landsteiner test. Antiserum with all three antibody types (A, B, and AB) is added to a blood sample. The matching antibody will bind to the sample, and excess antibody solution is then thoroughly rinsed. The sample temperature is raised (typically to 56 C) to release the bound antibodies. Each blood type can then be added to the solution and agglutination observed microscopically when the type matches the released antibody type. Although absorption-elution is more sensitive than the Landsteiner method and more useful for aged blood stains, the process can give false results when the excess antibody solution is not properly rinsed or cross-reactivity produces an incorrect result.

Histochemical: Histochemical tests cover a wide range of methods that use a chemically-labelled antibody to visualize antigen binding. Tests may involve fluorescent labels, enzymes, or catalysts (such as ferritin in methods similar to blood detection). In the 1990's ELISA (enzyme-linked immunosorbent assay) tests were developed that could perform many tests in parallel with excellent sensitivity. Histochemical tests and ELISA are subject to cross-reactivity and interferences that should be accounted for in the analysis of results.

Common ABO serology errors include:

1. Misreporting/misinterpretation of population statistics.
2. Use of ABO or other serological evidence to make an individualization.
3. Failure to obtain reference samples from crime scene evidence; for example, samples from clothing should include crime scene stains and areas that do not appear to have biological material. This procedure elucidates whether the presence of ABO antigens is related to background contamination.
4. Failure to consider masking effects. If the victim secretes antigens that match antigens associated with a suspect, then ABO analysis does not provide probative information. Note that some individuals, such as type AB, will secrete both A and B antigens, if they are secretors.
5. Sensitivity: failure to address or report data limitations relating to the sensitivity of tests for the presence of antigens.
6. Failure to determine or account for secretor status. For example, if a suspect is a non-secretor, that individual's antigens will not be found in any biological evidence other than blood.

Other serological markers

Routine forensic investigations may include determination of additional protein markers, such as the Rhesus and Lewis factors. Although many such systems exist, forensic laboratories may or may not attempt to conduct possible tests because of resource constraints or sample availability. Laboratory methods are similar to those used for ABO typing, with similar considerations regarding test limitations and possible error sources. As in the case of ABO typing, there are many more variants of other genotypes than can be routinely determined in a forensic laboratory, so the population statistics in a particular case will depend on the sophistication of the test method. Details and population statistics are presented in Gaensslen and Culliford, and reliable forensic interpretations should reflect the considerations outlined there.

Other serological methods are detailed in Gaennslen and Culliford, including G6PD, PGM, and polymorphic protein systems. In many cases, reliable population estimates are provided concerning common phenotypes and racial categories prevalent in the United States.

Hair Microscopy

For the purposes of the current study, the FBI 1985 symposium on hair microscopy provides the most definitive guide for hair microscopy examination prior to the advent of mitochondrial DNA testing in 2000. (The Laboratory Division, Federal Bureau of Investigation, 1985) After the publication of the 2002 Houck and Budowle paper on the use of mitochondrial DNA testing in conjunction with microscopic hair comparison, it is expected that all hair comparisons be verified with mitochondrial DNA tests in felony crime cases that go to trial. (Houck & Budowle, 2002)

Koch has published a useful guide for the conduct of hair examination that may be used to review an analyst's work (including to the present day). (Koch, 2017)

The 1985 symposium included a report from the Committee on Forensic Hair Comparison Subcommittee on Report Writing, Conclusions and Court Testimony. The committee recognized the following issues:

- Some hair types provide difficult comparisons because of the lack of identifying characteristics or the complexity of identifying characteristics, including blonde Caucasian hair, grey hair, and hair from African American, native American, or Asian American populations. These limitations do not imply that such comparisons are impossible, but the committee recommended that uncertainties be reflected in any conclusions reported by an examiner.
- Probability estimates are not possible in most cases. The committee recognized that Gaudette's work can be a basis—for Caucasian hair—to state that "an estimate of the average odds against that one questioned hair having originated from another specific Caucasian individual would be about 4500 to 1 (800 to 1 for pubic hairs)." The committee recommended that these statistics be cited only upon direct examination in court and with the note that further research would be required.

The committee recommended the following conclusion types:

1. The questioned hair is consistent with having come from John Doe.

This conclusion must be based upon a strong association between the questioned hair and the known sample. There are several factors, such as the questioned hairs having intrasample variation which is found to be microscopically similar to the comparison sample, or the presence of unusual hair characteristics or hair treatment such as dying, bleaching, etc., which strengthen the association. Other tests such as sex determination or enzyme typing may provide additional support for this conclusion.

2. The questioned hair could have come from John Doe.

When only a limited association can be made between the questioned hair and the comparison sample, this conclusion would apply. Some of the factors which influence arriving at this conclusion are the presence of hair fragments, or the lack of any distinctive features in the questioned hair, for example, white hair.

3. John Doe qualifies as being the donor of the questioned hair.

This statement indicates to the reader that John Doe cannot be eliminated as a possible source of the questioned hair. It does not take into account whether there is a strong or limited association.

4. The questioned hairs could not have originated from John Doe.

This conclusion is based upon finding significant macroscopic and/or microscopic dissimilarities between the questioned hair and the comparison sample. The examiner must ensure that several factors are satisfied before this statement can be made, for example, the possibility that the hairs are atypical is remote, the known sample is adequate and representative with little intrasample variation.

5. The questioned hair is not consistent with having come from John Doe.

This conclusion applies when there are dissimilarities between the questioned hair and the comparison sample; however, there are factors present which do not allow the examiner to conclude categorically that John Doe could not be the donor of the questioned hair, for example, inadequate comparison samples, incomplete questioned hairs, a lengthy time lapse between the offense and the collection of the comparison sample.

6. No conclusion as to origin could be reached.

As stated earlier, there are cases in which no conclusion can be reached. One might find both similarities and differences between the questioned hair and the comparison sample or the questioned hair may be very minute. In these situations, the results are uninterpretable and no conclusion can be made as to whether the questioned hair could have originated from a specified source.

The committee recommended qualifying statements to clarify conclusions, such as unusual characteristics or featureless hairs. The committee recognized the limitations of research concerning the statistical characteristics of hair morphology and recommended against testimony that cited research as a basis for any statistical estimate of a source association. They recognized that the examiner could cite research as a basis for the methods in the discipline, especially on cross-examination. When examiners exceeded these recommendations, their testimony was coded as an error in exceeding the limits of validated science.

It should be noted that some aspects of the 1985 guidelines would be considered errors under current practice, especially under the DOJ ULTR. In these cases, the testimony should be considered correct with respect to standards in practice at the time of the trial.

2020 Standard

The DOJ ULTR specifies the following standards for hair microscopy testimony (US Department of Justice, 2020):

The examiner may offer any of the following conclusions:

1. Inclusion (i.e., included)
2. Exclusion (i.e., excluded)
3. Inconclusive

Inclusion

'Inclusion' is an examiner's conclusion that the source of the known hair sample can be included as a possible source of the questioned hair. The questioned hair could also have originated from additional individuals whose known hair sample encompasses the range of macroscopic and microscopic characteristics observed in the questioned hair.

The basis for an 'inclusion' conclusion is an examiner's decision that all assessed macroscopic and microscopic characteristics in a questioned hair are exhibited in the known hair sample with no meaningful differences.

Exclusion

> 'Exclusion' is an examiner's conclusion that the source of the known hair sample cannot be included as a possible source of the questioned hair based on the known hair sample provided.
>
> The basis for an 'exclusion' conclusion is an examiner's decision that all assessed macroscopic and microscopic characteristics in a questioned hair are not exhibited in the known hair sample.

Inconclusive

> 'Inconclusive' is an examiner's conclusion that no determination can be reached as to whether the source of the known hair sample can be included as a possible source of the questioned hair based on the known hair sample provided.
>
> The basis for an 'inconclusive' conclusion is an examiner's decision that the questioned hair contains both similarities to and differences with the known hair sample, or the questioned hair is of limited value for meaningful microscopical comparison such that the examiner is unable to determine whether or not a questioned hair could have originated from the source of the known hair sample.

An examiner shall not offer an 'inclusion' conclusion unless he or she explains that the questioned hair could also have originated from additional individuals whose known hair sample encompasses the range of macroscopic and microscopic characteristics observed in the questioned hair. Forensic hair examination is not a basis for personal identification.

- An examiner shall not assert that a questioned hair came from a particular individual to the exclusion of all other individuals.
- When comparison of a questioned animal hair to a known animal hair sample results in an 'inclusion,' an examiner shall explain that animal hairs do not typically possess sufficient differences in macroscopic and microscopic characteristics to distinguish between animals of similar breed and color.
- An examiner shall not assert that forensic hair examinations are infallible or have a zero error rate.
- An examiner shall not provide a conclusion that includes a statistic or numerical degree of probability except when based on relevant and appropriate data.
- An examiner shall not cite the number of forensic hair examinations performed in his or her career as a direct measure for the accuracy of a proffered conclusion. An examiner may cite the number of forensic hair examinations performed in his or her career for the purpose of establishing, defending, or describing his or her qualifications or experience.

• An examiner shall not use the expressions 'reasonable degree of scientific certainty,' 'reasonable scientific certainty,' or similar assertions of reasonable certainty in either reports or testimony unless required to do so by a judge or applicable law.

Fire Debris

The standard for fire debris testimony is the version of NFPA 921 in use at the time of the trial. The most up-to-date version of the standard is NFPA 921-2017. (National Fire Protection Association, 2017) For cases prior to 1992, NFPA 921-1992 should be used as the standard for testimony. (National Fire Protection Association, 1992) NFPA 921-1992 provided a consensus view of the results of scientific research in 1992, though most of the research referenced in that standard predates the publication significantly. A well-informed, trained investigator would have used the research cited in NFPA 921-1992 prior to its publication.

Key considerations include:

1. Alligator char (large shiny blisters observed in char) is not an exclusive indicator of the presence of liquid accelerants and may originate in many different types of fires.
2. Spalling is the breakdown in the surface tensile strength of concrete. Spalling is not necessarily due to the presence of liquid accelerant. Spalling may be caused by exposure to any high rate of heating by flame, high levels of radiation from any fuel, or rapid cooling of a mass of concrete, brick or masonry. A finding concerning the presence of liquid accelerant associated with spalling should include an evaluation of the condition of the surface before the fire.
3. “V” shaped patterns may be found on vertical surfaces. The patterns are caused by radiant heat energy from above the vertical surface and by the upward and outward movement of flames and hot gases that encounter a horizontal obstruction such as a ceiling. The angle of lines of a “V” pattern are associated with the size of a fire, burning rate, ventilation, and wall combustibility. The “V” angle is influenced by the relative rate of heat release but this variable should not be considered in isolation from other factors. The “V” angle provides substantive information concerning the direction of spread of a fire.
4. Inverted cone patterns, or inverted “V’s”, are triangular patterns, wider at their base than at the top. Truncated cone patterns may be caused by failures of wall structures (such as ceilings or plaster lath) or the influence of a horizontal obstruction on a fire movement pattern. Inverted cones and truncated cone patterns are different phenomena and should be identified appropriately during an investigation. Inverted cone patterns are manifestations of short-lived fires that do not fully evolve into floor-to-ceiling flame plumes. The pattern may be due to the lack of combustible material at the location, not the rate of heat release. Inverted cone patterns are not necessarily proof of flammable liquid fires.
5. The opinion of a fire investigator testifying as an expert witness should be based on the soundness of the investigator’s rationale for each conclusion. The evidence that forms the basis of any opinion or conclusion must be relevant and reliable.

DNA

Forensic DNA analysis relies on the examination of fragments of nuclear or mitochondrial DNA to determine the source of biological evidence. There are several, definitive sources of information concerning DNA standards and practices, including:

- Scientific Working Group on DNA Analysis Methods (SWGDAM). SWGDAM establishes the scientific basis for standards that are enforced through the FBI Quality Assurance

Standards and Combined DNA Index System. Numerous SWGDAM publications are available at www.swgdam.org/publications.

- Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, a SWGDAM publication. (Scientific Working Group on DNA Analysis Methods, 2017); this publication contains sections on the accepted approach to the interpretation of DNA mixtures.
- ANSI/ASB Standard for Forensic DNA Interpretation and Comparison Protocols. (AAFS Standards Board, 2019). This document contains a list of requirements with regard to considerations in the interpretation of DNA data.
- Interpretation Guidelines for Mitochondrial DNA Typing by Forensic DNA Testing Laboratories, a SWGDAM publication. (Scientific Working Group on DNA Analysis Methods, 2019); this publication includes the appropriate testimony framework for mitochondrial DNA analysis.
- Uniform Language for Testimony and Reports (ULTR) (US Department of Justice, 2020). The ULTR includes testimony standards for autosomal (nuclear) DNA, mitochondrial DNA, and Y-STR DNA analysis.
- National Academy of Science 1996 report, The Evaluation of Forensic DNA Evidence. (National Research Council, 1996) This report contains information concerning the accepted calculation of random match probabilities that continue to be used in autosomal DNA analysis. It also includes a discussion of VNTR DNA analysis, which was the accepted method of DNA analysis at the time of the report.
- National Institute of Standards and Technology STRBase, https://strbase.nist.gov/.

VNTR/RFLP DNA analysis

The first application of DNA analysis to forensic science used Restriction Fragment Length Polymorphisms, RFLP's. In RFLP, restriction enzymes break DNA into fragments, the size of which can be measured using gel electrophoresis. Variable Number Tandem Repeats (VNTR's) are common fragments identified in RFLP analysis. RFLP requires a large amount of sample, repeated gel electrophoresis analyses for each fragment type, southern blotting of the gel onto a polymer membrane, radioactive tagging of fragment products, careful reading of fragment patterns, and careful analysis of population statistics. Radioactive tagging and development requires weeks for each enzyme/fragment combination, and multiple combinations were required to conduct analyses of multiple DNA markers. This limitation extends the time for analysis to many months and in practice limited the number of VNTR's examined to four. The interpretation of the fragment patterns could be subject to subjective decisions concerning the presence or position of fragment patterns. RFLP analysis is limited in reliability when dealing with sample mixtures or contamination.

DQ Alpha

The DQ Alpha gene exhibits several polymorphisms that vary across the human population. The analysis of DQ Alpha provided the one of the first uses of PCR amplification in forensic science. Although the gene provides less discriminating information than VNTR's, it is much more sensitive because of the use of PCR. Amplified DQ Alpha fragments were identified using a reverse blot method similar to that used in RFLP, with similar issues with respect to mixtures and interpretation of fragment patterns. The probability of two randomly chosen persons with identical DQ Alpha genotype is about 0.053. Full probability estimates are provided in Budowle et al. (Budowle, et al., 1995)

Mitochondrial DNA (mt DNA)

Cells contain hundreds of copies of mt DNA. As a result, mt DNA is present in many cases in which nuclear DNA is not detectable or identifiable. Mitchondrial DNA is inherited along the maternal lineage. There are limited mt DNA haplotypes, and in some cases millions of individuals may share a mt DNA profile that is indistinguishable using typical methods. Therefore, mt DNA is useful for excluding suspects, not identification.

SWGDAM recommends that comparisons be reported as follows:

1. Exclusion: If samples differ at two or more nucleotide positions (excluding length heteroplasmy), they can be excluded as coming from the same source or maternal lineage.
2. Inconclusive: The comparison should be reported as inconclusive if samples differ at a single position only (whether or not they share a common length variant between positions 302-310).
3. Cannot Exclude: If samples have the same sequence, or are concordant (sharing a common DNA base at every nucleotide position), they cannot be excluded as coming from the same source or maternal lineage.

The DOJ ULTR further stipulates:

- An examiner shall not offer a 'cannot be excluded' conclusion unless he or she also explains that 1) all relatives from the same maternal lineage are expected to have the same or a concordant mtDNA haplotype and would also be included as potential contributors; and 2) unrelated individuals may also exhibit the same or a concordant mtDNA haplotype.
- An examiner shall not assert that a mtDNA haplotype is unique to a particular individual or is the basis for personal identification.
- An examiner shall provide a quantitative statement describing the weight of the evidence for all inclusions regardless of the magnitude of the resulting quantitative value.
- An examiner shall not assert that a mtDNA haplotype can be used to predict the specific population, racial, or ethnic group to which a person belongs.

General issues

For considerations regarding nuclear (autosomal) DNA and Y-STR's, the references at the beginning of this section should be consulted. All DNA analyses and testimony should reflect consideration of general issues:

- Population studies. The random match probability must be related to reliable databases that reflect the population of the suspect. For example, if the defendant is of east Asian ancestry, then the population of east Asians—or the closest analogue—should be the basis for a population estimate.
- Contamination may arise at any time during the collection and processing of evidence. Laboratories must have procedures that include sample controls and policies supported by validation studies for interpreting data potentially affected by contamination.
- Evaluation controls may include a positive amplification control, a positive sequencing control, a negative amplification control, and a reagent blank control.

- Mixture interpretation has been implicated in numerous wrongful conviction claims. A complete summary of the literature is available from NIST at https://strbase.nist.gov/mixture.htm.
- Other issues include: stutter, peak height ratios, thresholds, stochastic effects, allele dropout, low-template DNA, estimation of the number of contributors to a biological stain, cell separation, and new technologies. The examination of these issues in a DNA case should be documented in the Case Review Form.

Modern DNA mixture interpretation uses validated, probabilistic genomics software to calculate Combined Probability of Inclusion statistics. These systems address dropout and other issues effectively and provide an objective and reliable mixture interpretation framework when appropriately applied.

Bitemarks

The American Board of Forensic Odontology (ABFO) promulgates standards concerning the investigation, analysis, and reporting of bitemarks. The ABFO issued its first comprehensive guidelines in 1995. (American Board of Forensic Odontology, 1995) The guidelines have been described in various literature of the field. (Manas, Nilesh, Manika, & Betina, Bite Marks -- Revisited, 2016) (Bowers, 2002) (Wright, 2011) (Saks & et al, Forensic bitemark identification: weak foundations, exaggerated claims, 2016) In general, the 1995 guidelines permitted a bitemark examiner to testify concerning the following possible conclusions concerning whether an injury is a bitemark:

- Bitemark: Teeth created the pattern; other possibilities were considered and excluded.
- Suggestive: The pattern was suggestive of a bitemark, but there is insufficient evidence to reach a definitive conclusion at this time.
- Not a bitemark: Teeth did not create the pattern.

The guidelines permitted the following terms to relate a bitemark to a suspected biter:

- Biter
- Probable biter
- Cannot exclude
- Exclusion
- Inconclusive

The guidelines did not permit statements concerning the statistical probability of these conclusions. Statements that a particular biter could be matched to a bitemark to the exclusion of all other individuals should be considered an error.

ABFO updated the guidelines in 2018. (American Board of Forensic Odontology, 2018) The guidelines state, “An ABFO Diplomate shall not express conclusions unconditionally linking a bitemark to a dentition.” The guidelines also recommend that an independent, blind verification be obtained from at minimum one ABFO Diplomate. The guidelines permit the following statements concerning the origin of a pattern or patterned injury:

- Human Bitemark
- Not a Human Bitemark
- Inconclusive

The guidelines permit the following terms relating or linking a dentition to a human bitemark:

- Excluded as Having Made the Bitemark
- Not Excluded as Having Made the Bitemark
- Inconclusive

In some jurisdictions, including the state of Texas, no bitemark testimony is permitted at the present time.

Toxicology

The primary references for testimony related to toxicological analysis have been adopted by the AAFS Standards Board: the ANSI/ASB Best Practice Recommendation 037, First Edition, 2019, Guidelines for Opinions and Testimony in Forensic Toxicology (AAFS Standards Board, 2019)

Toxicological interpretation is a complex field that covers a wide variety of drugs, poisons, medical conditions, and biological matrices (e.g., blood, urine, oral fluid). In general, this study will use guidance established by the OSAC Toxicology Subcommittee, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/toxicology-subcommittee, to establish the consensus of approaches to toxicological analysis and reporting. ISO has established laboratory standards that detail specific issues (such as statistical error analysis) for the toxicology or drug laboratory, including ISO 17025. Additional references include:

- Principles of Forensic Toxicology 5th ed. 2020 Edition (Levine & Kerrigan, 2020)
- Flanagan, R. J., Cuypers, E., Maurer, H. H., & Whelpton, R. (2020). Fundamentals of analytical toxicology. John Wiley & Sons. (Flanagan, Cuypers, Maurer, & Whelpton, 2020)

Typical errors in forensic toxicology include:

- Incorrect quantitative interpretation or statistical error analysis
- Changes in samples due to aging or matrix effects
- Changes in samples due to stabilization or other reagents
- Isomers or other structural or chemical issues that contribute to misinterpretation of results
- Failure to maintain adherence to relevant laboratory standards
- Misinterpretations related to medical history (such as the use of prescribed medications)
- Misinterpretations related to postmortem effects

Scientific research has elucidated many issues related to forensic toxicology. The comprehensive examination of changes in testimony or interpretation is beyond the scope of this report. Specific issues in wrongful convictions will be examined in light of the accepted standards of 2020. If it is determined that the understanding of an error in toxicology may have changed since the time of the trial, that should be examined and documented on a case-by-case basis.

Seized Drugs

The primary source for the interpretation of seized drugs is the OSAC Seized Drugs Subcommittee, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/seized-drugs-subcommittee. The subcommittee references standards promulgated through SWGDRUG, the predecessor to the OSAC committee. The SWGDRUG website includes archival versions of some previous standards, http://swgdrug.org/archived.htm.

The DOJ has issued Uniform Language for Testimony and Reports for General Forensic Chemistry and Seized Drug Examinations (US Department of Justice, 2020), which provides six possible conclusions:

1. Identification (i.e., identified)
2. Consistent with
3. Not identified
4. Cannot be identified
5. Excluded
6. Inconclusive

Many wrongful convictions have been associated with seized drug interpretation, according to the National Registry of Exonerations. Some cases relate to "dry labbing" and other fraudulent practices in specific laboratories. Some cases relate to the use of presumptive tests to obtain a plea agreement or conviction in the absence of a confirmatory test. (Gabrielson & Sanders, 2016) According to OSAC, a presumptive test may be defined:

> "A screening test that indicates the presence of a material of interest although the test result does not constitute the identification of that material. A negative presumptive test indicates that the material of interest was not detected; it is not confirmation of its absence."

A confirmatory test is:

> "A test that is specific for a biological material or substance of interest and that is used for the conclusive identification…" or "A test used to confirm the identity of a substance. Generally used after a screening test."

The OSAC Seized Drugs Subcommittee addresses this distinction through the use of three categories of analyses:

Category A	Category B	Category C
Infrared Spectroscopy	Capillary Electrophoresis	Color Tests
Mass Spectrometry	Gas Chromatography	Fluorescence Spectroscopy
Nuclear Magnetic Resonance	Ion Mobility Spectrometry	Immunoassay
Raman Spectroscopy	Liquid Chromatography	Melting Point
X-Ray Diffractometry	Microcrystalline Test	Ultraviolet Spectroscopy
	Pharmaceutical Identifiers	
	Thin Layer Chromatography	

	Supercritical Fluid Chromatography Micro/macroscopic examination of cannabis	

Category C includes presumptive field tests. All tests are subject to misinterpretation, interferences, and analytical errors. For this reason, OSAC recommends that at least two tests are performed, one of which is a Category A test. When a Category A test is not performed, at least three tests are needed, two of which must be Category B. In practice, many cases may rely on Category C tests that are not confirmed in accordance with OSAC guidelines prior to adjudication. In addition, there is wide (and largely untested) variation in reliability among Category C presumptive tests. Issues may include drug mixtures, adulterants, contamination, and environmental factors. Presumptive testing errors may include:

- Use of a presumptive test as a confirmatory test
- Failure to account for known issues relating to the use of a presumptive test (such as a color indication related to cobalt thiocyanate's reaction with household cleaners to produce a positive result for cocaine)
- Failure to reexamine a plea-bargained case when a confirmatory test conflicts with a presumptive test result

Scientific research has elucidated many issues related to the analysis of seized drugs. Specific issues in wrongful convictions are assessed with regard to the accepted standards at time of adjudication and in 2020.

Firearms and Toolmarks

Current practice for firearm and toolmark analysis is established by the OSAC Firearms and Toolmarks Subcommittee, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/firearms-toolmarks-subcommittee. Testimony was assessed with regard to the draft standard, Range of Source Conclusions and Criteria in Toolmark Examinations (Firearms and Toolmarks Subcommittee of the Organization of Scientific Area Committees) and two testimony standards of the DOJ Uniform Language of Testimony and Reports (US Department of Justice, 2020), one dealing with Pattern Match Examination and another with Fracture Match Examinations.

For pattern match examinations relating to the bulk of firearms identifications, the ULTR provides for three possible conclusions:

1. Source identification (i.e., identified)
2. Source exclusion (i.e., excluded)
3. Inconclusive

The OSAC committee recognizes the work of its predecessor, SWGGUN, for more detailed guidelines, including analyses related to gunshot residue and projectile path reconstruction. The Association of Firearm and Toolmark Examiners provides a SWGGUN Admissibility Resource Kit, https://afte.org/resources/swggun-ark, that serves as a guide to current and historical issues in testimony related to firearms and toolmarks.

Friction Ridge

Current practice for friction ridge examination is established by the OSAC Friction Ridge Subcommittee, https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/friction-ridge-subcommittee. The committee's draft standards and the DOJ ULTR provide multiple documents to guide the assessment of testimony:

- Guideline for the Articulation of the Decision-Making Process Leading to an Expert Opinion of Source Identification in Friction Ridge Examinations (Friction Ridge Subcommittee of the Organization of Scientific Area Committees)
- Standard for Friction Ridge Examination Conclusions (Friction Ridge Subcommittee of the Organization of Scientific Area Committees)
- Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline (US Department of Justice, 2020)

The OSAC committee also recognizes standards associated with its predecessor, SWGFAST. The ULTR provides for three possible conclusions in friction ridge skin impression examinations:

1. Source identification (i.e., came from the same source)
2. Source exclusion (i.e., came from different sources)
3. Inconclusive

The ULTR further clarifies:

> An examiner shall not assert that two friction ridge skin impressions originated from the same source to the exclusion of all other sources or use the terms 'individualize' or 'individualization.' This may wrongly imply that a 'source identification' conclusion is based upon a statistically-derived or verified measurement or actual comparison to all other friction ridge skin impression features in the world's population, rather than an examiner's expert opinion.
>
> An examiner shall not assert that forensic latent print examination is infallible or has a zero error rate.
>
> An examiner shall not provide a conclusion that includes a statistic or numerical degree of probability except when based on relevant and appropriate data.

The accepted standard for friction ridge comparisons in the pre-2009 period has been discussed in various references, including a DOJ/FBI summary. (Peterson, et al., 2009) SWGFAST established standards similar to the ULTR, including issues related to sufficiency, error rates, and the ACE-V process. In many pre-2000 cases, examiners testified that the error rate for friction ridge identification was zero. Peterson states that "the discipline has no scientifically supported error rate." Therefore, statements concerning an error rate should have indicated the lack of scientific research.

Other Disciplines

It is expected that many wrongful convictions were associated with forensic analyses not covered in this coding guide. When such cases arose, the forensic evidence was judged on the basis of documents related to generally accepted practices and references, if they are available, and documented in the Case Review Form. When available, judicial opinions concerning the acceptance of evidence were used as a guide to the assessment of testimony.

Variable Definitions

TABLE: FORENSIC ERROR ARCHIVE

Archive Record Number is a randomly generated case identifier unique to each defendant in the dataset. The range is 1 to 1000000.

Guilty plea is recorded as “1” when the defendant pled guilty in the conviction that led to the wrongful conviction. It is recorded as “0” in all other instances.

DNA exoneration is recorded as “1” when the defendant was exonerated using postconviction DNA analysis, “2” when postconviction DNA analysis was performed but was not dispositive, and “0” in all other instances.

Drug case is a variable that is coded “Harris County drug case” in cases involving only seized drug analysis errors related to field test kits in Harris County, Texas; “Drug case only (not Harris County)” in cases involving only seized drug analysis errors in other jurisdictions, “Marijuana manufacture” in cases involving allegations of illegal production of marijuana, “Toxicology case” in cases involving allegations relating solely to toxicological analyses, and “No” in all other instances.

Crimes alleged is a variable that lists all crimes alleged against the defendant during the case. Possible values include: Accessory to Murder, Arson, Attempted Murder, Burglary/Unlawful Entry, Child Abuse, Child Sex Abuse, Conspiracy, Driving While Intoxicated (DWI), Drug Possession or Sale, Fraud, Gun Possession or Sale, Illegal Use of a Weapon, Kidnapping, Manslaughter, Military Justice Offense, Misdemeanor, Murder, Obscenity, Other Violent Felony, Perjury, Robbery, Sex Offender Registration, Sexual Assault, Sodomy, Supporting Terrorism, Tax Evastion/Fraud, Theft, Traffic Offense, Weapon Possession or Sale, Wrongful Use of Cocaine under UCMJ.

Forensic evidence is a variable that includes all types of forensic evidence that were considered during the adjudication of the defendant’s case. Types include: Accident reconstruction, Ballistic trajectory (crime scene), Biological evidence, Bitemark, Blood spatter (crime scene), Crime scene investigation, Digital evidence, DNA, Dog scent, Fiber/trace evidence, Fingernail comparison, Fire debris chemical analysis, Fire debris investigation (not chemical analysis), Firearms identification, Forensic medicine (non-pediatric), Forensic medicine (pediatric physical abuse), Forensic medicine (pediatric sexual abuse), Forensic pathology (cause and manner), Forensic pathology (firearm trajectory, blast, distance, handedness), Forensic pathology (time of death), Forensic pathology (wound assessment, toolmark), Gunshot residue, Hair comparison, Handwriting, Latent fingerprint, Latent palmprint, No forensic evidence, Other, Seized drug analysis, Serology, Shoe/foot impression, Soil comparisons, Tire/tread impression, Toolmark (not firearms or autopsy), Toxicology, Voiceprint.

Examination Code is a unique identifier for the examination. It contains the Archive Record Number concatenated with a letter from A to Z to permit the coding of up to 26 examinations. The dataset includes up to eight examinations in a particular case in the actual coding.

Forensic evidence required? Is a variable that specifies whether the forensic evidence was required in the balance of evidence to convict. Possible values are “Yes” and “No”.

Examiner characteristics is a variable that describes relevant aspects of the forensic analyst responsible for the examination. Possible values are: Bachelors degree, Masters degree, Doctoral degree, Degree unknown, Certified examiner, Uncertified examiner, Defense expert, Examiner outside forensic context (e.g., physician), Postconviction expert, Examiner lacked sufficient training, Unknown, Blank.

FSO characteristics is a variable that describes relevant aspects of the forensic science organization affiliated with the forensic analyst responsible for the examination. Possible values are: Commercial forensic organization, Public forensic science laboratory, State laboratory, Federal laboratory, Accredited laboratory, Accreditation status unknown, Law enforcement entity, Independent but reports to law enforcement, Independent of law enforcement, Private and retained by law enforcement or prosecution, Fire/arson investigation unit, Undetermined, Medical examiner's office, Coroner's office, Independent consultant, Medical provider (e.g., hospital), Blank.

Case error related to this examination is a variable that records whether any error of any type was associated with the forensic examination. Possible values are: “Yes” and “No”

NIST error typology is a variable that describes errors related to the National Institute of Standards and Technology error typology described in the Proceedings of the 2015 International Symposium on Forensic Science Error Management. (National Institute of Standards and Technology, 2015) Possible values are: Human bias, Forensic examiner variability, Improperly collected or labelled evidence from crime scene, Break in the chain of custody, Contamination and mislabeling of evidence, Mishandling (losing sample or sample mixups), Misinterpretation of evidence, Misinterpreting data, Failure to follow best practices/processes/methods, Poor documentation and transcriptions, Inadequate training, Analyst incompetence, Failure to review analysis of original analyst, Misinterpretation of postmortem artifacts, Measurement error (systematic or random); Error rates in technology solutions, Unvalidated method, Inadequate scientific foundation, Inaccurate or misleading statistics. Note: Errors related to fraud and other issues were coded under the Error Typology for this study, not the NIST typology.

Claims and other issues is a variable that codes additional considerations about forensic examiantions. Possible values are: Better technology available with an improved scientific foundation and probative value, DNA would provide more probative result, Possible cognitive or confirmation bias, Honest mistake, No error by this forensic examiner, Possible errors related to postconviction expert, A presumptive test was used to adjudicate case without confirmatory testing, This forensic evidence was exculpatory.

Type 1 Error is a variable that provides the coding of type 1 errors as described in Section 4 of this coding guide.

Type 2 Error is a variable that provides the coding of type 1 errors as described in Section 4 of this coding guide.

Type 3 Error is a variable that provides the coding of type 1 errors as described in Section 4 of this coding guide.

Evidence Standard at Time of Trial is a variable that specifies the testimony standard for the discipline as it should have been applied at the time of the trial of the defendant that resulted in the wrongful conviction.

Evidence Standard 2020 is a variable that specifies the testimony standard for the discipline as it should have been applied in 2020 at the time of the initiation of the research project.

Type 4 Error is a variable that provides the coding of type 4 errors as described in Section 4 of this coding guide.

Type 5 Error is a variable that provides the coding of type 5 errors as described in Section 4 of this coding guide.