JUDGING FIREARMS EVIDENCE

BRANDON L. GARRETT,* ERIC TUCKER† & NICHOLAS SCURICH‡

ABSTRACT

Firearms violence results in hundreds of thousands of criminal investigations each year. To try to identify a culprit, firearms examiners seek to link fired shell casings or bullets from crime scene evidence to a particular firearm. The underlying assumption is that firearms impart unique marks on bullets and cartridge cases, and that trained examiners can identify these marks to determine which were fired by the same gun. For over a hundred years, firearms examiners have testified that they can conclusively identify the source of a bullet or cartridge case. In recent years, however, research scientists have called into question the validity and reliability of such testimony. Judges largely did not view such testimony with increased skepticism after the Supreme Court set out standards for screening expert evidence in Daubert v. Merrell Dow Pharmaceuticals, Inc. Instead, the surge in judicial rulings came more than a decade later, particularly after reports by scientists shed light on limitations of the evidence.

In this Article, we detail over a century of case law and examine how judges have engaged with the changing practice and scientific understanding of firearms comparison evidence. We first describe how judges initially viewed firearms comparison evidence skeptically and
thought jurors capable of making firearms comparisons themselves—without an expert. Next, judges embraced the testimony of experts who offered more specific and aggressive claims, and the work spread nationally. Finally, we explore the modern era of firearms case law and research.

Judges increasingly express skepticism and adopt a range of approaches to limit in-court testimony by firearms examiners.

In December 2023, Rule 702 of the Federal Rules of Evidence was amended, for the first time in over twenty years, specifically due to the Rules Committee’s concern with the quality of federal rulings regarding forensic evidence, as well as the failure to engage with the ways that forensic experts express conclusions in court. There is perhaps no area in which judges, especially federal judges, have been more active than in the area of firearms evidence. Thus, the judging of firearms evidence has central significance for the direction that scientific evidence gatekeeping may take under the revised Rule 702 in federal, and then state courts. We conclude by examining lessons regarding the gradual judicial shift toward a more scientific approach. The more-than-a-century-long arc of judicial review of firearms evidence in the United States suggests that, over time, scientific research can displace tradition and precedent to improve the quality of justice.

TABLE OF CONTENTS

INTRODUCTION........................................................................................................ 103
I. FIREARMS METHODS AND THE FIRST HALF-CENTURY
   OF JUDICIAL RULINGS............................................................. 110
   A. A PRIMER ON FIREARM AND TOOLMARK IDENTIFICATION..... 111
   B. THE RECEPTION OF FIREARMS EXPERTS IN U.S.
      COURTS: 1902–1930 ..................................................... 113
   C. A NATIONAL BODY OF FIREARMS RULINGS: 1930s TO
      1960s................................................................................. 120
   D. PRE-DAUBERT CASES.......................................................... 124
II. MODERN SCIENTIFIC ASSESSMENTS AND GROWING
    JUDICIAL SKEPTICISM OF FIREARMS EVIDENCE........... 126
    A. POST-DAUBERT CASES.................................................... 131
    B. GROWING JUDICIAL SKEPTICISM.................................... 132
       1. Evaluation of the Scientific Studies ......................... 138
    D. LITIGATING THE ERROR RATE STUDIES ...................... 141
    E. TESTIMONIAL LIMITATIONS AND POST-NAS AND
       PCAST RULINGS............................................................ 146
On November 11, 2016, a police officer recovered a forty-caliber Smith & Wesson cartridge casing from the scene of a homicide in Washington D.C. A police officer reported seeing a person discarding a Smith & Wesson semiautomatic pistol shortly after the homicide occurred. Police sent a recovered cartridge casing to the crime lab where an examiner identified it—conclusively—"as having been fired" by the pistol recovered from the defendant, charged with first-degree murder. As the case approached trial, the defense challenged the admissibility of this proffered expert testimony, arguing it should be excluded because it was not the "product of reliable principles and methods." In other words, the method lacked "scientific validity." After hearing from several experts and reviewing published studies, Washington D.C. Superior Court Associate Judge Edelman found that there was insufficient evidence that firearms examiners can reliably make an identification. The judge ruled an expert could—at most—opine that "the recovered firearm cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting." As we will describe, this is a powerful new limit on firearms evidence, a field in which experts have confidently concluded for decades that one and only one firearm—to the exclusion of all other firearms in the world—can produce the ammunition found at a given crime scene.
While this case represented just one trial judge’s ruling, it not only forms a part of a sea change in judicial review of firearms evidence, but also the local repercussions point to more fundamental problems in our criminal system. Consider a later case before Judge Edelman, this one with charges brought against two men for two killings involving firearms evidence. Prosecutors were understandably concerned. In this case, D.C.’s Metropolitan Crime Lab had reported that the same weapon fired the cartridge casings found at each crime scene. Perhaps because they feared that the judge might view the evidence with renewed skepticism, the prosecutors took an unusual step: they asked independent examiners to take a look at the evidence.

The independent experts definitively concluded that two different firearms were involved—the opposite of what the D.C. crime lab examiners had concluded. Internally, the lab examiners reexamined the evidence and agreed the cartridges came from different weapons. After meeting with lab managers, however, they instead reported an altered finding of “inconclusive,” meaning that no conclusion could be reached. The management notified the ANSI National Accreditation Board (“ANAB”), which accredited the lab, that an internal review resulted in an “inconclusive” finding, but the audit that followed found that the lab managers had acted to conceal the errors in the case. In April 2020, ANAB suspended the lab’s accreditation, and as a result, the lab was shut down. Prosecutors then opened a new probe into its firearms unit, the lab director resigned, the lab disbanded, and the firearms unit remains closed as of this writing.

10. Id.
11. Id.
12. Id.
13. Id.
14. See id. at 2–3 (“DFS management not only failed to properly address the conflicting results reported to the DFS by the USAO, but also engaged in actions to alter the results reached by the examiners assigned to conduct a reexamination of the evidence.”).
17. Jack Moore, D.C. Abruptly Disbands Crime Lab’s Firearms Unit, WTOP NEWS (Sept. 16,
This rapidly unfolding crisis began with a spot-check in a single case prompted by a judge asking a fundamental question: How often do firearms examiners get it right versus wrong? For decades, few judges asked the question, but as we detail in this Article, judges have become increasingly engaged with the underlying science and have transformed a backwater area of forensic evidence into a subject of complex litigation. Indeed, in no other area have judges engaged in such a detailed manner with the limits of the testimony expressed by examiners—making firearms evidence the most prominent testing ground for the 2023 amendments to the Federal Rules of Evidence, designed to tighten judicial review of experts more generally, but with a focus on forensic evidence more specifically.18

Firearms examination is in great demand, with more than a hundred thousand requests for a forensic firearm examination each year in the United States.19 Firearms violence is a major problem in the United States—more than ten thousand homicides and almost five hundred thousand other crimes, such as robberies and assaults, are committed using firearms.20 When conducting these comparisons, examiners seek to link crime scene evidence—such as spent cartridge casings or bullets—with a firearm. These examiners assume that the manufacturing processes used to cut, drill, and grind a gun leaves distinct and identifiable markings on the gun’s barrel, breech face, firing pin, and other components. When the firearm discharges, those components in turn contact the ammunition and leave marks on it. Experts have long assumed, as we will describe, that firearms leave identifiable markings on the ammunition.21 They believe that they can definitively link spent ammunition to a particular firearm using these toolmarks.22 And for over a hundred years, examiners have offered criminal trial testimony relying

---

21. See infra Section I.A.
22. See id.
on this assumption.  

In recent years, the consequences of the uncritical judicial acceptance of firearms comparison testimony have come into sharper focus. Indeed, we now know that firearms evidence played a central role in numerous high-profile wrongful convictions. In the 2014 per curiam opinion in *Hinton v. Alabama*, for example, the U.S. Supreme Court reversed a conviction due to the defense lawyer’s inadequate performance in failing to develop firearms evidence at a capital murder trial. The central evidence was a State Department of Forensic Sciences examiner’s conclusion that six bullets were fired from the same gun: “[T]he revolver found at Hinton’s house.” The defense did not hire a competent and qualified expert, and the Court emphasized that “the only reasonable and available defense strategy require[d] consultation with experts or introduction of expert evidence.” Hinton was subsequently exonerated, and he commented: “I shouldn’t have [sat] on death row for thirty years . . . . All they had to do was to test the gun.”

This Article presents the results of a comprehensive review of all judicial rulings in the United States concerning firearms comparison evidence. Our database of more than 300 judicial rulings is available as a resource online. The database consists of written, published decisions (largely appellate opinions but also some trial rulings). We describe the three-part story of the path of firearms evidence: (1) initial skepticism of a novel set of methods, then moving to; (2) national acceptance of increasingly powerfully stated conclusions regarding firearms; and finally (3) a surge in

23. *See infra* Part I.  
25. *Id.* at 265.  
29. The cases that are included in this database were: [G]athered using searches of the Westlaw legal database, across all fifty states and the federal government, with rulings dating back over one hundred years. Where possible, trial rulings were obtained, but generally these cases reflect reported, written decisions containing the keywords used, and therefore largely reflect appellate rulings. The cases are searchable across a range of characteristics, including basic information concerning the state, year, type of court, and parties, but also details concerning the basis of the rulings and the factors relied upon by each court. The database describes whether the ruling employed a *Daubert* or *Frye* standard, or a ruling regarding local rules of evidence, and what the result of that ruling was. *Id.*
judicial opinions and skepticism of firearms comparison evidence that followed, not Daubert and the new reliability-focused standards for judicial review of scientific evidence, but rather a series of scathing reports by the scientific community calling into question the reliability of firearms evidence.

First, we describe how in the earliest cases, judges were actually quite skeptical of firearms comparison evidence, particularly when presented by self-styled experts, and often concluded that jurors were capable of making the comparisons themselves, without a need for expert testimony. However, particularly due to the influence of the flamboyant Major Calvin Goddard and his disciples, courts gradually embraced the firearms comparison evidence as the subject of expert testimony.

Second, we document how the claims made by experts became more specific and aggressive as the work spread nationally. Rather than simply describing a comparison between two sets of objects, firearms experts testified by making “uniqueness” claims: the theory that “no two firearms should produce the same microscopic features on bullets and cartridge cases such that they could be falsely identified as having been fired from the same firearm.” By the 1960s, this expert testimony was offered and accepted across the country. Professional groups emerged and set standards for the field, which courts took note of. Written judicial opinions became quite uncommon, and any judicial skepticism was largely limited to more unusual applications of the methods rather than the underlying methodology itself.

Third, we explore the modern era of firearms case law and research, with increasingly intense judicial interest and written opinions on the topic in the last two decades. In 1993, the Supreme Court decided Daubert v. Merrell Dow Pharmaceuticals and, along with its progeny and the revision to Federal Rule of Evidence 702 (“Rule 702”) and state-law analogues, judges now bear clearer and more rigorous gatekeeping responsibilities to assess the reliability of scientific evidence. Accompanying this shift in the courts, by the late 1990s, experts premised testimony on a “theory of

30. See infra Part I.
31. See infra Part I.
32. See infra Part I.
34. See infra Part I.
35. See infra Part II.
identification” set out by a professional association, the Association of Firearms and Tool Mark Examiners (“AFTE”). The AFTE instructs practitioners to use the phrase “source identification” to explain what they mean when they identify “sufficient agreement” of markings when examining bullets or cartridge cases.

In recent years, scientists have called into question the validity and reliability of this testimony—contributing to an explosion of judicial rulings. In a 2008 report, the National Academy of Sciences (“NAS”) found that “[t]he validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.” In its 2009 report, the NAS concluded “[s]ufficient studies have not been done to understand the reliability and repeatability of the methods.” The report also noted that “the lack of a precisely defined process . . . [that] does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.”

Over half of the judicial rulings that we identified have occurred since 2009, the year that the NAS issued its pathbreaking report. We detail dozens of opinions that have limited testimony of firearms experts in increasingly stringent ways.

Solidifying this trend, in 2016, the President’s Council of Advisors on Science and Technology (“PCAST”) reviewed in detail all of the firearm examiner studies that had been conducted to date, finding, with only one deemed appropriately designed, that “the current evidence falls short of the scientific criteria for foundational validity.” Most recently—beginning in the aforementioned 2019 case before Judge Edelman—scientists have testified about the research base of firearm examination.

---

38. See infra Part II.
42. Id. at 155. Judges have also raised concerns about the lack of specificity in the examination process. See, e.g., United States v. Green, 405 F. Supp. 2d 104, 114 (D. Mass. 2005) (stating the method is “either tautological or wholly subjective”); United States v. Shipp, 422 F. Supp. 3d 762, 779 (E.D.N.Y. 2019) (“[T]he sufficient agreement standard is circular and subjective.”).
43. President’s Council of Advisors on Sci. and Tech., Forensic Science in the Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods X (Sept. 2016) [hereinafter PCAST Report].
44. Id. at 111.
include psychologists, statisticians, and other academics with training in conducting science, rather than applying a forensic technique. As one judge put it, “[R]arely do the experts fall into such cognizable camps, forensic practitioners on one side and academic researchers on the other.”

The impact of these modern critiques on the admissibility of firearm examination has borne concrete results, but gradually. Comforted by more than a century of long-standing precedent, judges were slow to react to scientific concerns raised regarding firearms comparison evidence, even after the *Daubert* ruling. Yet in more recent years, as lawyers have increasingly litigated the findings of scientific reports and error rate studies, we have seen a dramatic rise in a judge’s willingness to engage with scientific limitations of the methods. That said, most judges have responded by imposing limits on how experts phrase conclusions in testimony, but we note there are reasons to doubt that this compromise solution will sufficiently inform lay jurors of the limits of the method.

For the first time since 2000, Federal Rule of Evidence 702 was amended, as of December 1, 2023. The Advisory Committee notes emphasize that these revisions are “especially pertinent” to forensic evidence. Further, for forensic pattern-comparison methods like firearms evidence, the committee noted that opinions “must be limited to those inferences that can reasonably be drawn from a reliable application of the principles and methods.” The amended Rule 702 specifically directs judges to (1) more carefully consider that the proponent of an expert bears the burden to show that the various reliability requirements are met and (2) underscore that the opinions that the expert formed are reliably supported by the application of the methods to the data. The rule changes squarely address the issues that judges have grappled with in the area of firearms evidence, perhaps more prominently than in any other area of scientific evidence. The rule changes target the two main concerns that judges have raised: the reliability of the methods and the overstatement of conclusions.

Thus, the body of case law regarding firearms evidence may only grow, and it may be a harbinger for how judges will engage with scientific evidence.

---

47. See infra Section II.D.
48. Regarding effectiveness of such measures, see Garrett et al., supra note 8, at 421–22. For further discussion, see infra Part III.
49. 2022 COMM. ON RULES OF PRAC. AND PROC., supra note 18, at 891–93.
51. Id. at 230.
52. 2022 COMM. ON RULES OF PRAC. AND PROC., supra note 18, at 891–93.
more broadly after the rule change. In a 2023 ruling, the Supreme Court of Maryland ruled that an expert can only opine on whether spent bullets or cartridges are “consistent or inconsistent” with those known to have been fired by a particular weapon.53 In perhaps a sign of things to come, a trial judge in Cook County, Illinois recently excluded firearms expert testimony entirely, based on scientific concerns with reliability, after conducting an extensive evidentiary hearing. There, the judge concluded that the probative value of the evidence was a “big zero” and raised the concern of “yet another wrongful conviction” based on such evidence if the jurors viewed “[t]he combination of scary weapons, spent bullets, and death pictures without even a minimal connection” to expertise that is repeatable and reproducible.54

These developments more fundamentally suggest that for judges and lawyers to carefully engage with the reliability rules set out in Daubert and in Rule 702, it takes engagement by the scientific community. Prominent scientific reports and studies have helped judges and lawyers apply scientific criteria to firearms examinations. The result has limited unsupported use of these firearms comparisons and may promote better methods in the future that can prevent errors and wrongful convictions.55 The changes to Rule 702 can cement these developments and ensure more careful review of scientific expert evidence more broadly. We conclude by examining the lessons to be learned from this more-than-a-century-long arc of judicial review of firearms evidence in the United States for future judicial engagement with science.

I. FIREARMS METHODS AND THE FIRST HALF-CENTURY OF JUDICIAL RULINGS

In this Part, we begin by describing the basic approach used by firearms and toolmark examiners. The approach has been in use for over a hundred years, and its origins trace to a single pioneering examiner, Major Calvin H. Goddard, who powerfully transformed courts’ early skepticism toward firearms comparison evidence to near-universal acceptance.56 Considered the “father” of modern forensic firearms examination, Goddard assembled databases of information from gun makers and pioneered a “comparison microscope,” a device with side-by-side eyepieces, to make comparing firearms evidence more convenient.57 While quite primitive compared with modern technology, Goddard introduced the use of the microscope in

55. See infra Section II.E.
57. Id.
firearms comparison, which was seen as permitting a level of sophisticated visual analysis that a layperson lacked access to. We describe how, in the 1930s, Goddard often testified in trials about the comparison microscope, further cementing the method’s legitimacy to courts. Over time, other practitioners and crime laboratories adopted similar methods and began to testify as experts. We describe in this Part what reasoning courts used through the 1930s as they moved from early skepticism to acceptance of this expert testimony.

A. A PRIMER ON FIREARM AND TOOLMARK IDENTIFICATION

Toolmark identification is the practice of human observers opining on whether toolmarks were produced by a particular tool. A tool is considered any device that serves a mechanical purpose (for example, screwdrivers, pliers, knives, pipe wrenches). As the tool contacts softer material, it sometimes leaves marks on the softer object’s surface. The resulting marks are called “toolmarks.” A firearm consists of many tools that perform mechanical functions to fire a bullet. Therefore, firearm identification is considered a subspecialty of toolmark identification. The goal of firearm identification is to determine whether two bullets or cartridge cases were fired by the same firearm.

Firearm identification typically involves the examination of features or marks on either bullets or cartridge cases. A piece of unfired ammunition contains four components: (1) a cartridge case, (2) a primer, (3) propellant (gun powder), and (4) a bullet. The cartridge case holds the unit of ammunition together with the bullet in its mouth. When an individual pulls the trigger of a firearm, a firing pin strikes the primer, which is at the head of the cartridge case. Striking the primer creates a spark that ignites the propellant. The ignition of the propellant forces the bullet to detach from the cartridge case and exit the barrel of the firearm. All of these operations have the potential to impart marks on the cartridge case, on the bullet, or on both. For example, manufacturers use firing pins with different shapes, which are often readily apparent on a fired cartridge case. Similarly, the barrel of the gun has grooves machined into it to impart a spiral spin on the bullet (akin

58. *Id.*

59. One text gives the following example: “For example, when a butter knife is dragged along the surface of butter, one may observe a series of lines across the top of the butter. In this case, the mark in the butter is a toolmark and the knife is the tool that made the mark.” RONALD NICHOLS, FIREARM AND TOOLMARK IDENTIFICATION: THE SCIENTIFIC RELIABILITY OF THE FORENSIC SCIENCE DISCIPLINE 1 (2018).

60. United States v. McCluskey, No. 10-2734, 2013 U.S. Dist. LEXIS 203723, at *7 (D.N.M. Feb. 7, 2013) (“Firearm identification is a specialized area of toolmark identification dealing with firearms, which involve a specific category of tools.”).
Practitioners call these types of features “class characteristics.” Class characteristics are the result of design features selected by the manufacturer. For example, a manufacturer may choose to use an elliptical-shaped firing pin or a barrel with six right-hand twisting grooves. The ammunition’s size is also a class characteristic. Class characteristics are a useful first step in firearm examination since observing differences in class characteristics can immediately rule out the possibility that two bullets or cartridge cases were fired by the same gun.

Agreement in class characteristics alone, however, is not sufficient to determine that bullets or cartridge cases were fired by the same gun. To draw that inference, examiners must identify and evaluate “individual characteristics,” which are defined by the AFTE as:

Marks produced by the random imperfections or irregularities of tool surfaces. These random imperfections or irregularities are produced incidental to manufacture and/or caused by use, corrosion, or damage. They are unique to that tool to the practical exclusion of all other tools.

Examiners rely on training and experience to assess whether striations are uniquely the result of a particular firearm (in other words, individual characteristics), as opposed to incidental striations that occurred during production and may be apparent in many different firearms of the same class. Examiners following the AFTE protocol can reach one of several conclusions based on their evaluation of the individual characteristics: identification, elimination, inconclusive, or unsuitable for comparison.

There are no numeric thresholds for how many individual characteristics must be observed before the examiner can declare that two bullets or cartridge cases were fired by the same gun (that is, “an identification”). Rather, the AFTE protocol states that an identification can be reached “when the unique surface contours of two toolmarks are in agreement.”

---

61. The official definition used by the professional Association of Firearms and Tool Mark Examiners is “[m]easurable features of a specimen which indicate a restricted group source. They result from design factors and are determined prior to manufacture.” GLOSSARY OF THE ASSOCIATION OF FIREARM & TOOL MARK EXAMINERS 38 (6th ed. 2013).
62. Id. at 65.
63. These incidental striations are often called “subclass characteristics,” or features that may be produced during manufacture that are consistent among items fabricated by the same tool in the same approximate state of wear. These features are not determined prior to manufacture and are more restrictive than class characteristics. Subclass characteristics can easily be confused with individual characteristics. See Gene C. Rivera, Subclass Characteristics in Smith & Wesson SW40VE Sigma Pistols, 39 AFTE J. 247 (2007).
As defined by the AFTE:

This “sufficient agreement” is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. . . . The statement that “sufficient agreement” exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.65

This criterion of “sufficient agreement” has been roundly criticized by numerous commentators and courts for being “circular.”66 It is, however, the criterion adopted by the AFTE and widely used by practicing firearm examiners who conduct casework.67


While there is increasingly voluminous scholarship regarding the early origins of gun control in the United States, we are not aware of scholarship exploring the early use of experts seeking to link firearms to particular shootings.68 In this Section, we detail what we learned from assembling our database of firearms rulings, collected using searches of legal databases and supplemented with unpublished trial court orders where available.69 As we will describe, twenty-nine of the earliest rulings predated Frye v. United States, a 1923 case that formed the basis for the federal standard for judicial review of novel expert evidence: a requirement of “general acceptance” within the relevant scientific community.70 Further, none of the eleven rulings decided from 1923–1930 cited to Frye—we did not see courts relying on the Frye standard until many decades later. Many of these rulings, absent clear rules of evidence concerning expert testimony, instead focused on

65. Id.
66. See, e.g., PCAST REPORT, supra note 43, at 60 (“More importantly, the stated method is circular. It declares that an examiner may state that two toolmarks have a ‘common origin’ when their features are in ‘sufficient agreement.’ It then defines ‘sufficient agreement’ as occurring when the examiner considers it a ‘practical impossibility’ that the toolmarks have different origins.”).
69. See supra note 28 for a description of the database and a link to it.
70. Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).
whether experts could assist or inform the jury.\textsuperscript{71} The earliest rulings date back to the 1870s and they were quite mixed on whether it was erroneous or correct to have admitted expert testimony concerning firearms.\textsuperscript{72}

One of earliest reported cases discussing firearms comparison evidence, \textit{Commonwealth v. Best},\textsuperscript{73} was written in 1902 by none other than Oliver Wendell Holmes, then the Chief Justice of the Massachusetts Supreme Judicial Court. Best was convicted of murder, and on appeal, argued that certain firearms comparison evidence offered at the trial was erroneous.\textsuperscript{74} The State argued at trial that Best shot a milkman twice with a Winchester rifle found in Best’s kitchen.\textsuperscript{75} To prove this, the State fired a third bullet through the gun, took a photograph of it, and published photographs of this bullet and the bullets found in the victim’s body as evidence.\textsuperscript{76}

In conjunction with these photographs, the State called an expert witness to “testify that [the bullets] were marked by rust in the same way that they would have been if they had been fired through the rifle at the farm, and that it took at least several months for the rust that he saw in the rifle to form.”\textsuperscript{77} In other words, the bullets found at the crime scene were rusted only because they were fired through the rusty barrel of Best’s rifle.\textsuperscript{78} Best’s counsel argued—at trial and on appeal—that the evidence was inadmissible because “the conditions of the experiment did not correspond accurately with those of the date of the shooting,” that “the force impelling the different bullets were different in kind,” that “the rifle barrel might be supposed to have rusted more in the little more than a fortnight that had intervened, and that it was fired three times on [the murder date], which would have increased the leading of the barrel.”\textsuperscript{79} To wit: environmental factors called the expert’s conclusion into question.

\textsuperscript{71} Today, such a standard is reflected in Federal Rule of Evidence 702(a). See FED. R. EVID. 702(a) (asking whether “the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue”).


\textsuperscript{73} \textit{Commonwealth v. Best}, 62 N.E. 748 (Mass. 1902).

\textsuperscript{74} \textit{Id.} at 749–50.

\textsuperscript{75} \textit{Id.} at 750.

\textsuperscript{76} \textit{Id.}

\textsuperscript{77} \textit{Id.}

\textsuperscript{78} \textit{Id.}

\textsuperscript{79} \textit{Id.}
In his quintessentially succinct style, Justice Holmes swiftly disposed of these arguments, concluding that expert testimony was the only way “the jury could have learned so intelligently how that gun barrel would have marked a lead bullet fired through it,” and “the sources of error suggested were trifling.”

Indeed, despite this being one of the first published opinions that we could find on the admissibility of firearms toolmark evidence, Justice Holmes found “no reason to doubt that the testimony was properly admitted.”

Rejecting the other arguments that Best made on appeal, the court upheld the conviction.

On the West Coast, two years later, the California Supreme Court decided People v. Weber, a 1906 case that also involved crude firearms comparison evidence. Four members of the Weber family had been killed on their property, three from gunshots and one from blunt force trauma. Police found a .32-caliber revolver in the basement of the Weber barn with dried blood on it, along with five discarded cartridges. The defendant was tried and convicted of one of the murders, and he appealed. During the trial, the State called “an expert in small arms” who testified that he “compared the markings on the bullets taken from the bodies with the markings on the bullets which he had fired from the pistol,” concluding that these bullets were all fired from the alleged murder weapon. While the trial court initially admitted this testimony, the next day, the court struck it, concluding “the comparison of the . . . bullets . . . is not a matter of expert testimony, but one within the ordinary capacities of the average juror or citizen.” Thus, the testimony was excluded, but the bullets were all admitted into evidence for the jury to compare during deliberations. On appeal, the California Supreme Court did not disturb the trial court’s ruling, but it did reject the defense’s argument that admitting the bullets into evidence was erroneous.

The court instead held that admitting the evidence to help the jury identify the murder weapon “was pertinent and important.”

In the 1920s, courts gradually moved toward considering firearms examiners as expert witnesses. In State v. Clark, the Oregon Supreme
Court considered a criminal appeal of a manslaughter conviction. Charles Taylor, a worker in Oregon’s National Cascade Forest Reserve, was part of a group assigned to bridge maintenance. Each worker brought a .30-30 Winchester rifle, hoping to hunt “camp meat.” One night, Clark and Taylor began hunting, and each fired an initial shot so they could use the spent cartridges as communication whistles. Taylor then left to hunt, but he was never seen alive again. The subsequent search party found a shell near Taylor’s body and an empty shell in the barrel of Taylor’s gun. According to the court, both shells “bore on the brass part of the primer a peculiar mark evidently caused by a flaw in the breechblock of the gun from which they had been fired.” This design flaw “caused a very slight, almost microscopic protuberance in the primer of the shell, which enlarged photographs made very clear to the naked eye.” Law enforcement fired several shots from Clark’s gun, and the cartridges produced the same mark. Additionally, Clark’s gun created a “sort of double scratch” on the inside of the rim of each shell fired, while Taylor’s gun “made only a single scratch.” Because of this, the court eliminated “the theory that deceased might have been accidentally shot with his own gun.” The court held these tests had produced “strong evidence that [Clark] was present and fired the shot that killed Taylor.” This evidence was presented in trial by the sheriff who described the marks but does not appear to have made more specific conclusions. Clark’s counsel objected to this testimony and admission of the photographs, but no specific reason for the objection was provided—unsurprising because rules surrounding lay and expert witnesses were less formal in this era. The court held that the testimony was proper and the evidence was admissible.

In a 1922 case, the Alabama Supreme Court explicitly held—unlike in the cases discussed so far—that firearms comparison examiners could testify as expert witnesses. A person was convicted for killing a man and his dog

91. *Id.* at 362.
92. *Id.* at 362–63.
93. *Id.* at 367.
94. *Id.*
95. *Id.*
96. *Id.*
97. *Id.*
98. *Id.*
99. *Id.*
100. *Id.* at 370.
101. *Id.* The only specific objection regarding the shells was that the photographs were impermissibly enlarged, which the court rejected. *Id.* at 371.
102. *Id.* at 370–71.
via gunshot. Police had found a revolver near the victim’s body and the revolver had one cartridge in the chamber that had been discharged. The State called someone “familiar with such things, [who] had used pistols and shells a good deal,” to testify as an expert. This expert claimed the casing in the empty chamber and the barrel of the revolver demonstrated it “had not been discharged recently.” The defense unsuccessfully objected, arguing that the person was not an expert. On appeal, the Alabama Supreme Court upheld the testimony: “A witness may have expert knowledge of some of the more ordinary affairs of life.”

In a 1923 case, however, the Illinois Supreme Court powerfully objected to expert evidence on firearms comparison. The court reversed the conviction on appeal for multiple reasons, but it particularly took issue with the State’s use of a police officer as an expert. At trial, a police officer testified for the State that a gun in evidence was the one fired at the victim because it “was the identical revolver from which the bullet introduced in evidence was fired on the night [the victim] was shot.” The officer was “asked to examine the Colt automatic .32 aforesaid, and gave it as his opinion that the bullet introduced in evidence was fired from the Colt automatic revolver in evidence.” The court emphasized:

He even stated positively that he knew that that bullet came out of the barrel of that revolver, because the rifling marks on the bullet fitted into the rifling of the revolver in question, and that the markings on that particular bullet were peculiar, because they came clear up on the steel of the bullet.

The court elaborated:

The evidence of this officer is clearly absurd, besides not being based upon any known rule that would make it admissible. If the real facts were

---

104. *Pynes*, 92 So. at 665.
105. *Id.*
106. *Id.*
107. *Id.*
108. *Id.*
109. *Id.* For a case from the next year finding a similar expert “competent” and any error harmless, see *Laney v. United States*, 294 F. 412, 416 (D.C. Cir. 1923).
111. *Id.* at 94.
112. *Id.*
113. *Id.* The Court also questioned the qualifications of the officer:
The state sought to qualify [the officer] for such remarkable evidence by having him testify that he had had charge of the inspection of firearms for the last 5 years of their department; that he was a small-arms inspector in the National Guard for a period of 9 years; and that he was a sergeant in the service in the field artillery, where the pistol is the only weapon the men have, outside of the large guns or cannon.
114. *Id.* (emphasis added).
brought out, it would undoubtedly show that all Colt revolvers of the same model and of the same caliber are rifled precisely in the same manner, and the statement that one can know that a certain bullet was fired out of a 32-caliber revolver, when there are hundreds and perhaps thousands of others rifled in precisely the same manner and of precisely the same character, is preposterous.\footnote{115}

Finally, the court focused on lay versus expert opinions:

\begin{quote}
Mere opportunity does not change an ordinary observer into an expert, and special skill does not entitle a witness to give an opinion, when the subject is one where the opinion of an ordinary observer is admissible, or where the jury are capable of forming their own conclusions from the pertinent facts susceptible of proof in common form. . . . If any facts pertaining to the gun and its rifling existed by which such fact could be known, it would have been proper for the witness to have stated such facts and let the jury draw their own conclusions.\footnote{116}
\end{quote}

The court thus strongly rejected admitting an expert to opine on such firearms evidence.\footnote{117}

By the late 1920s, however, judicial rulings began to shift as the work of Major Goddard became more known. Goddard founded a private crime laboratory—"The Bureau of Forensic Ballistics"—and published the \textit{American Journal of Police Science}. Goddard became particularly well-known for assisting with the investigation in the Sacco and Vanzetti case in Massachusetts and in the St. Valentine’s Day Massacre in Chicago in 1929.\footnote{119} Before Goddard published his seminal article on ballistic evidence for the U.S. Army in 1925, \textit{Forensic Ballistics}, many judges, as described above, viewed firearms comparison as a crude technique that jurors could conduct themselves by visually examining the evidence.\footnote{120}

This began to change. For example, in a 1928 Kentucky case, \textit{Jack v. Commonwealth}, the state supreme court discussed firearms comparison testimony and found the evidence "important if competent, but highly prejudicial if incompetent."\footnote{121} The court discussed an article by Major Goddard in \textit{Popular Science Monthly}\footnote{122} and summarized the process:

\begin{quote}
\end{quote}
There is in use a special microscope consisting of two barrels so arranged that both are brought together in one eyepiece. The fatal bullet is placed under one of these barrels, and a test bullet that has been fired through defendant’s pistol is placed under the other barrel, and this brings the sides of the two bullets together and causes them to fuse into one object. If the grooves and other distinguishing marks on both bullets correspond, it is said to show that both balls were fired from the same pistol.123

The court concluded:

It thus appears that this is a technical subject, and in order to give an expert opinion thereon a witness should have made a special study of the subject and have suitable instruments and equipment to make proper test . . . Clearly the witnesses in this case were not qualified to give such opinions and conclusions and the admission of such evidence was erroneous and prejudicial.124

The court therefore rejected the testimony not because it doubted the method itself but because the proffered experts did not follow proper practices.

One year after Jack, the Kentucky Supreme Court again examined firearms comparison testimony in Evans v. Commonwealth.125 The defendant, Evans, was indicted for murder of the Pineville, Kentucky chief of police, and he was ultimately convicted of manslaughter.126 Six shots were fired in the murder, and police had dug up a bullet from the ground near the scene.127 Evans’s primary argument on appeal was that firearms comparison evidence was improper, so the court addressed it “with some degree of elaboration.”128 The court referenced Jack and noted that one month after Jack was published, Major Goddard—who wrote the article referenced by the court in Jack—offered to testify.129 Goddard was given the defendant’s automatic .45 pistol, seven cartridges taken from this pistol, six cartridges found at the scene of the crime, and the bullet that police had taken from the dirt.130 Goddard concluded “that he was convinced that the bullet that had been introduced into evidence had been fired through [Evans’s] pistol.”131

To justify this conclusion, Goddard gave a detailed account of how he compared the different bullets by putting “the two bullets under the two status of an exact science.” Id. at 963.

123. Id. at 963–64.
124. Id. at 964 (emphasis added).
125. Evans v. Commonwealth, 19 S.W.2d 1091 (Ky. 1929).
126. Id. at 1092.
127. Id.
128. Id. at 1093.
129. Id.
130. Id. at 1094.
131. Id. (emphasis added).
microscopes together, [so that] in the center . . . you see a single bullet. . . . [I]f these bullets were fired through the same pistol they will match . . . .” 132 Goddard testified that he “only required one single test to identify the bullet in evidence as having been fired through the Evans pistol.” 133

During Goddard’s cross-examination, the jury was allowed to examine the evidence using the microscope. 134 The defense objected that Goddard’s conclusion was one of fact that the jury should instead determine. 135 The court rejected this argument. 136 Interestingly, the court concluded that Goddard’s opinion was an ordinary lay opinion, not that of an expert. 137 Cross-examination was thus a sufficient safeguard, and “rigid adherence” to the rules of evidence “would be subversive of the ends for which they were adopted.” 138 The defense also objected to the jury looking through the microscopes which the court quickly dismissed as without “well-founded reason.” 139

These two Kentucky Supreme Court opinions formed the framework for the modern approach to firearms comparison evidence. 140 Jack demonstrates that courts would not always let a specific person testify as a qualified expert on firearms comparison. But Evans shows that the courts were not concerned about the underlying validity of the methodology of firearms comparisons. If the State could produce a witness in the mold of Major Goddard, following the now-respected comparison microscope methodology, then the testimony would routinely be admitted.

C. A NATIONAL BODY OF FIREARMS RULINGS: 1930S TO 1960S

Beginning in the 1930s, judges began to further develop case law in other parts of the country, with new experts testifying. We identified forty rulings from 1931–1970, each set out in our database. During this time period, rulings spread nationally, as judges appear powerfully influenced by Evans, 140 which became one of the lodestar cases for adoption of firearms comparison evidence. Use of toolmark evidence for firearms comparison

132. Id. at 1095.
133. Id. (emphasis added).
134. Id. at 1096.
135. Id. at 1097.
136. Id.
137. Id. The court compared Goddard’s testimony to that of a lay witness, saying that “he could smell gasoline,” even though “the average man would have great difficulty in telling just how coal oil or gasoline smells, though acquainted with their odors.” Id.
138. Id.
139. Id.
140. Evans v. Commonwealth, 19 S.W.2d 1091 (Ky. 1929).
began to be called “accepted” and “well-recognized” as a methodology. As
time went on, judges simply cited to Evans and other prototypical early cases
to admit expert testimony, and discussion of the merits of firearms
comparison methods diminished. Further, defendants increasingly did not
challenge the evidence but rather focused on the preservation of evidence or
the qualifications of the testifying experts. These challenges were almost
always unsuccessful.

In 1937, for example, the Florida Supreme Court briefly concluded that
a firearms comparison expert was “fully qualified to testify as an
expert . . . and to draw a reliable conclusion as to whether or not the bullet
found in the body of the deceased was fired from the pistol introduced in
evidence.”141 In another Missouri case, the expert himself explained that he
“was not a ballistic expert,” but he still argued he had “much experience in
the work of identifying firearms.”142 Despite this concession, the court
concluded that “he was an expert in the identification of firearms and bullets
by the comparison method by means of a microscope.”143 In 1938, an
Oklahoma appellate court further explained:

There were few decisions with reference to the introduction of expert
testimony to identify the weapon from which a shot was fired until recent
years, but the science of ballistics is now recognized as one of the best
methods in ferreting out crime that could not otherwise be detected. Expert
testimony to identify the weapon from which a shot was fired is generally
admitted under the rules covering other forms of expert testimony, and it
is the modern tendency of the courts to allow the introduction of such
testimony, where the witness’ preparation as shown by experience and
training qualifies him to give expert opinion on firearms and ballistics
tests.144

By 1940, experts could cite fifteen years of experience in “the firing of
different caliber pistols,” which was enough to qualify a person as a firearms
comparison expert.145 In a 1941 case in Virginia, an expert from the FBI
testified that he had twenty years of experience, “six of which had been
devoted to the examination of firearms.”146 The expert testified that the
cartridge he examined was fired by the defendant’s shotgun.147 The
reviewing court cited to Evans,148 as courts continued to do. For example, in

142. State v. Couch, 111 S.W.2d 147, 149 (Mo. 1937).
143. Id.
146. Ferrell v. Commonwealth, 14 S.E.2d 293, 295 (Va. 1941).
147. Id. at 296.
148. Id. at 297.
State v. McKeever,\(^{149}\) the expert testified this was his 191st trial—the court allowed the evidence to be admitted without discussion, simply citing to Evans.\(^{150}\) Increasingly brief opinions found “no error” in introduction of such testimony.\(^{151}\)

There were some outliers. For example, a 1948 New Mexico Supreme Court ruling reversed the admissibility of “ballistic expert” testimony, which allegedly matched a specific gun to the bullet that killed the victim.\(^{152}\) After being qualified, the expert testified about his methodology, calling the firearm’s marks “absolutely identical.”\(^{153}\) The court was concerned that the expert had concluded with statements such as: “I will state positively that the evidence bullet (death bullet) was fired out of State’s Exhibit No. 2, this [defendant’s] gun.”\(^{154}\) The court emphasized that while firearms comparison is “almost, if not an exact science,” and “judicial notice may be taken” of the method, ballistic experts still must, “like . . . experts generally,” only provide “opinion testimony.”\(^{155}\) While “[i]t may be true that such witnesses as Colonel Goddard, who testified in Evans v. Commonwealth and other reported cases, are so skilled in the science of forensic ballistics that the chance of error is negligible,” they are the exception.\(^{156}\) Yet, “[t]he belief of a witness that his skill is so transcendent that an error in judgment is impossible, may itself be false or a mistake, assuming that the science is exact.”\(^{157}\)

In a rare 1951 Georgia Supreme Court case, Henderson v. State, the court excluded firearms comparison testimony due to concerns with the specific expert. The defense attorney asked the expert “why he did not measure the distance and depth of the grooves, and the witness explained by giving the reply that the microscope was the highest and best evidence.”\(^{158}\) The court held that the answer was not a “response to the question propounded,”\(^{159}\) that the right to a “thorough and sifting cross-examination”

\(^{149}\) State v. McKeever, 101 S.W.2d 22 (Mo. 1936).

\(^{150}\) Id. at 29.

\(^{151}\) See, e.g., Pilley v. State, 25 So.2d 57, 60 (Ala. 1946) (“In the introduction of this evidence there was no error.”); Kyzer v. State, 33 So.2d 885, 887 (Ala. 1947) (finding no error without explanation). In Collins v. State, 33 So.2d 18, 20 (Ala. 1947), the court overruled objections to the expert testimony, stating: “We have had occasion several times to consider questions of this sort, and the principles of law applicable to the same have been repeated frequently, so that it will not be necessary to do so again . . . .” Yet, in none of those prior opinions did the court actually repeat or state its reasoning.


\(^{153}\) Id. at 257–58.

\(^{154}\) Id. at 260 (emphasis added).

\(^{155}\) Id.

\(^{156}\) See id. at 261 (citation omitted).

\(^{157}\) Id.

\(^{158}\) Henderson v. State, 65 S.E.2d 175, 177 (Ga. 1951).

\(^{159}\) Id.
was violated, and that the judgement should be reversed for a new trial. In a Maryland case, the defendant also attacked the State’s firearms comparison testimony. The court emphatically rejected this position:

For many years ballistics has been a science of great value in ferreting out crimes that otherwise might not be solved. When a pistol is fired, a pressure is developed within the shell which drives the bullet out of the barrel, and the shell is driven back against the breech of the pistol with similar force. The markings on the hard breech of the pistol are thereby stamped on the soft butt of the shell. Testimony to identify the weapon from which a shot was fired is admissible where it is shown that the witness offering such testimony is qualified by training and experience to give expert opinion on firearms and ammunition.

The court cited back to Best and Evans to justify this result, despite the faint marks and acknowledgment that the marks could have been explained by a different type of weapon.

In a 1964 Florida case, the court provided the following explanation about the recognition of firearms comparison testimony:

It is now well established that a witness, who qualifies as an expert in the science of ballistics, may identify a gun from which a particular bullet was fired by comparing the markings on that bullet with those on a test bullet fired by the witness through the suspect gun. An expert will be permitted to submit his opinion based on such an experiment conducted by him. The details of the experiment should be described to the jury.

Finally, a 1969 Illinois appellate case offers some of the earliest descriptions of class and individual characteristics, the predominant terminology in modern firearms comparison testimony:

When a weapon is received at the laboratory it is classified as to type, caliber, make and model. Each gun has class characteristics common to its particular make and model. In addition, each gun has its own individual characteristics. . . After the gun is received at the laboratory, if operable, it is fired into a bullet recovery box. The bullet in question is then compared with the test bullet under a comparison microscope.

160. Id.
162. Id.
163. See id. at 635–36 (noting that “it was admittedly possible that the bullets could have been fired from a Luger” rather than the defendant’s gun).
164. Roberts v. State, 164 So. 2d 817, 820 (Fla. 1964).
During this time, courts routinely rejected challenges to firearms experts’ qualifications.\(^1\) And expert qualifications only increased: by now, some experts testified that they had worked on “approximately three to four thousand cases of ballistics.”\(^2\) Judicial review of forensic evidence in the following decades involved significant deference, with trial courts deferring to the expert witnesses, and then the appellate courts deferring to the trial courts. Often, courts focused on the specific examiner’s experience rather than assessing the field’s foundational validity.\(^3\)

D. PRE-DAUBERT CASES

In the 1970s and 1980s, leading up to the Daubert ruling in 1993, courts routinely admitted firearms expert testimony, often without discussion.\(^4\) We located only twenty-four such rulings, perhaps because unpublished rulings became far more common given the broader acceptance of such expert testimony. While challenges to expert qualifications typically failed—with courts citing to the experience of the examiner—courts generally expected examiners to also possess specialized training and credentials.\(^5\)

Some courts excluded firearms testimony based on other issues.\(^6\) In a federal case, the defendant was denied access to an expert to examine the evidence and testimony, which was found particularly problematic given the quality of the evidence itself, as “seventy-five percent of this slug was...”\(^7\)

---

3. This, however, is not universal. For a more recent ruling, see State v. Raynor, 254 A.3d 874, 887–88 (Conn. 2020) (noting that refusing to consider new information as a scientific field evolves “would transform the trial court’s gatekeeping function . . . into one of routine mandatory admission of such evidence, regardless of advances in a particular field and its continued reliability”).
4. See, e.g., Hampton v. People, 465 P.2d 394, 400 (Colo. 1970) (stating there was no abuse of discretion for admitting a firearm comparison expert’s testimony). For perhaps the first case referring to the discipline as a type of toolmark comparison, see United States v. Bowers, 534 F.2d 186, 193 (9th Cir. 1976).
5. See, e.g., State v. Hunt, 193 N.W.2d 858, 867 (Wis. 1972) (stating “the witness had great experience in the field of ballistics”); Acoff v. State, 278 So. 2d 210, 217 (Ala. 1973) (concluding expert testimony of witness with “more than six years” of firearms comparison training was “properly allowed”); People v. McKinnie, 310 N.E.2d 507, 510 (Ill. App. Ct. 1974) (finding examiners’ “considerable practical experience” was sufficient, despite lack of “scientific” training). But see State v. Seebold, 531 P.2d 1130, 1132 (Ariz. 1975) (affirming exclusion of proffered experts at trial in which one admitted “he was not a scientist or a criminalist” and the second was a gunsmith and gun shop owner who “had no formal education in the field of ballistics and had never testified before in this field”); Cooper v. State, 340 So. 2d 91, 93 (Ala. Crim. App. 1976) (“The State, in attempting to establish Charles Wesley Smith as an expert in ballistics, elicited some general information on his background, but failed to establish many specific facts to support his expertise in the field of ballistics.”); Bowden v. State, 610 So. 2d 1256, 1258 (Ala. Crim. App. 1992) (affirming trial court’s exclusion of firearms expert’s testimony because it was not a “clear abuse of . . . discretion”).
destroyed and the identification was made on the remaining 25%.”\textsuperscript{172} Other cases relied on the Confrontation Clause, including one in which a police officer testified about a report by an examiner who was not present at trial.\textsuperscript{173} Still other courts considered whether experts sufficiently described their work.\textsuperscript{174}

In general, experts continued to reach highly aggressive conclusions that were permitted by courts. For example, the expert in a 1981 Wyoming case resolved, “The markings on the bullets from the home of appellant’s brother matched the markings found on the bullet removed from [the defendant], establishing that they had been fired from the same gun.”\textsuperscript{175} In a leading Virginia case, an expert testified he was “certain” one of the bullets removed from the victim’s body was fired from the defendant’s pistol, and there was “no margin of error.”\textsuperscript{176} The defendant argued on appeal that this “no margin of error” statement was impermissible.\textsuperscript{177} The court rejected this argument, simply concluding that the statement went toward the weight of the testimony, not its admissibility.\textsuperscript{178}

Pre-\textit{Daubert}, some defendants did contest whether firearms experts relied on sufficient facts and data. In an exemplar Utah case, the expert testified at a preliminary hearing that a bullet fired from the alleged murder weapon matched a bullet taken from the victim’s body.\textsuperscript{179} But while he gave this conclusion, he was not able to give “an exact description of the striations, nor did he have photographs of them available with him in court.”\textsuperscript{180} The court rejected arguments that the expert did not have sufficient foundation for his conclusion, holding that the testimony was within the expert’s specialized knowledge.\textsuperscript{181}

\textsuperscript{172}. Barnard v. Henderson, 514 F.2d 744, 746 (5th Cir. 1975).  
\textsuperscript{173}. Stewart v. Cowan, 528 F.2d 79, 82–83 (6th Cir. 1976).  
\textsuperscript{174}. People v. Miller, 334 N.E.2d 421, 429 (Ill. App. Ct. 1975). Other cases found it sufficient to admit testimony finding similar class characteristics, even when there was not enough information to compare any individual characteristics. See, e.g., State v. Bayless, 357 N.E.2d 1035, 1058–59 (Ohio 1976).  
\textsuperscript{177}. Id.  
\textsuperscript{178}. Id.  
\textsuperscript{179}. State v. Schreuder, 712 P.2d 264, 268 (Utah 1985).  
\textsuperscript{180}. Id.  
\textsuperscript{181}. Id. at 268–69.
II. MODERN SCIENTIFIC ASSESSMENTS AND GROWING JUDICIAL SKEPTICISM OF FIREARMS EVIDENCE

Following the U.S. Supreme Court’s ruling in 1993 in Daubert v. Merrell Dow Pharmaceuticals, Inc., federal courts began to more carefully scrutinize firearms evidence, although exclusion remained rare.182 Daubert led to the revision of Federal Rule of Evidence 702 in 2000 that established new standards to assess the reliability of scientific expert testimony. Many of the defendants’ objections shifted from concerns about the experts’ qualifications to concerns about the reliability of the methodology and conclusions,183 and about the use of inadmissible hearsay evidence as a basis for the experts’ conclusions.184 At the state level, there was not any immediate difference in how courts approached firearms expert testimony post-Daubert; methodology and expert qualifications were more explicitly mentioned, but the overall analysis largely remained the same.185

In the late 1990s and early 2000s, courts began rejecting expert firearms comparison testimony as unreliable, largely by relying on Daubert. In our database, we include just seven cases from 1993–2000. However, the number of rulings begins to dramatically increase after 2000, with 188 rulings from 2000 to 2022. We turn next to that rich body of modern case law.


Figure 1 illustrates this remarkable trend—one can see a fairly steady number of twenty or fewer reported judicial rulings regarding firearms comparison evidence through the 1990s. Yet, beginning in the early 2000s, these rulings began to dramatically increase in number.

The Supreme Court in *Daubert* revolutionized judicial review of scientific evidence by setting out five factors for courts to consider in evaluating expert testimony: whether the theory or technique relied on (1) can be (and has been) tested, (2) has been subjected to peer review and publication, (3) has a known or potential rate of error, (4) includes the existence and maintenance of standards controlling its operation, and (5) is generally accepted within the relevant scientific community. We provide an overview of each factor and how courts generally have reviewed them in the context of firearms comparison testimony.

First, courts generally have not questioned the “testability” of firearms forensics, a “key question” when examining reliability. A series of courts have held that the propositions that “firearms leave discernible toolmarks on bullets and cartridge casings fired from them, and that trained examiners can conduct comparisons to determine whether a particular gun has fired particular ammunition... can be, and have been, tested.”

---

187. *Id.* at 593.
Second, many courts have determined the AFTE method of toolmark identification has been subject to sufficient peer review and publication, largely through the AFTE Journal. However, courts are beginning to more rigorously inspect the validity of the peer review process at that journal. Prior to January 2020, the AFTE Journal used a highly unusual “open-review” process whereby the identities of the authors and the reviewers were disclosed and direct communication was encouraged. Furthermore, all of the reviewers were members of AFTE who “ha[d] a vested, career-based interest in publishing studies that validate their own field and methodologies.”

These factors led a D.C. Superior Court judge to conclude in 2019: “[T]he vast majority of [firearms comparison] studies are published in a journal that uses a flawed and suspect review process, [which] greatly reduces its value as a scientific publication.” Therefore, the peer review factor “on its own does not, despite the sheer number of studies conducted and published, work strongly in favor of admission of firearms and toolmark identification testimony.” Nevertheless, courts have cited to other studies or reports to validate the soundness of toolmark comparison—one federal court curiously cited to the 2009 NAS and 2016 PCAST reports as evidence of peer review, despite how damning those reviews are of the method.

Third, courts have tended to view the error rate for forensics firearms testing as low, though they also sometimes acknowledge that the error rate

---

189. See, e.g., Ashburn, 88 F. Supp. 3d at 245–46 (finding AFTE method has been subjected to peer review through the AFTE Journal); Otero, 849 F. Supp. 2d at 433 (describing the Journal’s peer reviewing process and finding the methodology subject to peer review); United States v. Taylor, 663 F. Supp. 2d 1170, 1176 (D.N.M. 2009) (finding AFTE method subjected to peer review through AFTE Journal and two articles submitted by the government in peer-reviewed journal about the methodology); Monteiro, 407 F. Supp. 2d at 366–67 (describing AFTE Journal’s peer reviewing process and finding it meets peer review element).


191. Id. at *35.

192. Id. at *36.

193. See Romero-Lobato, 379 F. Supp. 3d at 1119 (D. Nev. 2019) (“Of course, the NAS and PCAST Reports themselves constitute peer review despite the unfavorable view the two reports have of the AFTE method. The peer review and publication factor therefore weighs in favor of admissibility.”). But see United States v. Tibbs, No. 2016 CF1 19431, 2019 D.C. Super. LEXIS 9, at *29 (D.C. Super. Ct. Sept. 5, 2019) (“If negative post-publication commentary from an external reviewing body can satisfy this prong of the Daubert analysis, then the peer reviewed publication component would be more or less read out of Daubert, leaving behind only the requirement of some type of publication.”).
is “presently unknown.” One federal court concluded that “it is not possible” to calculate an absolute error rate for firearms analysis because “the process is so subjective and qualitative.” This third factor is particularly important for rigorous assessment because “an expert witness’s ability to explain the methodology’s error rate—in other words, to describe the limitations of her conclusion—is essential to the jury’s ability to appropriately weigh the probative value of such testimony.”

Faced with numerous studies purporting extremely low error rates, many courts have simply accepted the validity of these conclusions that forensics firearms testing does have a nominal error rate or has “a false positive rate of 1.52%.”

In more recent years, as we discuss in a later Section in more detail, courts have begun to reexamine the validity of the error studies and rates presented. Citing basic design flaws of most studies in the field and the studies’ failure to address a large number of “inconclusive” results, one court, for example, found “it difficult to conclude that the existing studies provide a sufficient basis to accept the low error rates for the discipline that these studies purport to establish.” Other courts noted concerns with the lack of rigorous testing but did not find this sufficiently persuasive to exclude the evidence outright.

Fourth, many judges have focused on how the AFTE methodology lacks clearly defined, objective standards. Judges have variously described the AFTE method as “inherently vague,” more of a description of the process of firearm identification rather than a strictly followed charter for the

---


197. See Ashburn, 88 F. Supp. 3d at 246 (“[T]he error rate, to the extent it can be measured, appears to be low, weighing in favor of admission.”); United States v. Otero, 849 F. Supp. 2d 425, 433–34 (D.N.J. 2012) (summarizing several studies indicating a low error rate); United States v. Taylor, 663 F. Supp. 2d 1170, 1177 (D.N.M. 2009) (“[T]his number [less than 1%] suggests that the error rate is quite low.”); Monteiro, 407 F. Supp. 2d at 367–68 (summarizing relevant studies and finding that the known error rate is not “unacceptably high.”).


199. See infra Section II.D; State v. Terrell, No. CR170179563, 2019 Conn. Super. LEXIS 827, at *3 (Conn. Super. Ct. Mar. 21, 2019) (“[The toolmark field] is also not static. A methodology may at one time be viewed as reliable by the scientific community and later fall out of favor.”).


201. Romero-Lobato, 379 F. Supp. 3d at 1120 (“While the Court is cognizant of the PCAST Report’s repeated criticisms regarding the lack of true black box tests, the Court declines to adopt such a strict requirement for which studies are proper and which are not. Daubert does not mandate such a prerequisite for a technique to satisfy its error rate element.”).

field,203 and “merely unconstrained subjectivity masquerading as objectivity.”204 And as many courts have pointed out, “the AFTE standard is circular—an identification can be made upon sufficient agreement, and agreement is sufficient when an identification can be made.”205 The inherent subjectivity has weighed against admissibility of firearms comparison evidence for many courts.206 Courts, however, have often also noted that they find such subjectivity “not fatal” to admissibility.207 Thus, courts often note that subjectivity alone does not make a method unreliable and they are focused on evaluating reliability.208

Finally, the last Daubert factor hinges on general acceptance within the relevant scientific community. Who constitutes the “relevant” scientific community has never been defined with precision, yet it is often determinative. Because the AFTE method is accepted within the organization’s own community of firearms examiners, courts frequently find the requisite general acceptance.209 But other judges have pointed out that this narrow definition is comprised exclusively of individuals “whose professional standing and financial livelihoods depend on the challenged discipline.”210 Thus, perhaps the relevant scientific community should be

---

204. Tibbs, 2019 D.C. Super. LEXIS 9 at *69.
206. See, e.g., Romero-Lobato, 379 F. Supp. 3d at 1121 (“With the AFTE method, matching two tool marks essentially comes down to the examiner’s subjective judgment based on his training, experience, and knowledge of firearms. This factor weighs against admissibility.”); United States v. Ashburn, 88 F. Supp. 3d 239, 246–47 (E.D.N.Y. 2015) (discussing subjectivity); Ross, 129 N.Y. S. 3d at 633 (describing testimony that “there is no across-the-board standard as to what is ‘sufficient agreement’ in his field”); United States v. Sebbern, No. 10 Cr. 87(SLT), 2012 U.S. Dist. LEXIS 170576, at *11 (E.D.N.Y. Nov. 30, 2012) (“[T]he standards employed by examiners invite subjectivity.”).
208. See, e.g., Romero-Lobato, 379 F. Supp. 3d at 1120 (“The mere fact that an expert’s opinion is derived from subjective methodology does not render it unreliable.”); United States v. Otero, 849 F. Supp. 2d 425, 431 (D.N.J. 2012) (“[E]xpert testimony on matters of a technical nature or related to specialized knowledge, albeit not scientific, can be admissible under Rule 702, so long as the testimony satisfies the Court’s test of reliability and the requirement of relevance.”).
209. See, e.g., United States v. Shipp, 422 F. Supp. 3d 762, 782 (E.D.N.Y. 2019) (“Most courts have, in cursory fashion, identified toolmark examiners as the relevant community, and have summarily determined that the AFTE Theory is generally accepted in that community.”).
210. United States v. Tibbs, No. 2016 CF1 19431, 2019 D.C. Super. LEXIS 9, at *73 (D.C. Super. Ct. Sept. 5, 2019); see also Shipp, 422 F. Supp. 3d at 783 (“The AFTE Theory has not achieved general acceptance in the relevant community.”). One court notes, “It is self evident that practitioners accept the validity of the method as they are the ones using it. Were the relevant scientific community limited to
broadened to include nonpractitioner research scientists.

While acknowledging the discipline’s weaknesses, most federal courts have balanced the Daubert factors and found testimony admissible. As one federal court put it: “[T]his lack of objective criteria is countered by the method’s relatively low rate of error, widespread acceptance in the scientific community, testability, and frequent publication in scientific journals.”

Further, as noted, Rule 702 was revised in 2000 to incorporate Daubert, but it specified additional factors, including asking courts to examine the application of a method to the facts in a case. Courts vary in whether they simply consider Daubert factors alone, or whether they also discuss Rule 702—as will be discussed next, litigants have increasingly focused on the as-applied language in Rule 702, critiquing how the method was used, as well as on the language an expert used to express conclusions.

A. POST-DAUBERT CASES

As a federal district court noted in 2005, for over a decade after the Daubert ruling, “every single court post-Daubert has admitted [firearms identification] testimony, sometimes without any searching review, much less a hearing.” When courts did examine firearms evidence, early post-Daubert challenges often focused on if the expert’s qualifications were sufficient under Rule 702, even if they did begin to discuss questions regarding reliability of methods and principles used. Further cases

practitioners, every scientific methodology would be deemed to have gained general acceptance.” State v. Terrell, No. CR170179563, 2019 Conn. Super. LEXIS 827, at *14 (Conn. Super. Ct. Mar. 21, 2019). In other forensics fields, acceptance among only practitioners has been deemed unreliable and has led to the exclusion of the evidence under Daubert. See, e.g., United States v. Saelee, 162 F. Supp. 2d 1097, 1104 (D. Alaska 2001) (“[G]eneral acceptance of the theories and techniques involved in the field . . . among the closed universe . . . proves nothing.”).


212. FED. R. EVID. 702.


216. See, e.g., State v. Samonte, 928 P.2d 1, 26–27 (Haw. 1996) (discussing the defendant’s argument that prosecution’s firearms expert was not qualified); Whatley v. State, 509 S.E.2d 45, 50 (Ga. 1998) (rejecting the defendant’s argument that evidence used was “inherently unreliable” and noting the
discussed—and rejected—the question of whether an expert’s conclusions were based on inadmissible hearsay, rather than their own observations and conclusions.\textsuperscript{217} And many courts, both state and federal, continued to admit the testimony without serious discussion.\textsuperscript{218}

In additional cases, judges dismissed objections to firearms experts whose testimony was said to reach “ultimate issues,” with the judges noting that the experts only opined regarding an acceptable “reasonable scientific certainty.”\textsuperscript{219} Thus, judges have emphasized the flexibility of the \textit{Daubert} and \textit{Kumho Tire}\textsuperscript{220} standards. As a Southern District of New York ruling explained:

> The Court has not conducted a survey, but it can only imagine the number of convictions that have been based, in part, on expert testimony regarding the match of a particular bullet to a gun seized from a defendant or his apartment. It is the Court’s view that the Supreme Court’s decisions in \textit{Daubert} and \textit{Kumho Tire}, did not call this entire field of expert analysis into question. It is extremely unlikely that a juror would have the same experience and ability to match two or more microscopic images of bullets.\textsuperscript{221}

\section*{B. GROWING JUDICIAL SKEPTICISM}

The federal courts took the lead in beginning to scrutinize firearms comparison testimony more closely. Judges began to write opinions with detailed examinations of the underlying methods experts used. Federal courts then imposed partial exclusions regarding either (1) the methods or qualifications of the particular experts or (2) the language the expert was permitted to use to describe the conclusion. In more recent years, state courts, including trial courts, have joined federal courts in asking more detailed questions and limiting uses of firearms expert testimony.

\textsuperscript{217} See \textit{State v. Montgomery}, No. 94CA40, 1996 Ohio App. LEXIS 1361, at *14 (Oh. Ct. App. Mar. 29, 1996) (“While it is true that other colleagues provided [the expert] with information . . . the major part of his opinion was based on his own observations and expertise.”).


\textsuperscript{219} \textit{State v. Riley}, 568 N.W.2d 518, 526 (Minn. 1997).

\textsuperscript{220} \textit{Kumho Tire Co. v. Carmichael}, 526 U.S. 137, 152–53 (1999) (setting out the application of \textit{Daubert} to expert testimony by nonscientists).

A turning point was the District of Massachusetts ruling in *United States v. Green*. Then-judge Gertner described that the firearms expert had planned to testify about individual characteristics which the expert stated could be matched “to the exclusion of every other firearm in the world.” At the opinion’s outset, the court stated that this conclusion was “extraordinary.” The court also gave one of the earliest detailed descriptions of the exactness—or lack thereof—of the toolmark comparison methodology:

In firearm toolmark comparisons, exact matches are rare. The examiner has to exercise his judgment as to which marks are unique to the weapon in question, and which are not. In fact, shell casings have myriad markings, some of which appear on all casings from the same type of weapon (“class characteristics”) or those manufactured at the same time (“sub-class characteristics”). Others are arguably unique to a given weapon (“individual characteristics”) or are unique to a single firing (“accidental characteristics”).

Judge Gertner then explained:

The task of telling them apart is not an easy one. Even if the marks on all of the casings are the *same*, this does not necessarily mean they came from the *same gun*. Similar marks could reflect class or sub-class characteristics, which would define large numbers of guns manufactured by a given company. Just because the marks on the casings are *different* does not mean that they came from *different guns*. Repeated firings from the same weapon, particularly over a long period of time, could produce different marks as a result of wear or simply by accident.

Judge Gertner emphasized that in “distinguishing class and sub-class characteristics from individual ones,” the examiner “conceded, over and over again, that he relied mainly on his subjective judgment. There were no reference materials of any specificity, no national or even local database on which he relied.” Despite these concerns, the court candidly acknowledged that “the problem for the defense is that *every single court* post-*Daubert* has admitted this testimony, sometimes without any searching review, much less a hearing.” Judge Gertner ultimately allowed the expert testimony because “any other decision [would] be rejected by appellate

---

223. *Id.* at 107.
224. *Id.*
225. *Id.*
226. *Id.*
227. *Id.*
228. *Id.*
courts, in light of precedents across the country.”

Nevertheless, the court did not “allow [the expert] to conclude that the match he found by dint of the specific methodology he used permits ‘the exclusion of all other guns’ as the source of the shell casings.”

In a second Massachusetts case, United States v. Monteiro, Judge Gertner next held—for the first time—that firearms comparison evidence was inadmissible on an as-applied challenge under Rule 702. Because of “the extensive documentary record,” the court held that the “underlying scientific principle behind firearm identification—that firearms transfer unique toolmarks to spent cartridge cases—is valid under Daubert.”

At the same time, Judge Gertner noted that the “process of deciding that a cartridge case was fired by a particular gun is based primarily on a visual inspection” that is “largely a subjective determination.” Because of this subjectivity, a testifying examiner must “follow the established standards for intellectual rigor in the toolmark identification field with respect to documentation of the reasons for concluding there is a match (including, where appropriate, diagrams, photographs or written descriptions), and peer review of the results by another trained examiner in the laboratory.” Ultimately, the court concluded that even though the methodology could be reliable and even though the examiner was qualified based on his training and experience, the expert’s opinion was inadmissible because the expert did not sufficiently comply with proper peer review and documentation requirements.

Other federal courts began to follow the approach of Judge Gertner. The Northern District of California in 2007 held that an expert could only testify to a “reasonable degree of certainty in the ballistics field.” But the court commented:

[It is important to note that—at least according to this record—there has never been a single documented decision in the United States where an incorrect firearms identification was used to convict a defendant. This is not to say that examiners do not make mistakes. The record demonstrates that examiners make mistakes even on proficiency tests. But, in view of

229. Id. at 109.
230. Id. at 124.
232. Id. at 355.
233. Id. (emphasis added).
234. Id.
235. Id. The Government, however, was allowed—without prejudice—to resubmit evidence of the test results that complied with the standards in the field. Id.
the thousands of criminal defendants who have had an incentive to challenge firearms examiners’ conclusions, it is significant that defendants cite no false-positive identification used against a criminal defendant in any American jurisdiction.\textsuperscript{237}

Other federal courts, however, instead continued to admit conclusions given with “100% degree[s] of certainty.”\textsuperscript{238} The next shift occurred after the scientific community produced substantial reports raising new reliability questions.


Over half of the rulings in our database occurred after 2009 when the National Academy of Sciences released a groundbreaking report concerning forensic evidence. To be sure, commercial legal databases may have a greater concentration of more recent appellate rulings. But one might have expected a similar outpouring of judicial rulings after the Daubert ruling in 1993—a fairly modern opinion. Instead, we observe change following an intervention by the scientific community over a decade and a half later.

During this time, the separate field of comparative bullet lead analyses—in which examiners claimed to use chemistry to identify unique elemental makeup of a bullet—was discredited and abandoned by the FBI after the NAS found it lacked any scientific foundation.\textsuperscript{239} The NAS is “a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare.”\textsuperscript{240} Indeed, even before the report, courts had begun to exclude such evidence.\textsuperscript{241} While it was a very different discipline, those developments may have raised further concerns in the judiciary regarding the work of firearms examiners.

In a 2008 report focused on the feasibility of a national ballistic imaging database, the NAS concluded that underlying assumptions of firearms comparisons were not yet validated.\textsuperscript{242} Furthermore, “a significant amount

\begin{itemize}
  \item \textsuperscript{237} Id. at *41.
  \item \textsuperscript{239} NAT’L Rsch. Council, Forensic Analysis: Weighing Bullet Lead Evidence 6 (2004).
  \item \textsuperscript{240} See 2008 NAS REPORT, supra note 40, at iii.
  \item \textsuperscript{241} See, e.g., Clemons v. State, 896 A.2d 1059, 1074–79 (Md. 2006); Ragland v. Commonwealth, 191 S.W.3d 569, 574–80 (Ky. 2006).
  \item \textsuperscript{242} 2008 NAS REPORT, supra note 40, at 3 (“The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.”).
\end{itemize}
of research” would need to be done to determine what characteristics might allow one to determine a probative connection between pieces of firearms evidence.  

In 2009, the NAS released its landmark report, *Strengthening Forensic Science in the United States*, after Congress directed NAS to undertake the study recognizing that substantial improvements were needed in the field of forensic science. The 2009 NAS Report contains a scientific assessment of a variety of forensic science disciplines along with recommendations for improvements in each discipline and to the forensic system as a whole. The Committee assembled by the NAS included prominent forensic scientists, research scientists, lawyers, and judges. The Report identified a wide range of methodological issues with the practices of forensic firearm and toolmark identification.

Although the NAS Report did acknowledge that class characteristics are helpful in narrowing the pool of firearms that may have fired a particular bullet or cartridge case, it recognized that firearm examiners necessarily go beyond class characteristics when making an identification. The Report noted that a “fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process” and that the AFTE methodology “does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.” The Report concluded that “[b]ecause not enough is known about the variabilities among individual tools and guns, [firearm examiners are] not able to specify how many points of similarity are necessary for a given level of confidence in the result.”

Building on this work by the NAS, the 2016 President’s Council of Advisors on Science and Technology (“PCAST”) published its 2016 report on the use of forensic science in criminal proceedings. The report was a response to President Obama’s question “whether there [we]re additional steps on the scientific side, [in addition to those identified in the 2009 NAS Report], that could help ensure the validity of forensic evidence used in the Nation’s legal system.” The advisory group consisted of “leading

243. *Id.; see also* United States v. Taylor, 663 F. Supp. 2d 1170, 1175 (D.N.M. 2009) (describing the scope of that report which focused on feasibility of a ballistics database but noting that the question “was inextricably intertwined with the question of ‘whether a particular set of toolmarks can be shown to come from one weapon to the exclusion of all others’”).
244. *See 2009 NAS REPORT, supra note 41, at xix.
245. *See id. at xix–xx.
246. *Id. at 155.
247. *Id.
248. *Id. at 154.
249. *PCAST REPORT, supra note 43, at x.*
scientists and engineers, appointed by the President to augment the science and technology advice available to him from inside the White House, and from cabinet departments and from other Federal agencies." The group focused on six feature-comparison methods including firearms-comparison evidence.

Consulting with forensic scientists, PCAST reviewed more than two thousand studies from various disciplines. The field had responded to the NAS reports by conducting new studies, and PCAST undertook a deep examination of them. As the NAS had done in its 2009 Report, PCAST asked whether each discipline met basic requirements for scientific validity, which consists of both “foundational validity”—whether the method can, in principle, be reliable—and “validity as applied”—whether the method has been reliably applied in practice.

To be foundationally valid, a method must have been subject to “empirical testing by multiple groups, under conditions appropriate to its intended use.” Specifically, “the procedures that comprise it must be shown, based on empirical studies, to be repeatable, reproducible, and accurate, at levels that have been measured and are appropriate to the intended application.” The studies must also provide “valid estimates of the method’s accuracy,” demonstrating how often an examiner is likely to draw the wrong conclusion even when applying the method correctly (that is, a scientifically valid error rate). As PCAST explained, “Without appropriate estimates of [the method’s] accuracy, an examiner’s statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact.

Ultimately, as described below, PCAST concluded that all but one of the existing studies did not use appropriate designs to truly test the ability of a firearm examiner to make accurate identifications. PCAST went on to conclude that “[b]ecause there has been only a single appropriately designed study, the current evidence falls short of the scientific criteria for foundational validity.” Much like the NAS report that preceded it, PCAST pointed to the necessity for additional, appropriately designed studies to test

250. Id. at iv.
251. Id. at 7.
252. Id. at 2.
253. Id. at 47–48, 56–58.
254. Id. at 5.
255. Id. at 47.
256. Id. at 5.
257. Id. at 6.
258. Id. at 111.
the validity of firearm examination.\textsuperscript{259}

1. Evaluation of the Scientific Studies

PCAST divided the firearms identification studies it reviewed into two different types: set-to-set studies and sample-to-sample studies. In a set-to-set study, examiners are given two sets of bullets and then asked to link the first set of bullets to the second set of bullets. In a sample-to-sample study, examiners are given two bullets to compare and are asked to judge whether the bullets were fired by the same gun or not. This process is then repeated for other test sets of bullets. PCAST concluded that “set-based studies are not appropriately-designed black-box studies from which one can obtain proper estimates of accuracy.”\textsuperscript{260}

The principal problem of set-to-set studies is that test takers can leverage the design to gain inferences about other comparisons, making the task totally unlike real-world comparison work.\textsuperscript{261} For example, if an examiner identifies a match between bullets one and two, and then determines that bullet one and bullet A match, then bullet two and bullet A must also be a match by implication. Thus, a test taker would get a correct response for linking unknown bullet two to known bullet A despite never directly comparing the bullets. PCAST noted that: “[t]he Director of the Defense Forensic Science Center analogized set-based studies to solving a ‘Sudoku’ puzzle, where initial answers can be used to help fill in subsequent answers.”\textsuperscript{262} Because of this, set-to-set studies typically yield errors rates of zero and very few inconclusive responses.\textsuperscript{263}

At the time PCAST conducted its analysis, there was only a single sample-to-sample study available for firearms identification. The unpublished study was conducted by researchers at the Ames Laboratory in Iowa.\textsuperscript{264} In this first Ames Lab study, 218 firearm examiners were mailed a test packet that contained cartridge cases to examine. Each test packet totaled 15 separate comparisons for the examiners to evaluate. Unbeknown to the participants, 10 of the comparisons were different-source comparisons, for

\begin{thebibliography}{9}
\bibitem{} 259. Id.
\bibitem{} 260. Id. at 106.
\bibitem{} 261. United States v. Cloud, 576 F. Supp. 3d 827, 842–43 (E.D. Wash. 2021) (“Such studies lack external validity, as examiners conducting real-world comparisons have neither the luxury of knowing a true match is somewhere in front of them nor of making process-of-elimination-type inferences to reach their conclusions.”).
\bibitem{} 262. PCAST REPORT, supra note 44, at 106.
\bibitem{} 263. Id.
\bibitem{} 264. DAVID P. BALDWIN, STANLEY J. BAJIC, MAX MORRIS & DANIEL ZAMZOW, A STUDY OF FALSE-POSITIVE AND FALSE-NEGATIVE ERROR RATES IN CARTRIDGE CASE COMPARISONS (2014) [hereinafter AMES I].
\end{thebibliography}
which the correct response was elimination, and 5 were same-source comparisons, for which the correct response was identification. Examiners were instructed to work alone on the test and to follow the AFTE protocol.

The study reported a 1.01% false positive error rate. What was not stated explicitly in the study is that 33.7% of the responses were deemed inconclusive—a pattern of results wildly at odds with the results from the set-to-set studies. PCAST noted that “the closed-set studies show a dramatically lower rate of inconclusive examinations and of false positives. With this unusual design, examiners succeed in answering all questions and achieve essentially perfect scores. In the more realistic open designs, these rates are much higher.”

One federal district court, in extensively discussing the PCAST report findings, noted, “Based on the above information, the court finds that the potential rate of error for matching ballistics evidence based on the AFTE Theory does not favor a finding of reliability at this time.” The court noted, however, that the FBI and the Ames Laboratory were “currently conducting a second black box study on the AFTE Theory.” That study was posted online in early 2021 (and subsequently removed from the Internet).

The FBI/Ames Laboratory study (hereinafter “Ames II”) utilized a design ambitious in size and scope. First, the study contained both cartridge case and bullet comparisons. The vast majority of previous firearms comparison studies examined only cartridge cases. Second, the study consisted of three rounds that attempted to measure accuracy (round one), repeatability (round two), and reproducibility (round three). Repeatability refers to “the ability of an examiner, when confronted with the exact same

265. Id. at 3.
266. Id. at 16. There were 2,180 different source comparisons of which 735 were inconclusive (735 / 2,180 = 33.7%). Compare that figure to a well-known set-to-set study by Hamby which reported only 8 inconclusives—0.1%—out of 7,605 comparisons. J.E. Hamby, David J. Brundage & James W. Thorpe, The Identification of Bullets Fired From 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries, 41 AFTE J. 99 (2009).
267. PCAST REPORT, supra note 44, at 110. PCAST was not the first group to point out the shortcomings of set-to-set studies. The Ames study, for example, stated, Several previous studies have been carried out to examine this and related issues of individualization and durability of marks [1-5], but the design of these previous studies, whether intended to measure error rates or not, did not include truly independent sample sets that would allow the unbiased determination of false-positive or false-negative error rates from the data in those studies.
269. Id. at 779.
comparison once again, to reach the same determination as when first examined.”

Reproducibility refers to “the ability of a second examiner to evaluate a set previously viewed by a different examiner and reach the same conclusion.” No other study had attempted to measure repeatability and reproducibility of firearm examiner judgments.

In round one of the study, 256 active firearm examiners were sent test packets—each test packet contained 15 comparison sets of bullets and 15 comparison sets of cartridge cases. For each comparison, participants were instructed to make a judgement according to the AFTE Range of Conclusions. Participants were admonished not to discuss their results with anyone else. However, only 173 participants out of 256 returned their test packets. According to the authors, “the overall rate of false positive error rate was estimated as 0.656% and 0.933% for bullets and cartridge cases, respectively, while the rate of false‐negatives was estimated as 2.87% and 1.87% for bullets and cartridge cases, respectively.”

Here again, there was an enormous amount of inconclusive responses: over 50% of the bullet comparisons were deemed inconclusive, and over 42% of the cartridge comparisons were deemed inconclusive.

In round two of the study, participants were sent the same test packet they examined previously. Only 105 participants completed this round. The percentage of time that examiners reached the same conclusion in round one and round two ranged from 79% to 62%. This does not necessarily mean the examiner reached the correct conclusion about two-thirds of the time; rather, it only suggests she reached the same conclusion about two-thirds of the time. According to the authors, a statistical test comparing the “observed agreement” between conclusions reached in round one and in round two to the “expected agreement” “indicat[ed] ‘better than chance’ repeatability.” However, two different statisticians concluded that: “[t]he level of repeatability and reproducibility as measured by the between rounds consistency of conclusions would not appear to support the reliability of

272. Id. at 11.
273. The Range of Conclusions includes the following options: (1) Identification, (2a) Inconclusive-A, (2b) Inconclusive-B, (2c) Inconclusive-C, (3) Elimination, and (4) Unsuitable. See AMES I, supra note 264, at 7.
274. AMES II, supra note 276, at 2.
275. Id. at 35.
276. Id. at 39.
277. Id. at 39.
278. Id. at 45.
firearms examination.”

Only 80 participants completed round three of the study. The percentage of time that 2 different participants examined the same test set and reached the same conclusion ranged from 68% to 31%. These latter results are striking. Less than one-third of the time, 2 different participants looked at the same bullets and reached the same conclusion. This means that over two-thirds of the time (69.1%), 2 different participants reached different conclusions when examining the same set of bullets. A statistical test revealed “better than chance” agreement for same-source bullet comparisons but not different-source bullet comparisons.

D. Litigating the Error Rate Studies

The conclusion reached by PCAST that “firearms analysis currently falls short of the criteria for foundational validity” did not go unnoticed by the defense bar. Admissibility challenges to firearm examiner testimony surged—we include more than eighty such cases in our database. These challenges often summarized the PCAST analyses and conclusions in arguing that the field failed to pass Daubert’s muster. These challenges, however, almost universally failed. Critics of PCAST sought to characterize the report as authored by outsiders who failed to learn the fundamentals of firearm examination and who committed numerous errors in their own analysis. But the tides have recently begun to shift, as courts imposed new,
albeit still limited, restrictions on the type of testimony firearm examiners may offer and how they express conclusions. We have identified thirty-seven judicial rulings imposing limitations on firearms comparison testimony and set out each in Appendix A.

Two factors have contributed to the shifting tides. First, in addition to citing the NAS and PCAST reports, attorneys have called mainstream research scientists to testify generally about scientific methods and principles and specifically about the discipline of firearm examination. These experts are not firearm examiners and typically have never conducted a firearm examination. Much like the practitioner/researcher distinction in medicine, these experts are researchers who study whether the methods employed by the practitioners are effective. These experts are poised to evaluate claims made in court regarding scientific practices.

The second major factor concerns additional examination of the PCAST-reviewed studies that potentially undermine the reported error rates and the utility of the validation studies. As noted, one-third of the responses in the Ames I study were inconclusive.

What ought to be done with those responses? PCAST ultimately calculated the error rate without considering them. Other firearm studies actually count inconclusive responses as correct responses, based on the logic that “an inconclusive response is not an incorrect response [so they are] totaled with the correct response and figured into the error rate as such.” But what if those responses are errors? The error rate would be as high as 35% in the Ames I study. Other sample-to-sample studies conducted after the PCAST analyses have reported rates of inconclusive responses over 50%. Clearly, determining how to count over half of the responses in a validation study is critical.

There are many legitimate reasons to count the inconclusive responses in the Ames I study, including the fact that “[t]he fraction of samples reported as inconclusive cannot be attributed to a large fraction of poorly marked

---

286. See, e.g., Tibbs, 2019 D.C. Super. LEXIS 9; Ross, 129 N.Y.S.3d 629.
287. See generally, e.g., Faigman et al., supra note 45.
288. For example, judges are supposed to consider whether research appears in a “peer-reviewed” scientific journal. See supra notes 189–192 and accompanying text. Most research on firearm examination is published in the AFTE Journal which is touted in court as a “peer-reviewed scientific” journal. See AFTE J., https://afte.org/afte-journal. Upon closer inspection, however, the peer-review process used by the AFTE Journal is highly dissimilar to the usual process that occurs at scientific journals. See Tibbs, 2019 D.C. Super. LEXIS 9, at *25.
289. See supra note 266 and accompanying text.
291. See AMES II, supra note 271, at 35.
knowns or questioned samples in this group\textsuperscript{292} and an inconclusive response is also defined by AFTE as an absence of insufficient quality of marking to reach an identification or elimination.\textsuperscript{293} As noted in a 2020 scientific article, a proper study design would include inconclusive test items so that inconclusive responses could be evaluated and incorporated into the error rate.\textsuperscript{294} No study has yet done so and, as a result, error rates observed in the studies span a range so large as to be wholly unhelpful—anywhere from one percent to over fifty percent, depending on whether the responses are dropped or considered as erroneous. Thus, as one district court recently put it,

But providing examiners in the study setting the option to essentially “pass” on a question, when the reality is that there is a correct answer—the casing either was or was not fired from the reference firearm—fundamentally undermines the study’s analysis of the methodology’s foundational validity and that of the error rate.\textsuperscript{295}

This crucial issue of inconclusive responses was never considered prior to \textit{Tibbs}, discussed earlier in this Article,\textsuperscript{296} in which a defense expert raised the concern during an admissibility hearing. The judge in \textit{Tibbs} called it “perhaps [the] most substantial issue related to the studies proffered to support the reliability of firearms and toolmark analysis”\textsuperscript{297} and noted that “the methods used in the proffered laboratory studies make a compelling case that inconclusive should not be accepted as a correct answer in these studies.”\textsuperscript{298} To be sure, in one 2020 Washington, D.C. case, a judge discounted those findings for which no defense expert was presented to explain these error rate issues.\textsuperscript{299}

Then again, another 2020 case in Oregon limited the admissibility of firearms testimony without the benefit of a defense expert witness.\textsuperscript{300} This judge expressed major concerns about inconclusive responses in firearms comparison studies and their impact on reported error rates:

\textsuperscript{292}. Ames I, supra note 264, at 19.
\textsuperscript{297}. Id. at 56–57.
\textsuperscript{298}. Id. at 57–58.
It appears to be the case that the only way to do poorly on a test of the AFTE method is to record a false positive. There seems to be no real negative consequence for reaching an answer of inconclusive. Since the test takers know this, and know they are being tested, it at least incentivizes a rate of false positives that is lower than real world results. This may mean the error rate is lower from testing than in real world examinations.\footnote{301}

A litany of other concerns besides the inconclusive response issue have been raised about the error rate studies. We mention four important issues here.

First, and most fundamentally, none of the studies were test-blind—the participants knew that they were being tested. There is powerful evidence that human subjects are predictably biased—and behave differently—when they know that they are being tested. The PCAST report emphasized the need for blind testing of forensic techniques.\footnote{302} So have a host of researchers based on a large body of research documenting the manner in which cognitive biases can lead forensic examiners to make errors.\footnote{303} Although blind testing is standard in medicine, it has never been standard in error rate studies in forensics.

Second, many of the volunteer participants in both of the Ames studies simply dropped out or participated but did not complete the test. In the Ames II study, “32% of the 256 examiners receiving their first packets failed to report any results, and another 32% of the 256 dropped out before completing all six mailings.”\footnote{304} No analysis of the participants who initiated the study but declined to complete it was conducted.\footnote{305} Attrition bias due to nonrandom dropout is a serious concern that has an unknown impact on the reported error rates. Although one court has noted that the “use of volunteers . . . does not provide the clearest indication of the accuracy of the conclusions that would be reached by average toolmark examiners,”\footnote{306} courts have not focused on issues related to participant dropout.

Third, there are also questions about whether the materials being used in the studies, such as the types of firearms and the quality of the fired items, are sufficiently representative to draw inferences about the field writ large.

\footnotesize
\begin{itemize}
\item \footnote{301}{Id. at 1265.}
\item \footnote{302}{PCAST REPORT, supra note 43, at 58–59.}
\item \footnote{303}{See generally, e.g., Itiel E. Dror, Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias, 92 ANALYTICAL CHEMISTRY 7998 (2020).}
\item \footnote{304}{Alan H. Dorfman & Richard Valliant, Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives, 5 FORENSIC SCI. INT’.L: SYNERGY 1, 5 (2022).}
\item \footnote{305}{Id.}
\end{itemize}
By design, studies should be of varying degrees of difficulty, but unfortunately, “[w]ith a few exceptions, each of the forensic firearms studies to date focuses on a single firearm,” and the exceptions are telling, whereas studies that use different types of firearms have resulted in very different error rates for each type.³⁰⁷ Further, if in a study, “an examiner is over and over comparing bullets or cartridge cases from the same brand and model, then he or she can be expected to be picking up nuances along the way. A later comparison will have an advantage over the first. We can expect this to lead to a reduction in sample error rates.”³⁰⁸ Unlike other forensic identification fields, none of these studies have used technology or databases to ensure the test items are challenging.³⁰⁹ Nor has there been any careful analysis of how representative or challenging these studies are, and this basic problem has not received the judicial attention that it should.

Finally, judges have not focused on the appalling levels of nonrepeatability and nonreproducibility of firearms work in the Ames II study: “[E]xaminers examining the same material twice, disagree[d] with themselves between 20% and 40% of the time.”³¹⁰ They disagreed with other examiners even more, up to 69% of the time for nonmatching bullets and up to 60% of the time for nonmatching cartridges.³¹¹ Although there is spirited debate about inconclusive results and whether or not they constitute errors in a study, these rates of intra- and interparticipant consistency should eclipse that entire discourse—they set a limit on validity and cannot be dismissed as a disagreement about the interpretation of inconclusive responses. Yet, likely because Daubert explicitly mentions error rates, not rates of consistency, courts have yet to grapple with these findings and how they can be reconciled with professed error rates of one percent or less.

All of this said, it is not uncommon for judges to respond to these studies, the PCAST report, and critiques from research scientists dismissively. Judges have commonly relied on precedent to make conflated arguments against the invalidity of the studies. For example, one judge in New York state—a Frye jurisdiction where the standard for expert evidence admissibility is the “general acceptance” of the method within the relevant

---

³⁰⁷. See Dorfman & Valliant, supra note 304, at 5 (“The few studies that have carried out comparisons over a variety of guns have displayed marked differences in the case of coming to correct conclusions.”).
³⁰⁸. Id.
³⁰⁹. Nicholas Scurich, Inconclusives in Firearm Error Rate Studies are Not “a Pass,” L. PROBABILITY & RISK (2022) ("[R]esearchers should intentionally select challenging test items, in a manner similar to Professor Koehler’s exemplary fingerprint examiner study involving ‘close nonmatches.’").
³¹¹. Dorfman & Valliant, supra note 304, at 6.
scientific community—recently emphasized that the acceptance of firearms comparison methods within the community of practitioners is “nearly universal”312 and that “the Appellate Division . . . has repeatedly upheld the admission of ballistics expert testimony without the need for a Frye hearing.”313 But the judge then went on to hold that “the PCAST report has been thoroughly discredited”314 and the “the very type of study called for by PCAST—a ‘black box study’—has, since the time of the PCAST report, been repeatedly utilized to validate firearm and toolmark comparison methodology.”315 There was no engagement with the results of those studies or their limitations. Unfortunately, it is common for judges to rely on precedent as a form of “general acceptance” by the courts and not carefully examine the reliability of scientific evidence.316

E. TESTIMONIAL LIMITATIONS AND POST-NAS AND PCAST RULINGS

In recent years, courts have more rigorously evaluated the field of firearms examination, in contrast to over fifty years in which claims made by firearm examiners regarding the foundational validity were uncritically accepted.317 These more searching evaluations have led judges to note limitations and knowledge gaps that had rarely been discussed in judicial opinions. Despite increasing awareness of the limitations of the field, almost all courts have nevertheless found it admissible.318 This created a new conundrum for courts: how to admit firearms identification evidence in a

312. State v. Vasquez, No. 2203/2019, at 3 (N.Y. Sup. Ct., July 24, 2022). According to this judge, the relevant scientific community is not “experts in ‘scientific methodology,’ which is to say, scientists,” id. at 2, but rather “trained and accredited experts in the field of microscopic ballistics and forensic firearm and toolmark examination” as well as “non-firearm practitioners enumerated in the multiple validation studies that have been conducted to demonstrate the reliability of the discipline and its examination results,” id. at 3 (emphasis added). Conducting a study to demonstrate a result is not good science.

313. Id. at 5.

314. Id. at 4.

315. Id. at 5.

316. Stephanie L. Damon-Moore, Trial Judges and the Forensic Science Problem, 92 N.Y.U. L. REV. 1532, 1564 (2017) (“Ironically, the ultimate safeguard against judicial error—appellate review—may actually discourage judges from gatekeeping effectively.”).

317. See, e.g., United States v. Shipp, 422 F. Supp. 3d 762, 775 (E.D.N.Y. 2019) (“Even though prior decisions have found toolmark analysis to be reliable, it is incumbent upon this court to thoroughly review the critiques of the AFTE Theory found in the NRC and PCAST Reports.”); United States v. Adams, 444 F. Supp. 3d 1248, 1266 (D. Or. 2020) (concluding that it could not “find that the AFTE method enjoys ‘general acceptance’ in the scientific community”); People v. Ross, 129 N.Y.S.3d 629, 641 (N.Y. Sup. Ct. 2020) (“By beyond comparing class characteristics forensic toolmark practice lacks adequate scientific underpinning and the confidence of the scientific community as whole.”).

way that does not overstate its value or cause the fact finder to be misled. In the Sections that follow, we report the four tacks that courts have taken when admitting firearm examination evidence: (1) limiting the language that experts can use when testifying to their conclusions, (2) limiting conclusions to class characteristics only, (3) ruling that evidence concerning the proficiency of firearms experts is relevant to the preliminary question whether to qualify the expert, and (4) examining the as-applied question whether the method was reliably used in the particular case.

1. Limiting Conclusion Testimony

While many courts have continued to admit firearms examiner testimony, “[m]any of these courts admitted the proffered testimony only under limiting instruction restricting the degree of certainty to which firearm and toolmark identification specialists may express their identifications.” The case law that has resulted is diverse, sometimes inconsistent, and reflects a gradual evolution of judicial approaches. As we will describe, in general, a range of courts have limited testimony based on the concerns about toolmark identification methodology.

The earlier decisions had held that an examiner could only testify to a milder degree, forbidding aggressive statements of a match, “the exclusion of all other firearms in the world,” and instead imposing a more cautious formulation, such as a “reasonable degree of ballistic certainty.” Other courts have taken a different approach, using more familiar standards of proof as a frame of reference—courts have ruled that the examiner can only opine that it is “more likely than not” that the bullet recovered from the crime scene came from the defendant’s firearm. The table below summarizes some of the main approaches that courts have taken toward limiting such testimonial conclusions. Appendix A summarizes all thirty-seven opinions.

320. See, e.g., Shipp, 422 F. Supp. 3d at 783 (preventing a toolmark expert from testifying “to any degree of certainty, that the recovered firearm is the source of the recovered bullet fragment or the recovered shell casing”); Adams, 444 F. Supp. 3d at 1266–67 (same); United States v. Monteiro, 407 F. Supp. 2d 351, 373 (D. Mass. 2006) (same); Davis, 2019 U.S. Dist. LEXIS 155037, at *24 (“[W]itnesses may not testify as to a ‘match,’ that the cartridges bear the same ‘signature,’ that they were fired by the same gun, or words to that effect.”); United States v. Glynn, 578 F. Supp. 2d 567, 575 (S.D.N.Y. 2008) (limiting testimony to “be stated in terms of ‘more likely than not,’ but nothing more”).
323. See Glynn, 578 F. Supp. 2d at 574–75 (limiting testimony to “more likely than not” conclusion).
that we have located, through 2022, including unpublished trial court rulings.

TABLE 1. Testimonial Limitations on Firearms Examiners

<table>
<thead>
<tr>
<th>Court-ordered Conclusion Language</th>
<th>Citations from selected examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>“qualitative opinions” can only be offered on the significance of “class characteristics”</td>
<td>People v. Ross, 129 N.Y.S.3d 629 (N.Y. Sup. Ct. 2020)</td>
</tr>
</tbody>
</table>

The approach toward firearms testimony has evolved over the past two decades. The consensus approach, early on, was shared by a series of courts that adopted the formulation, “a reasonable degree of ballistic certainty.” Thus, the court in *Diaz* allowed the examiner to testify “that cartridge cases or bullets were fired from a particular firearm ‘to a reasonable degree of ballistic certainty,’ ” as did a series of other federal courts. In *Monteiro*, the district court ruled that the examiner could testify that the “class characteristics were in complete agreement,” but aside from observing that consistency, to a “reasonable degree of ballistic certainty,” no further probabilistic statement could be offered. The court reasoned, “Allowing

the firearms examiner to testify to a reasonable degree of ballistic certainty permits the expert to offer her findings, but does not allow her to say more than is currently justified by the prevailing methodology.\textsuperscript{327}

It is not clear what a reasonable degree of certainty consists of—as a result, the U.S. Department of Justice has barred examiners in federal cases from using that or similar terminology.\textsuperscript{328}

An examiner shall not assert that two toolmarks originated from the same source with absolute or 100% certainty, or 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.\textsuperscript{329}

The Department also barred examiners from making assertions of a “zero error rate” or infallibility.\textsuperscript{330} Those requirements marked a real charge from prior practice.

Second, during this time, some judges, like the Department of Justice itself, began to focus on the probabilistic claims by experts and limited toolmark experts’ testimony about conclusions that claim infallibility or the lack of any error rate—courts rejected assertions of zero error rates.\textsuperscript{331} Thus, courts rejected assertions of being “100% sure” or “certain.”\textsuperscript{332} In \textit{Monteiro}, the judge rejected the use of the phrase “a match to an exact statistical certainty.”\textsuperscript{333} Similarly, in \textit{United States v. Gardner}, the judge held that the opinion could not be made with “unqualified” certainty.\textsuperscript{334}

A growing group of judges then offered intermediate approaches. Another District of Columbia judge held that an expert can testify that ammunition is “consistent with” being fired from the same firearm.\textsuperscript{335} The

\textsuperscript{327} \textit{Id.} at 372.

\textsuperscript{328} \textit{U.S. DEP’T OF JUST., UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS FOR THE FIREARMS/TOOLMARK DISCIPLINE PATTERN ANALYSIS} 3 (2020).

\textsuperscript{329} \textit{Id.} at 3.

\textsuperscript{330} \textit{Id.}

\textsuperscript{331} See, e.g., \textit{United States v. Romero-Lobato}, 379 F. Supp. 3d 1111, 1117 (D. Nev. 2019) (acknowledging that the “general consensus” of the courts “is that firearm examiners should not testify that their conclusions are infallible or not subject to any rate of error, nor should they arbitrarily give a statistical probability for the accuracy of their conclusions”); \textit{State v. Terrell}, No. CR170179563, 2019 Conn. Super. LEXIS 827, at *3 (Conn. Super. Ct. Mar. 21, 2019) (same); \textit{United States v. Glynn}, 578 F. Supp. 2d 567, 574 (S.D.N.Y. 2008) (limiting testimony in part because when experts “make assertions that their matches are certain beyond all doubt, that the error rate of their methodology is ‘zero,’” there is a risk of “giving the jury the impression . . . that [the methodology] has greater reliability than its imperfect methodology permits.”).

\textsuperscript{332} \textit{United States v. Parker}, 871 F.3d 590, 600 (8th Cir. 2017).


\textsuperscript{335} \textit{United States v. Sutton}, No. 2018 CF1 009709, at *5 (D.C. Super. Ct. May 9, 2022) (permitting the examiner to opine “that the ammunition at issue is consistent with being fired from the same firearm”).
district court in *United States v. Shipp* ordered that the expert “may not testify, to any degree of certainty, that the recovered firearm is the source of the recovered bullet fragment or the recovered shell casing.”\textsuperscript{336} That court carefully examined the findings of the PCAST Report, and while it did not permit characterization of the level of certainty, the examiner could offer a statement of consistency.\textsuperscript{337} Other courts have taken this approach.\textsuperscript{338}

Going further to limit the testimony, in more recent cases, judges have barred any certainty-based statements at all. Thus, in the *Tibbs* ruling, the court held that the examiner could not offer any probability that the firearm in question could be included, but only that “the recovered firearm cannot be eliminated as the source of the cartridge casing found on the scene of the alleged shooting.”\textsuperscript{339} In the *Goodwin-Bey*\textsuperscript{340} ruling, the trial court did the same.\textsuperscript{341} In *United States v. Willock*, the district judge ordered “a complete restriction on the characterization of certainty.”\textsuperscript{342} Other courts have taken the same approach.\textsuperscript{343} Still other cases permitted the examiner to point to features and their similarities but not describe any level of agreement or consistency.\textsuperscript{344}

None of these approaches adopt the approach of the American Statistical Association, which explains that to assert any degree of probability of an event, an established statistical basis must exist for that asserted degree of probability.\textsuperscript{345} Under this approach, an expert must be clear that no such statistical basis exists if none does exist.

\footnotesize{336. United States v. Shipp, 422 F. Supp. 3d 762, 783 (E.D.N.Y. 2019).}
\footnotesize{337. Id. at 778.}
\footnotesize{340. State v. Goodwin-Bey, No. 1531-CR00555-01, slip op. at 7 (Mo. Cir. Ct. Dec. 16, 2016).}
\footnotesize{341. Id. (limiting testimony "to the point this gun could not be eliminated as the source of the bullet").}
\footnotesize{343. See United States v. White, No. 17 Cr. 611, 2018 U.S. Dist. LEXIS 163258, at *3 (precluding expert from testifying "to any specific degree of certainty as to his conclusion that there is a ballistics match").}
\footnotesize{344. See, e.g., United States v. Green, 405 F. Supp. 2d 104, 124 (D. Mass. 2005); People v. Ross, 129 N.Y.S.3d 629, 642 (N.Y. Sup. Ct. 2020) ("The People may call an expert to testify as to whether there is evidence of class characteristics that would include or exclude the firearm at issue...[T]he examiner may not opine on the significance of any marks other than class characteristics, as the reliability of that practice in the relevant scientific community as a whole has not been established. Moreover, any opinion based in unproven science and expressed in subjective terms such as ‘sufficient agreement’ or ‘consistent with’ may mislead the jury and will not be permitted.").}
2024] JUDGING FIREARMS EVIDENCE 151

The opinion that has gone farthest of all, however, is one of the most recent, a so-far unpublished opinion in 2023 by a trial judge in Cook County, Illinois. As noted, the judge wholly excluded firearms expert testimony, based on a review of scientific concerns with reliability. 346

2. Limiting Non-Class-Based Opinions

Some jurisdictions under both Daubert and Frye have limited testimony to opinions offered on class characteristics only. 347 That is, an expert can explain that the same type of gun fired the bullets or cartridge cases, but the expert cannot say that the same gun fired the bullets or cartridge cases. For example, here is the limiting instruction given by one federal judge who restricted the testimony to class characteristics:

[Firearm examiner’s] expert testimony is limited to the following observational evidence: (1) the Taurus pistol recovered in the crawlspace of [defendant’s] home is a 40 caliber, semi-automatic pistol with a hemispheric-tipped firing pin, barrel with six lands/grooves and right twist; (2) that the casings test fired from the Taurus showed 40 caliber, hemispheric firing pin impression; (3) the casings seized from outside the shooting scene were 40 caliber, with hemispheric firing pin impressions; and (4) the bullet recovered from gold Oldsmobile at the scene of the shooting were 40/10mm caliber, with six lands/groves and a right twist. 348

Courts have reasoned that descriptions of class characteristics are objective and measurable, whereas linking bullets to a particular gun is not “the product of a scientific inquiry,” 349 and “any opinion based in unproven science and expressed in subjective terms such as ‘sufficient agreement’ or ‘consistent with’ may mislead the jury and will not be permitted.” 350

3. Qualification and Proficiency Rulings

Judges have also focused on the proficiency of the particular expert to answer the preliminary question of whether a person is qualified to be an expert under Rule 702. Rule 702 requires that an expert witness have sufficient “knowledge, skill, experience, training, or education.” 351

347. See, e.g., United States v. Adams, 444 F. Supp. 3d 1248, 1267 (D. Or. 2020); Ross, 129 N.Y.S.3d at 642 (“The People may proffer their NYPD ballistics detective as an expert in firearm and toolmark examination for the testimony on class characteristics as described above.”).
348. Adams, 444 F. Supp. 3d. at 1267.
349. Id. at 1266.
351. FED. R. EVID. 702 (requiring that an expert be “qualified as an expert by knowledge, skill, experience, training, or education”).
Typically, proficiency tests are administered by commercial test providers in which accredited labs are required to administer such tests annually. For example, one leading provider, Collaborative Testing Services (“CTS”), makes available on its website the results of its tests for each discipline. CTS has cautioned that no “error rate” can be generalized from such tests because they are designed to be elementary. Such tests are not proctored, can be taken in groups, have no time limit, include materials that are of unknown realism and difficulty, and are not “blind,” since participants know that it is a test. However, the results do highlight the types of errors that practitioners may make. For example, a 2022 test included 7 participants, or 2% of the examiners, that failed to correctly identify the bullet that the known firearm had in fact fired in a test with a very small number of items; far higher numbers of examiners reported inconclusive responses which were also not accurate (but which CTS noted may follow lab practices).

In United States v. Cloud, the judge emphasized that one of the two examiners in the case had failed a proficiency test and was allowed to return to work after a second proficiency test, in which the examiner had to do an “in-depth consultation” with a supervisor. The court found that it could not “in good conscience qualify [the examiner] as an expert with the requisite skill to perform fingerprint comparisons when her two most recent proficiency exams either contained an error or required a significant amount of assistance from her supervisor” and further, the finding was bolstered by the portions of “testimony and performance reviews that touch on her skill, willingness to take correction, and confidence performing her work.” In the Willock case, the examiner’s “qualifications, proficiency and adherence to proper methods [we]re unknown.”

353. COLLABORATIVE TESTING SERVS., INC., CTS STATEMENT ON THE USE OF PROFICIENCY TESTING DATA FOR ERROR RATE DETERMINATIONS 3 (2010).
355. See COLLABORATIVE TESTING SERVS., INC., FIREARMS EXAMINATION TEST NO. 22-5261 SUMMARY REPORT 3 (2022). CTS also noted that inconclusive responses were not counted as “outlier[]” errors, as “CTS is aware that many labs will not, as a matter of policy, report an elimination without access to the firearm or when class characteristics match.” Id.
357. Id.
Many courts traditionally focused on an expert’s credentials and self-professed expertise when conducting this inquiry into the qualifications of the witness. However, as one of the authors and Gregory Mitchell have argued, a careful inquiry into objective proficiency of the witness should be an integral part of the question whether a person should be qualified as an expert. Indeed, such proficiency issues can raise larger red flags concerning the reliability of a crime lab unit and not just an individual examiner. Years before the Metropolitan Crime Lab had its accreditation revoked, as described in our introduction, a firearms examiner had failed a proficiency test after two colleagues had verified the work, implicating their own proficiency as well. Perhaps more careful attention to those proficiency tests could have prevented subsequent errors and systems failures of the firearms unit and the entire laboratory.

4. As-Applied Challenges

Still additional challenges have focused on Rule 702(d), which used to provide that qualified expert testimony is admissible only when “the expert has reliably applied the principles and methods to the facts of the case.” These “as applied” challenges focus on the work that an expert does and not just whether they followed the right steps, but also whether their casework was actually supported by a valid method. Thus, some challenges have focused on, for example, the lack of documentation by firearms experts and the way they used their methods in a particular case. Some courts have found the presence of some documentation, such as “notes, worksheets, and photographs,” to be sufficient.

360. See id. at 940–49. Other courts have cited to the existence of proficiency testing as evidence of reliability, which as Garrett & Mitchell discuss, is not well supported. See, e.g., United States v. Johnson, No. (S5) 16 CR. 281 (PGG), 2019 U.S. Dist. LEXIS 39590, at *46 (S.D.N.Y. Mar. 11, 2019), aff’d, 861 F. App’x 483 (2d Cir. 2021) (“While these proficiency tests do not validate the underlying assumption of uniqueness upon which the AFTE theory rests, they do provide a mechanism by which to test examiners’ ability—employing the AFTE method—to accurately determine whether bullets and cartridge casings have been fired from a particular weapon.”).
362. FED. R. EVID. 702(d).
364. For a case rejecting an as-applied challenge because the expert would not testify that a bullet came from a specific firearm, see United States v. Tucker, 18 CR 0119 (SJ), 2020 U.S. Dist. LEXIS 3055, at *3 (E.D.N.Y. Jan. 8, 2020).
III. LESSONS FROM THE PATH OF FIREARMS EVIDENCE

The arc of judicial review of firearms evidence follows a pattern that is familiar in forensics more generally. Early judicial skepticism of a novel technique was overcome by claims of expertise relying on new technology (a microscope at the time), forceful claims to expertise by aggressive personalities (chiefly Major Goddard), some highly useful applications of the technique (to simply measure class characteristics), and steady accumulation of precedent. Then, as scientific critiques and evidence of error rates mounted, judges began to express some skepticism which has substantially increased in recent decades, producing a large body of law limiting firearms evidence in a range of ways.

That said, we underscore that other courts have not sought to introduce evidence concerning limitations of firearms evidence, much less imposed limitations. An appellate court in Missouri, for example, found no error in a judge’s refusal to allow defense attorneys to cross-examine the firearms expert concerning the findings of the NAS and PCAST reports. Further, even in recent years, “many courts have continued to allow unfettered testimony from firearm examiners who have utilized the AFTE method.”

The community of firearm examiners has mounted aggressive defenses of their work. In one memorable critique of how scientists and judges have raised questions concerning firearms comparison work, general counsel for the FBI wrote, “It is a lamentable day for science and the law when people in black robes attempt to substitute their opinions for those who wear white lab coats.” And yet it has been scientists—not judges—who have raised the deepest concerns about firearm examination. Statisticians, for example, criticize firearms comparison methods as having been “developed by insular communities of nonscientist practitioners” who, as a result, “did not incorporate effective statistical methods.” As one litigator colorfully

366. State v. Mills, 623 S.W.3d 717, 729–31 (Mo. Ct. App. 2021), transfer denied (June 29, 2021) (“The trial court excluded the reports and their contents but did not deny defense counsel from asking questions about the flaws in toolmark and firearm examination as Appellant argues.”).


wrote in a *Daubert* brief, “Astrologers believe in the legitimacy of astrology. . . . And toolmark analysts believe in the reliability of firearms identification; their livelihoods depend on it.”\(^{370}\)

The response to these scientific critiques has been to call them “flawed”\(^{371}\) and double down on the claim that error rates are extraordinarily low. The FBI, for example, asserted in a 2022 case that there is an error rate of “1%.”\(^{372}\) Federal prosecutors have repeatedly argued that “[f]irearms and toolmark identification meets all the *Daubert* criteria. Accordingly, there is no scientific or legal basis to exclude this evidence or even limit it.”\(^{373}\)

Indeed, then-Attorney General Loretta Lynch more broadly responded to the PCAST report, upon its release, as not affecting the work of the Department of Justice: “We remain confident that, when used properly, forensic science evidence helps juries identify the guilty and clear the innocent. . . . While we appreciate their contribution to the field of scientific inquiry, the department will not be adopting the recommendations related to the admissibility of forensic science evidence.”\(^{374}\)

Some reactions in the field have been less defensive. Apparently in response to criticism by Judge Edelman, AFTE has opened its publications to outside viewing—one judge “applauds the publication’s changes and encourages AFTE and similar organizations to continue to open their publications up for criticism and review from the larger scientific community if they wish to meet *Daubert*’s rigorous standard.”\(^{375}\) However, the judge nevertheless found that the quality of the studies did not provide strong support for admissibility under *Daubert*.\(^{376}\)

One response by judges has been, as described, to limit the verbal formulations that firearms experts use when reaching conclusions. There are reasons to doubt that this compromise solution has been effective in


\(^{371}\) See Agar, *supra* note 368, at 166 (“Accreditation, widespread proficiency testing, the success of ATF’s NIBIN database, the Commerce Department’s recognition of firearms identification, and the reliance of the U.S. government on firearms identification to investigate and solve the assassination of a U.S. president serve as cornerstones for the ‘general acceptance’ of the firearms identification discipline.”).


\(^{373}\) Gov’t’s Response to Defendant’s Motion in Limine to Exclude Ballistics Evidence, or Alternatively, for a *Daubert* Hearing at 23, United States v. Hunt, No. 5:19-cr-00073-R, 2020 WL 3549386 (W.D. Okla. April 27, 2020).


\(^{375}\) Cloud, 576 F. Supp. 3d at 842.

\(^{376}\) Id.
communicating to jurors the limitations of firearms evidence. Two of us collaborated on a mock jury study examining how laypersons evaluate different firearms expert conclusions.\textsuperscript{377} None of the limitations on firearms testimony adopted by courts, such as reasonable scientific certainty or more likely than not, had any impact on conviction rates except for the most far-reaching language, imposed in \textit{Tibbs}, that barred any conclusion linking the firearms in question but rather permitting only a statement that a firearm cannot be excluded.\textsuperscript{378}

To be sure, the more recent rulings that permit only testimony concerning class characteristics go further than ruling out any language of inclusion. They limit the expert to testimony concerning objective measurements (for example, the width of the cartridge or bullet) and prevent more speculative testimony concerning probabilities that something came from a particular firearm. These rulings return firearms comparison to its roots: measuring objects. This can be useful and provide valuable information.

We have not seen judges take the approach to reliability, which is codified in Rule 702, that PCAST did, for example, insisting that “[t]he only way to establish the scientific validity and degree of reliability of a subjective forensic feature-comparison method—that is, one involving significant human judgment—is to test it \textit{empirically} by seeing how often examiners actually get the right answer.”\textsuperscript{379}

In fact, some judges have expressly rejected this approach, stating that PCAST’s requirement of empirical study “goes beyond what is required by Rule 702.”\textsuperscript{380} However, there are strong reasons to think that jurors will benefit from more information regarding error rates and the reliability of the firearms comparison method, just as PCAST recommends and as mock jury experts have found productive—even just the bare acknowledgement that errors occur can impact jurors who assume that these experts are infallible unless told otherwise.\textsuperscript{381}

\textsuperscript{377} See Garrett et al., \textit{supra} note 8.
\textsuperscript{378} \textit{Id.}
\textsuperscript{379} AN ADDENDUM TO THE PCAST REPORT ON FORENSIC SCIENCE IN CRIMINAL COURTS 1 (Jan. 6, 2017).
\textsuperscript{380} United States v. Harris, 502 F. Supp. 3d 28, 38 (D.D.C. 2020); \textit{see also} United States v. Hunt, 464 F. Supp. 3d 1252, 1258 (W.D. Okla. 2020) (“[T]he Court declines Defendant’s invitation to restrict judicial review to techniques tested through black-box studies.”).
We note that this guidance extends not just to black box-type studies of the method, but also proficiency testing and other assessments of how well experts do their work in case-work settings, as well as blind testing, in which they do not know that they are being tested. Given how cognitive biases can impact the work of examiners in forensic settings, the evidence from black box studies may substantially underestimate error rates in actual casework. Moreover, jurors are extremely receptive to such information as well.

Nor have we seen judges take the approach of the American Statistical Association, which would require examiners to affirmatively state that there is no statistical basis for any probabilistic conclusion in their field. Judges have, perhaps understandably, been far more comfortable with limiting conclusion language of experts than affirmatively requiring experts to explain limitations of their methods.

The 2023 amendments to Federal Rule of Evidence 702 encourage judges to more carefully consider that the proponent of an expert bears the burden to show that the various reliability requirements are met as well as that the opinions that the expert formed are reliably supported by the application of the methods to the data. That rule change, while reflecting prior law and not intended to change the substance of Rule 702, highlights the importance of judicial gatekeeping regarding the evidence that the proponent of the expert has that the work done, as well as the opinions reached, were grounded in reliable interpretation of data. The amendment supports the approach that we recommend: simply put, the exclusion of methods that are not demonstrated to be reliable. At a minimum, experts should also, as the American Statistical Association states, disclose all of the known limitations of their work.

Despite mounting scientific concerns and a limited response to the problem of firearms testimony by the Department of Justice, there has been a substantial federal investment in increasing the use of firearms

---

384. ASA REPORT, supra note 345, at 4–5.
386. We also note proposed standards from a different group that are in progress and largely restate the AFTE identification-based approach. See Firearms & Toolmarks Subcommittee, Standards: At an SDO for Further Development & Publication, NAT’L INST. OF STANDARDS & TECH. (Mar. 1, 2022), https://www.nist.gov/osac/firearms-toolmarks-subcommittee [https://perma.cc/6ZWD-R7ED].
comparison work. The federal database, the National Integrated Ballistic Information Network (“NIBIN”), has been supported by extensive federal grants, including regarding the expensive imaging equipment used on firearms evidence, to enter it into the database. Interestingly, the algorithms used to search that database remain a black box—the federal government has sponsored research on increasing the speed and efficiency of searches but not on how reliable “hits” are using the database.\(^3\)

Technology may eventually supply reliable means to provide quantitative information about the probability that a bullet or shell casing came from a particular firearm. Statistical approaches to this problem are under development, and one has been piloted by researchers with some promising initial results.\(^4\) It may be that this is a scientific challenge that can be met. But for many decades, courts were willing to allow examiners to claim expertise that they lacked, based on assertions of experience, training, and proficiency that were not tested. Fortunately, now that those assertions have been minimally tested, some courts are stepping back to assess whether this expertise should be permitted. It is an object lesson in the acceptance and use of expert evidence in criminal courts, however, that has taken over a century for that shift to occur.

We end by emphasizing two other points. In this Article, we have focused on firearms comparison work, but it is only one specialty in the area of forensic toolmark comparison. It is among the most commonly used and has attracted sustained scientific and judicial attention, but as David Kaye and colleagues have importantly pointed out, “there is less research into the accuracy of associating impressions from tools such as screwdrivers, crowbars, knives, and even fingernails.”\(^5\) There is every reason to think that those other types of toolmark comparison raise similar or far larger reliability concerns.

Further, in this Article we have focused on criminal cases that proceed to a trial and evidentiary rulings at trial and on appeal. Yet, courts often do not have a Daubert hearing or issue written rulings regarding expert evidence


\(^5\) YALE LAW SCHOOL FORENSIC SCIENCE STANDARDS PRACTICUM, TOOLMARK-COMPARISON TESTIMONY: A REPORT TO THE TEXAS FORENSIC SCIENCE COMMISSION 10 (2022) (“There are fewer limiting opinions involving source attribution to other tools, probably because fewer of these examinations are performed, and fewer reports bubble up to the courts.”).
questions. There have been high-profile wrongful convictions in cases involving firearms evidence, like that of Curtis Flowers who had six criminal trials, and no reported decisions discussing the firearms evidence involved. In a very interesting 2020 case, a judge found it appropriate for an exonerated person to introduce experts to show that the firearms evidence should have been exculpatory at the time of trial. And most criminal cases are not tried. Lawyers may plea bargain cases based in part on the perceived power of a firearms comparison. Indeed, courts have regularly rejected application of Daubert reliability standards in other pretrial contexts, such as an application for probable cause relying on a firearms comparison. Further, laboratory audits have occurred based on revelations regarding errors in firearms work, which have not generated any written opinions in court, but which highlight the importance of forensic science commissions and other bodies tasked with investigating quality control failures in crime laboratories. Thus, while there may be increasingly careful judicial review of firearms expertise in trial settings, much of the use of forensic evidence may remain largely unreviewed by judges.

CONCLUSION

We do not know how often people have been wrongly convicted based on erroneous firearms comparison conclusions. But we do know of people convicted based on firearms evidence testimony who have since been exonerated. For example, on January 16, 2019, Patrick Pursley was convicted based on firearms evidence testimony who have since been exonerated, in part because “evidence in 1993 was scant by today’s standards, and when you start with scant evidence you’re not in a good


position to reevaluate it years later.”

In that case, the judge found that defense experts demonstrated conclusively that the cartridge cases in question were not fired by the gun attributed to Pursley.

We have described how over the past one-hundred-plus years, judges’ initial skepticism of early firearms experts transformed into growing judicial acceptance, in large part because confident experts displayed new terminology, techniques, and technology like the comparison microscope. The result was—and still remains—“an overwhelming acceptance in the United States and worldwide of firearm identification methodology.” But despite a mountain of long-standing precedent, judicial acceptance of this testimony has eroded in recent years. After many decades of rote acceptance of the assumptions underlying the methodology, judicial interest in firearms expert evidence has exploded. Over half of the judicial rulings that we identified have occurred since 2009, the year that the NAS issued its pathbreaking report. Dozens of opinions limit testimony of firearms experts in increasingly stringent ways.

This sea change has occurred because of the work of lawyers, judges, and particularly scientists, who have played a key role in generating a new body of precedent. Scientists have demanded studies to examine questions of reliability, and they have exposed how the resulting studies uncovered deep concerns regarding error rates in firearms analysis. Firearms experts may have testified with confidence in the past. But today, they increasingly face defense experts who turn the microscope to the scientific flaws underlying firearms identification. In turn, judges have increasingly engaged closely with scientific research, error rate studies, and defense expert witnesses.

The Daubert revolution did not result in an immediate shift in how judges reviewed firearms evidence, but over time, judges have begun to grapple with the reliability standards. The scientific community continues to inform that work with detailed critiques. In turn, defense lawyers have launched more precise challenges that have shaped precedent.

The December 2023 revisions to Rule 702, designed to address both the burden to show that an expert is reliable and the manner in which experts reach and express conclusions, will solidify the focus—sharpened in firearms evidence rulings—on both of those important aspects of the judicial


396. Id.

gatekeeping role. The resulting body of law has already reshaped how firearms evidence is received in criminal cases, and it provides important lessons regarding the slow, but perhaps steady, reception of science in our precedent-bound halls of justice.
# APPENDIX

## APPENDIX A. Judicial Rulings Limiting Firearms Evidence, 2005–2022

<table>
<thead>
<tr>
<th>Citation</th>
<th>Limitation on Testimony</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States v. Felix, No. CR 2020-0002, 2022 U.S. Dist. LEXIS 213513 (D.V.I. Nov. 28, 2022)</td>
<td>Limiting testimony to conclusions regarding class characteristics and whether individual toolmarkings were “consistent”</td>
</tr>
<tr>
<td>United States v. Adams, 444 F. Supp. 3d 1248 (Or. 2020)</td>
<td>Observational evidence permitted but no methods of conclusions relating to whether casings “matched” to be admitted</td>
</tr>
<tr>
<td>People v. Ross, 129 N.Y.S.3d 629 (Sup. Ct. 2020)</td>
<td>Ruling that “qualitative opinions” can only be offered on the significance of “class characteristics”</td>
</tr>
<tr>
<td>State v. Raynor, 254 A.3d 874 (Conn. 2020)</td>
<td>Permitting “more likely than not” testimony</td>
</tr>
<tr>
<td>United States v. Harris, 502 F. Supp. 3d 28 (D.D.C. 2020)</td>
<td>Instructed expert to abide by DOJ limitations, including not using terms like “match” and not claiming to exclude all firearms in the world</td>
</tr>
<tr>
<td>Williams v. United States, 210 A.3d 734 (D.C. 2019)</td>
<td>Finding error to permit expert to testify that there was not “any doubt” in conclusion</td>
</tr>
<tr>
<td>State v. Gibbs, 2019 Del. Super. LEXIS 639 (Del. Sup. Ct. 2019)</td>
<td>May not testify to a “match” with any degree of certainty, and may not testify to a “reasonable degree” or “practical impossibility”</td>
</tr>
</tbody>
</table>


United States v. Medley, No. PWG 17-242 (D. Md. April 24, 2018) — Permitting “consistent with” but no opinion fired by same gun

State v. Terrell, 2019 Conn. Super. LEXIS 827 (Conn. 2019) — Prohibiting testimony regarding likelihood so remote as to be practical impossibility


United States v. White, 2018 U.S. Dist. LEXIS 163258 (S.D.N.Y. 2018) — Holding that expert may not provide any degree of certainty unless pressed on cross-examination and may then present “personal belief”

State v. Jaquwan Burton, Superior Court, No. CR14-0150831 (Conn. Super. Ct. Feb. 1, 2017) — Permitting “consistent with” but no opinion that it was fired by same gun

Missouri v. Goodwin-Bey, No. 1531-CR00555-01 (Mo. Cir. Ct. Dec. 16, 2016) — Limiting to “the recovered firearm [that] cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting”

Gardner v. United States, 140 A.3d 1172 (D.C. 2016) — Error to admit “unqualified” testimony with “100% certainty”

United States v. Cazares, 788 F.3d 956 (9th Cir. 2015) — Limiting to “reasonable degree of scientific certainty”

United States v. Black, 2015 U.S. Dist. LEXIS 195072 (D. Minn. 2015) — Limiting to “reasonable degree of ballistics certainty” and barring “certain” or “100%” conclusions

United States v. Ashburn, 88 F. Supp. 3d 239 (E.D.N.Y. 2015) — Limiting to “reasonable degree of ballistics certainty” and precluding “certain” and “100%” sure statements

United States v. McCluskey, 2013 U.S. Dist. LEXIS 103723 (D.N.M. 2013) — Limiting testimony to “practical certainty” or “practical impossibility”

United States v. Mouzone, 687 F.3d 207 (4th Cir. 2012) — Approving trial ruling limiting any expression of certainty

casing found on the scene of the alleged shooting”
<table>
<thead>
<tr>
<th>Case</th>
<th>Restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commonwealth v. Pytou Heang, 942 N.E.2d 927 (Mass. 2011)</td>
<td>Limiting to “reasonable degree of ballistics certainty”</td>
</tr>
<tr>
<td>United States v. Diaz, 2007 U.S. Dist. LEXIS 13152 (N.D. Cal. 2007)</td>
<td>Limiting to “reasonable degree of certainty in the ballistics field” and no testimony “to the exclusion of all other firearms in the world.”</td>
</tr>
</tbody>
</table>