Reliability and Error Rates in Forensic Techniques

VARSTVOSLOVJE
Journal of Criminal Justice and Security
year 2026
volume 28
pp. 1–19

Željko Karas

Purpose:

It is easy to consider the results of forensic techniques as reliable given they were developed in the natural and technical sciences. The aim of the article is to research error rates with several forensic techniques (DNA, fingerprints, firearms etc.) and identify the influence on any errors made by various people in certain phases of examining evidence. Noting that the frequency of errors varies between the European and American systems, the article also presents each system's main features.

Methods:

The methodology used to meet the above aims is analysis of research on error rates in different forensic techniques. The most important elements of forensic examinations, their main criticisms, and adjustments in laboratory practice are described.

Findings:

It is shown that the large number of erroneous verdicts in the American system has attracted criticism, especially from US scientific bodies, which has seen some techniques being re-examined. Although errors can be made in any forensic examination, protocols should be introduced to improve the current situation. Survey results reveal Croatian student police officers hold different perceptions compared to findings in other countries.

Research Limitations:

Despite the survey results being limited to the Croatian system, they can be compared with other systems. Theoretical conclusions and explanations do not refer to individual countries.

Practical Implications:

The results could influence the way forensic evidence in criminal procedures as well as the work of all persons who assist with collecting forensic evidence are viewed.

Originality/Value:

Explanations of differences in the reliability of certain techniques and legal systems are provided.

Key words: forensics, error rate, reliability, police investigation

UDC: 343.98

Zanesljivost in stopnje napak v forenzičnih tehnikah

Namen prispevka:

Rezultate forenzičnih tehnik je lahko obravnavati kot zanesljive, saj so bile razvite na področjih naravoslovnih in tehničnih znanosti. Namen članka je raziskati stopnje napak pri različnih forenzičnih tehnikah (DNK, prstni odtisi, strelno orožje itd.) ter prepoznati vpliv različnih oseb na morebitne napake v posameznih fazah preiskovanja dokazov. Ker se pogostost napak razlikuje med evropskim in ameriškim sistemom, predstavlja članek tudi glavne značilnosti obeh sistemov.

Metode:

Metodologija, uporabljena za dosego navedenih ciljev, temelji na analizi raziskav o stopnjah napak pri različnih forenzičnih tehnikah. Opisani so najpomembnejši elementi forenzičnih preiskav, njihove glavne kritike ter prilagoditve laboratorijske prakse.

Ugotovitve:

Prikazano je, da je veliko število napačnih sodb v ameriškem sistemu povzročilo številne kritike, zlasti s strani ameriških znanstvenih institucij, zaradi česar so bile nekatere tehnike ponovno preučene. Čeprav lahko do napak pride pri kateri koli forenzični preiskavi, bi bilo treba uvesti ustrezne protokole za izboljšanje trenutnega stanja. Rezultati ankete razkrivajo, da imajo hrvaški študenti policijske šole drugačna stališča v primerjavi z ugotovitvami iz drugih držav.

Omejitve / uporabnost raziskave:

Čeprav so rezultati ankete omejeni na hrvaški sistem, jih je mogoče primerjati z drugimi sistemi. Teoretični sklepi in razlage se ne nanašajo na posamezne države.

Praktična uporabnost:

Rezultati bi lahko vplivali na način razumevanja forenzičnih dokazov v kazenskih postopkih ter na delo vseh oseb, ki sodelujejo pri zbiranju forenzičnih dokazov.

Izvirnost / pomembnost prispevka:

Podana so pojasnila o razlikah v zanesljivosti posameznih tehnik in pravnih sistemov.

Ključne besede: forenzika, stopnja napak, zanesljivost, policijska preiskava

UDK: 343.98

1 INTRODUCTION

In popular culture, forensic techniques are often portrayed as very fast, straightforward and providing infallible results (the 'CSI effect'). Since they stem from the natural and technical sciences, the reliability of these techniques is portrayed as significantly higher than statements gathered from personal sources whose credibility cannot be verified by repeatable scientific techniques. The results of such techniques are presented as objective facts arising from technical methods, and whose veracity is not affected by any individual in the process. In casual conversations, it is common to hear that some forensic methods can be trusted because they are based on verifiable facts and empirical methods. Similar claims by reputable forensic institutions can be found in various earlier publications. For example, the Federal Bureau of Investigation [FBI] discussed the infallibility of fingerprint comparisons (U.S. Department of Justice, Federal Bureau of Investigation, 1984). In criminal proceedings, some forensic examiners have declared their techniques to be 100% reliable. Claims like this appeared in several professional publications back then (as described in Cole, 2004). Some judges also declared in their rulings that falsely identifying a person as the source of biological evidence is not possible with DNA analysis (Roberts, 2017).

To ensure that the community still trusts police laboratories and provide transparency in the relationship, it must be acknowledged that in every police activity several factors can lead to errors. In cases of wrongful conviction, the police often launched the criminal investigation and established the initial suspicion against an individual who later ended up erroneously convicted. Following the efforts of the police, various other entities can become involved in investigation procedures, and their activities can also impact the accuracy of the results. When a large number of subjects participates in all stages of the processing of forensic evidence, viewing just the final stage of forensic analysis as an independent technical process of discovering facts is insufficient. Here, it is necessary to take account of the human factor in processing from the outset of the investigation at the crime scene through to the final presenting of evidence in court.

In relation to the frequency of wrongful convictions, the majority of articles about errors in forensic techniques come from the United States, where 'innocence projects' also rely on multiple ways of detecting errors. Analyses of wrongful conviction cases show that errors in forensic methods are responsible for a very high share. According to the Innocence Project, 46% of errors in verdicts entailed the incorrect application of forensic techniques, while the National Registry of Exonerations states the share in exonerations accounted for by forensic technique errors is about 29% (Saber et al., 2022). While some authors report a share in this respect of around 63% in wrongful convictions, this refers to a group of judgments that saw defendants being subsequently acquitted based on DNA analysis (Saks & Koehler, 2005). In 74% of cases of wrongly convicted persons who later had their judgment overturned based on DNA analysis, the prosecutor's office had used some type of forensic evidence (Garrett, 2011, p. 90). Even though errors can be made in any system, in European continental systems the number of wrongful convictions compared to in the USA allows greater room for optimism, which might explain the less pressing need for additional extrajudicial projects to prove some individuals are actually innocent (Huff & Killias, 2013).

2 RELIABILITY OF FORENSIC TECHNIQUES

2.1 Criticism of forensic techniques

Errors in forensic techniques might happen at a general level due to the unreliability of a method/ technique itself, or after the final analysis when presenting the results in court. Before then, errors may arise from irregularities in the early collection of evidence at the crime scene, the way it was stored or delivered, and similar preparatory actions. As to the reliability of forensic techniques themselves, a few types of forensic techniques have already attracted criticism, especially in the US system. The best-known and most extensive critique was made in 2009 by the National Research Council [NRC] of the National Academy of Sciences, when it asserted that, except for DNA analysis, no forensic method can reliably and consistently link evidence to individuals. The NRC also stated that forensics is not sufficiently connected to the main branches of science, fundamental research is lacking, and it is not consistent with the spirit of open discussion, while being beyond reproach (NRC, 2009).

The council added that a few methods may be questioned in terms of individualisation; that is, they do not establish individualisation at the appropriate level of proof in criminal proceedings. Most errors were detected in the areas of serology, hair comparison, bite marks, and shoe prints (Garrett, 2011). For the NRC, many methods commonly used in criminal proceedings lack scientific testing and validation to be viewed as reliable in criminal proceedings. Although similar extensive criticism cannot be found in other forensic circles, over time aspects of individual methods have been subjected to scrutiny. For instance, individual cases of error led to the English May Inquiry (United Kingdom Home Office, 1994), the Canadian Kaufman Report (Kaufman, 1998), or German rules on psychological examinations (“Der Bundesgerichtshof« [Federal Court of Justice], 1999). As shown already by the report’s title and contents, the NRC’s views pertain solely to the situation in the United States.

One example of an unreliable technique is the Compositional Bullet Lead Analysis used in the USA in over 2,500 court cases. It was claimed the chemical composition of bullets could be determined with the technique, where it was assumed that each box of bullets is unique (Giannelli, 2011). On this basis, it was concluded that a bullet came from a particular box of ammunition. The method was employed, for example, to assess whether one or more shooters were involved the assassination of President Kennedy. In 2004, the NRC declared the method to be unreliable for the intended purposes (NRC, 2004). The FBI withdrew it from its laboratory in 2005, mentioning its relative probative value and high costs (Tobin et al., 2017). It is noted that there is no record of this method ever having been used in systems in Europe for the purpose of bullet classification.

Unlike with more general criticism, tests of individual methods have a long tradition. One example of a method that undergoes occasional re-testing is used to determine voice matching ('voice spectrograph'), which in 1979 the US National Academy of Sciences expressed criticism of for not having sufficient support in scientific data, despite it still being accepted by some American judges in criminal proceedings (Garrett, 2011, p. 55). A review of error rates associated with individual methods revealed voice spectrography has an error rate of around 63%, and bite mark comparisons of around 64% (Covey & Beety, 2022). Microscopic hair comparison techniques have been proven to be unreliable, and the FBI has reviewed all old cases that employed the method (Garrett, 2020, p. 248).

2.2 Reliability of fingerprint comparisons

The criticism by scientific bodies in the USA led to numerous forensic methods being re-examined with the goal of again assessing their reliability and error rate, even though no new data had suggested any issues. A well-known forensic technique in use for over a century is the comparison of fingerprints. In the USA, in 2016 the Presidential Council of Advisors on Science and Technology (2016) concluded that, notwithstanding that it is a fundamentally valid method, in court proceedings it should always be accompanied with evidence about the level of error and reliability.

The earliest case that saw a perpetrator being identified based on a fingerprint comparison involved the technique being applied by Vučetić in 1892 in Argentina. He discovered that a mother had murdered her two children and then falsely accused a neighbour of doing that (the Rojas case). Use of the method led to a wrongly arrested suspect being released.

In the US system, a sample of 50,000 fingerprints was compared to determine the likelihood of similarity among the population. As part of the analysis, Meagher, Budowle and Ziesig (1999) compared prints to establish the level of coincidence among individual features. Each print was compared with the other 49,999 prints to detect the number of minutiae in the same position. No prints that matched with a larger number of minutiae were found. Relative to the number of fingers of all mankind, the probability of two samples being identical in four minutiae is expressed as $59x10^{18}$ (Stoney, 2001, p. 389). It is thus quite unlikely to find two similar or matching prints in five or six minutiae (Stoney, 2001, p. 341).

The focus on determining the minimum number of minutiae is part of the history of European criminalistics and the subject of discussions among expert bodies for a long time. In most European countries, 12 matching minutiae are required (16 in some countries). The mentioned threshold is considerably higher than with the American model, which does not define the minimum number of positions. In the American system, seven or eight matching minutiae is acceptable in criminal proceedings if supported by a professional expert's assessment, and nine or more minutiae is satisfactory.

2.3 Reliability of DNA analysis

DNA analysis is performed at predetermined locations where possible variations have been known. A sequence can be repeated a certain number of times at a certain position. In the mentioned 2009 NRC report, DNA analysis was declared to be the 'gold standard' and a thoroughly tested method that should act as a benchmark for other techniques. Yet, it was not always like this. The NRC had previously criticised DNA methods on a multiple occasions. In 1992, the NRC stressed that the error rate in proficiency tests was about 2% in false positives, meaning that continuous improvements to the DNA method were needed (Koehler et al., 2023). The NRC also expressed criticism of DNA analysis in 1996. Years later, after DNA analysis had been considerably improved and tested, the NRC highlighted other techniques not researched in the same way as DNA: fingerprints, shoe prints, toolmarks, and hair comparisons.

When DNA analysis was first relied on in criminal proceedings, the level of error was not disclosed (Thompson, 1993; Koehler, 2017). In early uses of DNA analysis in the English courts, its role in detection was seen as more valuable than in establishing evidence. The reliability of DNA was made difficult by the fact that most forensic testing was performed by the very manufacturers that had supplied the equipment used, and those business links may well have led them to present favourable results (Mnookin, 2001). In their evidence, manufacturers typically referred to wide extrajudicial uses in other areas of genetics, even though the purpose and type of samples are completely different there. At the time, several American judges had held that the method could not make a mistake in identifying a person as the source of biological material (Mnookin, 2001).

Nevertheless, by the early 1990s the probability of a coincidental matching of DNA profiles became the subject of debate. Intense discussions emerged in the scientific community due to objections concerning genetic differences in the population (Roberts, 1991), including in the most distinguished journals (Lewontin & Hartl, 1991). By 1993, several courts had ruled DNA analysis was inadmissible (Roberts, 2017, p. 201). The name "DNA wars" was given to this period in time, which nonetheless saw a number of new improvements being introduced.

As concerns the DNA method's reliability, back when DNA was analysed at six loci (it is 13–27 loci or more today), several people could have coincidentally matching allele sequences at the observed positions on the DNA (loci). This was demonstrated in several cases. In England, between 2001 and 2006, the result of analysis revealed that in 27.6% of cases several people may have been the source of the sample (Thompson, 2012, p. 244). In such cases, it is only when additional loci are considered that a difference can be shown. One example is the Easton case in 1999 where the DNA profile of a person matched the sample of the burglar at six DNA loci, despite the suspect living 400 km away from the crime scene and being disabled following Parkinson's disease. The DNA sample was then tested at the four additional DNA loci, and the person was excluded. Although the company that performed the tests had about 1 error in 50 cases at the time (Mnookin, 2001), most courts were impressed by the new method. Similarly, in 1989 a man was convicted of rape (the Ochoa case) after analysis of biological samples, irrespective of the fact the level of the DNA analysis then available meant the same type of genetic sequence could be found in about 16% of the population. Testing of more loci in 2002 made it clear that other sequences in his DNA differed from the perpetrator's biological sample (Covey & Beety 2022, p. 189).

2.4 Reliability of forensic firearms and tools analysis

Unlike with the biology of DNA, the morphology of bullet marks cannot be compared by way of predefined positions with known variations. The groove marks of a barrel are visible on a bullet in the form of striations (a series of parallel lines of different width, spacing, depth or profile). The examiner must compare the entire surface of the bullet to look for any similarities (here digital scanning tools have not made the process faster). The random mechanical pattern present on a bullet is in principle similar to the markings found on the edges of various tools (knifes, pliers etc.).

In the American system, ballistics and tool marks have been criticised for failing to meet rigorous scientific standards and not reaching the same level as for DNA. The first forensic examination of a barrel marking on bullet was made by Jeserich in 1898 in Germany using a test-fire and microphotograph comparison (Nickell & Fischer, 2014). One should not confuse this with Goddard, who in 1835 was able to link a bullet with a manufacturing mould (not a reference sample bullet) (Hueske & Bell, 2019). A considerable body of work on accuracy has since been published, as extensively reviewed by De Kinder and Bonfanti (1999). Tool mark analyses were performed on screwdriver marks many years ago by Burd and Kirk (1942). In an extended study of 201 forensic examiners and 10 serially manufactured barrels, not a single error in individualisation was found (Hamby, 2001). The study was then expanded to include 507 more examiners from 20 countries. The research continued for 10 years, without any error in matches being seen (Hamby et al., 2009). In Europe, forensic analysis in 1985 included a plan to include 7,862 police pistols in the German state of Baden-Württemberg following the discovery of a murder. The pistol in question was shown to be item 3,704 in the sample. This analysis, which is the most extensive forensic laboratory testing of weapons of a similar age and wear, confirmed the individuality of the bullet and cartridge marks.

2.5 Reliability of analysis of some other evidence

Shoeprints and vehicle tyre marks do not have variations always in the same positions. Use of them can cause damage, lines, notches, scratches and similar shapes. The position, direction, dimensions and shape of these are unique to a particular piece of footwear or vehicle tyre. The shape and position of such markings is an individual characteristic given that people wear shoes in varying ways and on different surfaces. The error rate is less than 0.48% for a sample of 3,500 analyses (Richetelli et al., 2020).

However, it is different in forensic fibre analysis where individuality can generally only be proven exceptionally if the subject fibres possess characteristics that are rare. Forensic hair analysis can also be reliable only as an exception. The usual structure of hair does not contain individual morphological characteristics.

While some techniques, like forensic geology, first appeared early in the 20th century, by the end of the century they had completely vanished from textbooks after it was argued that there were not enough individual features (all mud is the same). The first reported analysis of soil composition was in 1908 following a murder in a field in Germany, when Popp analysed layers of mud and minerals on the suspect's trousers and compared them with samples found on the route leading away from the crime scene. Despite the method being overlooked for some time, in the last decade a few serious cases were solved based on samples with unique characteristics (Woods, 2016). Such examples reveal that there are differences between techniques and that not all can be grouped in the same category as DNA profiling.

2.6 Error rates in forensic techniques

Regardless of the general reliability of individual techniques, errors can occur in any technique due to the numerous activities entailed in collecting and preparing evidence. With error rates in fingerprint analyses, in experimental studies under controlled conditions Ulery et al. (2011) found a rate of 0.1% false positives and 7.5% false negatives in an experiment involving 100 pairs of fingerprints (selected from 744 pairs). Tangen et al. (2011) established a level of 0.68% false positives and 7.8% false negatives in an experiment involving 37 experts. It was shown that accuracy here is affected by the number of visible lines, the area of the print, the visibility of the delta, the contrast of the lines, and similar factors. Kellman et al. (2014) determined there were errors in 9.6% of comparisons in an experiment, or about 200 errors in 2,292 comparisons.

In experiments to assess the reliability of firearms forensics, analyses have shown error rates below 1% (Petraco et al., 2012, p. 5), with similar results presented by other authors (Riva & Champod, 2014). In analysis by Peterson and Markham (1995), the error rate was reported to be about 0.6% for firearms and 1.5% for tool marks. Analysis of 173 examiners in 8,640 comparisons revealed a 0.65% level of error in firearms analysis for bullets and 0.93% for cartridges, and a false negative rate of 2.87% for bullets and 1.87% for cartridges (Monson et al., 2023). An extensive body of research of this kind can be found.

3 ERRORS IN THE ACTIVITIES OF DIFFERENT SUBJECTS

3.1 General remarks

Apart from the general reliability of forensic methods, it is necessary to observe the role of the people involved in the processes and who can contribute to errors. The aforementioned 2009 NRC Report also lists, besides the validity of the techniques, problems like errors in the testimony of forensic examiners or exaggeration in their statements. Garrett (2011) found 61% incorrect statements made by forensic experts in wrongful conviction cases, with the most incorrect statements concerning serology, hair, bite and shoe mark comparisons (Garrett, 2011, p. 90). Morgan (2023) researched 732 cases of wrongful conviction in the USA and established the majority of errors did not refer to the work of forensic scientists but to other factors. According to the National Registry of Exonerations, in cases where there were no false reports or similar sources of errors caused by other entities, only irregularities among some government bodies, around 30% of the contribution came from prosecutorial misconduct and about 35% from the police (Gross et al., 2020). In a few cases, the wrongful conviction was partly due to incompetent lawyers who had failed to notice evidence that would have benefited the innocent person (Covey & Beety, 2022, p. 379). In other cases, it was discovered the defendants had themselves participated in wrongful convictions in order to protect other perpetrators or take the blame (Huff & Killias, 2013).

3.2 Errors in the activities of forensic examiners

Errors made while presenting evidence can include exaggerating statistical data or presenting non-existent statistical data, or inaccurately describing the results of forensic techniques. Laboratory quality standards have accordingly been broadened to the phase of giving testimony. At the same time, numerous examples exist of intentional irregularities in the activities of laboratory employees. The biggest review of forensic errors was made by Garrett and Neufeld (2009) with respect to cases where DNA analysis led to exoneration. In their sample, they established that forensic techniques were responsible for a share of 62%, with the largest share in the way results had been presented in criminal proceedings. DNA tests of a condom performed by analyst Mills at the USACIL in 2002 were later tested again when it was concluded that the suspect had no link with any evidence in the case. Subsequent retest of three-year results of DNA analysis by the same forensic expert showed 55% of them were incorrect. Such serial violations have not been reported in literature on continental European systems.

3.3 Problems with samples

In relation to DNA analysis, instances of test samples being accidentally mixed together have been discovered in several federal states (Las Vegas, Pennsylvania, Washington). In England, a serial rapist who committed 140 crimes was not caught in time (the Night Stalker case) since, even though the police had initially suspected him, DNA testing had erroneously excluded him (Wailoo et al., 2012). In Melbourne, 7,000 cases of DNA analysis had to be re-checked in 2008 following the discovery of an error. In Houston, a laboratory had to be closed in 2003 after a quality audit found several errors and deficiencies. Dozens of the laboratory's cases that were then tested in other laboratories did not confirm the initial results (Garrett, 2011).

A well-known error in the collection or storage of samples happened in 2009 when the sample of another suspect was on the following day mixed together with materials from a rape case. After discovery of the error, retesting of all DNA tests going back 5 years was ordered.

3.4 Errors in the work of the police or the prosecutor's office

Police officers conducting a crime scene investigation who do not follow the instructions about protection while handling different types of evidence, separate packaging, and labelling can be a cause of contamination. According to Morgan (2023), in cases of DNA testing errors in the United States, 18 cases of irregularities were found in the work of the police or prosecutors, and 7 cases of irregularities in securing the chain of evidence. Morgan states errors in procedures included the misplacement or loss of exculpatory evidence in police storage facilities, deliberate ignoring of exculpatory evidence due to police bias and, in some cases, unskilled technicians deliberately linking evidence to suspects they believed to be guilty.

4 FREQUENCY OF ERRORS: SOME COMPARISONS

Official reports reveal differences in the frequency of wrongful convictions. In European countries, data on about 25 cases were publicly available by 2008. Two cases had by then been reported in Switzerland and Italy, three in France and cases in Austria (Otto, 2004; Schiffer, 2009), while three cases in the Netherlands met the conditions for a review of the procedure (Robins, 2023). More recent German analysis shows 31 cases of wrongful conviction between 1990 and 2016, a large number of which were due to false testimony or a false criminal charge in sexual offences, inaccurate confessions or identifications, and in a few cases the false testimony of a forensic expert on psychological or psychiatric issues (Leuschner et al., 2020). If the wrongful conviction rates were the same as in the USA, there would be about seven cases of wrongful conviction per year in Germany, about five in France. However, significantly fewer were registered, which leads Killias to conclude that errors are universal yet appear with different frequencies in some countries (Huff & Killias, 2013, p. 61).

Although the lower level of forensic transparency in European systems is sometimes cited as a circumstance that contributes to the statistical difference, such a view seems to have no basis. It is known that numerous forensic methods stem from European systems and have been extensively researched in the last century, and one can expect errors made with them to also be investigated if they were frequent. The wider professional public and the media would certainly not allow cases that are controversial to be concealed. For example, the Zwahlen case (Kehrsatz Mord) in Switzerland regarding a murder in 1985 (the body of a wife was found in a freezer) was called by the media the most famous criminal case in the country (Born, 1989). Several newspaper articles about the case have been published, along with books and films, while a citizens' association was founded, which reveals the strong interest in wrongful convictions. Elsewhere in Europe, the public has shown considerable interest as well, with errors having been discussed for a long time, such as a paper published in Germany in 1928, and a similar one in France to do with the Dreyfus affair (Grunewald, 2023, p. 133). At the end of the 19th century, Gross started to develop a scientific approach to criminalistics in Central Europe, motivated by the need to minimise errors and the incompetence of criminal investigators at the time (Gross, 1908).

Variation in the ways that certain forensic techniques are implemented can be observed in the rare cases where the same evidence was assessed by forensic experts from different countries. The most famous example is the Mayfield case from 2004 dealing with fingerprints on a bag of detonators found in Madrid following terrorist attacks on a railway (Cole, 2016). Based on a Spanish request sent via Interpol, the FBI identified 10 identical positions of fingerprint details and arrested lawyer Mayfield based on that comparison. Assessments were made by four forensic experts, and later by a fifth expert. However, after considering their findings, the Spanish experts concluded the two fingerprints did not match. The Spanish authorities, in cooperation with Algeria, later discovered the real suspect, whose prints did match (Daoud), and the FBI conducted an extensive analysis of the error and established that it was a problem of unclear prints and the presence of similar details in the mentioned 10 positions (U.S. Department of Justice, 2006). This case shows that large differences can exist even when it comes to a reliable forensic method and experienced examiners.

The case was covertly subjected to scientific analysis 5 years later by the very same five experts who again compared the same prints (they did not know they were from the old case). Together, the experts had about 85 years of experience in fingerprint analysis. Three of them determined there was no match with the prints from the Mayfield case (opposite to their previous conclusion), one declared the print was not sufficiently clear for a decision to be made, and another found a positive match (Dror et al., 2006).

With respect to European systems, Kloosterman et al. (2014) noted errors observed in biological samples in the Netherlands. In several DNA analyses, the profile of a female person from a narrow ethnic group was found (the Avenger Zuuk case), only for it to be later established that the match was due to contamination by a forensic laboratory employee. Nevertheless, the contamination did not lead to erroneous verdicts.

In relation to DNA analysis, there is a known case of a contamination error in 2007 in Germany near Stuttgart (the Phantom of Heilbronn). In around 40 criminal offences, a comparison of DNA profiles showed biological samples of the same person at all crime scenes. It was subsequently found that the reason for this was factory contamination of cotton swabs (Taupin, 2019).

5 INQUISITORIAL FEATURES IN CRIMINAL PROCEEDINGS

The inquisitorial features of European criminal proceedings mean the procedure for proposing and examining experts in court is different than in the American system since the court also plays a major role and is active in seeking facts. The many advantages of such inquisitorial features include reducing the level of wrongful convictions (Lupária & Greco, 2020). The American system is more focused on the parties in the procedure and erroneous verdicts are more likely than in continental proceedings.

In Germany, the notable variation in the number of wrongful convictions vis-à-vis the American system is explained by the different role played by the court, which actively investigates incriminatory and exculpatory evidence, and thus does not have just a passive role like in the American model (Grunewald, 2023). No forensic experts work for prosecution or defence, and instead they are appointed by the court from a list of authorised experts, who can be punished if they give false testimony (Huff & Killias, 2013, p. 142). Further, European courts must provide an explanation of the conviction, which means setting out in detail the evidence it found to be trusted and have been established.

The US partisan model of criminal procedure has been criticised for the considerable share of erroneous verdicts. By 2020, 2,372 cases of acquittal had been reported. American case-law did not respond to this by changing the rules of criminal procedure governing the method of determining forensic examiners or the role of the court, and instead sought new conditions for evidence to be admissible in criminal proceedings (the Daubert case in 1993 overturning Frye's "general acceptance" standard from 1923).

In England, following the error in the Maguire case (false statements about the source of alleged explosive chemicals), it was proposed to introduce a court expert similar to that seen in the French system. A court expert in France is not a witness for a particular party but holds the status of a court officer, selected from the court's official list. Their report and opinion is available to both the court and the parties before the proceedings begin (Spencer, 2017).

6 SHORT SURVEY ON THE PERCEPTION OF ERRORS

6.1 Research on the topic

Individual forensic techniques in surveys reveal general differences, with handwriting analysis being rated the least reliable (65%) and DNA analysis the most reliable with about 89% accuracy (Ribeiro et al., 2019). Garrett and Mitchell (2016) used a Likert scale on a sample ($N = 251$), among which 25.9% of respondents rated fingerprints to be very reliable and another 50.6% as reliable. In a survey by Lawson (2014), forensic analysis tools were rated the least reliable and DNA the most reliable.

Koehler (2017) conducted an online survey ($N = 210$) of individuals qualified to serve as a juror and collected their evaluations of the accuracy of DNA, fingerprint, bite mark, hair and handwriting examinations. They rated DNA as the most reliable method, and handwriting analysis as the least reliable. Koehler estimated there was 1 error per million DNA analyses, 1 error per five million fingerprints, and 1 error per 100,000 fingerprint analyses (Koehler, 2017). In surveys where forensic examiners themselves estimated errors in their fields, the numbers were very low (Murrie et al., 2019). Ramsey (Ramsey & Frank, 2007) found differences in the estimated error rate depending on the authority being surveyed (police, judiciary, prosecutors). Error estimates generally ranged from 0.5% to 15% depending on the forensic method.

6.2 Method and research

In order to research the perception of errors in forensic methods, an online survey was conducted of police officers who studied at the University of Applied Sciences in Criminal Investigation and Public Security in Zagreb. The form of the survey (questionnaire scales) was similar to what was used for this topic while researching the perception of reliability (Koehler, 2017). The following methods were listed in the survey: DNA, fingerprint, firearm, handwriting, hair and fibre analysis. In addition to errors with positive matching, respondents could assess errors with negative results (falsely exclusions). After these enquiries came a question about the sources on which the estimation was based. A further question asked about the type of activities respondents believed could contribute to erroneous results.

6.3 Results in brief

The survey was completed by 167 police officers (from a total of about 350 police officer students). The majority of police officers who completed the survey were in their senior years of study (3rd–5th year). Most respondents were men (64%), with an average age of 21 years. Fingerprint analysis was assessed to be a technique that could not involve a mistake by 28% of respondents, and DNA by 19% of them. These figures show that a considerable share of the respondents regarded the results of those forensic methods to be infallible. Fingerprint analysis was revealed to be perceived as more reliable than DNA analysis, with such responses being statistically significant (chi, $p = .007$).

According to the median responses in the survey ($p = .002$ – $p = .03$), the error rate of DNA analysis was estimated at 3.6%, for fingerprints at 4.1%, for hair and fibre analysis at 6.8%, for weapons analysis (cartridges or bullets) at 8.7%, and for handwriting analysis at 15.7%. With regard to false negatives (incorrect sample exclusions), the survey responses show only slightly larger proportions: the median for DNA analysis was 6.5%, 7.2% for fingerprints, 10.3% for firearms analysis, and 16.3% for handwriting analysis.

The median perception of the error rate in DNA or fingerprint analysis is somewhat higher than the perception established in some of the previously mentioned surveys, while the estimates for handwriting analysis are a little lower. Data from this survey include polarised responses because, on one hand, a large part of the responses describe some techniques as infallible, while other respondents estimate the error rate at level higher than found in comparable research studies. Only the ranks of forensic techniques are similar to those established in other surveys. It seems the assessment was to some extent influenced by the presentation of errors in foreign professional circles. To be clear, no case of an erroneous verdict (final judgment) for an innocent person based on forensic evidence in Croatia has ever been reported. A few distinguished scholars started the Innocence Project Croatia (Derenčinović et al., 2017), yet not a single case of a wrongful conviction has been found to date (Primorac et al., 2024). There are known cases of errors made in stages before the judgment, where were discovered during police inquiries.

7 CONCLUSIONS

The results of forensic methods are often considered to be more reliable than other types of evidence. Still, there is a possibility of an error in forensic techniques given that many people are involved in the various stages, from collecting evidence at the crime scene to conducting forensic analysis and presenting the results in court. Early in the 21st century, numerous criticism was expressed about forensic techniques in the US system. An important fact while considering the American criticisms is the status of the body that made these remarks (NRC), because it is part of the scientific community. Since it is not a body that defines standards for laboratories or makes legal rules in the criminal procedure, its influence on the authorities involved is limited.

The NRC's highlighting of DNA analysis as a method that should be a role model for other methods cannot be justified in terms of implementation given that some other types of evidence are significantly more complex in the morphology, variability and method of analysis. DNA analysis is quite simple in a technical sense as it is performed on devices which automatically carry out most of the activities. Criticisms by American bodies about the insufficient analysis of other techniques except for DNA may have led to trust in other methods being undermined, albeit they have been associated with any issues. Extensive criticisms have too broadly questioned otherwise reliable methods and placed them in the same category as the controversial ones. It would have been more beneficial to categorise techniques by the level of possible issues.

Despite it being sometimes claimed that the results of forensic techniques are objective facts, one must be aware that a large number of experts participate in all stages and they can influence the results of these actions with various incidental activities. Therefore, the results can actually sometimes reflect just the technical processing of various intentional or unintentional irregularities of individual persons. Some forensic experts were convinced about the infallibility of their techniques or even rejected the hypothetical existence of errors. Instead of such an absolute approach, it is more realistic to take the circumstances into account and check the procedures with the goal of improving them. In terms of criminal procedures, it seems that European systems which use a *court expert* – not an *expert witnesses* – could have impact on impartiality. The manner in which they are appointed, and the role of the court in verifying the results, may be responsible for a smaller number of wrongful convictions.

REFERENCES

Burd, D. Q., & Kirk, P. L. (1942). Tool marks. Factors involved in their comparison and use as evidence. *Journal of Criminal Law and Criminology (1931-1951)*, *32*(6), 679–686.

Der Bundesgerichtshof [Federal Court of Justice]. (1999, July 30). *BGH 1 StR 618/98* (BGHSt 45, 164). HRRS-Datenbank. https://www.hrr-strafrecht.de/1/98/1-618-98.pdf

Born, H. (1989). *Mord in Kehrsatz: Wie aus einer Familientragödie ein Justizskandal wurde.* Weltwoche.

Cole, S. A. (2004). Grandfathering evidence: Fingerprint admissibility rulings from Jennings to Llera Plaza and back again. *The American Criminal Law Review, 41,* 1189–1276.

Cole, S. A. (2016). Scandal, fraud, and the reform of forensic science: the case of fingerprint analysis. *West Virginia Law Review*, *119*(2), 523–548.

Covey, R. D., & Beety, V. E. (2022). *The wrongful convictions reader*. Carolina Academic Press, LLC.

De Kinder, J., & Bonfanti, M. (1999). Automated comparisons of bullet striations based on 3D topography. *Forensic Science International*, *101*(2), 85–93. https://doi.org/10.1016/S0379-0738(98)00212-6

Derenčinović, D., Roksandić Vidlička, S., & Dragičević Prtenjača, M. (2017). "Innocence Projects" and subsequent DNA testing in Croatia: a possible reality or an unattainable desire? *Zbornik Pravnog fakulteta u Zagrebu*, *67*(3–4), 373–404.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74–78. https://doi.org/10.1016/j.forsciint.2005.10.017

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Harvard University Press.

Garrett, B. L. (2020). Wrongful convictions. *Annual Review of Criminology*, *3*, 245–259. https://doi.org/10.1146/annurev-criminol-011518-024739

Garrett, B. L., & Mitchell, G. (2016). Forensics and fallibility: Comparing the views of lawyers and jurors. *West Virginia Law Review, 119*(2), 621–636.

Garrett, B. L., & Neufeld, P. J. (2009). Invalid forensic science testimony and wrongful convictions. *Virginia Law Review*, *95*(1), 1–97.

Giannelli, P. C. (2011). Comparative bullet lead analysis: A retrospective. *Criminal Law Bulletin*, *47*, 306–315.

Gross, H. (1908). *Handbuch für Untersuchungsrichter als System der Kriminalistik. Teil 1* [Handbook for investigating judges as a system of criminology. Part 1]. De Gruyter. https://doi.org/10.1515/9783112351567

Gross, S. R., Possley, M. J., Jackson Roll, K., & Huber Stephens, K. (2020). Government misconduct and convicting the innocent, the role of prosecutors, police and other law enforcement. *University of Michigan Public Law School Scholarship Repository*. https://dx.doi.org/10.2139/ssrn.3698845

Grunewald, R. (2023). *Narratives of guilt and innocence: The power of storytelling in wrongful conviction cases*. NYU Press.

Hamby, J. E. (2001). *Forensic firearms examination* [Doctoral thesis, University of Strathclyde]. ProQuest Dissertations & Theses. https://www.proquest.com/dissertations-theses/forensic-firearms-examination/docview/301564207/se-2?accountid=28931

Hamby, J. E., Brundage, D. J., & Thorpe, J. W. (2009). The identification of bullets fired from 10 consecutively rifled 9mm Ruger pistol barrels: A research project involving 507 participants from 20 countries. *AFTE Journal*, *41*(2), 99–110.

Hueske, E., & Bell, S. (2019). *Firearms and fingerprints* (Rev. ed.). Chelsea House. https://www.perlego.com/book/3769196/firearms-and-fingerprints-revised-edition-pdf

Huff, C. R., & Killias, M. (Eds.). (2013). *Wrongful convictions and miscarriages of justice: Causes and remedies in North American and European criminal justice systems*. Routledge.

Kaufman, F. (1998). *The commission on proceedings involving Guy Paul Morin*: *Report*. Ontario Ministry of the Attorney General.

Kellman, P. J., Mnookin, J. L., Erlikhman, G., Garrigan, P., Ghose, T., Mettler, E., Charlton, D., & Dror, I. E. (2014). Forensic comparison and matching of fingerprints: using quantitative image measures for estimating error rates through understanding and predicting difficulty. *PLOS ONE*, *9*(5), Article 94617. https://doi.org/10.1371/journal.pone.0094617

Kloosterman, A., Sjerps, M., & Quak, A. (2014). Error rates in forensic DNA analysis: definition, numbers, impact and communication. *Forensic Science International: Genetics*, *12*, 77–85. https://doi.org/10.1016/j.fsigen.2014.04.014

Koehler, J. J. (2017). Intuitive error rate estimates for the forensic sciences. *Jurimetrics*, *57*(2), 153–168. https://doi.org/10.2139/ssrn.2817443

Koehler, J. J., Mnookin, J. L., & Saks, M. J. (2023). The scientific reinvention of forensic science. *Proceedings of the National Academy of Sciences*, *120*(41), e2301840120. https://doi.org/10.1073/pnas.2301840120

Lawson, V. Z. (2014). *The influence of naive and media-informed beliefs on juror evaluations of forensic science evidence* [Doctoral dissertation, CUNY Graduate Center]. CUNY Academic Works. https://academicworks.cuny.edu/gc_etds/242/

Stoney, D. A. (2001). Measurement of fingerprint individuality. In: H. C. Lee, & R. E. Gaensslen, (Eds.), *Advances in fingerprint technology* (pp. 337–398). CRC Press. https://doi.org/10.1201/9781420041347

Leuschner, F., Rettenberger, M., & Dessecker, A. (2020). Imprisoned but innocent: Wrongful convictions and imprisonments in Germany, 1990-2016. *Crime & Delinquency*, *66*(5), 687–711. https://doi.org/10.1177/0011128719833355

Lewontin, R. C., & Hartl, D. L. (1991). Population genetics in forensic DNA typing. *Science*, *254*(5039), 1745–1750.

Lupária, L., & Greco, C. (2020). Unveiling wrongful convictions between the US and Italy: Cross-learning from each other's mistakes. *The Wrongful Conviction Law Review*, *1*(1), 101–123. https://doi.org/10.29173/wclawr12

Meagher, S., Budowle, B., & Ziesig, D. (1999). *50K vs. 50K fingerprint comparison test*. (Government Exhibit 500, United States v. Mitchell). https://chance.dartmouth.edu/chance_news/for_chance_news/ChanceNews12.05/government.pdf

Mnookin, J. L. (2001). Fingerprint evidence in an age of DNA profiling. *Brooklyn Law Review*, *67*(1), 13–70.

Monson, K. L., Smith, E. D., & Peters, E. M. (2023). Accuracy of comparison decisions by forensic firearms examiners. *Journal of Forensic Sciences*, *68*(1), 86–100. https://doi.org/10.1111/1556-4029.15152

Morgan, J. (2023). Wrongful convictions and claims of false or misleading forensic evidence. *Journal of Forensic Sciences*, *68*(3), 908–961. https://doi.org/10.1111/1556-4029.15233

Murrie, D. C., Gardner, B. O., Kelley, S., & Dror, I. E. (2019). Perceptions and estimates of error rates in forensic science: A survey of forensic analysts. *Forensic Science International*, 302, Article 109887. https://doi.org/10.1016/j.forsciint.2019.109887

National Research Council (NRC). (2004). *Forensic analysis: Weighing bullet lead evidence*. National Academies Press. https://doi.org/10.17226/10924

National Research Council (NRC). (2009). *Strengthening forensic science in the United States: A path forward.* National Academies Press. https://doi.org/10.17226/12589

Nickell, J., & Fischer, J. F. (2014). *Crime science: Methods of forensic detection*. University Press of Kentucky.

Otto, H.-D. (2004). *Das Lexikon der Justizirrtümer: Skandalöse Fälle, unschuldige Opfer, hartnäckige Ermittler* [The encyclopedia of judicial errors: Scandalous cases, innocent victims, persistent investigators] (Orig.-Ausg., 1. Aufl.). Ullstein. https://www.deutsche-digitale-bibliothek.de/item/CLKRLC4FDE35YJUFFME2D34PPWJV2ZPU

Peterson, J., & Markham, P. (1995). Crime laboratory testing results, 1978 – 1991, II: Resolving questions of common origin. *Journal of Forensic Sciences, 40*(6), 1009–1029. https://doi.org/10.1520/JFS13871J

Petraco, N. D. K., Chan, H., De Forest, P. R., Diaczuk, P., Gambino, C., Hamby, J., Kammerman, F. L., Kammrath, B. W., Kubic, T. A., Kuo, L., McLaughlin, P., Petillo, G., Petraco, N., Phelps, E. W., Pizzola, P. A., Purcell, D. K., & Shenkin, P. (2012, March). *Application of machine learning to toolmarks: Statistically based methods for impression pattern comparisons.* National Institute of Justice. https://nij.ojp.gov/library/publications/application-machine-learning-toolmarks-statistically-based-methods-impression

President's Council of Advisors on Science and Technology (2016, September). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Executive Office of the President of the United States. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

Primorac, D., Sokanovic, L., & Bozhinovski, A. (2024). Correcting miscarriages of justice: Innocence project Croatia. In D. Primorac (Ed.), *Forensic science and molecular anthropology: Topics selected from 12th ISABS Conference on Forensic and Anthropological Genetics* (p. 83). IntechOpen. https://doi.org/10.5772/intechopen.113954

Riva, F., & Champod, C. (2014). Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases. *Journal of Forensic Sciences*, *59*(3), 637 –647. https://doi.org/10.1111/1556-4029.12382

Ramsey, R. J., & Frank, J. (2007). Wrongful conviction: Perceptions of criminal justice professionals regarding the frequency of wrongful conviction and the extent of system errors. *Crime & Delinquency*, *53*(3), 436–470. https://doi.org/10.1177/0011128706286554

Ribeiro, G., Tangen, J. M., & McKimmie, B. M. (2019). Beliefs about error rates and human judgment in forensic science. *Forensic Science International, 297*, 138–147. https://doi.org/10.1016/j.forsciint.2019.01.034

Richetelli, N., Hammer, L., & Speir, J. A. (2020). Forensic footwear reliability: Part III—Positive predictive value, error rates, and inter-rater reliability. *Journal of Forensic Sciences, 65*(6), 1883–1893. https://doi.org/10.1111/1556-4029.14552

Roberts, L. (1991). Fight erupts over DNA fingerprinting: A bitter debate is raging over how the results of this new forensic technique are interpreted in court. *Science, 254*(5039), 1721–1723.

Roberts, P. (2017). *Expert evidence and scientific proof in criminal trials*. Routledge. https://doi.org/10.4324/9781315094205

Robins, J. (2023). *Murder, wrongful conviction and the law*. Routledge. https://doi.org/10.4324/9781003251484

Saber, M., Nodeland, B., & Wall, R. (2022). Exonerating DNA evidence in overturned convictions: Analysis of data obtained from the National Registry of Exonerations. *Criminal Justice Policy Review*, *33*(3), 256–272. https://doi.org/10.1177/08874034211033327

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science, 309*(5736), 892–895.

Schiffer, B. (2009). *The relationship between forensic science and judicial error: A study covering error sources, bias, and remedies* [Doctoral thesis, University of Lausanne]. IRIS: Institutional Repository of the University of Lausanne. https://api.unil.ch/iris/server/api/core/bitstreams/d2efc012-d3fc-40db-b06f-77c8fcee0115/content

Spencer, J. R. (2017). Court experts and expert witnesses: Have we a lesson to learn from the French? In P. Roberts (Ed.), *Expert evidence and scientific proof in criminal trials* (1st ed., pp. 569–592). Routledge.

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science, 22*(8), 995–997. https://doi.org/10.1177/0956797611414729

Taupin, J. M. (2019). *Interpreting complex forensic DNA evidence*. CRC Press. https://doi.org/10.4324/9781351023788

Thompson, W. C. (1993). Evaluating the admissibility of new genetic identification tests: Lessons from the "DNA War". *The Journal of Criminal Law and Criminology, 84*(1), 22–104. https://doi.org/10.2307/1143886

Thompson, W. C. (2012). Forensic DNA evidence: The myth of infallibility. In S. Krimsky & J. Gruber (Eds.), *Genetic explanations: Sense and nonsense* (pp. 227–255). Harvard University Press.

Tobin, W. A., Sheets, H. D., & Spiegelman, C. (2017). Absence of statistical and scientific ethos: The common denominator in deficient forensic practices. *Statistics and Public Policy*, *4*(1), 1–11.

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

United Kingdom Home Office. (1994). *Report of the inquiry into the circumstances surrounding the convictions arising out of the bomb attacks in Guildford and Woolwich in 1974: Final report (HC 449).* GOV.UK. https://assets.publishing.service.gov.uk/media/5a7c290540f0b645ba3c7137/0449.pdf

U.S. Department of Justice, Federal Bureau of Investigation. (1984). *The science of fingerprints: Classification and uses*. U.S. Government Printing Office. https://dn760008.eu.archive.org/0/items/scienceoffingerp00unit/scienceoffingerp00unit.pdf

U.S. Department of Justice. (2006). *A review of the FBI's handling of the Brandon Mayfield case*. Office of the Inspector General. https://oig.justice.gov/sites/default/files/legacy/special/s0601/final.pdf

Wailoo, K., Nelson, A., & Lee, C. (Eds.). (2012). *Genetics and the unsettled past: The collision of DNA, race, and history*. Rutgers University Press.

Woods, B., Lennard, C., Kirkbride, K. P., & Robertson, J. (2016). Soil examination for a forensic trace evidence laboratory–Part 3: A proposed protocol for the effective triage and management of soil examinations. *Forensic Science International*, *262*, 46–55. https://doi.org/10.1016/j.forsciint.2016.02.034

About the Author

Željko Karas, Ph.D., is a Professor of Criminal Investigation at the University of Applied Sciences in Criminal Investigation and Public Security, Ministry of the Interior in Zagreb, Croatia. E-mail: zkaras@fkz.hr