How to solve a crime using forensic statistics: the role of data analysis to support the Justice

Franco Taroni[1] and Silvia Bozza[1,2]

[1] School of Criminal Justice, University of Lausanne, Switzerland

[2] Department of Economics, Ca' Foscari University of Venice, Italy

Forensic science and forensic medicine rely on a body of scientific principles and technical methods to help with issues in legal proceedings, such as criminal, civil or administrative investigations. These disciplines seek to help demonstrate the existence or past occurrence of events of legal interest, such as a crime. Forensic science, in particular, assists the various participants in the justice system, such as investigators, public prosecutors and decision-makers at large, in examining events related to persons of interest and recovered traces. This may involve the analysis of the nature of fluids and other materials, such as textile fibers, handwritten documents, glass and paint fragments. Forensic medicine, in turn, assists the judicial system by offering information in a variety of domains, such as the cause of death and the estimation of the age of living persons. More generally, forensic disciplines thus take a major interest in aspects such as the investigation of crimes and the direct examination of living or death persons (i.e., victims and suspects) and the vestiges of actions. In other words, forensic disciplines assist in the reconstruction of past events of judicial relevance that are unknown to us. Thus, the domain must deal with the fundamental notion of *uncertainty*. The natural response to uncertainty is the search of more information. Naturally, this involves the examination and comparative analysis of so-called 'evidential material' (i.e., DNA traces, toxic substances, crime scene findings, data imaging, etc.) followed by an assessment of the evidential strength of these scientific results within the particular context of the relevant event under investigation.

However, throughout the history of forensic science, including more recent periods, challenges arose from the discovery of cases of miscarriage of justice in which scientific findings played a major role. These cases generate a continuing and serious stream of debate about the status of some areas of forensic practice with respect to scientific standards of reliability. At the same time, numerous courts across legal systems have repeatedly emphasized the need for practicing scientists to continually monitor the performance of their domain of specialization.
Most importantly, scientists need to scrutinize both the rationale underlying the various areas of practice and the ways in which scientific results are evaluated and presented in context.

Today, many of the so-called traditional forensic identification practices (e.g., those involving questioned documents, dentition, X-ray images) are systematically compared to purportedly better-founded and better-researched fields, in particular forensic DNA analysis, to point out the lack of fundamental research and the predominant reliance upon arbitrary expert opinions. Many legal and scientific researchers and practitioners invoke this observation to call for a revision of research agendas, towards a more systematic collection of data on agreed measurable factors and the development of sound probabilistic methods for evidence evaluation and interpretation under uncertainty.

The fundamental constraint in forensic science, in much the same way as in science in general, is that available information is limited and incomplete. This means that categorical conclusions about events of judicial interest are impossible.
Reasoning in the light of uncertainty thus represents the regular case. The inevitability of uncertainty implies the necessity to determine the degree of belief that may be assigned to a particular uncertain event or proposition, such as '*Is the suspect the donor of the recovered trace?*', '*Is the toxic substance the cause of the patient's death?*', etc. It is in this context that inferential sciences, including statistics, can offer a valuable and substantial approach, in particular the Bayesian paradigm. In fact, whenever uncertainty is recognized as an inherent aspect of a given inferential problem, and a statistical approach is feasible, then this approach represents a normative reference in that it captures and indexes uncertainty based upon a precise and logical line of reasoning[1].

Scientific progress relies on past experience, but how exactly such experience is to be used to inform future directions and decision-making represents a fundamental challenge. On the basis of what one sees, combined with any existing knowledge, one seeks to assess, if possible in a quantitative way, one's uncertainty about a particular event of interest, yet the reality is that this kind of reasoning for extending knowledge provides only an incomplete basis for a conclusion. It follows naturally from this that scientific discussions and public debates on science should focus on uncertainty explicitly: that is, our perspective should seek to distinguish between what is more likely and what is less likely, rather than attempt to endorse a concept of certainty that cannot be warranted by the limited and imperfect evidence that arises in practical proceedings.

With that being said, the very relevant need of all science is a way to deal quantitatively with what is commonly known as the 'probabilities of causes', with the term cause being understood as an uncertain proposition (e.g., '*The suspect is the source of the DNA trace found on the victim*'). The fundamental task thus consists of discriminating between events of interest, or causes (e.g., in forensic medicine, the patient's cause of death), in the light of particular acquired information (i.e., scientific findings).

From a practical perspective, the ability to deal with reasoning under uncertainty represents a core aspect of operational procedures that seek to qualify as rational. By offering an explicit way for specifying and articulating uncertainties, they help recipients of expert information introduce results of scientific examinations into a coherent whole, along with multiple other items of evidence. The latter aspect reveals a further level of challenge. Daily inference tasks encountered by investigators, scientists and other participants in legal proceedings (judges, prosecutors and lawyers) are characterized not only by single and isolated items of evidence, but multiple items associated with a possibly complicated mutual dependency structure. It is therefore natural to enquire about logical procedures that can deal with items of evidence that occur in combination and, in particular, the way in which multiple items of evidence stand in relation to each other. Such analyses reach further levels of complication essentially because they need to be conducted in the light of intricate frameworks of circumstances, that is situations involving many variables.

Probabilistic reasoning alone is not, however, an endpoint of forensic or medical applications in the legal process. Clearly, at the end of the day, decisions must be made. Once that uncertainty is recognized and formalized, the combination of uncertainty with the ultimate decision represents the core feature of legal proceedings. For example, a Court of Justice may have to

[1] Lindley D.V. (2014). *Understanding Uncertainty*. John Wiley & Sons, Hoboken.

decide if it finds a defendant guilty of the offence for which he has been charged. While probabilistic reasoning under uncertainty can be considered a topic to be studied in its own right, systematic research on how probability is coherently applied in the wider context of rational decision-making under uncertainty, in particular with regard to forensic science applications, is still a largely unexplored field.
Forensic statistics is the name of a discipline that addresses questions of inference and decision-making in the judicial system and represents an important branch of modern statistics. Forensic statistics engages in the challenge of guiding people in reasoning under uncertainty, which is common to law and forensic science. Reasoning under uncertainty raises not only legal and scientific questions of technical difficulty and practical importance, but also fundamental questions in a wide variety of domains.
It is important to make people aware of the academic and societal importance of this field, not only for the administration of justice but also for its impact on societal civilization.

…

The interested reader can refer to the recent Aitken C., Taroni F., Bozza S. (2021). *Statistics and the evaluation of evidence for forensic scientists*. John Wiley & Sons, Chichester, 3rd edition.