Marjan Sjerps
Netherlands Forensic Institute
Volmerlaan 17
2288 GD Rijswijk
m.sjerps@nfi.minjus.nl

Research

Forensic Statistics

Lately, forensic science has been receiving much attention in the media. Popular programs such as 'Crime Scene Investigation' show how forensic investigations can contribute to obtaining and reporting results in criminal law. These investigations cover a broad range of areas of expertise, varying from DNA profiling to examining shoe prints, and from toxicology to handwriting. Experts in other scientific fields are also regularly called upon to appear in court. Statistics and probability theory frequently underlie statements made by experts and therefore play an important role in the judicial system. Nevertheless, the fields themselves are seldom placed in the limelight. Forensic statistics is a broad and dynamic field, in which there is still much work to be done. For the mathematician it is a fascinating applied field with exciting examples. Marjan Sjerps works as a statistician at the Netherlands Forensic Institute. This text is a translation of a 2004 prize winning article from the Nieuw Archief voor Wiskunde.

According to those in the know, the role of experts in criminal law is on the rise (Nijboer, 2002). Since statistics forms an important component of many areas of expertise, it also enjoys increasing attention. Outside of the Netherlands a number of publications and websites are dedicated to statistics in (criminal) law. Examples are Finkelstein and Levin 1990, Aitken and Stoney 1991, Gastwirth 2000, Good 2001, websites of the Federal Judicial Centre and of Charles Brenner. There are only a few references in Dutch; see for example van Koppen et al. 2002 and Broeders 2003, which mentions it briefly. In 2005 different authors contributed to a book (Sjerps and Coster van Voorhout). Since very diverse areas of expertise such as pathology, biology, physics, chemistry and psychology are involved, the statistical techniques that are used are also very diverse. Nevertheless, these areas have a number of aspects in common when applied to criminal law. For example, there is usually a suspect and a chosen characteristic of this suspect is compared to trace evidence associated to the crime. The DNA profile of the suspect can be compared to the DNA profile of a trace found in the underwear of a victim of sexual abuse (see figure 1), or a rifle found in the suspect's house can be compared to a bullet found at the scene of an armed robbery. In another case a dog may compare the suspect's scent to a scent that is thought to be on an object touched by the perpetrator, after which the expert must report the evidential value of the dog indicating (or not) a similarity between the scents. In this type of research, the important factor is the evidential value of observed similarities and differences. Up to approximately twenty years ago, there was no general usable definition of the term *evidential value* and every area of expertise had its own method to reflect the evidential value of its research results. This void was filled by the emergence of the so-called Bayesian approach in forensic statistics (also called the 'logical' or 'likelihood ratio' approach). A model was developed that was based on a well-known rule of Bayes from probability theory. It gives a quantitative expression for the evidential value of different types of evidence but can also be used as a guiding principle for experts reporting their conclusion verbally. The standard texts in this field are Aitken (1995) and Robertson and Vignaux (1995). This approach is applied to an increasing number of subfields, sometimes giving rise to interesting questions (see the frames). In this article we will briefly explain the Bayesian model. We will also discuss three pitfalls in which the average lawyer easily falls when interpreting evidence, and many mathematicians with

Formulating hypotheses

A still unsolved problem concerning the foundations of statistics is the following. Suppose that at the site of a crime committed by two perpetrators two blood traces A and B are found, possibly originating from the two perpetrators. The blood traces are tiny and of terrible quality, making only a partial DNA-profile possible. The profiles show that the two blood traces were made by different persons. The 'match probability' is one in a million for trace A and one in a hundred for trace B. The police catch one suspect, whose DNA profile corresponds to that of trace A. The evidential value, measured by the *likelihood ratio* (LR), now depends on the wording of the hypotheses. The LR for the two hypotheses (1a) The suspect is the donor of one of the two traces (1b) The suspect is not the donor of either of the two traces is half a million; the LR for the two hypotheses (2a) The suspect is the donor of trace A (2b) The suspect is not the donor of trace A is one million. One difference between the pairs of hypotheses is that the first pair can be stated without knowing the DNA profiles of the traces and of the suspect, while the second is based on this information. In both cases the computation of the LR is based on the assumption that the traces A and B were made by different persons, but this is only known after DNA profiling is carried out. The discussion now centres on the following. May the expert formulate his hypotheses using the available evidence and then present the LR for these hypotheses? Or is it illegal to use available evidence because the statements of the hypotheses should be independent of the evidence (see also Meester and Sjerps 2004 and the discussion with Dawid).

him. See Broeders (2003) for a fairly complete overview of this type of pitfall. A completely different branch of forensic statistics concentrates on the way data is collected. This can for example concern environmental crime, with questions like how to sample a landfill, or barrels containing XTC production waste. Designing experiments for investigations also belongs to this subfield. I will touch upon this briefly. To indicate why forensic statistics is more than just the application of ordinary statistics to an unusual field, I will also briefly discuss the specific forensic aspects. Finally, I will look beyond my own sphere of activities, the statistical aspects of the technical forensic research done by the Netherlands Forensic Institute, and briefly mention some other areas in which statistics is used for criminal law.

How strong is the evidence: the Bayesian model

To express the value of evidence, the Bayesian approach in forensic statistics uses a general model based on Bayes' theorem:

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B]} .$$

In this expression, $P[A]$ is the probability of A being true and $P[A|B]$ is the conditional probability of A being true given that B is true. If we apply this rule to two hypotheses H_p and H_d and evidence E, we can deduce (using a subjective definition of probability) that

$$\frac{P[H_p|E]}{P[H_d|E]} = \frac{P[H_p]}{P[H_d]} \cdot \frac{P[E|H_p]}{P[E|H_d]} .$$

We interpret this formula as meaning that the ratio of the probability of two hypotheses is modified by introducing evidence. The new odds (on the left) arise from the old odds (in the middle) by multiplication by the *likelihood ratio* LR (the ratio on the right). In words: *the posterior odds equal the prior odds times the likelihood ratio.*

In criminal law this rule can be applied as follows: the prosecutor has a hypothesis H_p, for example 'the suspect wrote this threatening letter', and the defender has another hypothesis H_d, for example 'someone else wrote the letter'. The judge will convict the suspect if the prosecutor's hypothesis is much more probable than the defender's, that is, if the posterior odds of these two hypotheses are large. At the start of the trial, the judge has a certain estimate of the prior odds. During the trial, evidence is brought forward: some of it in favour of the suspect and some of it against. Every time a piece of evidence is brought forward, the judge will adjust his estimate of the probability ratio according to Bayes' rule. That is, he multiplies it by an LR greater than one when the evidence is against the suspect and with a LR smaller than one when the evidence is in favour of the suspect. Once all evidence has been presented, the judge makes a decision based on his final estimate of the probability ratio between the prosecutor's hypothesis and the defender's.

Figure 1 A suspect's DNA profile is compared to the DNA profile of a trace from the victim's underwear in a case of sexual abuse.

Although this appears to be a simple application, this thought process has consequences for the way experts must report their results. The expert cannot make a pronouncement about the posterior odds of the prosecutor's and defender's hypotheses without making an assumption on the prior odds. For example, the forensic handwriting expert cannot say anything about the probability that the suspect has written the threatening letter without making assumptions about the probability before he studies the letter.

Such assumptions are pre-eminently part of the judge's tasks, not the expert's. It follows that based on his knowledge, the expert can only make a pronouncement concerning the LR of the evidence for the hypothesis H_p versus the hypothesis H_d. According to the Bayesian model, this is exactly the expert's task: reporting this LR so that the judge can subsequently adjust his estimate of the probability ratio of H_p versus H_d.

An example based on a true case

At the scene of a burglary the perpetrator cuts himself on the glass of a window that was smashed. The DNA expert determines the DNA profile of the blood on the glass, which is then compared to that of the suspect. In the courtroom the prosecutor claims that the blood comes from the suspect (H_p), while the suspect's lawyer says that his client has nothing to do with the case and that the blood is someone else's (H_d). Suppose that the DNA profiles are a perfect match; what is then the LR? The numerator is the probability of a

The value of a DNA database match

A bank is robbed by a masked person. The car that is used to leave the crime scene is dumped and set on fire. The mask is found inside. The traces of saliva found on the partially burnt mask result in a DNA profile that has a probability of one in ten thousand of matching the profile of an random person (due to the low quality of the traces only a partial DNA profile is obtained). The saliva profile is compared to a DNA database of 3000 suspects, one of whom has the same DNA profile. The trace was compared to 3000 persons instead of only one. Question: does comparison with the DNA database increase or decrease the value of the DNA match as evidence? *Answer 1*: by making several comparisons the probability of a chance hit increases. If there were ten thousand people in the database, you would already expect a match to occur simply by chance. The value of the match as evidence therefore decreases. *Answer 2*: the information that the 2999 'non-matching' persons in the database can be excluded as perpetrator increases the value of the evidence. If the whole world population were in the database and there were only a single match, we would be certain that he is the donor of the saliva. The evidential value of the match therefore increases. For the correct answer, see the frame at the end of this article.

match when the blood sample comes from the subject. If no mistakes are made, this probability is one. The denominator is the probability of a match when the blood sample comes from an random person.

Based on reference material, the expert has some knowledge of this 'match probability': it is less than one in a billion. The LR is therefore one billion, which means that the probability ratio of H_p versus H_d is increased by a factor of one billion. However, the expert cannot say how large this ratio becomes since this depends on the prior odds. This way, the Bayesian approach provides both a definition of the expert's task and a quantitative definition of the notion of 'evidential value'. Indeed, it follows from the model that the size of the LR reflects exactly how much more probable the prosecutor's hypothesis becomes with respect to the defender's when evidence is added. In other words, the LR exactly reflects the value of the evidence. In the example given above, the evidential value is at least one billion. If the blood sample were of such low quality that only a partial DNA profile could be obtained with a match probability as large as one in a million, the LR would have been a million. The Bayesian model can also be used to combine several pieces of evidence. If the hypotheses remain unchanged, it easily follows that the LR of the combination of two independent pieces of evidence is the product of the LRs of the individual pieces. This means that, in principle, when analyzing hair a morphological study can be combined with DNA profiling. In less simple cases, the analysis rapidly becomes very complex. At present, research is being done into the use of 'Bayesian belief networks' for combining evidence.

The LR approach has been applied successfully in cases where numbers are available. One example is DNA profiling (see for example Evett and Weir 1998, and Sjerps and Kloosterman 2003), where the LR is used not only for simple comparisons of a trace with a suspect but also in more complex situations where the numerator of the LR is not one. DNA testing to establish relatedness is such an area, where, for example, an unknown victim is compared with puatitve parents, or a foetus with the mother and her putative rapist. Less well-known examples are automatic speaker recognition, where a computer program is used to compare the voice of someone on a telephone tap to a suspect's voice (see Broeders 2003 and references found therein), and the comparison of glass (Curran et al. 2000). LRs can also be computed for scent identification tests with dogs, face recognition by witnesses and polygraph tests (see Van Koppen et al. 2002 and references found there; here the LR is usually referred to as the 'diagnostic value'). Even if there is some discussion concerning the formulation of hypotheses and the reporting of an LR in areas where the LR can be calculated effectively, nowadays most experts agree that it is their task to report an LR. In areas where numbers are less readily available, the Bayesian model has raised a great deal of controversy and it is not generally accepted that an expert must report an LR. For a long time, verbal statements have been made in these areas concerning the probability of the prosecutor's hypothesis and few experts are inclined to change this. For example, an fire arms expert may conclude that 'the probability that the bullet was fired by the suspect's gun is very high'. In the Bayesian model he cannot make such a conclusion based on his expertise because the probability depends on the prior odds. In Bayesian terms, the expert should state his conclusion in another way; for example: 'our investigation gives a very strong indication that the bullet came from the suspect's gun'. Few lawyers will recognize the difference between these two statements. The discussion concerning reporting is therefore still ongoing in such areas.

Fallacies in the appreciation of evidence

In addition to providing the definition of the expert's role and of evidential value, the Bayesian way of thinking has also revealed a number of fallacies. These are based on common errors of thinking and have been known for some time in psychology (Kahneman et al. 1982). Within forensic literature, however, they have only been seriously considered in the last thirteen years (Evett 1995, Broeders 2003). The most well-known is the *prosecutor's fallacy*. Suppose, for example, that the perpetrator's blood is found at a crime scene. The (partial) DNA profile is compared to that of suspect J. Smith and matches. The probability that an arbitrary person would have this profile is, let's say, one in a million. The prosecutor can present the following reasoning:

1. The probability that the profiles match while the blood is not J. Smith's is one in a million.
2. The profile matches, so the probability that the blood is not J. Smith's is one in a million.
3. The probability that the blood is J. Smith's is therefore 99.9999%.

Obviously the prosecutor makes a mistake when he concludes (2) from (1). Indeed, the probability that the samples match given that the blood is not J. Smith's is not equal to the probability that the blood is not J. Smith's given that the samples match. Nevertheless, the literature shows that this error is made by many. My own restricted experience and that of my colleagues also shows that Dutch lawyers sometimes make this mistake. The consequences can be considerable in cases where the 'match probability' is relatively large and there is barely any other evidence. In the same case, the defender can reason as follows:

1. The Netherlands has about 16 million inhabitants other than J. Smith.
2. We can expect 17 persons, including J. Smith, to have the same DNA profile.
3. The probability that the blood is J. Smith's is about 1 in 17, that is, approximately 6%.

In itself, this reasoning is correct but it is based on a number of hidden assumptions, for example that the perpetrator is Dutch and that all Dutch (including all women, babies, elderly, handicapped, etc.) have the same prior probability of being the perpetrator. This reasoning is also sometimes used in Dutch courtrooms. In the literature it is called the *defence fallacy*. Finally, there is also the *base rate fallacy*. In this case, the LR is not scaled by the prior odds. Suppose, for example, that a masked robbery takes place where the perpetrator's saliva is found on a cap lying in the car used to leave the crime scene. If there aren't any suspects yet, the expert can compare the frequency of the DNA profile of the saliva in different populations.

Suppose that in population A the frequency is one in a million and in population B it is one in a billion. The LR for the hypothesis that the perpetrator comes from population A and not B is then one thousand. The base rate fallacy is now that one tends to think that this means that the probability that the perpetrator comes from population A is much higher than that he comes from B. However, the probability that the perpetrator comes from population A of course also depends on a great number of other factors, of which the size of the population is the most obvious.

The above makes it clear that it is not sufficient for a forensic expert to decide what his statement is about, he must also consider how he can prevent fallacies and misunderstandings when formulating his conclusions. However, the ultimate solution for the prevention of fallacies and misunderstandings is yet to be found; the importance of psychological research in this is increasing.

Figure 2 Environmental investigation is an area where sampling is of great importance for the final conclusions. Up to now laboratories have always spent much time and money optimizing their analytical methods, while little attention is payed to the taking of samples. Nowadays it has been recognized that the biggest source of error is not in the lab but in the field.

Gathering data

The above concerns situations where conclusions must be made based on research results. However, a large part of the questions my colleagues and I were asked recently concerned situations where observations still had to be carried out. Environmental investigation, for example, is an area where sampling is of great importance for the final conclusions. Often the question is whether the concentration of certain substances exceeds the legal limit. Sampling and sample analysis are very expensive, while the batches that are concerned can be large and heterogeneous (see figure 2). Consequently, in the sampling protocols that are legally mandated, a large batch can be approved or rejected based on only a few ianalytical results. In the Dutch regulation concerning the sampling and analysis of building materials (execution decree 1998), for example, a 2000 tonne lot is approved or rejected based on three analytical results, each obtained by mixing four samples. In practice, it turns out that the lots do not always satisfy the assumptions, for example because illegal substances were mixed in, making the lot more heterogeneous than was assumed. Furthermore, 'hot-spots' in the lot can cause outliers during analysis. The effects of this and the correct statistical treatment are still unclear. For statisticians, this area forms an enormous challenge (see among others Keith 1996, Gilbert 1987, Patil and Rao 1994 and Stelling and Sjerps 1999). Sampling often still does not receive the attention it deserves. Laboratories spend much time and money optimizing their analytical methods, while they pay remarkably little attention to the samples they receive. In practice, the most important source of error is not in the lab but in the field. The differences between samples are usually much bigger than the differences between repeated analyses of the same sample. The legislature also pays little attention to this fact: the suspect is allowed a counter analysis in an independent laboratory but the samples used there are taken right next to the samples that were analyzed before.

In setting up a forensic investigation, one always comes across the question of how observations must be collected. An example is a case where the suspect claimed that his gun went off accidentally while he made a certain manoeuvre. The fire arms experts of the NF determined the probability of such an event through an experiment in which they repeated the suspect's manoeuvre with his weapon approximately 400 times. Many days and blisters later the experts found that the gun had not gone off a single time. In general, however, experiments are done within the framework of a research project, for example to introduce a new method. A third situation where the gathering of data is important concerns the sampling of a batch of discrete units, such as numerous bags of narcotic substances or pills, a set of barrels with unknown contents that has been dumped somewhere in an abandoned area, or a great number of fibres found on a victim's clothing, of which only a limited number can be analyzed more closely. There are different methods of sampling in such cases. At present a directive for sampling in batches of narcotic substances is being developed on a European level, which includes a discussion of the advantages and disadvantages of these methods (ENFSI Drugs WG 2003).

Specific forensic aspects

Forensic statistics is characterized by a number of particular aspects. First of all, of course, we have the multifaceted field in which it is applied, which asks for a broad range of statistical techniques, e.g. experimental design, classification, reliability intervals, Bayesian methods, simulation techniques, non-parametric methods, control charts, regression and population genetics. Moreover, the objects that are studied are not commonplace. In DNA-profiling it can, for example, concern blood, saliva or sperm traces, which are sometimes available only in very small quantities, partially decayed or mixed together. In physical investigations, it can, for example, concern garbage bags, screwdrivers, shoes (see figure 3), bullets, handwriting, voice or noise; in chemical investigations, it can be explosives, gunshot residue, glass fragments, tape, ink or car paint. Even in comparing ear prints

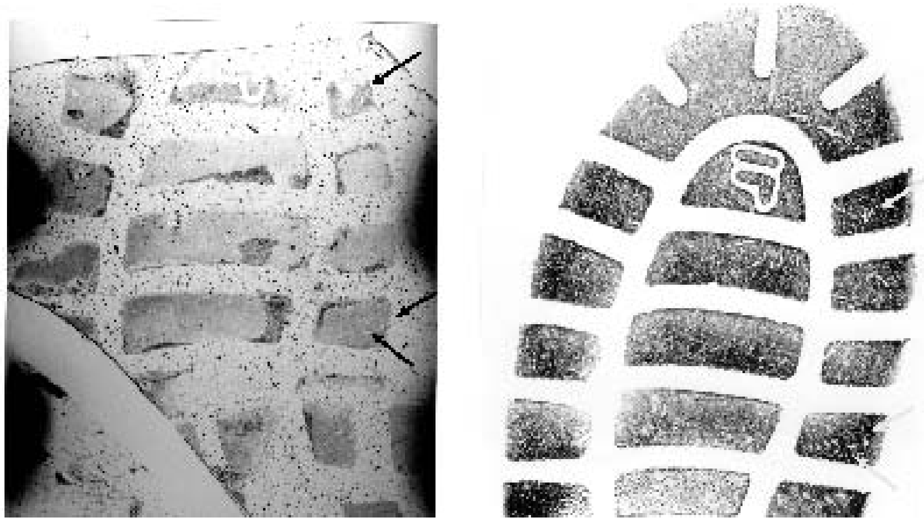

Figure 3 In legal cases it is often important to know the probability that, for example, a suspect's shoe caused a given print. Many forensic statisticians believe that the Bayesian, or subjective, notion of probability is better adapted to this type of question than the mathematical notion of probability that is based on Kolmogorov's axioms.

quite a bit of statistics is involved. Casework aimed at individualisation, where a trace is compared to a certain characteristic or object associated to the suspect, has the particular property that the attention is turned towards the relation between the trace and one specific object or individual. The lawyer, for example, wants to know the probability that this particular shoe of suspect Jones has caused the shoe print. Because of this the Bayesian or subjective notion of probability is much better suited to the situation than the frequentist or mathematical notion of probability based on Kolmogorov's axioms. In a forensic context it is very artificial to see a probability as the ratio of the number of times that Jones' shoe made the print when one hundred thousand prints are made, or as some abstract probability measure. Often the police find a number of pieces of forensic evidence in one case. At a burglary, for example, fingerprints, shoe prints, DNA traces and tool marks can be found. A specific forensic question is then not only what the evidential value of the different pieces of evidence are but also what the evidential value is of the combination of all these pieces. In this case the Bayesian model offers a directive, as described above. The legal context in which the investigation is always conducted also determines a number of characteristic properties. For example, in general the expert knows what hypothesis the prosecutor will present during the trial but usually not exactly what hypothesis the defender will bring forward. Moreover, the hypotheses can change at any time if new information becomes available or if the defence unexpectedly comes up with an alternative scenario. The expert can of course himself come up with alternative hypotheses. Once the hypotheses have been stated, the assessment of the prior odds is left to the judge. While most scientists want to make statements concerning the posterior odds of the hypotheses they are studying (for example: this is the probability that the patient has a certain illness, given the symptoms), according to the Bayesian theory described above, the forensic expert should restrict his statements to reporting the evidential value. The legal context also causes the forensic expert to be interested in a different type of error when reviewing hypotheses in the traditional manner. When comparing the refractive index of glass fragments found in the clothing of the suspect to the refractive index of samples of the smashed-in window at a burglary site, the usual null hypothesis is that the two populations do not differ (Rudin and Inman 2003). In general, the statistician is then interested in a type I error, the probability of rejecting the null hypothesis when it is actually true, and the procedure is aimed at controlling this error, for example keeping it smaller than 5% or 1%. This error is associated with the probability that the suspect is wrongfully acquited. Indeed, this is the probability that the expert concludes that the glass fragments do not come from the window while they in fact do. For experts this is not the most important error. It is more important to control the type II error, the conclusion that the glass fragments come from the window while this is not the case, which can lead to a wrongful conviction.

Other research in forensic statistics

Statistics is used much more often in legal applications than people realize. Sometimes it forms the basis for an expert's conclusion, for example in criminal psychology, which considers, among other things, the evidential value of recognition by witnesses, or of the result of a polygraph test (Van Koppen et al. 2002). The evidential value of scent identification tests with specially trained dogs is also in part based on statistical considerations (Schoon and Van Koppen 2002). A completely different area where statistics plays a role is in assessing the probability of recidivism in TBS patients (persons who have been admitted, involuntarily, to a forensic psychiatric hospital in the Netherlands) (De Ruiter 2002, Brand and Diks 2001). A great variety of questionnaires can be used for this, which combined with a clinical evaluation by the psychiatrist or psychologist lead to an estimation of this probability. Subsequently, this is used by the lawyer in his judgment of whether TBS must be extended. Probability theory also contributes to criminal law, for example in determining whether a game must be regarded as a game of skill or one of chance (Van der Genugten et al. 2001). Other examples are the probability analysis carried out in cases such as that of Bianca K., the child care worker who was accused of causing a special form of suffocation in a number of children whom she was responsible for at a daycare centre. A similar well-known case is that of Lucia de B., the nurse suspected of being responsible for a number of enigmatic deaths in her department. In this type of case, the size of the probability that certain observations are based only on chance plays an important role (Elffers 2003).

Solution: answer 2 is correct. The evidential value of the DNA match is increased by the database search. However, the probability that the matching person is the donor of the saliva trace may still be relatively small in such cases. Indeed, if the suspect is identified by a database search, there need not be any other evidence against the suspect. In other words, the prior odds may be extremely small, much smaller than in the situation where DNA is compared from a suspect who is identified by other evidence. Hence, the only thing special about a database search is that the other evidence may be completely missing, or only in favour of the suspect. See Meester and Sjerps (2003, 2004) and the references mentioned there.

References

1 Aitken C.G.G., *Statistics and the evaluation of evidence for forensic scientists*, Wiley, Chichester, 1995.

2 Aitken C.G.G. and Stoney D.A. (eds) *The use of statistics in forensic science*, Ellis Horwood Limited, Chichester, 1991.

3 Brand E.J.F.M. and Diks G.J.M. 'Richtlijnen voor risicotaxatie in de forensische diagnostiek: theorie en praktijk'. *Tijdschrift voor psychiatrie* (**43**) 10 (2001), p. 693–704.

4 Broeders, A.P.A. *Op zoek naar de bron — over de grondslagen van de criminalistiek en de waardering van het forensisch bewijs*, (PhD thesis) Kluwer, Deventer, 2003.

5 Coster van Voorhout J.A. and Sjerps M. (eds) *Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht*, Kluwer, 2005.

6 Curran J.M., Hicks T.N., Buckleton J.S. *Forensic interpretation of glass evidence*, CRC Press, Boca Raton, 2000.

7 Elffers H. 'Bij toeval veroordeeld? — statistische analyse van dienstroosterdata in het strafproces', *Nederlands Juristen Blad* (**34**), (2003), p. 1812–1814.

8 ENFSI (European Network of Forensic Science Institutes) *Drugs Working Group, Guidelines on representative drug sampling*, 2003.

9 Evett I.W. , 'Avoiding the transposed conditional', *Science & Justice* (**35**), (1995), p. 127–131.

10 Evett I.W. and Weir B.S., *Interpreting DNA-evidence*, Sinauer associates, Sunderland MA, 1998.

11 Finkelstein M.O. and Levin B. *Statistics for lawyers*, Springer-Verlag, New York, 1990.

12 Gastwirth J.L.(ed.) *Statistical science in the courtroom*, Springer-Verlag, New York, 2000.

13 Genugten B.B. van der, Das M. and Borm P.E.M., *Behendig gokken in en rond het casino* (+cd-rom), Academic Service, 2001.

14 Gilbert R.O. *Statistical methods for environmental pollution monitoring*, Van Nostrand Reinhold, New York, 1987.

15 Good P.I. *Applying Statistics in the Courtroom: A New Approach for Attorneys*, CRC Press, Boca Raton, 2001.

16 Kahneman D., Slovic P. and Tversky A., *Judgement under uncertainty: heuristics and biases*, Cambridge University Press, Cambridge, 1982.

17 Keith, L.H. (ed.) *Principles of environmental sampling, 2nd ed.*, American Chemical Society, USA, 1996.

18 Koppen P.J. van, Hessing D.J., Merckelbach H.L.G.J., Crombag H.F.M., *Het recht van binnen — psychologie van het recht*, Kluwer, Deventer, 2002.

19 Meester R and Sjerps M. 'The evidential value in the DNA-database search controversy and the two-stain problem', *Biometrics* (**59**), 2003, p. 727–732.

20 Meester R. and Sjerps M. 'Why the effect of prior odds should accompany the likelihood ratio when reporting DNA-evidence', *Law, Probability and Risk*, 2004.

21 Nijboer J.F. *Het gekooide denken — over de rol van discipline binnen feitenonderzoek en bewijs in het recht*, Inaugural lecture, Universiteit Leiden, 18–10–2002.

22 Patil G.P. and Rao C.R. (eds.), *Handbook of statistics 12: Environmental statistics*, North Holland, Amsterdam, 1994.

23 Robertson B. and Vignaux G.A. *Interpreting evidence*, Chichester, Wiley, 1995.

24 Rudin N. and Inman K. 'Articulating Hypotheses-The Null Hypothesis and Beyond', *News of the California association of criminalists*, 1st Quarter, (2003), p. 7–8.

25 Ruiter C. de, 'De terbeschikkingstelling', in: Koppen P.J. van, Hessing D.J., Merckelbach H.L.G.J., Crombag H.F.M., *Het recht van binnen — psychologie van het recht*, Kluwer, Deventer, 2002, p. 991–1001.

26 Schoon A. and van Koppen P.J. 'Identificatie door honden', in: Koppen P.J. van, Hessing D.J., Merckelbach H.L.G.J., Crombag H.F.M., *Het recht van binnen — psychologie van het recht*, Kluwer, Deventer, 2002, p. 597–622.

27 Sjerps M. and Kloosterman A.D. 'Statistical aspects of interpreting DNA-profiling in legal cases', *Statistica Neerlandica* (**57**) 3, (2003), p. 368–389.

28 Stelling, M. and Sjerps M. 'A sampling and hypothesis testing method for the legal control of solid hazardous waste', *Environmetrics* (**10**), (1999), p. 247–259.

29 Uitvoeringsregeling bouwstoffenbesluit bijlage F; Handhavingsprotocol schone grond, Staatscourant 30 januari 1998.

30 Websites: Federal Judicial Centre: www.fjc.gov Charles Brenner: www.DNA-view.com.