



---

Legal and Statistical Aspects of Some Mysterious Clusters

Author(s): Stephen E. Fienberg and D. H. Kaye

Reviewed work(s):

Source: *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 154, No. 1 (1991), pp. 61-74

Published by: [Blackwell Publishing](#) for the [Royal Statistical Society](#)

Stable URL: <http://www.jstor.org/stable/2982695>

Accessed: 23/05/2012 12:28

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing and Royal Statistical Society are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society. Series A (Statistics in Society)*.

<http://www.jstor.org>

## Legal and Statistical Aspects of some Mysterious Clusters

By STEPHEN E. FIENBERG†

and

D. H. KAYE

*Carnegie Mellon University, Pittsburgh, USA*

*Arizona State University, Tempe, USA*

[*Read at the International Conference on Forensic Statistics held in Edinburgh, April 2nd–4th, 1990*]

### SUMMARY

Criminal and civil trials often involve events that appear to cluster together in time or space, and the existence and size of the cluster often is interpreted as implying that the occurrence of the events could not be a coincidence. This paper examines the statistical evidence introduced in several cases to show how such mysterious clusters should be interpreted. The paper considers this form of evidence in the context of legal views on the admissibility of evidence about ‘similar events’, and it suggests a more formal statistical argument that might be used to justify admissibility in one category of cases.

*Keywords:* CLUSTERS OF EVENTS; EXPERT WITNESSES; LATENT VARIABLES; LAW; SIMILAR EVENTS; STATISTICAL TESTIMONY

### 1. INTRODUCTION

The finely chiselled lips parted. He said, ‘Mr Bond, they have a saying in Chicago: “Once is happenstance. Twice is coincidence. The third time it’s enemy action.”’ These words of the arch villain in Ian Fleming’s novel *Goldfinger* capture a principle firmly etched in the Anglo-American law of evidence: where a purely accidental occurrence would not create liability, proof of similar, prior incidents may be admitted to show that the occurrence is not mere happenstance.

In this paper, we describe the situations under which courts allow proof of similar prior incidents to disprove a claim of coincidence. Then, we consider some recent American cases that follow the trend towards greater reliance on forensic statistical assessments described in a National Research Council panel report (Fienberg, 1989) by allowing expert testimony on the probability of an accidental string of similar incidents. Finally, we examine some of the statistical aspects of these calculations and their legal relevance. Our intent is to raise legal and statistical questions regarding the role of a particular form of statistical evidence in actual legal cases, to examine the fit between legal theory and the statistical methodology that has been invoked in such cases and thereby to stimulate further legal and statistical thinking about apparent clusters in legal settings.

### 2. SIMILAR HAPPENINGS AND TRANSACTIONS: ADMISSIBILITY OF EVIDENCE OF CLUSTERS

Common law prefers each party to a lawsuit to prove its case strictly by evidence of what happened on the occasion in question. Although proof of happenings and

†*Address for correspondence:* Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

transactions from other places and other times may shed light on the events that are the subject of the litigation, this proof also may be distracting or prejudicial. The resulting 'extrinsic evidence limitation', as we may call it, is frequently litigated in criminal cases. Much of the early precedent is described in Eggleston (1983). Lord Herschell LC described the general principles in a famous dictum in *Makin versus Attorney General for New South Wales* (1894) App. Cas. 57, 65 (P. C. 1893):

'It is undoubtedly not competent for the prosecution to adduce evidence tending to show that the accused has been guilty of criminal acts other than those covered by the indictment, for the purpose of leading to the conclusion that the accused is a person likely from his criminal conduct or character to have committed the offence for which he is being tried. On the other hand, the mere fact that the evidence adduced tends to show the commission of other crimes does not render it inadmissible if it be relevant to an issue before the jury, and it may be so relevant if it bears upon the question whether the acts alleged to constitute the crime charged in the indictment were designed or accidental, or to rebut a defence which would otherwise be open to the accused.'

That is, where the past crimes are relevant *only* as evidence of the defendants' disposition toward such crimes, they may not be proved in the prosecution's case in chief. Where past crimes are relevant on some other theory, such as refuting a claim that the allegedly criminal acts at bar were performed innocently, they may be proved as long as their probative value for this purpose is sufficient. Thus, 'other crimes' evidence is not forbidden because it lacks probative value in showing the conduct in the particular case. Rather, it is excluded despite its probative value—either because the proof of the similar happenings would be unduly time consuming or uncertain, because the extrinsic events prove less than the jury might think, or because the jury might be too willing to impose liability on an innocent person who has engaged in other blameworthy conduct.

In the *Makin* case, the Privy Council acknowledged that the application of these 'clear' principles to the facts 'is by no means free from difficulty'. The remains of 13 infants were discovered in places where the Makins were living or had lived, and the Crown charged the Makins with the murder of two of these children. One was identified by his clothing and hair. His mother testified that the Makins had agreed to adopt her son in exchange for only £3. The evidence also included some damaging admissions and suspicious behaviour plus testimony of other mothers whose children had disappeared after these women had left them with John and Sarah with payments that were too small to support them for very long.

The jury convicted the Makins of murdering the boy whose remains had been identified. On appeal, the couple argued that all the evidence concerning other missing children should not have been admitted. The Privy Council rejected this argument. Although its opinion did little to explain the basis for this conclusion, counsel for the Crown had stressed that 'the recurrence of the unusual phenomenon of bodies of babies having been buried in an unexplained manner in a similar part of premises previously occupied' implied that the deaths were 'wilful and not accidental'.

The general principles enunciated in the *Makin* case, as well as the 'not accidental' reasoning, continue to be applied, with varying success, on both sides of the Atlantic (for example, see Cross (1979), Imwinkelreid (1984) and Cleary (1984)). Perhaps the most famous application of the *Makin* case occurred in *Rex versus Smith*, 11 Cr. App.

R. 229, 84 L.J.K.B. 2153 (1915). In this 'brides in the bath' case, George Joseph Smith was accused of drowning Bessie Mundy in the small bathtub of their quarters in a boarding house. Mundy had left all her property to Smith in a will executed after their 'marriage' (Smith was already married). The trial court allowed the prosecution to prove the deaths of two other women who had gone through marriage ceremonies with Smith and to argue that the circumstances surrounding their deaths in their bathtubs were remarkably similar (see Marjoribanks (1929)). The Court of Criminal Appeal affirmed the resulting conviction on the ground that the evidence in connection with Mundy's death alone made out a *prima facie* case, and the other incidents were properly admitted 'for the purpose of shewing the design of the appellants'.

Undoubtedly, the probabilities in the Makin and Smith cases are not easily computed, but the effect of proof of similar events on the posterior probability can be shown in an idealized situation. Let  $G = 1$  be the event that the accused is guilty of act 0; let  $X_0 = 1$  be the event that act 0 occurred and  $X_i = 1$  be the event that act  $i$  occurred for  $i = 1, 2, \dots, k$ . Then under suitable conditions on the positive correlations among  $G$  and the  $\{X_i\}$  we show in Appendix A that

$$P(G = 1 | X_0 = 1, X_1 = 1, \dots, X_k = 1) > P(G = 1 | X_0 = 1). \quad (1)$$

Thus, there is inferential value (see Fienberg and Kadane (1983) and Fienberg and Schervish (1986)) in the knowledge that  $k$  other acts similar to act 0 occurred, even if we have no direct link between them and  $G$  or between them and the accused's commission of acts 1, 2,  $\dots, k$ .

### 3. ESTIMATING PROBABILITIES FOR CLUSTERS

In some situations, it has been feasible to quantify related probabilities such as  $P(X_0 = 1, X_1 = 1, \dots, X_k = 1 | G = 0)$ . Although courts often express misgivings about 'probability evidence' in criminal cases, several recent cases have allowed a conditional probability for a cluster of mysterious events to be presented.

#### 3.1. 'Crib Deaths' at a Baby-sitter's Home

A modern incarnation of the Makin case is *State versus Pankow*, 144 Wis. 2d 23, 422 N.W.2d 913 (App. 1988). Sandra Pankow provided baby-sitting in her home in Appleton, Wisconsin. Some babies were kept in a crib or play-pen in the basement. Older children whom she watched testified that the infants sometimes were confined by sheets, blankets or boards placed over the crib or play-pen and that, when some babies cried, they were put in the basement with towels tied over their mouths. One baby died in 1980 and another in 1982, apparently of sudden infant death syndrome (SIDS). When a third baby died in Pankow's home in 1985, the county coroner arranged for an autopsy. University pathologists tentatively determined the cause of death to be SIDS, but ultimately attributed the death to asphyxia. The bodies of the other two infants were exhumed. The pathologists concluded that these deaths also resulted from asphyxia. Pankow was charged with three second-degree murders, and a jury convicted her of two of them.

The prosecution adduced considerable medical testimony. One pathologist opined that there was a 95–97% probability that homicidal asphyxiation caused the death of the third child and a probability in excess of 99% for another. A paediatric

pathologist from Sheffield characterized the pattern of deaths as most likely an 'abnormal psychosocial reaction to a crying situation of the child'. A medical examiner reviewed the autopsy reports and supporting materials and said that 'I would cluster them all together as asphyxial deaths due to external asphyxiation. . . . I would call it homicide.'

The statistical testimony in the case came from Robert Hauser, a professor of sociology at the University of Wisconsin. Hauser estimated the probability of occurrence of three or more deaths attributable to SIDS in the same household during a 5-year period, assuming that the occurrences of deaths are independent and that there is no common cause. He was supplied with certain generally accepted data on SIDS: two SIDS deaths occur per 1000 live-births; 90% of SIDS deaths occur in children under 6 months of age; 90% of SIDS deaths occur between midnight and 9 a.m. Then, using a binomial argument with  $n = 20$  (for the number of children in the Pankow home) and  $p = 0.00002$  (obtained from the fact that the three deaths occurred at 6 months or more of age during daylight hours and the assumption that time of day and age are independent of death), he determined that the probability of three or more deaths would be  $0.91 \times 10^{-13}$  and observed that:

'There are about 3,600,000 infants born in the United States every year. Suppose that we took each one of these infants [as] surviving to 6 months and purely by chance assigned them to baby sitters in groups of 20. That means that for each new set of births in the course of a year, we would have about 180,000 baby sitters each with 20 infants in their care. Let's suppose further that those infants were cared for full time, all the time, 24 hours a day from the age of 6 months onward. Then the rate of 9.1 in a trillion means that we would expect to observe as many as three deaths among the charges of one baby sitter about once in 600 years.'

His report included several other probability calculations that varied the values of  $p$  (0.002, 0.0002 and 0.00002) and  $n$  (20 and 25), and his testimony introduced a Poisson approximation to the binomial probabilities.

The defence presented a videotaped deposition of a physician with a Master's degree in epidemiology. This expert criticized Hauser's written report and stated that a Poisson calculation was more appropriate, but he did not give any details. When asked to write down the probability function of the Poisson calculation, he was unable to do so; nor could he explain why it was the proper distribution to use.

On appeal, Pankow argued that 'in all cases it is error to permit an expert witness to testify as to mathematical probabilities that are offered to show that the defendant was the person who committed the crime'. The Appellate Court's rejoinder that '[s]tatistical evidence is not inadmissible *per se*' finds support in the decisions of every jurisdiction (save one) that has passed on the admissibility of probabilities in criminal cases (Kaye, 1987). Nevertheless, more discriminating objections to the statistical modelling might have been attempted. Is the incidence of SIDS deaths uniform across the country and across ethnic groups? Does confining a child to a play-pen with a wooden board that prevents the child from standing up increase the risk of SIDS? Is the time of day at which these deaths occur independent of the age of the affected infants (as the use of  $0.002 \times 0.1 \times 0.1$  for the binomial probability presupposes)? Although we can always question the precise figures introduced, the probability of a cluster of SIDS deaths like that in the Pankow case seems to be quite small, thus casting substantial doubt on the 'accidental' explanation of the deaths.

### 3.2. *Cardiac Arrests in Surgical Intensive Care Units*

#### 3.2.1. *United States versus Narciso*

*p*-values associated with suddenly elevated death rates in hospitals have triggered prosecutions of nurses in recent years. The earliest such case that we have located is *United States versus Narciso*, 446 F. Supp. 252 (E. D. Mich. 1977). In the summer of 1975, cardiopulmonary arrests at a Veterans Administration hospital in Ann Arbor, Michigan, rose to four times their usual rate (Stross *et al.*, 1976). An analysis of the hospital's records found no changes in the patient population that could account for the upsurge, and it established that the incidents were concentrated in the intensive care unit (ICU). A grand jury indicted two nurses, Filipina Narciso and Leonora Perez, for murder and related offences.

At a trial lasting 3 months, the government presented 89 witnesses, including 17 experts. By the end of the expert medical testimony, there was little doubt that many of the patients had received a muscle relaxant without prescription (446 F. Supp. at 307). The government sought to show that the defendants were the only people present when the drug must have been injected, but the testimony of the lay witnesses introduced to prove that the accused nurses were always present was 'confusing' and 'inconsistent'. Apparently, the epidemiological study was not used, and the government was prohibited from proving any respiratory arrests not charged in the indictment (446 F. Supp. at 322). After 13 days of deliberation, the jury found the nurses guilty of some of the poisonings (446 F. Supp. at 310).

Cataloging various instances of prosecutorial misconduct, the trial court set this verdict aside. In the closing argument the prosecutor asked: 'What are the odds, ladies and gentlemen, what is the chance, what is the probability that these Defendants have engaged in these activities and that all these factors that are incriminating could exist and the Defendants would still nevertheless be innocent?' (446 F. Supp. at 323). This argument, the court held, was a 'most egregious error', for it invited the jury 'to engage in a speculative combination of the charges' in the face of the court's instruction that '[e]ach charge, and the evidence pertaining to it must be considered separately. You may not consider evidence introduced as to one count in arriving at a verdict on any other count'.

This atomized treatment of the evidence precludes the legitimate use of the clustering of cardiopulmonary arrests to suggest that some criminal misconduct has taken place. However, the medical testimony about muscle relaxants proves this point more directly, and the clustering here does not show that the two nurses were responsible. In any event, the United States attorney chose to have the indictment dismissed rather than to undertake a second trial (United Press International, 1978).

#### 3.2.2. *Rachals versus State*

When the number of cardiac arrests of patients at a hospital in Georgia surged in late 1985, Adelle Franks, an epidemiologist at the US Center for Disease Control (CDC) in Atlanta examined the records for that year. Looking at the frequency of these incidents in most months of the year, she determined that the usual incidence of cardiac arrests at the hospital ranged from 0 to 4, but that, in November, 11 cardiac arrests occurred on the 3–11 o'clock shift. According to the Court of Appeals in

Rachals *versus* State, 184 Ga. App. 420, 61 S.E.2d 671 (1987), *aff'd*, 364 S.E.2d 671, cert. denied, 108 S. Ct. 2909, she reported that the probability of this occurring 'by chance alone is less than one in a trillion'. The cardiac arrests were concentrated among the patients under the care of one surgical nurse, Terri Rachals.

Rachals was charged with six counts of murder and 20 counts of aggravated assault. The state contended that she administered potassium chloride (KCl) to patients in intensive care, causing every cardiac arrest that occurred while she was on duty during the period under investigation. The jury acquitted Rachals of all the murders and 19 of the 20 alleged assaults. However, Rachals had confessed to injecting 20 ml of KCl into blood plasma for one 'very, very ill' patient who had asked her to 'let him die'. For this patient, the jury found her 'guilty but mentally ill' of an aggravated assault.

Rachals appealed on various grounds, including the admissibility of the epidemiologist's testimony. The Court of Appeals summarized this testimony as follows:

'In the month of November, five cardiac arrests had occurred in one day and one patient had a total of eight cardiac arrests in that one month. Dr. Franks listed all cardiac arrest patients for the period investigated and the primary nurse on duty with that patient. Rachals was the primary nurse for 11 cardiac arrest patients in the month of November. No other nurse was the primary care nurse for more than one cardiac arrest patient. Dr. Franks charted all 24 nurses for that month and the number of cardiac arrests that occurred when they were not on shift to calculate a "rate ratio." The "rate ratio" for most nurses was around one, while the "rate ratio" for Rachals was 26.6, which means that in 26.6 times, it was more likely that a cardiac arrest would occur while she was on duty than when she was not on duty. . . . [T]he rate ratio show infinitely large and unmeasurable [sic] because all of the cardiac arrests that occurred on the 3:00 o'clock to 11:00 o'clock shift occurred while she was on duty. 361 S.W. 2d at 674.'

We suspect that this rendition is garbled, and it appears that there was more to the testimony than this, for the court notes that 'by inference [it] could be interpreted to mean that "Terri Rachals was, by odds of five out of nine, probably guilty"'. The court expressed 'serious reservations about mathematical computations as to the probability of guilt', but felt 'constrained' to uphold the admission of the testimony by virtue of a Georgia Supreme Court opinion allowing unspecified 'mathematical computations' concerning fibre evidence to be used in *Williams versus* State, 251 Ga. 749, 312 S.E. 2d 40 (1983). The Georgia Supreme Court affirmed Rachal's conviction without discussing the epidemiologist's testimony.

### 3.2.3. *State versus Bolding*

In late 1984, workers in the ICU of Prince George's Hospital Center in Maryland noticed an unusual number of cardiopulmonary resuscitation (CPR) incidents involving patients cared for by Jane Bolding. Some of these patients had multiple cardiac arrests and abnormally high levels of potassium. In March 1985, after one patient suffered six arrests on Bolding's shifts, the hospital suspended Bolding from work, and the patient recovered. Two weeks later, after 23 hours of intensive police interrogation, Bolding confessed to killing two patients. She was charged with murder, but the charge was dismissed because of doubts about the admissibility of the confession and the lack of corroborating evidence (Weaver, 1988).

The investigation, however, continued. Maryland authorities enlisted the aid of the CDC. The findings of the CDC epidemiologists are reported in Sacks *et al.* (1988) and a more detailed, unpublished CDC report (Centers for Disease Control, 1985). After seeing the CDC report, a grand jury indicted Bolding for two murders and seven attempted murders. A judge ruled the confession to have been coerced and the fruit of an illegal arrest (Weaver, 1988). Deprived of the confession, the State made the statistical analysis the linchpin of its case. Using a logistic regression of cardiac arrests on age, sex, severity of illness and post-operative status, Sacks *et al.* (1988) found that Bolding's patients were 47.5 times more likely to experience arrest than were those of other nurses and that the epidemic ceased when Bolding left. At the trial, Sacks testified that '[t]he chances of [the large number of cardiac arrests during the epidemic period] happening by chance is about one in 100 trillion'. This, he added, 'would be like picking out one second from all of time'. To establish further the fact of wrongdoing, a forensic pathologist testified that he was 99% certain that the high levels of potassium found in the alleged victims came from unauthorized injections (Harrison, 1988a).

At the same time, the original report cautioned that 'statistical analysis cannot answer whether or not intentional acts were committed against patients. No matter how strong, epidemiologic associations of cardiac arrests with nurse attendants cannot address factors such as exclusive access to patients or intent.' During five hours of cross-examination, Sacks stated that the association between cardiac arrests with abnormal levels of potassium and Bolding's was 'consistent with intentional actions', but in response to a defence study suggesting that a physician's assistant who had testified against Bolding could have administered KCl to Bolding's patients, Sacks conceded that '[i]t's not as plausible and consistent as [Bolding] being the greater risk factor, but it is plausible and consistent' (Harrison, 1988b).

At the close of the State's case, the court granted a defence motion for acquittal (Harrison, 1988c). This result does not depart from the legal doctrine on the admissibility of evidence of other crimes. The no accident logic of the Makin case justifies only the introduction of a generally disfavoured type of evidence. It does not require that this evidence be believed or that it be dispositive. The evidence on the many cardiac arrests and their association with Bolding as opposed to other nurses was admitted.

### 3.3. *Deaths in a Paediatric Intensive Care Unit*

An increased mortality rate in a paediatric ICU in a San Antonio, Texas, hospital led to the conviction of another nurse. When 82 patients died in a 2½-year period, Gregory Istre, an epidemiologist with the CDC, pored over the charts. According to the Court of Appeals in *Jones versus State*, 751 S.W.2d 682, 683-84 (Tex. App. 1988), he was 'able to eliminate a number of variables such as age, race, sex, medical history, severity of illness, procedures, surgery, surgeons and therapeutic intervention as explanation . . .'. The timing of the deaths pointed to only one nurse, Genene Jones. 'The investigators determined that a child would have 10.7 times the risk of dying while appellant was working than at the times she was not working'. Likewise, 'a CPR event was 25 times more likely to occur when appellant was working', and 'as to 8 of the 9 patients who had *recurrent* CPRs on different shifts in the epidemic period, appellant was assigned to their care during each CPR episode'. A more complete description of the study appears in Istre *et al.* (1985).



Jones was indicted for the offence of injury to a child in connection with one of these events. The case was tried without a jury. The court convicted Jones and sentenced her to 60 years confinement. The evidence showed that Jones was familiar with the anticoagulant Heparin, that a 4-week-old baby who had been admitted to the hospital because of pneumonia experienced repeated cardiopulmonary arrests and overdoses of Heparin, that a nurse who suggested to her supervisor that Jones was connected to the unusual CPRs found a note in her mail-box in Jones's writing stating 'You're dead', and that Jones had identified herself to another prisoner in the county jail as 'the nurse that killed the babies'.

On appeal, Jones argued that evidence of the other incidents was erroneously admitted because there was no proof that she was responsible for any of those CPRs. One judge accepted this argument, insisting that '[u]nless we accept mere presence as evidence of guilt, the elaborate statistical structure built by Dr. Istre proves nothing concerning the guilt of appellant' (751 S.W. 2d at 687). This, however, overlooks or rejects the logic of the no accident reasoning of the Makin case and many other cases. The defence counsel in the Makin case likewise had insisted that the 'bodies were not shown to have been the bodies of children committed to the care of [the Makins]'. The evidence was admitted in the Makin case precisely because the 'mere presence' of so many bodies in places associated with the Makins was indicative of guilt. The logic of the no accident theory is not of the form that 'if X deliberately committed act A, then X deliberately committed act B'. The reasoning is that, even if there is a reasonable doubt that X committed act A (when this event is viewed in isolation) and a reasonable doubt that X committed act B (when that event is seen in isolation), the jury may reasonably believe that X committed both acts. Even the majority of the Texas court seemed ignorant of this well-established exception to the extrinsic evidence limitation. These judges felt compelled to resort to an *ad hoc* balancing of probative value and prejudice to justify the trial court's admission of the evidence.

The upsurge in deaths and cardiopulmonary arrests extending into 1982 in San Antonio was not the only mysterious cluster in which Genene Jones was implicated. In August of that year, Jones worked in the office of a paediatrician in Kerrville, Texas. Within a 1-month period, six patients under Jones's care suffered seven respiratory arrests. The state charged Jones with murdering the first of these children. The prosecution produced toxicological and circumstantial evidence that Jones had repeatedly injected the child with succinylcholine chloride, a curare-like muscle relaxant, producing respiratory arrest leading to death. It argued that Jones injected all six children to dramatize the need for a paediatric ICU in the community. No statistical analysis was undertaken, but evidence of all the incidents was admitted. A conviction and a 99-year sentence followed. The Court of Appeals in Austin upheld the admission of evidence about the entire cluster under several theories, including the one that we have emphasized here: 'the pattern of offenses tends to negate the explanation that the incidents were other than deliberate actions on her part and that they were the result of natural causes or negligence'; Jones *versus* State, 716 S.W. 2d 142, 161 (Tex. App. 1986).

#### 4. SERIAL CRIMES AND OTHER CLUSTERS: SOME ILLUSTRATIONS

In this section we briefly examine some links between the topics considered in the previous section and those in two other legal domains—serial crimes and toxic torts.

The unifying feature is the focus on a series of events with common features grouped together in space or time.

#### 4.1. *Commonwealth versus Jamieson*

In a 1987 criminal trial in the Common Pleas Court in Pittsburgh, Pennsylvania, Joseph Jamieson was accused of committing seven rapes over a 10-month period during 1985–86. The rapes were linked by a seemingly common pattern and by genetic analysis of secretions on the victims, their bed sheets and clothing. A forensic expert testified for the prosecution that the fragmentary evidence from each case was consistent with there being a single perpetrator, that 0.32% of the population could have deposited seminal stains consistent with the composite evidence and that the defendant's blood and enzyme markers put him in this group. Despite criticisms of this statistical evidence presented by a statistician (Stephen Fienberg) called by the defendant, the jury convicted on all seven charges, and a juror noted later that the genetic evidence was especially compelling (Fienberg, 1990; Fienberg and Straf, 1991).

In this case, the choice of which rapes to link is similar to the selection of events in the clusters in the cases considered in previous sections. However, the court and the parties paid little or no attention to the evidentiary theory that might justify the admission of the similar acts evidence, and the no accident logic does not seem to apply. Another exception to the ban on extrinsic evidence, allowing proof of a common scheme or plan (Cleary (1984), p. 559), may apply to this, and to many other serial crimes, but discussion of this point is beyond the scope of the present paper.

#### 4.2. *Woburn Water Case*

In January 1972, 3-year old Jimmy Anderson was diagnosed as having acute lymphocytic leukaemia, and his parents began to search for a cause. Ultimately, they identified a 'cluster' of 12 childhood leukaemia cases in the East Woburn, Massachusetts, neighbourhood in which they lived, and they 'linked' these to contaminated water in two of the city's eight wells. In May 1982, 11 Woburn families filed suit in the Federal District Court against W. R. Grace (through its Cryovac division) and Beatrice Foods claiming that poor waste disposal led to ground-water contamination through the two wells that caused the fatal cases of leukaemia in their families.

A statistical study (Lagakos *et al.*, 1986) found a positive association between access to the water from the two wells and the incidence rate of childhood leukaemia, using the results of analyses based on a proportional hazards model with time varying covariates, as well as positive associations with several other medical disorders. This study proved highly controversial. Several epidemiologists and statisticians presented critiques of its methodology and discussed whether the positive associations should be viewed as evidence for a causal link between the contaminated water and the occurrence of the disease.

The trial (Anderson *versus* Cryovac, Inc.) began in 1986 and was divided into three stages. When complications arose with the jury's special verdict at the first stage, the families reached a settlement of \$8 million with W. R. Grace (the case against Beatrice Foods already having been resolved in that company's favour). As a result, the statistical evidence regarding the mysterious cluster of leukaemia cases and the association with the contaminated wells was never presented in court. This evidence raises many of the same questions of selection and linkage that arose in such earlier mass exposure

cases as the Agent Orange litigation and United States *versus* Allen (for example, see Fienberg (1989), pages 131–136).

In many ways, these issues are related to those considered in Section 2. The principal difference is that in the similar facts cases there ordinarily is no need for independent proof of a causal relation between the similar events. Because a reasonable inference of causation can flow from the cluster itself, the unusualness of the cluster normally is sufficient to allow the evidence into the record. The issue then becomes the weight accorded to this evidence in assessing the guilt or innocence of the defendant in each specific instance. In contrast, in the Woburn water case, the law would have required evidence to support a causal link between the pollution and the leukaemia cases, in a collective sense. The issue of the possible selective identification of clusters provides a key statistical tie between the two types of case.

## 5. OVERVIEW OF SOME STATISTICAL ISSUES

In virtually every one of the cases described here involving evidence of other events linked to a specific criminal charge, the overriding statistical issue is selection bias. How were the similar events selected? From what population were the similar events chosen? In the Pankow case, the prosecution's expert defined the population in terms of the number of children cared for by the accused over a given period of time. We might question the choice of time frame and ask about the evidence to support the numbers 20 or 25 children used in the calculations. In the Bolding case, a cluster analysis isolated the period of the increase in cardiac arrests. How unusual would the cluster of events appear if set against a larger span of time? What about unusual clusters of other types of medical emergencies over the same period of time? In the Jamieson case, were there any other rapes during the period of time in question that were originally associated with those involved in this case but that were eliminated because the evidence did not match that of the common blood profile linked to the defendant? The calculation of a  $p$ -value as a measure of surprise regarding the occurrence of a cluster of unusual events inevitably triggers the suspicion of reporting bias. After all, every event in a large discrete sample space is rare, but some appear to us as rarer than others.

There is yet another way to look at issues of selection. Following the occurrence of several of the cases described in Section 3, many public health officials argued that clusters of mortality would occur less frequently if there were routine monitoring of in-hospital mortality. We then must ask what would happen if we looked at all the nurses and doctors in the country. How many of them would be associated with clusters of unusual deaths in a given year? What is the probability that a nurse or doctor will have one or more such clusters over the course of a career? Essentially, the question is whether the probative value of the evidence of a cluster depends on how it is collected. We believe that the answer is clearly yes. Although we do not maintain that selection bias vitiated the analyses in many of the cases surveyed here or that foul play was not apparent in these cases, we do believe that when the added evidence of similar events is needed to make out a convincing case against a defendant, as it often is, it is important to consider how this set of events came to be designated as similar.

Finally, we turn to the form of presentation of statistical evidence in cases involving similar events to rebut the suggestion of coincidence. The traditional legal role of such evidence to demonstrate the implausibility of the explanation that the occurrence of

all the events was accidental fits rather naturally into the frequentist calculation of a  $p$ -value under the null hypothesis that the events are independent. Thus, virtually all the statistical evidence presented in these cases involved the calculation of  $p$ -values associated with the occurrence of clusters at random. However, there is a rather large gap between providing such statistical evidence and determining how it should affect the calculation of the probability of guilt. As we noted, there is a more direct way to view the evidence of similar events and its probative value through calculation in a Bayesian framework. The evidence from cases such as Pankow, Bolding and Jones needs to be re-examined in this framework, and questions about the appropriate likelihood function for each individual case need to be addressed (cf. the discussions of the relevant likelihood functions for the Jamieson case in Fienberg (1990)). We plan to do so at a later time.

#### ACKNOWLEDGEMENTS

Preparation of the present paper was supported in part by the Hebrew University, where one of us (SEF) served as Berman Visiting Professor in the Department of Statistics, and by the University of Chicago School of Law, where DHK was a Visiting Research Fellow. We thank Robert Hauser for providing us with transcripts and other materials from State *versus* Pankow, Joel Tarr and Marvin Zelen, who provided information on the Woburn water case, and Zvi Gilula, Louis Gordon and Ester Samuel-Cahn for discussion relating to the arguments in Appendix A.

#### APPENDIX A: SOME PROBABILITY INEQUALITIES FOR SIMILAR EVENTS

In this appendix we present a more detailed description of probabilistic inequalities that we believe are consistent with the no accident rationale for admissibility of evidence regarding events similar to those linked to the act or acts under litigation. We present the arguments in a Bayesian-like context where the quantity of interest is the posterior probability of guilt (see Fienberg and Kadane (1983) and Fienberg and Schervish (1986)). The additional evidence is deemed to be relevant if the portion of the likelihood linked to it is such that the evidence changes the probability of guilt, i.e. if the likelihood ratio is different from unity.

We begin by defining various random variables linked to the legal setting involving similar happenings and transactions. We let the happening associated with the litigation be event 0 and assume that there are  $k \geq 1$  additional events. Then we let

$$G = \begin{cases} 1, & \text{if the accused is guilty,} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$X_i = \begin{cases} 1, & \text{if the } i\text{th event occurs,} \\ 0, & \text{otherwise,} \end{cases}$$

for  $i = 0, 1, 2, \dots, k$ .

We now introduce a latent (unobserved) random variable  $Y$  that is an indicator for the tendency of the accused to engage in behaviour that produces events such as those in question. We let

$$Y = \begin{cases} 1, & \text{if the tendency is present,} \\ 0, & \text{otherwise.} \end{cases}$$

Clearly,  $Y$  is positively correlated with each of the  $\{X_i\}$ , and  $Y$  is positively correlated with  $G$  if  $X_0 = 1$ .

*Assumption 1.* Given  $X_0$  and  $Y$ ,  $G$  is independent of  $(X_1, X_2, \dots, X_k)$ .

Thus, once we know that event 0 has occurred and that the accused has a tendency to commit such acts, our assessment of  $G = 1$  is not influenced by  $X_1, X_2, \dots, X_k$ . Since we never observe  $Y$ , however, we must use  $X_1, X_2, \dots, X_k$  given  $X_0$ .

*Assumption 2.*  $X_0, X_1, \dots, X_k$  are conditionally independent given the latent variable  $Y$ .

Assumption 2 is standard in the literature on latent variables (for example, see Gilula (1979) and Goodman (1974)). Although it might be possible to relax it somewhat, the conditional independence is what makes us believe that additional events add evidence about  $Y$  and thus about  $G$ .

Now we give a formal representation to the positive correlations described intuitively above. Let

$$\alpha_j = P(G=1 | X_0=1, Y=j), \quad j = 0, 1,$$

and

$$\beta_{ij} = P(X_i=1 | Y=j), \quad i = 0, 1, 2, \dots, k, \quad j = 0, 1.$$

Then we make the following monotonicity assumption.

*Assumption 3.*  $\alpha_1 > \alpha_0, \beta_{i1} > \beta_{i0}, i = 0, 1, \dots, k$ .

*Theorem.* Under assumptions 1-3,

$$P(G=1 | X_0=1, X_1=1, \dots, X_k=1) > P(G=1 | X_0=1, X_1=1, \dots, X_{k_1}=1)$$

where  $k > k_1 \geq 0$ .

*Corollary.*  $P(G=1 | X_0=1, X_1=1, \dots, X_k=1) > P(G=1 | X_0=1)$ .

The proof of the theorem is straightforward and follows from the representation

$$P(G=1 | X_0=1, X_1=1, \dots, X_k=1) = \frac{\theta \alpha_1 \prod_{i=0}^k \beta_{i1} + (1-\theta) \alpha_0 \prod_{i=0}^k \beta_{i0}}{\theta \prod_{i=0}^k \beta_{i1} + (1-\theta) \prod_{i=1}^k \beta_{i0}}.$$

The theorem is in accord with the legal argument that the occurrence of events similar to those at issue in the litigation increases the probability that the accused is guilty in a manner other than the forbidden 'because D committed acts  $X_1, X_2, \dots, X_k$ , D committed act  $X_0$ '. The word 'similar' refers both to the nature of the events themselves and to the link of the events to the accused in some way. The lack of such linkage would undercut the plausibility of assumption 3 and thereby undercut the relevance of evidence. A second implication of the theorem is that, the larger the number of similar events, the stronger the evidence in support of the hypothesis that the accused is guilty.

The inequalities in the theorem bear a strong resemblance to those that arise among *associated* random variables in the sense of Essary *et al.* (1967). Although we can derive a result similar to the corollary by assuming that  $G, X_1, X_2, \dots, X_k$  are *associated*, we have been unable to produce a proof of a result similar to that in the theorem, which allows every additional similar event to increase the probability of guilt.

It is important to understand that, although the inequality depends on the existence of the

latent variable  $Y$ , we neither observe  $Y$  nor do we attempt to infer whether the accused does or does not have a criminal disposition ( $Y = 1$  or  $Y = 0$ ). Rather the inequality in effect averages over the totality of possible states and is true no matter what the value of  $\theta = P(Y = 1)$  as long as  $1 > \theta > 0$ . The introduction of  $Y$  as a latent variable is therefore unlike the propensity inference that the law forbids.

Finally, as a formal representation of the no accident logic but not of the forbidden propensity logic, the value of the theorem is purely heuristic. It justifies the intuition that has influenced many of the judicial opinions discussed here. Like the intuition that it explicates, the theorem by itself does not reveal when evidence of similar offences will be admissible. For example, one referee of this paper correctly pointed out that the theorem applies where the accused on a charge of burglary had two prior convictions for burglary. But it would apply only so far as to permit the prosecution to prove the two prior burglaries (like the previous bathtub deaths in *Rex versus Smith*) to show that the third burglary was not an accident (just as the prosecution argued in the *Smith* case that the drowning of poor Bessie Mundy was no accident). It would not allow proof of the other crimes to show that  $Y = 1$ , i.e. that the accused is the type of person who commits burglaries and therefore is more likely to be the burglar in the case at bar, for the proof of the theorem does not use this inferential structure. Therefore, unless the accused had a theory about 'accidental burglaries', the theorem would not support the admission of any evidence of the prior crimes.

#### REFERENCES

- Centers for Disease Control (1985) Cluster of unexplained deaths at a hospital—Maryland. *Report EPI-86-17-1*. Centers for Disease Control, Atlanta.
- Cleary, E. (ed.) (1984) *McCormick on Evidence*, 3rd edn, pp. 549–592. St Paul: West.
- Cross, R. (1979) *Evidence*, 5th edn. London: Butterworth.
- Eggleston, R. (1983) *Evidence, Proof and Probability*, 2nd edn, pp. 88–102. London: Weidenfeld and Nicolson.
- Essary, J. D., Proschan, F. and Walkup, D. W. (1967) Association of random variables with applications. *Ann. Math. Statist.*, **38**, 1466–1474.
- Fienberg, S. E. (ed.) (1989) *The Evolving Role of Statistical Assessments as Evidence in the Courts*. New York: Springer.
- (1990) Legal likelihoods and *a priori* assessments: what goes where? In *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard* (eds S. Geisser, J. S. Hodges, S. J. Press and A. Zellner), pp. 141–162. Amsterdam: North-Holland.
- Fienberg, S. E. and Kadane, J. B. (1983) The presentation of Bayesian statistical analyses in legal proceedings. *Statistician*, **32**, 88–98.
- Fienberg, S. E. and Schervish, M. J. (1986) The relevance of Bayesian inference for the presentation of statistical evidence and for legal decisionmaking. *Bost. Univ. Law Rev.*, **66**, 771–798.
- Fienberg, S. E. and Straf, M. L. (1990) Statistical evidence in the US courts: an appraisal. *J. R. Statist. Soc. A*, **154**, 49–59.
- Gilula, Z. (1979) Singular value decomposition of probability matrices: probabilistic aspects of latent dichotomous variables. *Biometrika*, **66**, 339–344.
- Goodman, L. A. (1974) The analysis of systems of qualitative variables where some variables are unobservable: part I, A modified latent structure approach. *Am. J. Sociol.*, **75**, 1179–1259.
- Harrison, K. (1988a) Expert rules out chance in Bolding patient deaths. *Washington Post*, June 7th.
- (1988b) Bolding defense impugns study implicating nurse. *Washington Post*, June 8th.
- (1988c) Judge acquits Nurse Bolding. *Washington Post*, June 21st.
- Imwinkelreid, E. (1984) *Uncharged Misconduct Evidence*. Wilmette: Callaghan.
- Istre, G. R., Gustafson, T. L., Baron, R. C., Martin, D. L. and Orlowski, J. P. (1985) A mysterious cluster of deaths and cardiopulmonary arrests in a pediatric intensive care unit. *New Engl. J. Med.*, **313**, 205–211.
- Kaye, D. H. (1987) The admissibility of "probability evidence" in criminal trials—part II. *Jurimetr. J.*, **27**, 160–172.

- Lagakos, S. W., Wesson, B. J. and Zelen, M. (1986) An analysis of contaminated well water and health effects in Woburn, Massachusetts (with discussion). *J. Am. Statist. Ass.*, **81**, 583-614.
- Marjoribanks, E. (1929) *For the Defence: the Life of Sir Edward Marshall Hall*, p. 329. New York: Macmillan.
- Sacks, J. J., Stroup, D. F., Will, M. L., Harris, E. L. and Israel, E. (1988) A nurse-associated epidemic of cardiac arrests in an intensive care unit. *J. Am. Med. Ass.*, **259**, 689-695.
- Stross, J. K., Shasby, M. D. and Harlan, W. R. (1976) An epidemic of mysterious cardiopulmonary arrests. *New Engl. J. Med.*, **295**, 1107-1110.
- United Press International (1978) Poisoning charges dropped against two nurses. *New York Times*, Feb. 2nd, p. 16.
- Weaver, C. (1988) The chilling case of Nurse 14. *Regardie's*, May, 93-144.