Elffers, of the Netherlands Institute for the Study of Crime and Law Enforcement in Leiden, combined the Juliana data with data from another hospital in which de Berk had previously worked to get his figure. Gill and Grünwald insist that the analysis was misleading. "It makes little sense to do formal hypothesis testing when the data themselves have suggested the hypothesis," says Gill. "The only safe thing is to go out and collect new independent data."

Gill's own calculation estimates that the probability that the correlation arose by chance is not 1 in 342 million, but a much smaller 1 in 48, or even as low as 1 in 5 — figures that are unlikely to meet the 'beyond reasonable doubt' needed for a criminal conviction. But Elffers defends his original calculations, arguing that he applied a factor that corrected for his using some of the data twice. "Everyone is aware that I applied a correction," he says.

Fact or fallacy?

Aside from this debate, equally important is how the court interpreted the number. Philosopher of science Ton Derksen of the University of Nijmegen, who has written a book that criticizes de Berk's conviction, argues that the court made an elementary statistical error known as the prosecutor's fallacy.

The court needs to weigh up two different explanations: murder or coincidence. The argument that the deaths were unlikely to have occurred by chance (whether 1 in 48 or 1 in 342 million) is not that meaningful on its own — for instance, the probability that ten murders would occur in the same hospital might be even more unlikely. What matters is the relative likelihood of the two explanations. However, the court was given an estimate for only the first scenario. Without additional information, says Derksen, Elffer's number is meaningless — and could easily be misinterpreted as a very small chance that de Berk is innocent.

To raise further doubt, other important statistics were neglected by the court. When de Berk worked at Juliana between 1999 and 2001, there were six unexplained deaths in her unit. The same unit, in a similar period before de Berk started working there, had seven unexplained deaths. "It seems very strange," says Grünwald, "that fewer people die when there is a serial killer around." Derksen says that the statistics comparing deaths before and after de Berk started work at the hospital were mentioned by her defence lawyers, but were not sufficiently emphasized to have any influence on the court.

Due process

This neglect illustrates a difference between legal and scientific processes. Although science aims to bring together all relevant evidence, this is not necessarily true with the law. David Kaye, an expert in statistics and the law at Arizona State University in Tempe, notes that lawyers have an incentive, and even a duty, to select the evidence that makes their case stronger. "What the judge ends up hearing often comes from the two extreme ends of the distribution," he says.

"The magical power of the big number led everyone at an early stage to be totally convinced of de Berk's guilt."
— Richard Gill

Procedures to correct such distortions are also lacking, even after a trial has reached a verdict. In the United States, written statistical arguments are often protected by court orders, and so are not available for review or correction. "The data pertaining to an individual deserve some protection," says statistical expert Joseph Gastwirth of George Washington University in Washington DC, but a summary of the expert reports should be made publicly available, he suggests.

Independent scientific comment of this kind occurred during the Clark case, but to unknown effect. In the 1999 Clark trial, Meadow testified that the chance of two infants from the same mother dying of Sudden Infant Death Syndrome (SIDS) was only 1 in 73 million. Two years later, after the first appeal, the Royal Statistical Society in London, condemned both this figure and its interpretation. The figure would be valid only if SIDS cases arise independently within families, the