From killer nurses to quantum entanglement, and back Part 1

This article is loosely based on the farewell lecture I gave on the occasion of my retirement as professor of Mathematical Statistics in Leiden, one and a half years ago. A talk on statistics is not complete without some data and a talk on mathematics is not complete without the proof of a theorem, so you will see both. But, while reviewing the content of the formal lecture, I took a new look at some of the data from one of my killer nurse cases, and made some exciting new discoveries. So the article grew, and now has to be published in two parts. Welcome to Part 1. In the lecture, I presented two of my recent research passions: one of which enjoyed some measure of success, the other seemingly ending in failure. There are, I think, amusing links between the two, at various levels. And working on this article, I made new statistical discoveries concerning the case of the English nurse Ben Geen – that's the research passion I was so unhappy about. I now feel some optimism.

Richard D. Gill

For me, the Ben Geen case starts with my earlier experience in the case of the Dutch nurse Lucia de B., and that starts with a small number of 2×2 tables (Figure 1). These tables played an enormous role in the conviction for serial murder of the Dutch nurse Lucia de Berk.

The data is also presented in a recent paper *Elementary Statistics on Trial (the case of Lucia de Berk*) from R. D. Gill, P. Groeneboom and P. de Jong (2018). This data was analysed for the court in 2002 by my old colleague and old friend Henk Elffers, who had been contacted in 2001 by police investigators at the beginning of what became a ten year saga. I only got involved half way through.

JKZ *		Incident in shift: Yes	Incident in shift: No	
Lucia on duty	Yes	8	134	142
	No	0	887	887
		8	1021	1029

RKZ-41 **		Incident in shift: Yes	Incident in shift: No	
Lucia on duty	Yes	1	0	1
	No	4	361	365
		5	361	366

RKZ-42 ***		Incident in shift: Yes	Incident in shift: No	
Lucia on duty	Yes	5	53	58
	No	9	272	281
		14	325	339

* Juliana Kinderziekenhuis, Medium Care Unit 1, 1 Oct 2000 – 9 Sep 2001; ** Rode Kruis Ziekenhuis, afdeling 41, 6 Aug – 26 Nov 1997; *** Rode Kruis Ziekenhuis, afdeling 42, 1 Aug – 30 Nov 1997

Figure 1. Roster data from the case of Lucia de B.

Foto: sudok1

During nearly a year ending in September 2001 when police investigations started, on a medium care ward at the Juliana Children's Hospital (JKZ) in the Hague, there were in total 1029 (3×343) 8-hour shifts (3 shifts a day, 7 days a week). In 8 of them, an "incident" occurred. All 8 in the shifts of a certain nurse called Lucia. Several years earlier, at another hospital, the Red Cross Hospital (RKZ), during the same four months in two intensive care wards (Wards 41 and 42; the data from ward 42 misses 9 days at the beginning and end of the 4 month period) there were 5 and 14 incidents respectively. Lucia was only on duty once in RKZ-41 but on just that one occasion she netted one of the 5 incidents! She mainly worked on Ward 42, where she netted disproportionately many of the 14 incidents (she had one third of the incidents in only one sixth of the shifts). Does one need a statistician to interpret it to the board of judges of a criminal court where Lucia is being tried for serial murder? The data speak for themselves.

But do the data speak the truth, the whole truth, and nothing but the truth? As you perhaps already know, the Lucia case can be considered a success story. Her life sentence for ten murders of children and old people got reversed. At the retrial, the judges, in their summing up, congratulated the nurses on their devotion and their professional efforts to save the lives of their patients, lives which (they said) were unnecessarily shortened through medical errors. The errors were caused by misdiagnosis, chaotic management, ignorance of the content of the patients' medical dossiers ... they were committed by hospital specialists and hospital managers. A witch-hunt ensued and an intelligent nurse with a charismatic personality and a colourful past was a natural scapegoat.

Inspired by our success in getting Lucia a retrial (I was one of a group of people fighting for a retrial, from 2006) I got involved in several similar cases, in particular, in the case of one Ben Geen (*not* "Green"), who also got a life sentence for an enormous number of attacks on patients in a short time period in a small hospital in a provincial town in England ... at the same time as the Lucia case was playing out. I will show you some data from the Ben Geen case in a moment. But first, here are some more 2×2 tables.

Quantum entanglement

The data in the four 2 × 2 tables in Figure 2 were collected in an experiment performed in Delft in 2015 in two labs at opposite ends of the university campus. The results were rapidly published in the paper *Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres* from B. Hensen, H. Bernien, A. Dréau et al. (2015).

In the two labs, 1300 meters apart, we imagine some lab assistants at work, conventionally named Alice and Bob. The labs are connected by dedicated glass fibre cables and ordinary internet connections. It takes light four micro-seconds to traverse that distance (there are a million microseconds in one second). A lot will be done in both labs in many short time intervals each of length four microseconds: so short that there cannot be any influence on events in Alice's lab, from events taking place in that time interval in Bob's lab; or vice versa.

In each lab there is a diamond, and in each diamond there is a "nitrogen-vacancy" defect: two adjacent carbon atoms are missing and in their place are one nitrogen atom and one little hole. This "vacancy" is home of an electron whose "spin" in various directions can be addressed and manipulated and read-out with the help of state-of-the-art lasers and electronics. If asked, it only ever answers "up" or "down".

To start with, the two distant *spins* are prepared in a way which certainly can induce correlations between their properties. It is not necessary to explain how that is done, and indeed, perhaps it cannot be *explained*. We could run through the mathematics (a few lines of simple linear algebra of tensor-products), but that would not make things less mysterious. Then, at each of the two labs, a fair coin is tossed, heads marked "1" and tails marked "2". In each lab, the resulting "1" or "2" is used as an *input* or a *setting* to set a switch to position "1" or position "2" on a measurement device which actually houses the diamond. The spin is immediately measured and a result comes out. This *measurement outcome* or *output* is binary, we call it "+" or "–" (these are just labels). Let me emphasize: **this happens within such a small time interval that there can be no knowledge at Alice's side of Bob's *setting*, till after Alice's *outcome* is determined, and vice versa.**

All apparatus is then reset. The two spins are put into the "quantum entangled singlet state" anew. Repeat.

Call each repetition one *trial*. Stop when you have performed 245 trials. The measurement settings

Outcomes		Bob setting 1 +	Bob setting 1 –		Bob setting 2 +	Bob setting 2 –	
Alice setting 1	+	23	3	26	33	11	44
	–	4	23	27	5	30	35
		27	26	53	38	41	79
Alice setting 2	+	22	10	32	4	20	24
	–	6	24	30	21	6	27
		28	34	62	25	26	51

Success rate	0,80	=	196	÷	245

Figure 2. The Delft Bell experiment

correspond to directions in which "spin" is measured, and the outcome is "up" or "down" in the direction measured. It's not possible for humans to actually toss fair coins and switch switches in four microseconds, so the random settings are generated by arguably equivalent but much faster physical devices.

Anyway, each trial has two *inputs* taking values in the set of two symbols {"1", "2"}, let's call them *a* and *b*, and two *outputs* taking values in the set of two symbols {"+", "–"}, let's call them *x* and *y*. So the data generated in one trial are represented by one string of four symbols *abxy*. There are 16 possible strings of four binary symbols. They were counted and are tabulated in the spreadsheet above.

Each *trial* is classified as being either a "success" or a "failure". The definition of "success" is: "at least one input is "1" and the outputs are the same, or both inputs are "2" and the outputs are different". There were 196 successes out of 245 trials, which is a success rate of 80%, and 80% is somewhat larger than 75% – in fact, statistically significant, (one-sided test) at the 5% level.

This experimental outcome resulted in a scientific paper very rapidly published in the prestigious journal *Nature*, and the news was reported in all the serious science supplements of decent newspapers everywhere in the world. In fact, a big race had been on. Three other experimental groups, in Munich, Vienna and at NIST in Boulder, Colorado, had performed variants of the same experiment at about the same time, but only published their results (in full agreement with the Delft findings) a little bit later. Delft was actually dark horse in the race, an unexpected contender.

Imagine tossing two fair coins repeatedly a total number of 245 times (trials). Each trial, the chance of seeing two tails is one quarter. We would expect around 75% of the trials to show one or two heads. The exciting thing about the Delft experiment is that the chance of seeing one or two heads in 196 or more of such trials (i.e., at least 80% of the trials) is only 3.9%, which is less than the magic "significance level" 5%. Why that is exciting, in fact, perhaps earth-shattering, I will try to explain in a moment. Just significant with a one-sided test at the 5% level, though perhaps guaranteeing you a publication in regional economics or social psychology, would not usually be enough to guarantee a publication in *Nature* in the field of physics. Despite this, it is fair to say that this experiment represents a milestone reached – a milestone we had been trying 50 years to reach.

There is plenty of evidence that this was not just a lucky chance. Aside from the copious support from quantum physics itself, reported in the published paper, we have the results of the other three experiments, each on their own passing the same milestone while using quite different quantum technologies. Taken together, the statistical evidence seems to me to be overwhelming.

By the way, quantum theory says that it is impossible to get beyond 85%, or, more precisely $100 \times (2 + \sqrt{2}) / 4\,\%$. That's a bound which needs sophisticated mathematics to derive and from which, alas, one does not gain much intuition. I could drop the name of Grootendieck and I could say something about the geometry of Hilbert space ... This means that the question arises that maybe, there is a better theory than quantum mechanics which would allow a higher limit? There is surely no reason to suppose that quantum mechanics is a "last stop" for physics. There are arguments that the real bound could be 100%. Some progress has been made finding an intuitive reason behind the mysterious $(2 + \sqrt{2}) / 4$, but the problem is still essentially open.

I will try to explain how I had something to do with the whole thing (also with the results obtained in Vienna and Munich and at NIST). The connection has to do with *martingale theory*, which has a lot to do with statistics. In particular, it has a lot to do with randomised clinical trials and the whole idea of randomisation in statistics. (I'll also explain how Lucia's exoneration and release from jail also had something to do with my interests in quantum mechanics. Everything is connected!)

First a little bit of probability theory. Consider four random variables X_1, X_2, Y_1, Y_2 which take the values ±1. Consider the sum of four indicator variables

$$I(X_1 = Y_1) + I(X_1 = Y_2) + I(X_2 = Y_1) + I(X_2 \neq Y_2).$$

Obviously, this sum can only take the values 0, 1, 2, 3 and 4. It can only take the value 4 if all four events occur simultaneously. But

$$X_1 = Y_1 \;\&\; X_1 = Y_2 \;\&\; X_2 = Y_1 \;\Rightarrow\; X_2 = Y_2.$$

So, the value 4 cannot be attained. The expression can't exceed 3. Divide the resulting inequality throughout by 4. We obtain

$$\tfrac{1}{4}\,\mathrm{I}(X_1 = Y_1) + \tfrac{1}{4}\,\mathrm{I}(X_1 = Y_2) + \tfrac{1}{4}\,\mathrm{I}(X_2 = Y_1) + \tfrac{1}{4}\,\mathrm{I}(X_2 \neq Y_2) \leq \tfrac{3}{4}.$$

Now I connect this to the Delft experiment. Suppose that for each trial, counterfactual outcomes could be defined, X_1, X_2, Y_1, Y_2, which stand for: "the outcome which Alice would see were her setting equal to 1", "the outcome which Alice would see were her setting equal to 2", "the outcome which Bob would see were his setting equal to 1", "the outcome which Bob would see were his setting equal to 2". These could be constructed (perhaps in many ways) in any crypto-deterministic (i.e., having a behind-the-scenes mechanistic explanation) mathematical physical theory which does not have action at a distance: Alice's *possible* outcomes can't depend on Bob's *actual* setting; and vice versa. Four microseconds is too short for the information to have travelled from Bob's side to Alice's. Both of Alice's *counterfactual outcomes* (the outcomes she would have seen, if either of her two possible *inputs* had been supplied) are functions of the stuff on Alice's side of the experiment, and the stuff connecting it to the source and at the central source itself, but can't possibly depend on the *actual input* at Bob's side.

Everything *might* depend on everything which happened in the *past* in *both* wings of the experiment. Moreover, there might be correlated time trends and time jumps in the physics on the two sides of the experiment. But if the two settings are chosen at each trial, anew, completely at random, independently of the past, then for each trial, conditional on all preceding ones, the just proven inequality says that the conditional probability of *success* (as it was defined earlier), conditional on the past trials in both wings of the experiment, is less than or equal to ¾. We have found that if we define a stochastic process by taking its increments to equal I(success) – ¾, this process is a *supermartingale*.

It's not difficult to prove from this (by a recursive coupling argument, and taking account of the 0/1 nature of the indicator random variable) that the random variable #successes is *stochastically smaller* than the Binomial(#trials, ¾) distribution; in other words, right-tail probabilities are less than or equal to Binomial right-tail probabilities.

I developed these martingale ideas back in 2001, introducing the idea of exploiting the randomisation of the settings instead of making i.i.d. assumptions about the physics of the outcomes in order to get safe bounds on the probability of large deviations from the Bell-CHSH local realism bound. I wanted such bounds because I wanted to hold a bet, and indeed, to win a bet, against an opponent who claimed they could simulate the quantum mechanical correlations on a network of computers, where the computer network reflected the spatio-temporal constraints of a "loophole-free" Bell-type experiment. I wanted to be pretty sure that I would win! My opponent complained about the rules which I wanted to impose. The independent jury we had recruited chickened out of their scientific duty to agree that my experimental protocol was fair. These are the rules which are imposed in the experiments of 2015 and later, and which had been written down 35 years before then.

For a fairly recent survey, see my paper "Statistics, Causality and Bell's Theorem" (Gill, 2014). A new survey is badly needed, thanks to the 2015 experiments.

Killer nurses

I now go back to (probably innocent) convicted serial killer nurses. The dataset presented in Figure 3 helped get the young English nurse Ben Geen a life sentence for two counts of murder and 15 of grievous bodily harm (a 16th count of grievous bodily harm was not considered proven), in the three consecutive months of December 2003, January 2004 and February 2004. I am certain he is innocent, just as I am certain that Lucia is innocent. And for much the same reasons. The reasons have little to do with statistics. The reasons have to do with the social structures in a modern hospital and the facts that (a) sick people do die in hospitals, (b) doctors do make mistakes, (c) top hospital managers and top medical specialists need to protect the reputation of their hospital. A fourth reason is (d) the coincidence that this case occurred shortly after the Shipman Enquiry, which blamed health-care administrators for not earlier noticing serial killer doctor Harold Shipman, who maybe murdered 250 patients.

Obviously, one of the most important factors in a doctor-patient relationship is that the patient has trust in his or her doctor. The medical establishment consequently has a strong interest in patients collectively having trust in doctors, and more generally in their health-care system. Similarly, we need to have confidence in our judges and in our legal systems. The guardians of our legal system strongly believe that our judges and our legal systems *must* not be *seen* to make mistakes. Unfortunately, a system which cannot admit to making mistakes can never learn from mistakes and is doomed to repeat them.

Learning from mistakes is good, but a new danger then arises that by learning the *wrong* lessons from one kind of mistake, one might increase the chance of making the *opposite* mistake. If the rate of false convictions goes down but nothing else really changes, the rate of false acquittals will go up. The more easily a health-care system goes into alarm-mode because of suspicion that it harbours a health-care serial killer, the more often *innocent* health-care professionals will trigger an alarm.

This key data-set in the Ben Geen case was later presented to the court by Michelle Brock, head-nurse of the Accidents and Emergency department where Ben Geen worked, at Horton General Hospital, a rather small hospital in the provincial market town Banbury in North Oxfordshire. Together with a dossier of perhaps 30 incidents all from December 2003 onwards, it had initially been compiled in great haste before the case was reported to the police. Michelle and some colleagues based their work on patient records and nurse attendance records at the hospital, looking *only* at what happened during Ben's shifts, their investigation triggered and guided by recent memory and gossip. Ben, who was a trainee nurse, had won a higher qualification at the beginning of December, allowing him to work under less supervision than before. The trigger for their investigation had been two sudden and surprising collapses of patients who had just entered A and E (also known as ED: Emergency Department) on Thursday 5 February. Ben had reported sick on Friday, and had had a free weekend after that. He was arrested on Monday morning, 9 February 2004, as he arrived for work (one third of the way into the last bar of the bar-chart). The bar-chart was later presented to the court and was also known to the medical experts who were consulted on the 18 individual cases. There is no doubt it had a big impact on everyone involved in the trial, including journalists covering the trial.

We catch a glimpse from the chart of the fact that a lot of old people and people with existing serious health

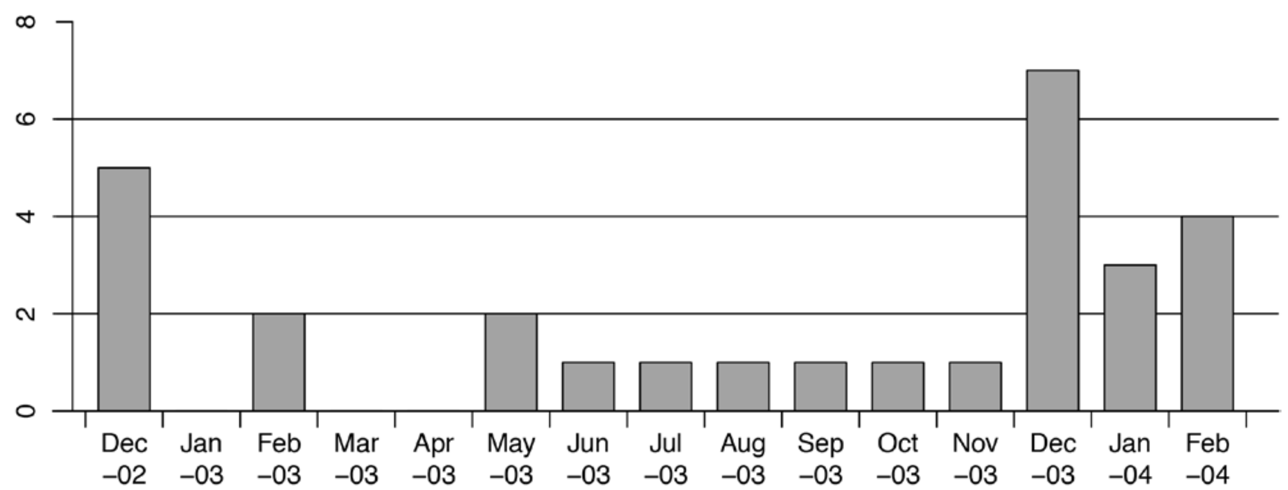


Figure 3. Admissions to critical care from the emergency department with a diagnosis of cardio-respiratory or respiratory arrest or hypoglycaemia, data: Head Nurse Brock

problems get brought to emergency care during the winter months of December, January and February with acute problems involving heart and lungs (a hot summer is also a danger period). A common diagnosis is *cardio-respiratory arrest* (the heart has stopped working and consequently the lungs too), much less common is "pure" *respiratory arrest* (the lungs have stopped working); fairly common is *hypoglycaemia*: a fall in blood glucose level. It *causes* fainting; breathing stops or is much suppressed. It can be *caused* by too much insulin or other glucose lowering diabetes tablets, delaying or missing a meal, not eating enough carbohydrate, unplanned physical activity, more strenuous exercise than usual, drinking alcohol – the risk of hypoglycaemia increases, the more alcohol you drink. In the bar-chart, nurse Brock has combined the three "standard" categories cardio-respiratory, respiratory, and hypoglycaemic arrest; but what is the correct category is hard to guess when a patient presents (arrives at the hospital). Past medical history, *and future medical events* will give clues as to what was actually going on. In an emergency situation, past medical history may be unknown.

Hospital nurses and authorities had been worried by the behaviour of the young male trainee nurse Ben Geen already before December 2003. His father was in the army, his mother was a nurse. He had been in the territorial army medical corps, and his ambition was to be qualified and then transferred as a combat medic to a military field hospital in Iraq. He was energetic and very "present", keen to get action and to get experience. He made some other nurses nervous. They were calling him "Ben Allitt" behind his back, not such a nice joke, since Bev Allitt is the very well-known name of a pre-Shipman famous English convicted serial killer nurse (an interesting case which, in my opinion, deserves fresh study).

In December 2003 the numbers of patients reaching an overstressed emergency ward in an underfunded minor hospital in a provincial English town, threatened by closure because of its small size, and with perhaps not the very best of medical specialists and not the very best of management, was bigger than ever. There were a couple of "surprising" events when patients who were initially thought to be in fairly good shape suddenly, and at the time unexplainably, worsened. Ben was usually around when anything happened simply because he was usually around: he was working double shifts in order to gain more and more experience as fast as possible, and also often fell in for absent colleagues.

On Thursday 5 February 2004, at the end of a hectic day, a chronic alcoholic diabetic was brought into hospital by his drinking mates, throwing up repeatedly and feeling very poorly, suffering fainting fits. Ben took a blood sample. The patient suddenly worsened and later had little idea what else Ben did to him. He certainly inserted a *canula* (a tube that can be inserted into the body, often for the delivery or removal of fluid or for the gathering of samples) and the patient was rapidly transferred to "critical care". The real problem for Ben came later: Ben went home with, unknown to himself (he said), a used plastic needle-less syringe containing some muscle relaxant in his nurse's smock. Such a syringe is used to administer necessary medications, including a muscle relaxant, through the *canula* prior to inserting breathing and feeding tubes into patients in the critical care ward. Ben stayed home sick on Friday, and then had the weekend free. His girlfriend, another nurse, doing the washing, told him off for this (she said) and told him to take it back as soon as possible. So, on Monday morning – with the syringe in his coat pocket – he was met by policemen as he entered the hospital. In some panic (he said) he stupidly further empted the remaining contents of the syringe into his pocket. Obviously, he tried to harm patients by injecting them with this stuff so that he could then play the hero, helping to resuscitate them! The so-called "Munchhausen by proxy" syndrome.

At his trial, the Crown secured the services of a famous and experienced expert (a highly distinguished professor of Anaesthesiology), who found a number of the events highly suspicious; another confidently swore that never ever in his long experience had he met with an *unexplained* respiratory arrest. They all agreed on that …

Of course they did. All respiratory arrests are "explainable", though different experts often give different explanations. Actually, whether a collapse is diagnosed as cardio-respiratory, respiratory or hypoglycaemic can be pretty arbitrary. When either heart or lungs get into difficulty, the other organ rapidly gets into difficulties too. Hypoglycaemic arrest (critically low blood glucose levels) always involves breathing problems (you faint when not enough oxygen is reaching your brain) and can trigger

further deterioration of heart and lung function. Reduced oxygen levels affect brain, heart, lung. Muscles burn oxygen, the brain burns oxygen. All arrests are explainable, but the categories which are ticked on forms in the patient's dossier and in the hospital's administrative records may differ and may be revised in the light of later events. The categories which tend to be chosen by nurses, doctors and administrators may depend on who is doing it, and may show trends and jumps as time goes by. Just one occurrence of an unusual diagnosis alerts people to its existence, and they start seeing it every day: the well known *Baader-Meinhof phenomenon*.

At the time each had actually occurred, each of the 18 cases in the criminal charges against Ben had been "normal". The last two had surprised some people (certainly not all), but because of earlier suspicion and gossip, they had triggered an emergency weekend-long internal hospital investigation, in which more than 30 dossiers of patients who had in recent months gone through Emergency *while Ben was on duty* were combed through, resulting in a dossier of 18 cases to hand over to the police on Monday. In fact the teams had access to 4000 patient medical records but were not interested in what happened when Ben was not there. Expert witnesses for the defence later explained how explainable each of the 18 was, though they were honest enough to admit that some cases were too complex to come to any clear conclusion. The prosecution had more expensive and more court-experienced experts than the defence. The prosecution experts were of course specifically hired to point out anomalies in each of the selected 18 cases, and tended to be rather confident of their diagnoses. Prosecution experts are "instructed" by the prosecution, defence experts are "instructed" by the defence.

Ben must have used a myriad of different techniques to cause all these unexplained medical emergencies and in many cases the expert witnesses called by the Crown in fact had conflicting ideas of what he might have done; though they did of course agree that he must have done something. All of the 18 patients were very sick, and what happened to each was what you may well have expected to happen in view of their existing severe and often complex conditions. But sometimes developments are fast, you do not "see them coming", and so a sudden worsening takes some nurses or some doctors by surprise. People, including Ben himself, did notice Ben often being there when such events took place. He had said, and said it in court again, that he thought he had been jinxed.

Ben's unemotional and careful account of what he could recall that he had seen and done in each case, the impression he gave that he knew the law better than the lawyers, the eminent professor's categorical statement that he had never seen an unexplained respiratory collapse in all his career, and the smoking gun which was the syringe, together clinched the matter for the jury. It mainly consisted of decent retired folk who had spent most of the trial napping during the presentation of interminable medical evidence (in 18 cases). The judge in his summing up made it very clear what verdict he expected from the jury.

Blood and urine samples from the trigger case showed traces of a muscle relaxant as well as of plenty of sedatives, but unfortunately the samples were not dated – one has no idea when they were taken nor by whom! Sedatives and muscle relaxant *should* have been present. The traces of muscle relaxant were of the same kind as was in the syringe. The consultant anaesthesiologist who had attended to the trigger patient as he went into intensive care said that she had asked (another nurse, later) for a different one. Ben said that he was not told to administer muscle relaxant, so, of course, had not done so. Hospital records were woefully incomplete. Since the earlier cases were not at the time thought to be suspicious, all samples of blood and urine had long ago been thrown away.

The annual pattern we see in that data can be seen in data which I analysed from many similar hospitals all over Britain. Of course, there is no data whatsoever about *unexplained* respiratory arrests. The data stored in a hospital database are administrative data. Every event has been put into a pre-existing category with an explanation, because it is not possible to enter it into the data base otherwise. The data in the database determines the fees of the medical consultants (the medical specialists) and the funding of the hospital. The data is not collected for scientific research or forensic investigation.

The three standard categories relevant to this case are *cardio-respiratory arrest*, *respiratory arrest*, and *hypoglycaemic arrest*. I already showed you the data supplied to the court by Ben's head nurse, combining

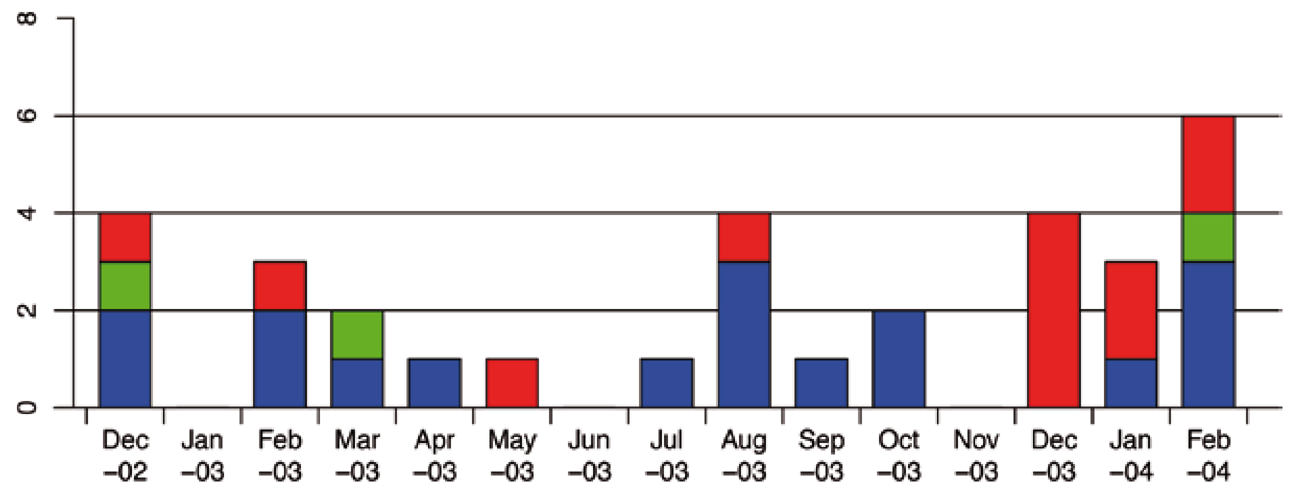


Figure 4. Admissions to CC from ED with CR, Hypo or Resp arrest, FOI data, Cardio-respiratory (blue), hypoglycaemic (green), respiratory (red)

those three categories. Much later, we got from Ben's hospital the data as presently archived in official hospital records. It was different! Also, the categories are still separate, and we have data from many more years.

The total numbers of relevant cases in December 2002 and in December 2003 are now equal to one another – both an unremarkable 4. Not 5 in 2002 versus an incredible 7 in 2003. The split between categories in the two periods of winter months is markedly different. In winter 2002 – 2003 it is normal, spread out over all three, but mostly *cardio-respiratory*. In winter 2003 – 2004 *almost everything* is being categorised as *respiratory*. The total number of cases in January, in both winters, is much less than in adjacent months, this is normal.

Normal case-mix (for the three categories of interest), both in this hospital and in all others (we have similar data from about 40 other hospitals all over England, for the thirteen year period 2000 – 2012), is a mix mainly of *cardio-respiratory*, with *respiratory* and *hypoglycaemic* normally each at roughly a fifth of the level of *cardio-respiratory* (Figure 4). They are both much less usual, but neither can be called *rare*.

There is also data in the official public enquiry held after Ben's conviction, held to find out why Ben wasn't caught earlier and to prevent such a tragedy from ever occurring again. "The number in December 2003 was six and this was only one more than in December 2002". Two different numbers, yet again. The inquiry did suggest that the very large numbers of incidents while Ben was carrying out his attacks might have been expected anyway, due to the winter season, perhaps masking incidents caused by Ben. It did heavily criticise the Emergency department for poor record keeping when updating patient medical notes and poor registration of withdrawals of dangerous medications.

The allocated space for this article has now run out. In Part 2 I will show results from a new analysis of some of the data which we have on the Ben Geen case, which I think could be part of the key to getting him a fair re-trial.

References

Gill, R. D. (2014). Statistics, Causality and Bell's Theorem. *Statistical Science, 29*(4), 512–528. doi:10.1214/14-STS490.

Gill, R. D., Groeneboom, P., & De Jong, P. (2018). Elementary Statistics on Trial (the case of Lucia de Berk). *Chance, 31*(4), 9–15.

Hensen, B., Bernien, H., Dréau, A. et al. (2015). Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature, 526,* 682–686, doi:10.1038/nature15759.

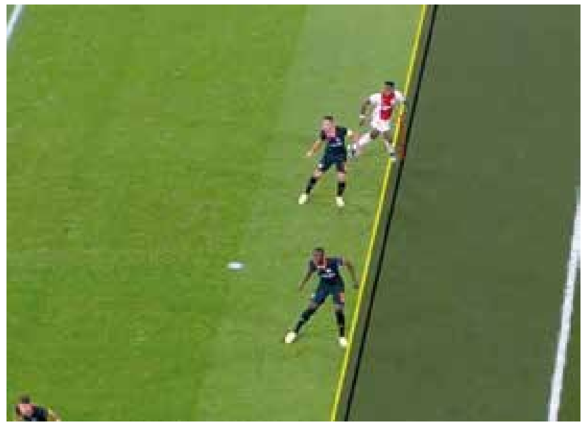
Richard Gill is professor emeritus of Statistics at Leiden University. This article is mainly based on his 'exauguration' speech. Apart from other distinctions he is an Honorary Member of the VVSOR.
E-mail: gill@math.leidenuniv.nl

Buitenspel, meetfouten en millimetergemier

Met de invoering van de Video Assistant Referee ofwel de VAR leek het er voor alles en iedereen een stuk eerlijker op te worden in de voetballerij. Met name de doellijntechnologie werd met gejuich begroet. Nooit meer eindeloze discussies over doelpunten die al of niet terecht waren goedgekeurd. En ook over buitenspel geen twijfel meer, zo dacht men optimistisch. Maar hoe goed de VAR ook werkt voor de eerstgenoemde situatie, buitenspel en de VAR is een minder gelukkige combinatie gebleken.

De gruwel van lange spelonderbrekingen

Hoewel ook het gebruik van de VAR bij buitenspel met hoge verwachtingen werd geïntroduceerd roept zij vooralsnog voornamelijk onvrede op. Denk alleen maar aan het afgekeurde doelpunt van Quincy Promes in de recente thuiswedstrijd van Ajax tegen Chelsea. De onvrede wordt veroorzaakt door het minutenlang VAR-overleg bij vermeend buitenspel en de vaak uiterst omstreden beslissing die daarop volgt. Het oog van de scheidsrechter blijkt niet altijd minder onrechtvaardig dan het oog van de computer, de hoge verwachtingen omtrent de VAR als snelle en eerlijke hulpscheidsrechter ten spijt.



Quincy Promes buitenspel?

Met de VAR kan worden gemeten, heel nauwkeurig tot op de millimeter. Maar helaas, meten is lang niet altijd zeker weten! Meten is weten met een foutmarge! En als de VAR dus aan het tijdrovende millimetermieren is over wel of niet buitenspel, dan is de terechte vraag of het gebruikte computerbeeld van de VAR wel de realiteit weergeeft en de beslissing dus wel correct en eerlijk is. Natuurlijk klopt het dat je met elkaar kunt afspreken dat de VAR-beslissing DE beslissing is en we ons niet bekommeren over onnauwkeurigheden van de metingen. Maar de terechte vraag is dan of we dat de sporters kunnen aandoen. Een foute beslissing van de scheidsrechter (denk aan de iconische handsbal van Thierry Henry), en dus ook van de VAR, heeft soms miljoenen euro-consequenties! Ook andere sporten dan voetbal kennen dit nijpende probleem dat alleen maar groter wordt omdat de prestatiedichtheid aan de top toeneemt en de onderlinge verschillen tussen de toppers steeds kleiner worden.

Zichtbaar teleurgestelde Koen Verweij mist goud op 3-duizendste seconde

Was het eerlijk dat Koen Verweij in Sotchi zilver kreeg terwijl het verschil met de nummer één slechts drieduizendste van een seconde was, een verschil zo klein dat die binnen de erkende foutmarges van de (tijd)metingsystemen lag? Zelfs de officiële tijdwaarnemers kennen die onnauwkeurigheidsmarges. Saillant punt hierbij was