Chance News (July-August 2005)

From ChanceWiki

Quotation

Numbers are like people; torture them enough and they'll tell you anything.

Forsooth

Frank Duckworth, editor of the Royal Statistical Society's newsletter RSS NEWS has given us permission to include items from their Forsooth column which they extract forsooth items from media sources.

Of course we would be happy to have readers add items they feel are worthy of a forsooth!

From the February 2005 RSS news we have:

> Glasgow's odds (on a white Christmas) had come in to 8-11, while Aberdeen was at 5-6, meaning snow in both cities is considered almost certain.

BBC website
22 December 2004

From the May 2005 RSS News:

> He tried his best--but in the end newborn Casey-James May missed out on a 48 million-to-one record by four minutes. His father Sean, grandfather Dered and great-grandfather Alistair were all born on the same date - March 2. But Casey-James was delivered at 12.04 am on March 3....

Metro
10 March 2005

> In the US, those in the poorest households have nearly four times the risk of death of those in the richest.

A Probability problem

A Dartmouth student asked his math teacher Dana Williams if he could solve the following problem:

> QUESTION: We start with n ropes and gather their 2n ends together.
> Then we randomly pair the ends and make n joins. Let E(n)
> be the expected number of loops. What is E(n)?

You might be interested in trying to solve this problem. You can check your answer here.

DISCUSSION QUESTIONS:

(1) There is probably a history to this problem. If you know a source for it please give the history here or on the discussion page above.

(2) Can you determine the distribution of the number of loops? If not estimate this by simulation and report you results on the discussion page.

Misperception of minorities and immigrants

Statistical Modeling, Causal Inference, and Social Science is a statistics Blog. It is maintained by Andrew Gelman, a statistician in the Departments of Statistics and Political Science at Columbia University.

You will find lots of interesting statistics discussion here. Andrew also gave a link to a cartoon in which Superman shows how he would estimate the number of beans in a jar. This also qualifies as a forsooth item.

In a July 1, 2005 posting Andrew continues an earlier discussion on misperception of minorities. This earlier discussion resulted from by a note from Tyler Cowen reporting that the March Harper's Index includes the statement:

> -Average percentage of UK population that Britons believe to be immigrants: 21%
> -Actual percentage: 8%

Harpers gives as reference the Market & Opinion Research International (MORI). We could not find this statistic on the MORI website but we found something close to it in a Readers Digest (UK) report (November 2000) of a study "Britain Today - Are We An Intolerant Nation?" that MORI did for the Readers Digest (UK) in 2000. The Digest reports:

- A massive eight in ten (80%) of British adults believe that refugees come to this country because they regard Britain as 'a soft touch'.
- Two thirds (66%) think that 'there are too many immigrants in Britain'.
- Almost two thirds (63%) feel that 'too much is done to help immigrants'.
- Nearly four in ten (37%) feel that those settling in this country 'should not maintain the culture and lifestyle they had at home'.

The Digest goes on to say:

- Respondents grossly overestimated the financial aid asylum seekers receive, believing on average that an asylum seeker gets £113 a week to live on. In fact, a single adult seeking asylum gets £36.54 a week in vouchers to be spent at designated stores. Just £10 may be converted to cash.
- On average the public estimates that 20 per cent of the British population are immigrants. The real figure is

around 4 per cent.

- Similarly, they believe that on average 26 per cent of the population belong to an ethnic minority. The real figure is around 7 per cent.

This last statistic is pretty close to the Harper's Index and the other responses give us some idea why they might over-estimate the percentage of immigrants.

In the earlier posting, the Harper's Index comments reminded Andrew of an article in the Washington Post by Richard Morin (October 8, 1995) in which Morin discussed the results of a Post/Keiser/Harvard survey "Four Americas: Government and Social Policy Through the Eyes of America's Multi-racial and Multi-ethnic Society"

The Keiser report includes the following data:

Table 1.1
## Misperceptions About the U.S. Population

What percentage of the U.S. population is ...	White Responses	African American Responses	Latino Responses	Asian Responses	Fact 1992 Census Data
...White?	49.9%	45.5%	46.7%	54.8%	74%
...African American?	23.8%	25.9%	22.7%	20.5%	11.8%
...Latino?	14.7%	16.3%	20.7%	14.6%	9.5%
...Asian?	10.8%	12.2%	10.8%	8.3%	3.1%

NOTE: Response percentages are averages of the estimates made by those
polled. Latinos can be of any race. Percentages do not total 100
because other races are not shown.

Source: Washington Post/Kaiser Family Foundation/Harvard University Survey on Race, October 1995

Note that, while it is true that the White population significantly underestimated the number of African Americans, Latinos, and Asians, the same is true for each of these groups.

John Sides sent Andrew the following data on the estimated, and actual percentage of foreign-born residents in each of 20 European countries from the the multi-nation European Social Survey :
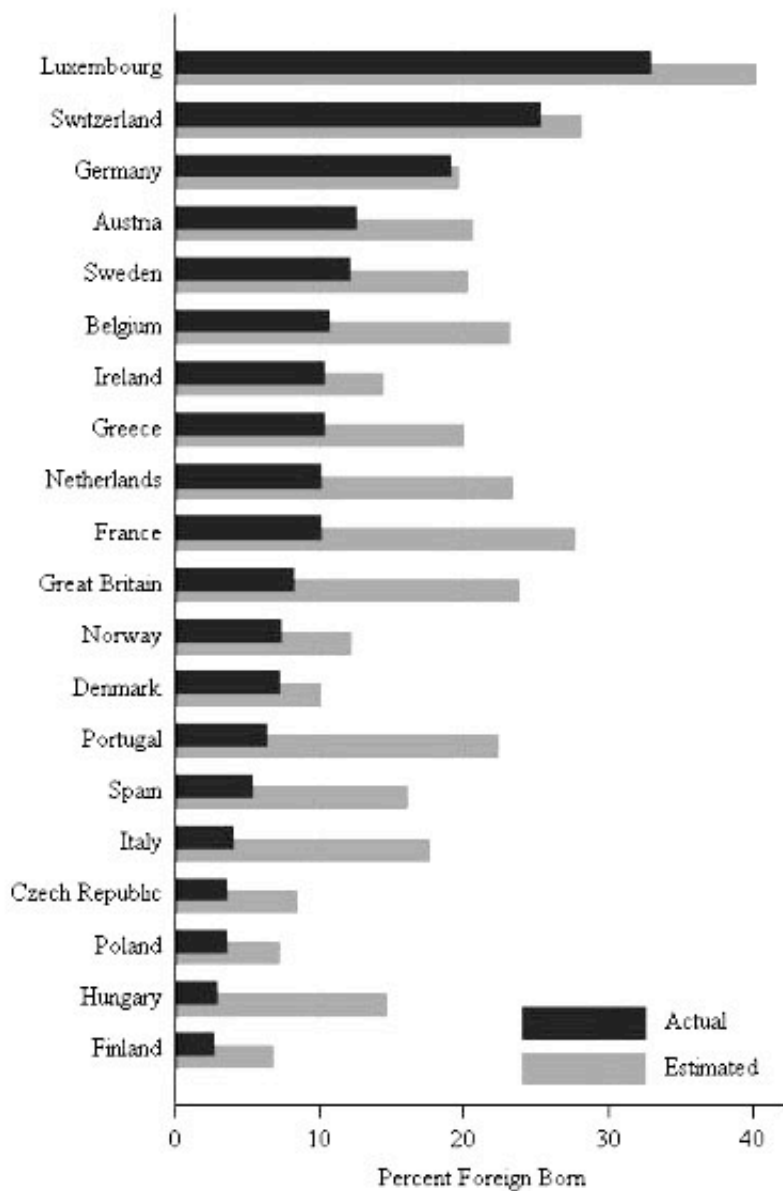
We see that we have signficant overestimation of the number of foreign-born residents, but Germany almost got it right. You will find further discussion on this topic by Andrew and John on the July1, 2005 posting on Andrew's blog.

DISCUSSION QUESTION:

(1) What explanations can you think of that might explain this overestimation? Can you suggest additional research that might clarify what is going on here?

I was quoting the statistics, I wasn't pretending to be a statistician

Sir Roy Meadow struck off by GMC
BBC News, 15 July 2005

Beyond reasonable doubt
Plus magazine, 2002
Helen Joyce

Multiple sudden infant deaths--coincidence or beyond coincidence

Sir Roy Meadow is a pediatrician, well known for his research in child abuse. The BBC article reports that the UK General Medical Council (GMC) has found Sir Roy guilty of serious professional misconduct and has "struck him off" the medical registry. If upheld under appeal this will prevent Meadow from practicing medicine in the UK.

This decision was based on a flawed statistical estimate that Meadow made while testifying as an expert witness in a 1999 trial in which a Sally Clark was found guilty of murdering her two baby boys and given a life sentence.

To understand Meadow's testimony we need to know what SIDS (sudden infant death syndrome) is. The standard definition of SIDS is:

> The sudden death of a baby that is unexpected by history and in whom a thorough post-mortem examination fails to demonstrate an adequate cause of death.

The death of Sally Clark's first baby was reported as a cot death, which is another name for SIDS. Then when her second baby died she was arrested and tried for murdering both her children.

We were not able to find a transcript for the original trial, but from Lexis Nexis we found transcripts of two appeals that the Clarks made, one in October 2000 and the other in April 2003. The 2003 transcript reported on the statistical testimony in the original trial as follows:

> Professor Meadow was asked about some statistical information as to the happening of two cot deaths within the same family, which at that time was about to be published in a report of a government funded multi-disciplinary research team, the Confidential Enquiry into Sudden Death in Infancy (CESDI) entitled 'Sudden Unexpected Deaths in Infancy' to which the professor was then writing a Preface. Professor Meadow said that it was 'the most reliable study and easily the largest and in that sense the latest and the best' ever done in this country.

> It was explained to the jury that there were factors that were suggested as relevant to the chances of a SIDS death within a given family; namely the age of the mother, whether there was a smoker in the household and the absence of a wage-earner in the family.

> None of these factors had relevance to the Clark family and Professor Meadow was asked if a figure of 1 in 8,543 reflected the risk of there being a single SIDS within such a family. He agreed that it was. A table from the CESDI report was placed before the jury. He was then asked if the report calculated the risk of two infants dying of SIDS in that family by chance. His reply was: 'Yes, you have to multiply 1 in 8,543 times 1 in 8,543 and I think it gives that in the penultimate paragraph. It points out that it's approximately a chance of 1 in 73 million.'

> It seems that at this point Professor Meadow's voice was dropping and so the figure was repeated and then Professor Meadow added: 'In England, Wales and Scotland there are about say 700,000 live births a year, so it is saying by chance that happening will occur about once every hundred years.'

> Mr. Spencer [for the prosecution] then pointed to the suspicious features alleged by the Crown in this present case and asked: 'So is this right, not only would the chance be 1 in 73 million but in addition in these two deaths there are features, which would be regarded as suspicious in any event?' He elicited the reply 'I believe so'.

> All of this evidence was given without objection from the defence but Mr. Bevan (who represented the Appellant at trial and at the first appeal but not at ours) cross--examined the doctor. He put to him figures from other research that suggested that the figure of 1 in 8,543 for a single cot death might be much too high. He then dealt with the chance of two cot deaths and Professor Meadow responded: 'This is why you take what's happened to all the children into account, and that is why you end up saying the

chance of the children dying naturally in these circumstances is very, very long odds indeed one in 73 million.' He then added:

'. . . it's the chance of backing that long odds outsider at the Grand National, you know; let's say it's a 80 to 1 chance, you back the winner last year, then the next year there's another horse at 80 to 1 and it is still 80 to 1 and you back it again and it wins. Now here we're in a situation that, you know, to get to these odds of 73 million you've got to back that 1 in 80 chance four years running, so yes, you might be very, very lucky because each time it's just been a 1 in 80 chance and you know, you've happened to have won it, but the chance of it happening four years running we all know is extraordinarily unlikely. So it's the same with these deaths. You have to say two unlikely events have happened and together it's very, very, very unlikely.'

The trial judge clearly tried to divert the jury away from reliance on this statistical evidence. He said: 'I should, I think, members of the jury just sound a word of caution about the statistics. However compelling you may find them to be, we do not convict people in these courts on statistics. It would be a terrible day if that were so. If there is one SIDS death in a family, it does not mean that there cannot be another one in the same family.'

Note that Meadow obtained the odds of 73 million to one from the CESDI report so there is some truth to the statement "I was quoting the statistics, I wasn't pretending to be a statistician" that Meadow made to the General Medical Council. Note also that both Meadow and the Judge took this statistic seriously and must have felt that it was evidence that Sally Clark was guilty. This was also true of the press. The Sunday Mail (Queenstand, Australia) had an article titled "Mum killed her babies" in which we read:

Medical experts gave damning evidence that the odds of both children dying from cot death were 73 million to one.

There are two obvious problems with this 1 in 73 million statistic: (1) Meadow assumed that in a family like the Clarks the events the "first child has a SIDS death" and "the second child has a SIDS death" are independent events. Because of environmental and genetics effects it seems very unlikely this is the case. (2) The 73 million to 1 odds might suggest to the jury that there is a 1 in 73 million chance that Sally Clark is innocent. The medical experts testimonies were very technical and some were contradictory. The 1 in 73 million odds were something the jury would at least feel that they understood. If you gave these odds to your Uncle George and asked him if Sally Clark is guilty he will very likely say "yes".

The 73 million to 1 odds for SIDS deaths are useless to the jury in assessing guilt unless they are also given the corresponding odds that the deaths were the result of murders. We shall see later that, in this situation, SIDS deaths are about 9 time more likely than murders suggesting that Sally Clark is innocent rather than guilty.

The Clarks had their first appeal in 2 October 2000. By this time they realized that they had to have their own statisticians as expert witnesses. They chose Ian Evett from the Forensic Science Service and Philip Dawid, Professor of Statistics at University College London. Both of these statisticians have specialized in statistical evidence in the courts. In his report Dawid gave a very clear description of what would be required to obtain a reasonable estimate of the probability of two SIDS deaths in a randomly chosen family with two babies. He emphasized that it would be important also to have some estimate of the variability of this estimate. Then he gave an equally clear discussion on the relevance of this probability, emphasizing the need for the corresponding probability of two murders in a family with two children. His conclusion was:

The figure *1 in 73 million* quoted in Sir Roy Meadow's testimony at trial, as the probability of two babies both dying of SIDS in a family like Sally Clark's, was highly misleading and prejudicial. The value of this probability has not been estimated with anything like the precision suggested, and could well be very much higher. But, more important, the figure was presented with no explanation of the logically correct use of such information - which is very different from what a simple intuitive reaction might suggest. In particular, such a figure could only be useful if compared with a similar figure calculated under the alternative hypothesis that both babies were murdered. Even though assessment of

the relevant probabilities may be difficult, there is a clear and well-established statistical logic for combining them and making appropriate inferences from them, which was not appreciated by the court.

These two statisticians were not allowed to appear in the court proceedings but only to have their reports read.

The Clarks' grounds for appeal included medical and statistical errors. In particular they included Meadow's incorrect calculation and the Judge's failing to warn the jury against the "prosecutor's fallacy".

Concerning the miscalculation of the odds for two SIDS in a family of two children, the judge remarks that this was already known and all that really mattered was that appearance of two SIDS deaths is unusual.

The judge then dismisses the prosecutor's fallacy with the remark:

He [Everett] makes the obvious point that the evidential material in Table 3.58 tells us nothing whatsoever as to the guilt or innocence of the appellant.

The judge concludes:

> Thus we do not think that the matters raised under Ground 3(a) (the statistical issues) are capable of affecting the safety of the convictions. They do not undermine what was put before the jury or cast a fundamentally different light on it. Even if they had been raised at trial, the most that could be expected to have resulted would be a direction to the jury that the issue was the broad one of rarity, to which the precise degree of probability was unnecessary.

The Judge dismissed the appeal.

After this the mathematics and statistical communities realized that it was necessary to explain these statistical issues to the legal community and the press. On 23 October Royal Statistics Society addressed these issues in a press release and in January 2002 they sent a letter to the Lord Chamberllor. Both of these are available here. Here is the letter to the Lord Chancelor:

> Dear Lord Chancellor,
>
> I am writing to you on behalf of the Royal Statistical Society to express the Society's concern about some aspects of the presentation of statistical evidence in criminal trials.
>
> You will be aware of the considerable public attention aroused by the recent conviction, confirmed on appeal, of Sally Clark for the murder of her two infants. One focus of the public attention was the statistical evidence given by a medical expert witness, who drew on a published study to obtain an estimate of the frequency of sudden infant death syndrome (SIDS, or "cot death") in families having some of the characteristics of the defendant's family. The witness went on to square this estimate to obtain a value of 1 in 73 million for the frequency of two cases of SIDS in such a family. This figure had an immediate and dramatic impact on all media reports of the trial, and it is difficult to believe that it did not also influence jurors.
>
> The calculation leading to 1 in 73 million is invalid. It would only be valid if SIDS cases arose Independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong reasons for supposing that the assumption is false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely than would be a case in another, apparently similar, family.
>
> A separate concern is that the characteristics used to classify the Clark family were chosen on the basis of the same data as was used to evaluate the frequency for that classification. This double use of data is well recognized by statisticians as perilous, since it can lead to subtle yet important biases.

For these reasons, the 1 in 73 million figure cannot be regarded as statistically valid. The Court of Appeal recognized flaws in its calculation, but seemed to accept it as establishing "... a very broad point, namely the rarity of double SIDS" [AC judgment, para 138]. However, not only is the error in the 1 in 73 million figure likely to be very large, it is almost certainly in one particular direction - against the defendant. Moreover, following from the 1 in 73 million figure at the original trial, the expert used a figure of about 700,000 UK births per year to conclude that "... by chance that happening will occur every 100 years". This conclusion is fallacious, not only because of the invalidity of the 1 in 73 million figure, but also because the 1 in 73 million figure relates only to families having some characteristics matching that of the defendant. This error seems not to have been recognized by the Appeal Court, who cited it without critical comment [AC judgment para 115]. Leaving aside the matter of validity, figures such as the 1 in 73 million are very easily misinterpreted. Some press reports at the time stated that this was the chance that the deaths of Sally Clark's two children were accidental. This (mis-)interpretation is a serious error of logic known as the Prosecutor's Fallacy . The jury needs to weigh up two competing explanations for the babies' deaths: SIDS or murder. The fact that two deaths by SIDS is quite unlikely is, taken alone, of little value. Two deaths by murder may well be even more unlikely. What matters is the relative likelihood of the deaths under each explanation, not just how unlikely they are under one explanation.

The Prosecutor's Fallacy has been well recognized in the context of DNA profile evidence. Its commission at trial has led to successful appeals (R v. Deen, 1993; R v. Doheny/Adams 1996). In the latter judgment, the Court of Appeal put in place guidelines for the presentation of DNA evidence. However, we are concerned that the seriousness of the problem more generally has not been sufficiently recognized. In particular, we are concerned that the Appeal Court did not consider it necessary to examine the expert statistical evidence, but were content with written submissions.

The case of R v. Sally Clark is one example of a medical expert witness making a serious statistical error. Although the Court of Appeal judgment implied a view that the error was unlikely to have had a profound effect on the outcome of the case, it would be better that the error had not occurred at all. Although many scientists have some familiarity with statistical methods, statistics remains a specialized area. The Society urges you to take steps to ensure that statistical evidence is presented only by appropriately qualified statistical experts, as would be the case for any other form of expert evidence. Without suggesting that there are simple or uniform answers, the Society would be pleased to be involved in further discussions on the use and presentation of statistical evidence in courts, and to give advice on the validation of the expertise of witnesses.

Yours sincerely
Professor Peter Green,
President, Royal Statistical Society.

Now that we all agree that we need to know the relative probability that a family with two babies loses them by SIDS deaths or by murder, what are these probabilities?. Roy Hill, Professor of Mathematics at the University of Salford tackled this question. His results were first given in an unpublished paper "Cot death or Murder-weighing the probabilities" presented to the Developmental Physiology Conference, June 2002 (available from the author). Hill published his results in the his article "Multiple sudden infant deaths--coincidence or beyond coincidence?, "Paediatric and Perinatal Epidemiology" 2004, 18, 320-326.

In the trial the news media made frequent references to "Meadow's law". This law is: "One cot death is a tragedy, two cot deaths is suspicious and three cot deaths is murder". This motivated Hill to test this law by estimating the relative probability of SIDS and murder deaths for the case of 1 baby, 2 babies, and 3 babies. This is a difficult problem since data from different studies give different estimates, the estimates differ over time etc. However Hill did a heroic job of combining all data he could find to come up with reasonable estimates. Here is what he found:

Regarding the issue of independence Hill concludes:

In the light of all the data, it seems reasonable to estimate that the risk of SIDS is between 5 and 10 times greater for infants where a sibling has already been a SIDS victim.

As to the relative probabilities of SIDS deaths and murders Hill provides the following estimates:

(1) An infant is about 17 times more likely to be a SIDS victim than a homicide victim.

(2) Two infants are about 9 times more likely to be SIDS victims than homicide victims.

(3) Three infants have about the same probability of being SIDS victims or homicide victims.

These estimates do not support Meadow's Law. Despite many references to Meadow's Law in the medical journals and the news media, the editor for Hill's article comments that it appears to be due to D.J. and F.J. M Di Maio and seems not to appear in any of Meadow's writings.

Hill's analysis is used in the very nice article on the Sally Clark case "Beyond reasonable doubt" by Helen Joyce in *Plus magazine*. This is a great article to have students read.

After the failure of their appeal, the Clarks started a campaign to get the news media to support their campaign. They also continued to search for medical explanations for their children's deaths. In the process they found that the prosecutor's pathologist who had performed the autopsies for the two children had withheld the information that their second child had been suffering from a bacterial infection which could have been the cause of a natural death. Recall that his first opinion had been that the first child also was a natural death. This information and the flawed statistics led the Criminal Cases Review Commission, which investigates possible miscarriages of justice, to refer the case back to the courts for another appeal.

In this appeal the judge ruled that if the bacterial infection information had been known in the original trial, the Sally Clark would probably not have been convicted, and so he allowed the appeal and quashed the convictions freeing Sally Clalrk after two and a half years in jail..

The judge also agreed that the statistical evidence was seriously flawed and conclued:

Thus it seems likely that if this matter had been fully argued before us we would, in all probability, have considered that the statistical evidence provided a quite distinct basis upon which the appeal had to be allowed.

Thus we cannot say that the famous 1 chance in 73 million statistic was responsible for Sally Clark being freed from jail but it is very likely the reason she spent two and a half years in jail.

For the complete Sally Clark story we recommend the book "Stolen Innocence" by John Batt available at U.K. Amazon. Batt is a lawyer and good friend of the Clarks. He attended the trials and his book tells the Sally Clark story from beginning to end. We also found it interesting to read the transcripts of the 2000 and 2003 appeal. These were not easy to find so we include at the end of this article the Lexis Nexis path to these transcrips.

Sir Roy Meadow was also the key prosecution witness in two other cases similar to the Sally Clark case: the Angela Canning case and the Trupti Patel case. Roy Hill also wrote an intersting article "Reflections on the cot death cases", *Significance*, volume 2 (2005), issue 1 in which he discusses the statistical issues in all three cases.

Finding the 2000 and 2003 transcripts in Lexis Nexis.

Open Lexis Nexis

Choose "Legal Research" from the sidebar

From "Case Law" choose "Get a Case"

Choose" Commonwealth and Foreign Nations" from the sidebar

Choose "Sally Clark" for the "Keyword"

Choose "UK Cases" for the "Source"

Choose "Previous five years" for the "Date."

The two "r v Clarks" are the appeals.

DISCUSSION QUESTIONS:

(1) Do you think it is easonable not to allow a doctor to practice because of a statistical error in a court case?

(2) Do you agree with the RSS that statistical evidence should only be provided by statistical experts?

(3) Give the 73 million to 1 odds to a few of your friends and ask them if they were on the jury, in the absence of any other information, would they think that this makes it is very likely that Sally Clark is guilty?

What women want

What women want
The New York Times, May 24, 2005, A 25
John Tierney

Are men more competitive than women?

In this OP-ED piece Tierney discusses and draws conclusions from a recent study that explores gender differences in competitive environments. Two researchers, Muriel Niederle of Stanford and Lise Vesterlund of the University of Pittsburgh, ran an experiment in which women and men had to choose to participate in either a competitive or a non-competitive task. They found that, among other things, women chose to compete less often than they should have, while conversely, men chose to compete more often than they should have. The researchers apparently anticipated this result, so in addition, Nielderle and Vesterlund designed their experiment to explore potential reasons for this difference.

Specifically, participants were paid to add up as many sets of five two-digit numbers as they could in five minutes. The experiment consisted of four tasks, each containing a potentially different scheme for how participants would be paid. After the tasks were completed, for actual payment one of the four tasks was chosen at random. Here is a brief description of the four tasks. (A more detailed description and analysis may be found at Muriel Niederle's website, in the draft article "Do Women Shy away from Competition? Do Men Compete too Much?"

1. (Piece-rate) Each participant calculates the given sums and is paid 50 cents per correct answer.

2. (Tournament) The participants compete within four-person teams consisting of two women and two men. The person who completes the most correct sums receives $2.00 per sum; the other members of the group receive nothing.

3. (Tournament Entry Choice) Each participant is given a choice of payment scheme: either by piece-rate or a tournament scheme in which the participant is paid $2.00 per correct sum if and only if she completes more correct sums than were completed by the other members of her group in task 2. (Thus it is possible for more than one member of the group to "win" the tournament.)

4. (Tournament Submission Choice) No new sums are calculated. Instead, each participant is given a choice: either receive the same piece-rate payment as was generated in task 1, or submit one's task 1 performance to a tournament in which the participant receives $2.00 per correct sum if and only if she completed more correct sums in task 1 than

did the other members of her group. (Again, it is possible for more than one member of the group to "win" the tournament.)

At the end of each task, each participant is only told her own performance on the task, and thus her decision to enter a tournament (in tasks 3 and 4) is not based on relative-ranking information. Also, after the tasks were completed, each subject was asked to guess the rank of her task 1 and task 2 performances. The main goal of the study is determine if men and women of the same ability on a task choose to compete at different rates, and if so, why.

Here are some highlights of their findings:

- Women and men performed equally well on both tasks 1 and 2.

- Women and men performed significantly better on task 2 (tournament) then on task 1 (piece-rate), and the size of the increase was independent of gender.

- Of the 20 tournaments in task 2, women won 11, men 9.
- 43% of the women, versus 75% of the men, ranked themselves first in their group.

- 35% of the women chose the tournament in task 3, versus 75% of the men.

- Women's task 2 performance does not predict tournament entry in task 3, and only does so marginally for men. In fact, women in the highest performance quartile for task 2 were less likely to enter the tournament than men in the lowest quartile.

- Approximately 27% of the gender difference in task 3 tournament entry can be explained by women and men forming different beliefs about their relative ranking. The remaining difference comes from a mix of both general factors (e.g. risk aversion) and tournament-specific factors (e.g. bias in estimating future performance.)

- Approximately 70% of women whose expected gain under a tournament scheme is favorable do not choose the tournament (tasks 3 and 4,) while approximately 63% of men elect a tournament when it is unfavorable to them (task 3.)

- 25% of the women submitted their task 1 results to a tournament in task 4, versus 55% of the men. Virtually all of this difference can be explained by men's over-confidence in their relative ranking.

DISCUSSION QUESTIONS:

1. Why do you think the task 3 and task 4 tournaments were designed the way they were?

2. In discussing the above gender differences, Tierney writes, "You can argue that this difference is due to social influences, although I suspect it's largely innate, a byproduct of evolution and testosterone. Whatever the cause, it helps explain why men set up the traditional corporate ladder as one continual winner-take-all competition-- and why that structure no longer makes sense." What do you think?

3. The researchers determined the probability of winning the tournament in Task 2 by "randomly creat[ing] four-person groups from the observed performance distributions." How exactly would one do this? They also determined, for each performance level (e.g., 15 correct sums) and each gender, the probability of winning a tournament with that score. How would this be done?

4. Niederle and Vesterlund also briefly discuss the cost to women for under-entry into tournaments and the costs to men for over-entry. They write, "While the magnitude of the costs is sensitive to the precise assumptions we make, the qualitative results are the same. The total cost of under-entry is higher for women, while the total cost of over-entry is higher for men. Since over-entry occurs for participants of low performance and under-entry for those with

high performance, by design the cost of under entry is higher than that of over entry." Explain and comment.

Rules of engagement - modelling conflict

The mathematics of warfare - Scientists find surprising regularities in war and terrorism

The mathematics of warfare
The Economist July 23, 2005 (Available from Lexis Nexis)

Is terrorism the next format for war?
Nature July 12, 2005
Phillip Ball

Academics Neil Johnson from the Univ. of Oxford and Michael Spagao from Royal Holloway College London are using the patterns of casualities to model the development of wars. They are attempting to monitor the casualties of the conflict in Iraq, using data from a database called IraqBodyCount.

The Nature article says:

> All wars and conflicts seem to generate a common and distinctive pattern of death statistics. Fifty years ago, the British mathematician Lewis Fry Richardson found that graphs of the number of fatalities in a war plotted against the number of wars of that size follow a relationship called a power law, where all the data points fall on a straight line if plotted logarithmically. This power law encodes the way in which large battles with large numbers of deaths happen very infrequently, and smaller battles happen more often.

The Economist article also gives a nice summary of power law relationships

> Power-law relationships are characterised by a number called an index. For each tenfold increase in the death toll, the probability of such an event occurring decreases by a factor of ten raised to the power of this index, which is how the distributions get their name.

The Johnson and Spagao paper suggests a difference between conflicts inside and outside G7-countries based on their index value.

A more worrying statistic comes from another paper on the same topic by Clauset and Maxwell at the British Institute of Physics who suggest that we can expect another attack at least as severe as September 11th within the next seven years.

How people respond to terrorist attacks

The rational response to terrorism
The Economist print edition July 21st 2005, Available from Lexis-Nexis

Nobel laureate Gary Becker from the University of Chicago and Yona Rubinstein from Tel Aviv University examine how the general public responds to the threat posed by suicide-bombers in Fear and the Response to Terrorism: An Economic Analysis.

A first analysis suggests an obvious response. The miles flown by passengers on US domestic airlines fell 30% between August and October 2001 and air travel hadn't regained its 2001 peak even two years after the attack of September 11th. According to Becker and Rubinstein, it is not the risk of physical harm that moves people; it is the emotional disquiet. People respond to fear, not risk.

They give an example of the effect of suicide-bombers on bus usage in Tel Aviv. There was one attack a month, for a

year, on average, from November 2001 and bus usage fell 30%. But this average masks material differences between different types of passengers. Casual users who bought tickets on the day of travel were much more likely to stay away with usage falling 40% after each attack. But regular passangers who used weekly or monthly tickets were largely undeterred.

The authors claim that the public responds to terrorism in a similar manner to its reaction to rare but deadly diseases, such as BSE or 'mad-cow disease', by avoiding beef en masse even though the probability of infection is very small.

They explain this reaction by saying that people can overcome their fear but they will only do so if it is worth their while. And overcoming their fear is a fixed cost, not a variable one, so people do not fight their fear each time they step on a bus; this only happens on their first journey. Once a person has come to terms with terror, it makes little difference whether he gets the bus twice a day or once a day. This choice may result in slightly higher risk of actual attack but a traveller is not adding anything to his fear of such a catastrophe. And it is fear, not the risk, that influences people.

Can you get fired over the wording of a questionnaire?

Researcher to be sacked after reporting high rates of ADHD
BMJ, Mar 26 1995, 330 (7493); 691
Jeanne Lenzer

This article is not currently available without a subscription, but will be available to the general public twelve months after the original publication date.

Dr Gretchen LeFever, a researcher who has claimed that attention deficit hyperactivity disorder in children has been overdiagnosed and overmedicated has been placed on administrative leave with the intent of terminating her employment. Her employer, the East Virginia Medical School, has accused her of scientific misconduct. In the article we read:

> Her work has been controversial. She first made headline news in 1999 when she reported that 8% to 10% of elementary school pupils in southeastern Virginia were being prescribed drugs for ADHD, a percentage two to three times the estimated national average (American Journal of Public Health 1999;89:1359-64).

> Criticism grew after she published the results of a 2002 study showing that the prevalence of the disorder among children in grades 2 to 5 had risen to 17% (Psychology in the Schools 2002; 39: 63-71).

> One of her main critics is Jeffrey Katz, a clinical psychologist in Virginia Beach and the local coordinator of the Children and Adults with Attention-Deficit/Hyperactivity Disorder group. Dr Katz questioned her claim that the condition had been diagnosed in 17% of children in grades 2 to 5.

> He said, 'When somebody like Dr LeFever makes these claims that are apparently not based on good research, it minimises a very real problem. Parents won't bring their children in for evaluation, because they are afraid that medication will be automatically prescribed. They think it's a bad thing and the sole treatment. But medication can have significant benefits.'"

An anonymous whistleblower accused her of scientific misconduct based on her 2002 publication. The survey question asked

> Does your child have attention or hyperactivity problems, known as ADD or ADHD?

but the publication reported the question as

> Has your child been diagnosed with attention or hyperactivity problems known as ADD or ADHD?

She has also been accused of conducting research on children without getting parental approval. The local IRB (Institutional Review Board) had originally determined that since only parents and teachers filled out surveys about the children that the children were not research subjects. This meant that the study was exempt from the normal parental approval requirements. After the allegations of scientific misconduct were raised, the medical school sought a second opinion from the national experts at the Office of Human Research Protections. This office ruled that the children were indeed research subjects. This meant that the research was not exempt from parental approval requirements.

Allegations of misconduct often degenerate into a "he said/she said" argument and it is difficult for an outsider to objectively evaluate the evidence. A web search on the name "Gretchen LeFever " will produce a wide range of opinions about her original research and the rationale for her firing.

Discussion questions:

(1) Is there a serious difference in the reported wording of the questionnaire? Would you expect the first wording to get a higher positive response? Why?

(2) Should a researcher be held responsible for ethical violations for a study that was approved by the local IRB if that approval was later found to be in error?

(3) When a parent fills out a survey about their child, are they implicitly giving permission for their child to be part of the research study? If not, what would constitute permission?

(4) Do you believe that Dr. LeFever is guilty of scientific misconduct? What would be an appropriate punishment?

You can't just go on telly and make up statistics, can you?

Take with a pinch of sodium chloride
The Guardian Aug 18th 2005
Margaret McCartney

It seems we can't buy anything unless it has the approval of boffins. But what does any of it mean? Margaret McCartney examines the suspect science that we swallow, apply and absorb every day in an online GuardianUnlimited article from .

When you read or hear something like '8 out of 10 people prefered X to Y', what are the details behind this sample survey result. The article give Pantene Pro-V as an example. They have recently been telling us, via shiny spreads in various magazines and TV ads, that its Anti-Breakage Shampoo, will lead to "up to 95% less breakage in just 10 days". It transpires that the sample size is just 48 and the survey was not a blind one.

Further investigations reveal that 10 samples of hair were tested, three times and the results were "significant". Furthermore the associated adds were vetted and approved by the Broadcast Advertising Clearing Centre.

In another example the UK's Advertising Standards Authority (ASA), with a staff of 100, look at all the major newspapers daily, but with an estimated 30m adverts printed every year in the UK, it is impossible for them to look at them all. In a recent case, a slimming pill advertising was withdrawn after making claims that were found to be based on a study on just 44 people. The ASA decided that this was too small a study to be valid.

The ASA director of communications "Talking generally, we may accept a small sample size as reasonable proof, but this would really depend on the statistical significance of whatever tests were done. Conditional claims lead to a host of different claims, especially when 'modal verbs' are used. We might ask them to change 'can' to 'could' if they didn't have 100% proof of the 'can'. But we would also expect them to hold proof relating to the 'could'."

The article lists other examples of dubious statistics such as "93% say their skin felt softer, and 79% say their skin

was firmer with each application" (of a skin care prodct) or a more serious example about medicinal benefits such as 'the drug "effectively reduces the risk of a heart attack" by "preventing build up of harmful plaques in your coronary arteries" and "reducing your risk of coronary heart disease"'.

"The key issue is that of evidence. If you don't have evidence to justify claims of benefit, then the whole argument begins to fall apart." says Dr Ike Iheanacho, editor of the Drug and Therapeutics Bulletin, a journal published by Which?.

The article finishes with the warning that marketing and science have got together and bred a weird hybrid form of sales-experiments that have taken over our advertising culture.

The more the merrier? First born do better at school

First born do better at school
The Guardian,Monday August 22, 2005
Rebecca Smithers, education editor,

This article highlights that younger children do less well in terms of overall educational attainment than their older brothers and sisters, regardless of family size or income. Futher, the impact of birth order was more pronounced in females in later life. This suggests that parents with limited financial resources may invest more time and money in the education of their eldest child.

The underlying data are based on the entire population of Norway aged 16-74, between the years 1986 and 2000. This unique data set collated using Norway's personal identity number system, allowed them to look across families and within families to distinguish the causal effect of family size on youngsters' education.

The authors comment that "there's a lot of psychological literature on why first-born children are most successful. The main suggestion is that the eldest child acts as a teacher for the younger children and learns how to organise information and present it to others." The research team followed the children through to adulthood and examined their earnings, full-time employment status and whether the individual had become a teenage parent. The findings are claimed to represent the first comprehensive analysis of the impact of family composition on educational achievement.

"In terms of educational attainment, if you are the fourth born instead of the first, you get almost one year less education, and that is quite a lot," Salvanes, the lead author, told Reuters. "And first-born children tend to weigh more at birth than their younger brothers and sisters, which is a good predictor for educational success. Children alone with two adults also tend to get more intellectual stimulation than children in large families who get less parental attention. First-born children seem to learn from teaching their younger siblings, contrary to the common notion that younger children benefit by learning from their elders", Salvanes said. So does that mean big sisters really are smarter? "Yes. It's hard to admit because I have older sisters," Salvanes said.

The research was carried out by Sandra Black and Paul Devereux in the Dept. of Economics at UCLA and Kjell Salvanes at the Norwegian School of Economics and Business Administration. It will be presented at the 2005 World Congress of the Econometric Society, and published in the Quarterly Journal of Economics. For now, the original paper is available on-line.

Racial Profiling

Profiling Report Leads to a Demotion
The New York Times, August 24, 2005
Eric Lichtblau

Lawrence Greenfield, head of the Bureau of Justice Statistics, was recently demoted after a dispute over a study of racial profiling.

The flashpoint in the tensions between Mr. Greenfeld and his political supervisors came four months ago, when statisticians at the agency were preparing to announce the results of a major study on traffic stops and racial profiling, which found disparities in how racial groups were treated once they were stopped by the police.

Political supervisors within the Office of Justice Programs ordered Mr. Greenfeld to delete certain references to the disparities from a news release that was drafted to announce the findings, according to more than a half-dozen Justice Department officials with knowledge of the situation. The officials, most of whom said they were supporters of Mr. Greenfeld, spoke on condition of anonymity because they were not authorized to discuss personnel matters.

What exactly, was in this report?

The April study by the Justice Department, based on interviews with 80,000 people in 2002, found that white, black and Hispanic drivers nationwide were stopped by the police that year at about the same rate, roughly 9 percent. But, in findings that were more detailed than past studies on the topic, the Justice Department report also found that what happened once the police made a stop differed markedly depending on race and ethnicity.

Once they were stopped, Hispanic drivers were searched or had their vehicles searched by the police 11.4 percent of the time and blacks 10.2 percent of the time, compared with 3.5 percent for white drivers. Blacks and Hispanics were also subjected to force or the threat of force more often than whites, and the police were much more likely to issue tickets to Hispanics rather than simply giving them a warning, the study found.

It's worth noting that the dispute was about the press release and not about the report itself. The [full report] is out on the web

The statistics described in the New York Times article appear in the followinng table (table 9 in the report).

Table 9. What police did during the traffic stop, by gender, race/Hispanic origin, and age of stopped drivers, 2002

| | | | What police did during traffic stop | | | | | | |
| | | | Searched — | | | | | Used force | Used exces- |
Characteristic of stopped drivers	Issued a warning	Ticketed the driver	Driver or vehicle	Driver	Vehicle	Handcuffed the driver	Arrested the driver	against the driver*	sive force against the driver*
Total	25.3%	58.6%	5.0%	3.6%	4.0%	2.8%	2.7%	1.1%	1.0%
Gender									
Male	24.2%	59.8%	7.1%	5.4%	5.6%	3.8%	3.7%	1.6%	1.4%
Female	26.9	56.8	1.8	0.9	1.6	1.1	1.1	0.4*	0.3*
Race/Hispanic origin									
White	27.4%	56.5%	3.5%	2.5%	2.9%	2.0%	2.0%	0.8%	0.7%
Black	18.3	58.4	10.2	8.1	7.1	6.4	5.8	2.7	2.5
Hispanic	18.2	71.5	11.4	8.3	10.1	5.6	5.2	2.4	2.0*
Other race	20.3	73.1	2.9*	1.3*	2.3*	0*	0*	1.1*	0.5*
Age									
16-19	31.0%	58.6%	8.9%	6.1%	8.2%	3.0%	2.9%	2.9%	2.9%
20-29	24.8	64.3	6.6	5.4	5.0	3.9	3.7	1.3	1.1
30-39	23.8	61.3	6.4	4.5	5.0	3.1	2.9	1.2	1.0
40-49	23.3	58.5	3.4	1.9	2.8	2.3	2.5	0.8*	0.6*
50-59	26.5	50.2	1.5	1.2*	1.1*	1.8	1.5	0.5*	0.4*
60 or older	26.7	44.9	0*	0*	0*	0*	0.5*	0*	0*
Estimated number	4,241,229	9,843,043	837,809	609,278	671,957	461,582	448,094	188,882	164,255

*Estimate is based on 10 or fewer sample cases.
*Force includes threat of force.

Critics of the Bush administration have accused them of burying the report, but if that was the intent, the publicity has only amplified the attention that this report has received. They also cite this report as evidence that the Bush

administration punishes those who publicize bad news.

There has always been concern about the independence of statistical estimates produced by U.S. Federal Agencies. If an administration could manipulate estimates of inflation and/or unemployment, then no one would trust those figures anymore.

On the other hand, politicians have always worried about unelected career government employees who may not be responsive or may even be openly hostile towards the goals of the to the elected President of the United States. Mr. Greenfield seems to appreciate the two sides of this issue in some quotes from him in the New York Times article.

> Mr. Greenfeld declined to discuss the handling of the traffic report or his departure from the statistics agency. But he emphasized in an interview that his agency's data had never been changed because of political pressure and added that "all our statistics are produced under the highest quality standards."

> As a political appointee named to his post by Mr. Bush in 2001, "I serve at the pleasure of the president and can be replaced at any time," Mr. Greenfeld said. "There's always a natural and healthy tension between the people who make the policy and the people who do the statistics. That's there every day of the week, because some days you're going to have good news, and some days you're going to have bad news."

This article has received a lot of coverage in the more liberal blogs. Run a web search on "blog Lawrence Greenfield racial profiling" to see some examples.

Bob Herbert, a writer on the editorial pages of the New York Times also commented on the racial profiling article on August 25. He offered his opinions, and then shared the following two anecdotes.

> Rachel Ellen Ondersma was a 17-year-old high school senior when she was stopped by the police in Grand Rapids, Mich., on Nov. 14, 1998. She had been driving erratically, the police said, and when she failed a Breathalyzer test, she was placed under arrest.

> An officer cuffed Ms. Ondersma's hands behind her and left her alone in the back seat of a police cruiser. What happened after that was captured on a video camera mounted inside the vehicle. And while it would eventually be shown on the Fox television program "World's Wackiest Police Videos," it was not funny.

> The camera offered a clear view through the cruiser's windshield. The microphone picked up the sound of Ms. Ondersma sobbing, then the clink of the handcuffs as she began maneuvering to free herself. She apparently stepped through her arms so her hands, still cuffed, were in front of her. Then she climbed into the front seat, started the engine and roared off. With the car hurtling along, tires squealing, Ms. Ondersma could be heard moaning, "What am I doing?" and, "They are going to have to kill me."

> She roared onto a freeway, where she was clocked by pursuing officers at speeds up to 80 miles per hour. She crashed into a concrete barrier, and officers, thinking they had her boxed in, jumped out of their vehicles. But Ms. Ondersma backed up, then lurched forward and plowed into one of the police cars.

> Gunfire could be heard as the police began shooting out her tires. The teenager backed up, lurched forward and crashed into the cop car again. An officer had to leap out of the way to keep from being struck.

> Ms. Ondersma tried to speed away once more, but by then at least two of her tires were flat and she could no longer control the vehicle. She crashed into another concrete divider and was finally surrounded.

> As I watched the videotape, I was amazed at the way she was treated when she was pulled from the cruiser. The police did not seem particularly upset. They were not rough with her, and no one could be

heard cursing. One officer said: "Calm down, all right? I think you've caused enough trouble for one day."

This is in contrast to a second incident in April 1998 where four young men in a van were pulled over.

> They were neither drunk nor abusive. But their van did roll slowly backward, accidentally bumping the leg of one of the troopers and striking the police vehicle.

> The troopers drew their weapons and opened fire. When the shooting stopped, three of the four young men had been shot and seriously wounded.

The woman in the first incident was white. In the second incident, three of the men were black and one was Hispanic.

Discussion questions:

(1) Do statisticians in the various U.S. Government Agencies need greater independence to protect them from political influences? If so, how could this be best achieved?

(2) Read the full report on racial profiling. What are the limitations to this study? How serious are the limitations?

(3) Do you find Mr. Herbert's two anecdotes to be more persuasive than the statistics on racial profiling?

Retrieved from "http://chance.dartmouth.edu/chancewiki/index.php/Chance_News_%28July-August_2005%29"