# A Case of Scientific Fraud? A Statistical Approach

Nora Bohossian

Master of Science

Department of Mathematics and Statistics

McGill University

Montreal,Quebec

2006-11-13

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Science.

*Your file Votre référence*
*ISBN: 978-0-494-32670-1*
*Our file Notre référence*
*ISBN: 978-0-494-32670-1*

"Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!"

**John Godfrey Saxe**
*"The Blind Men and the Elephant"*

## DEDICATION

This document is dedicated to all the young scientists striving for recognition in their field.

## ACKNOWLEDGEMENTS

This thesis reflects the scientist I am today, to which many professors have directly or indirectly contributed through the course of my education. I have an enormous amount of respect and gratitude for them all. However, I would like to acknowledge the profound effect Professor David Wolfson has had on shaping me, the statistician I am today. I thank him for suggesting this intriguing topic, for his incredible attention to detail, and the numerous long discussions when time was not on his side. I recognize that I have been particularly lucky to have had worked with him on this thesis. With its completion, I am taking with me a very precious experience, the benefits of which I will come to appreciate even more with time.

I would like to thank dearly Professor Terrence Speed of the Department of Statistics at the University of California at Berkeley, for making this thesis a possibility by sending me the statistical analyses, which led the Office of Research Integrity to find Imanishi-Kari guilty of scientific misconduct.

Professor Russell Steele's timely assistance in helping me understand the theory behind the mixture models was absolutely invaluable. I am particularly grateful to Professor Steele for sharing with me his R code on the Poisson mixture models.

I am especially indebted to Professor Marianna Newkirk for her valuable insight into how the case of Imanishi-Kari affected the field of immunology. For the same reasons, I would like to thank Professors Taff Jones, Malcolm Baines and Robert Murgita.

I am also very grateful to Dalia Halawani for sitting down with me one Sunday

afternoon and walking me through the complicated immunological theory behind the experiments that Imanishi-Kari carried out.

When I was just starting to research the topic on the misuse of statistics, Professor George P. H. Styan suggested several good readings and sources, which I appreciated very much.

I also thank Lory Ajamian and Celine Pilon for the translation of the thesis abstract in French.

My brother, Hrant Bohossian, has been very helpful in improving the quality of the images contained in this thesis. I thank him for his time and expertise.

The library service provided by Marika Asimakopulos has been impeccable and I would like to thank Marika for all her help.

A Masters thesis involves many administrative steps and I thank Carmen Baldonado for doing an incredible job in ensuring I follow these steps correctly and in a timely fashion.

This Masters thesis was primarily supported by the Department of Mathematics and Statistics and I would like to express my sincere gratitude for this financial assistance.

Last but certainly not least, I deeply thank my family and friends for their love and support always.

## ABSTRACT

In 1986 Thereza Imanishi-Kari, then an assistant professor at the Massachusetts Institute of Technology, was at the peak of her career. She had just coauthored a paper in the prestigious journal Cell with David Baltimore, a Nobel laureate. Their research was exciting and their findings promising.

Margot O'Toole, Imanishi-Kari's postdoctoral fellow at the time, was unable to reproduce some of the experimental results published in the paper and could not resolve this with her postdoctoral supervisor. Subsequently, O'Toole became convinced that there were serious errors in the paper and, shortly afterwards, the National Institutes of Health began officially investigating the questions she raised about it.

It may have been simply a character clash between Imanishi-Kari and O'Toole but partly due to the involvement of a figure such as Baltimore, this clash possibly ruined their careers, took 10 years to settle down, cost millions of dollars of public money, polarized the scientific community, and went down in history as one of the most widely followed cases of scientific fraud.

Based on statistical, forensic and other evidence, Imanishi-Kari was found guilty of scientific misconduct and banned from receiving public funding for 10 years. This was not the end of the matter, however, because Imanishi-Kari appealed the decision and was later exonerated.

In this thesis, we tell the statistical story by putting forward the statistical arguments that were used against Imanishi-Kari and the counterarguments to them.

# ABRÉGÉ

En 1986, la carrière de Thereza Imanishi-Kari, professeure adjointe à l'institut de Technologie du Massachusetts, était en plein essor. Elle venait de coécrire un article dans le prestigieux journal Cell avec David Baltimore, un lauréat du prix Nobel. Leur recherche était intéressante et leurs résultats prometteurs.

Margot O'Toole, contemporaine de Imananishi-Kari au postdoctorat l'époque, était incapable de reproduire certains des résultats expérimentaux publiés et ce, même avec l'aide de son directrice de recherche aux études postdoctorales; elle devint donc convaincue que le rapport comportait de sérieuses erreurs. S'ensuivit une enquête officielle par l'Institut national de la santé sur les doutes émis par O'Toole.

Ce qui aurait pu n'être qu'un malentendu ou un conflit de personnalité entre deux chercheurs devint un problème notoire dans la communauté scientifique à cause de l'implication de figures proéminentes comme Baltimore; le litige dura 10 ans et coûta des millions de dollars en fonds publics. La controverse aurait facilement pu détruire la carrière des deux femmes.

Sur la base de preuves statistiques, légales et autres, Imanishi-Kari fut reconnue coupable de mauvaise conduite scientifique et il lui fut interdit de recevoir des bourses publiques pour 10 ans. L'affaire ne s'arrêta toutefois pas là, puisque Imanishi-Kari en appela de la décision et fut exonérée.

La présente thèse vient analyser l'enchainement statistique des événements qui se sont déroulés, par le biais des arguments qui utilisés contre Imanishi-Kari et des arguments qui ont servi à la déculpabiliser.

TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 0

## The Road To This Thesis Topic

During my qualifying year in the Masters program, I took my first courses in statistics. Until then, I really did not know what statistics was about. I had gone through a Bachelor's degree without ever being exposed to statistics from the *academic* point of view; I only knew of statistics, as most of people know it, as pervading the news, on the weather channel, in sports, in fact in nearly every aspect of everyday life.

Because of the poor reputation that statistics (unjustly) carries, I acquired an interest in the misuse of statistics and I felt that my Masters thesis would be a great opportunity to learn more about the possible source of some of this suspicion about statistical practice. To provide greater focus, I decided to draw upon my other strong interest, health, and thus I came to "the misuse of statistics in the field of medicine."

With these key words in mind, I was quickly led, through an online search, to Professor Douglas Altman, who is currently the director of Centre for Statistics in Medicine and Cancer Research UK Medical Statistics Group. I was so determined to learn more about this questionable side of statistics that not even the fact that he was in England and I in Canada prevented me from contacting him to see if he would be interested in being my supervisor. Thankfully, he was wiser than I, and

after initial interest, he politely declined, explaining to me that overseas supervision would be too complicated.

Next, I entered into communication with Professor Chamont Wang at Trenton State College, New Jersey, whose book "Sense and Nonsense of Statistical Inference," I was reading at the time. He was quick to offer help and support which I shall not forget.

And, as if that was not enough, I was also communicating electronically with Professor Herbert F. Spirer at Columbia University, New York, whose book, "Misuse of Statistics," I was reading as well. For approximately the full summer, he kept "alerting" me to all the current misuses that were taking place in the news as well as to other interesting ones about which he knew.

At around that time, the person responsible for me taking up statistics in the first place sat me down, told me, essentially, "to calm down" and that he and I could look up some topics on the misuse of statistics and "take it from there." This person is my current supervisor, Professor David Wolfson, to whom I shall be forever thankful.

He gave me three ideas: the ongoing breast cancer screening controversy, the studies on therapeutic touch and the case of Imanishi-Kari. The goal, essentially, was to have the thesis comprise of three or four cases where statistics had been misused. All were intriguing and I spent most of the summer of 2005 reading up on them and accumulating knowledge.

While I was researching the case of Imanishi-Kari, I contacted Professor Terence Speed of the University of California at Berkeley, since I could not obtain the statistical analysis, which had played an important role in the initial finding of scientific misconduct. Professor Speed who had, as a statistical expert, defended Imanishi-Kari, had already communicated with a fellow student, Geva Maimon, a couple of years before. By the beginning of the academic year, I had the report on the statistical analysis in my possession, courtesy of Professor Speed.

It soon became clear to my supervisor and me that the full thesis could concentrate on the case of Imanishi-Kari alone.

It is my sincere hope that you will find this thesis topic as intriguing as I did.

Happy reading!

"Those were horrible years. I want to make sure that people know that I was exonerated - and that I am just like anybody else."

# CHAPTER 1

## Introduction

In 1986, David Weaver, Moema H. Reis, Christopher Albanese, Frank Constantini, David Baltimore and Thereza Imanishi-Kari published a paper in the 45th volume of the prestigious journal Cell. (Weaver et al., 1986) In 1994, the Office of Research Integrity (ORI) found Imanishi-Kari guilty of scientific misconduct for deliberately falsifying data, which appeared in the paper, and data, which she subsequently used to oppose the initial charges. (Grigg, 1994) Imanishi-Kari appealed the ORI's decision. In 1996, the Research Integrity Adjudication Panel, Departmental Appeals Board, Department of Health and Human Services, concluded that the evidence was not sufficient to support a case of misconduct. (Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel, 1996) In this chapter we briefly recount the turn of events that led to the initial indictment of Imanishi-Kari and, later, to her exoneration.

The results reported in the Cell paper and the main conclusion drawn by its authors had minor, though important, implications for the treatment of weakened immune systems caused by diseases such as AIDS. The conclusion was and still is controversial within the biological sciences community.

The immune system produces a large number of antibodies. Antibodies are proteins, which attack antigens, that is foreign substances, entering the body. The antibodies attack the antigens by binding to them. Once the immune system has produced antibodies to fight a specific type of antigen, those antibodies will always be present in the body and ready to counter that specific type of antigen the next time it enters the body. That is why people generally suffer from viral diseases such as mononucleosis (mono) or chickenpox, only once.

In certain mice, antibodies display a distinctive, genetically determined, chemical feature called an idiotype. These idiotypes can be used to study the inheritance of the genes that produce antibodies.

Imanishi-Kari had obtained a certain kind of antibody from hybridomas (fast growing cell-cultures) developed from an inbred strain of mice. Researchers at the laboratory supervised by Baltimore were able to extract the DNA that characterized these antibodies. Baltimore's idea was to engineer a gene containing this DNA, insert that gene into mice that were known to be missing it and then observe their immune response. Scientists call such gene a transgene and such mice transgenic mice.

Apart from the fact that Frank Constantini, a biologist at Columbia University, produced the transgenic mice, all experiments were conducted in the laboratories supervised by Imanishi-Kari and Baltimore at the Massachusetts Institute of Technology (MIT). Baltimore created, in his laboratory, a colony of the transgenic mice obtained from Constantini. His postdoctoral fellow, David Weaver, carried out the molecular analysis of the antibodies. Baltimore wanted to know more about the antibodies circulating in the blood of the mice but he had neither the skills nor the

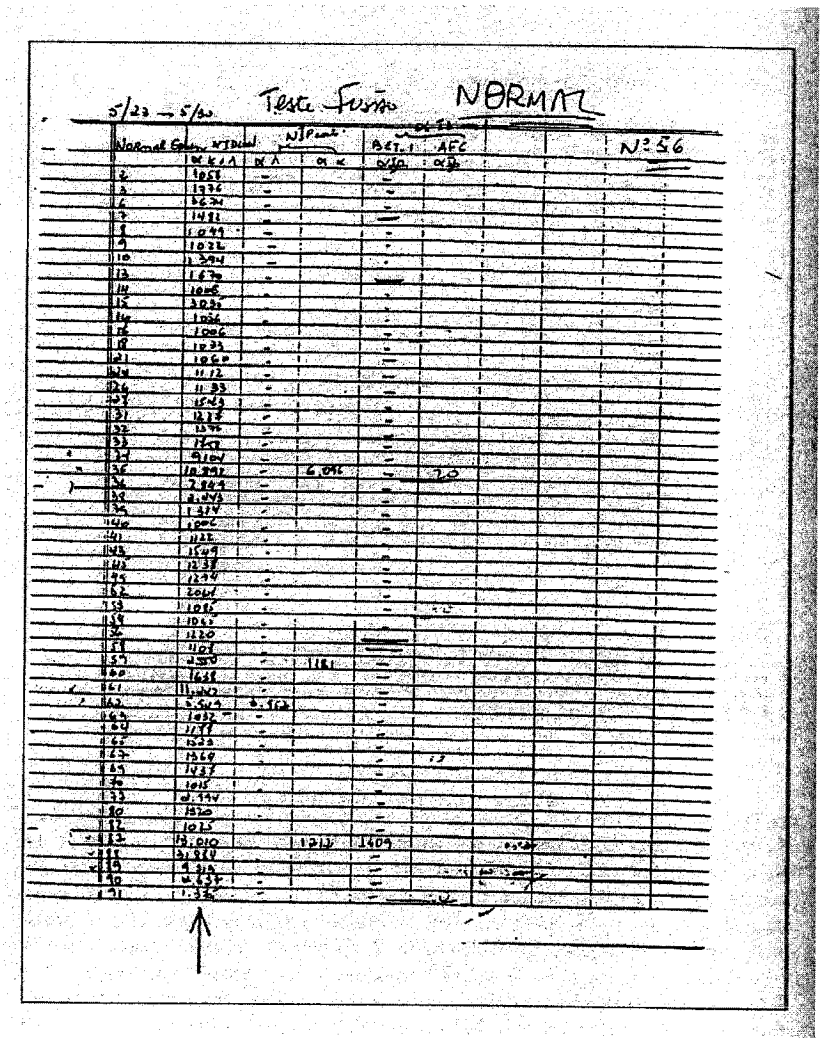
equipment for such analysis. Hence, he asked Imanishi-Kari to carry out complementary serological analysis of the antibodies.

Both independent analyses reached the same unexpected conclusion about the results of the experiment, namely, that antibody produced by the transgene appeared in the newly born mice. The central claim of the Cell paper was that this was due to "idiotypic mimicry." That is, the introduced gene did not cause the production of these foreign antibodies but rather it caused the immune system of the transgenic offspring to produce antibodies, which mimicked those foreign antibodies.

In the summer of 1985, Margot O'Toole was hired by Imanishi-Kari as a post-doctoral fellow to find out exactly how "idiotypic mimicry" worked. To understand better the experimental process, O'Toole attempted to repeat the experiments published in the Cell paper. She was unsuccessful at reproducing its results. Frustrated, O'Toole turned to Imanishi-Kari for explanation and requested to see the original laboratory notebooks.

In the spring of 1986, O'Toole found seventeen pages of lab notes that Imanishi-Kari's fellow Brazilian co-worker, Reis, and also co-author to the paper, had left behind. O'Toole found discrepancies between what was recorded and what was reported in the paper. For example, Figure 1-1 below shows a sample of one of these seventeen pages. This page recorded measures for antibody production of a normal mouse. The column, to which the arrow is pointing, indicates unusually high antibody production considering that this mouse was known to be missing the gene. The antibody count in mice with the foreign gene and in mice without it were similar. This implied that the presence of the transgene made no difference to the level of

antibodies produced by the experimental mice contrary to what was concluded in the paper.




*Reprinted by permission from W. W. Norton & Company, Inc. The Baltimore Case: A Trial of Politics, Science, and Character by Daniel J. Kevles. ©1998*

Figure 1–1: One of the 17 pages O'Toole found in Reis's notebook showing antibody production of a normal mouse.

Since Imanishi-Kari did not provide any of the original results to O'Toole, her discovery of these seventeen pages became one of the turning points in the history of the case. O'Toole began seriously doubting the veracity of the results published in the Cell paper. One event led to another and, eventually, in the spring of 1987, the National Institutes of Health (NIH) began its first official investigation into the case of Imanishi-Kari's scientific misconduct.

Ending early 1989, the first investigation of the NIH found that there were "significant errors of mis-statement and omission...[but] no evidence of fraud, conscious misrepresentation, or manipulation of data." (Anderson, 1991) A few months later, the NIH's newly formed Office of Scientific Integrity, which was later renamed to the Office of Research Integrity (ORI), reopened its investigation. As a result of this investigation, the ORI found Imanishi-Kari guilty of nineteen charges of scientific misconduct. The ORI used forensic analysis, statistical analysis, and scientific evidence to reach its conclusion.

All three tables of the Cell paper were doubted and believed to contain falsified or fabricated data. It was also believed that the results plotted in some of the figures in the paper did not pertain to true experiments. Examples of the paper's disputed tables and figures are given in Table 1-1 and Figure 1-2 below.

Table 2. Frequency of 17.2.25 Idiotype-Producing Hybridomas in Normal and Transgenic Mice

| Organ | 17.2.25 Idiotype-Positive | 17.2.25 Idiotype-Positive Plus: Anti-NP (κ) | Anti-NP (λ) | $\mu^a$ | $\mu^b$ |
|---|---|---|---|---|---|
| Normal Spleen | 1*/144 (<1%) | 1†/144 | 2/144 | 0/144 | 0/144 |
| Normal Lymph Nodes | 0/100 | 0/100 | 0/100 | 0/100 | 0/100 |
| Transgenic Spleen | 43/150 (28%) | 0/43 | 7‡/43 | 9§/43 | 1/43 |
| Transgenic Lymph Nodes | 129/190 (68%) | 6/129 | 12/129 | 33/129 | 10/129 |

B cell hybridomas were isolated from spleen and lymph nodes of transgenic and normal C57BL/6 mice. Secreted Ig from hybridomas was assayed for binding to the anti-17.2.25 idiotypic antibody or for NP-binding with either λ or κ light chains. The portion of the hybridomas that contained the μ isotype was assayed by the anti-allotype antibodies (see Figure 1) to associate the anti-17.2.25 idiotype binding with either the $\mu^a$ or $\mu^b$ allotype.

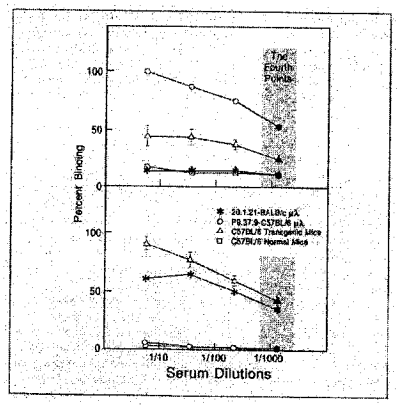
* Weak reactivity of hybridoma to anti-17.2.25.

† Not 17.2.25 idiotype.

‡ 2/7 were observed to have $\mu^a$ allotype.

§ 2/9 were found to be λ-bearing Ig, and 7/9 were found to be κ-bearing Ig.

Table 1–1: Table 2 of the Cell Paper (Weaver et al., 1986)




*Reprinted by permission from W. W. Norton & Company, Inc. The Baltimore Case: A Trial of Politics, Science, and Character by Daniel J. Kevles. ©1998*

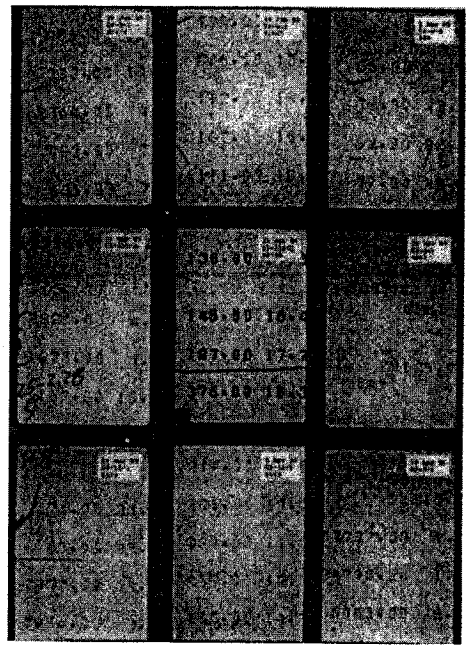
Figure 1–2: Figure 1 of the Cell paper.

We include these figures at this stage in order to aid the reader in better understanding the controversy that surrounded them. Table 1-1 illustrates the much disputed Table 2 of the Cell paper. The ORI believed that the data in this table did not exist at the time the paper was published. Rather, it believed that an entire data set had been falsified to provide data reported in this table after doubts had been raised about its validity. The next chapter examines in detail the ORI's statistical analyses of the evidence for the falsification of this data set. Figure 1-2 illustrates one of the seven figures in the Cell paper. The forth points on each of the curves, highlighted in the figure, were under dispute.

A standard practice for experiments like those performed in the Cell paper is to use a radiation counter to identify the presence or absence of antibodies. A reagent is used that reacts with a specific antibody. If the antibody is present, the reagent will react with it, emanating radiation. The intensity of that radiation is measured in counts per minute. The radiation counter prints out these counts per minute on so called counter tapes.

The Secret Service investigated Imanishi-Kari's experimental records. In particular, it obtained her counter tapes and compared them against other researchers' contemporaneous counter tapes. For instance, Figure 1-3 below exhibits a photograph of some of the tapes the investigators compared. The central tape, which stands out in clarity from the rest, belongs to Imanishi-Kari. The Secret Service concluded that Imanishi-Kari's tapes were more recent than they were actually dated to

be. This did not immediately suggest fraud but it provided evidence that Imanishi-Kari had not performed some of her experiments on the date that she had indicated.



*Reprinted by permission from Macmillan Publishers Ltd: NATURE, Volume 347, Issue 6293, page 507. Science meets forensic science by Christopher Anderson. ©1990*

Figure 1–3: A photograph of some of the radiation counter tapes, which the Secret Service investigated.

The ORI performed three statistical analyses using techniques of serial correlation, Poisson mixture models and distribution of digits. (Office of Research Intergrity Report, 1994) James Mosimann, an adjunct professor at the American University, was the ORI's statistical expert, and Austin M. Barron, an associate professor at the same university, was brought by the ORI from outside the NIH to analyze, in

part, the correctness of Mosimann's statistical arguments. (Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel, 1996; Kevles, 1998) After Imanishi-Kari appealed the ORI's decision, Terence Speed, a professor at the University of California at Berkeley, testified on her behalf for the statistical analyses performed by the ORI.

In the next three chapters we delve into what became known as "The Baltimore Case," by examining the statistical arguments put forward by the ORI, and the counterarguments to them.

# CHAPTER 2

## Serial Correlation

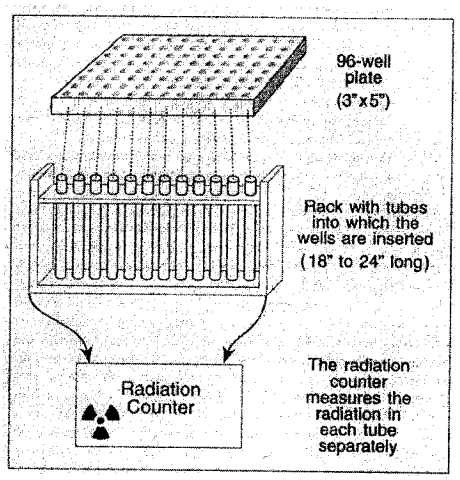
Table 2 of the Cell paper reported data for transgenic mice and normal mice. Much of the normal mouse data were taken from Imanishi-Kari's notebook. The rest of the data in the table were taken from Reis's notebook. Since a normal mouse in that experiment was later discovered to be actually transgenic, Imanishi-Kari confirmed during the investigation that she had used normal mouse data from another, earlier, experiment instead. The ORI believed that the entire data set, known as the January fusions [1] , was falsified by Imanishi-Kari in an attempt to produce data for Table 2 that otherwise did not exist.

The ORI analyzed data sets of radiation counts from comparable fusion experiments presented in three notebooks belonging to Imanishi-Kari, Reis and Weaver, respectively. The data from Imanishi-Kari consisted of a series of gamma radiation counter tapes. The purpose of the analysis carried out by the ORI was to determine if the radioactivity counts on these tapes were authentic records of a fusion experiment. In particular, the ORI analyzed disputed data from Imanishi-Kari's notebook

[1] **Fusion experiments (also referred to as fusions)** In this context, it refers to experiments measuring antibody production by cells, which are a fusion of a normal cell with a cancerous cell. A fusion of two cells refers to the union of two cell nuclei.

and also compared her data with data that were not in dispute from Reis's and Weaver's notebooks.

Antibody-growing cells from the transgenic mice were placed in well plates. Each of the wells was inserted into a tube in a rack, which was then placed in the radiation counter. Figure 2-1 below illustrates this process for a 96-well plate.




*Reprinted by permission from W. W. Norton & Company, Inc. The Baltimore Case: A Trial of Politics, Science, and Character by Daniel J. Kevles. ©1998*

Figure 2–1: Wells are put into tubes on a rack, which is then placed in the radiation counter. It measures the antibody production by cells in each well as radiation counts per minute (cpm).

The higher the number of antibodies produced in a well, the higher the radioactivity count for that well. Since the number of antibodies produced in any of the wells on a particular plate was assumed to be random, no pattern between high and low counts should have been observed. In other words, it was expected that if

the counts were authentic, then there would have been little or no serial correlation between high and low counts.

A sequence of autocorrelation coefficients was obtained by comparing each count with the count immediately following it, and with the count after that and so on up to the 30th subsequent count. Let $W_i$ be the radiation count for the $i^{th}$ well, $i = 1, 2, \ldots, n$. (Most of the plates used in the experiments were 96-well plates. Thus, for a single such plate, $n = 96$.) Further, if it is assumed that the process of counts $\{W_i\}$ is weakly stationary, that is

$$E(W_i) = \mu \text{ for } i = 1, 2, \ldots, n,$$

and

$$Cov(W_i, W_{i+j}) = \gamma_j \text{ for } i = 1, 2, \ldots, n \text{ and } j = 0, 1, 2, \ldots, \min(30, n - i),$$

then the correlation coefficient $\rho_{i,j}$ is only a function of the number of lags $j$, and is given by

$$\rho_{i,j} = \rho_j = \frac{Cov(W_i, W_{i+j})}{\sqrt{Var(W_i)Var(W_{i+j})}} = \frac{\gamma_j}{\gamma_0}. \tag{2.1}$$

The ORI report did not specify why the number of lags, $j$, was taken to be at most 30. However, it is important to note that, as Box et al. (1994) suggest is done in practice, in order to obtain a good estimate of the autocorrelation function, it is necessary that $n > 50$ and $j \leq n/4$.

Let $w_i$ be the radiation count *observed* for the $i^{th}$ well, $i = 1, 2, \ldots, n$. The authors of the ORI report refer to two estimators of $\gamma_j$. (Abraham and Ledolter,

1983; Kendall et al., 1983) The first estimator is given by

$$\hat{\gamma}_{j1} = \frac{1}{n}\sum_{i=1}^{n-j}(w_i - \bar{w})(w_{i+j} - \bar{w}) \tag{2.2}$$

and the second estimator by

$$\hat{\gamma}_{j2} = \frac{1}{n-j}\sum_{i=1}^{n-j}(w_i - \bar{w})(w_{i+j} - \bar{w}) \tag{2.3}$$

where $\bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$ and $j = 0, 1, 2, \ldots, min(30, n-i)$.

Kendall et al. (1983) assert that (2.2) is preferred over (2.3), because it has smaller mean square error in most applications. The estimator used by the ORI is not made clear in the report but its authors appear to have used (2.3) in their analysis. (We explore this issue further in Section 2.4.) Hence, their estimator of the correlation coefficient, $\rho_j$, as a function of the number of lags $j$ is given by

$$\hat{\rho}_j = \frac{\hat{\gamma}_{j2}}{\hat{\gamma}_{02}} \tag{2.4}$$

where $j = 0, 1, 2, \ldots, min(30, n-i)$.

In the next section, we describe the data that were examined in these serial correlation analyses.

## 2.1 Data Under Scrutiny

The disputed Table 2 of the Cell paper reported data from the experimental records of Imanishi-Kari and Reis. The ORI did not question the veracity of the data from Weaver and Reis, while the data from Imanishi-Kari's notebooks were the subject of close scrutiny. In fact, the ORI used Weaver's and Reis's data as the control data sets in their analyses.

Moreover, the ORI carried out computer simulations aiming to illustrate how various patterns in data can produce certain trends in the serial correlations. Consider, for example, environmental data, where serial correlation is very common. Suppose we are to compute the serial correlations of the average monthly temperatures for, say, a 10 year period starting January of Year 1 and ending December of Year 10. Since we have 12 observations per year over a 10 year period, we have a sample size of 120 observations. As discussed in the previous section, for a proper estimation of the serial correlations we should restrict the maximum number of lags, $j$, to at most 120/4 or 30.

Now, from one day to the next, high temperatures in summer tend to stay high and, similarly, low temperatures in winter tend to stay low. This behavior is very similar from year to year. Therefore, the average temperature in January of Year 1 is likely to be positively correlated with the average temperature in January of the subsequent years but negatively correlated with the average temperature in July of Year 1 as well as in July of subsequent years. Therefore, theoretically, we should have that $\rho_{12} = 1$, whereas $\rho_6 = -1$. Thus, if we are to plot the serial correlations estimated from our 120 average monthly temperatures against the number of lags, which we restrict to 30, we would expect to see a cyclical function and, in particular, a function with a cycle of length 12.

The first three sets of graphs displayed in Figure 2-2 below are examples of the computer simulations carried out by the ORI, while the forth set of graphs illustrates the serial correlations computed from Imanishi-Kari's data. Although it is not explicitly specified in the ORI report, it appears that the data sets for the simulations

exhibited in Figure 2-2 were generated reflecting the pattern the ORI claimed to have observed in Imanishi-Kari's original data. That is, the ORI's analysts observed a cyclical pattern of length 12 in the serial correlations computed from her data and wanted to show that inducing a specific pattern in a randomly or deterministically generated data can cause this type of pattern in the serial correlation function.

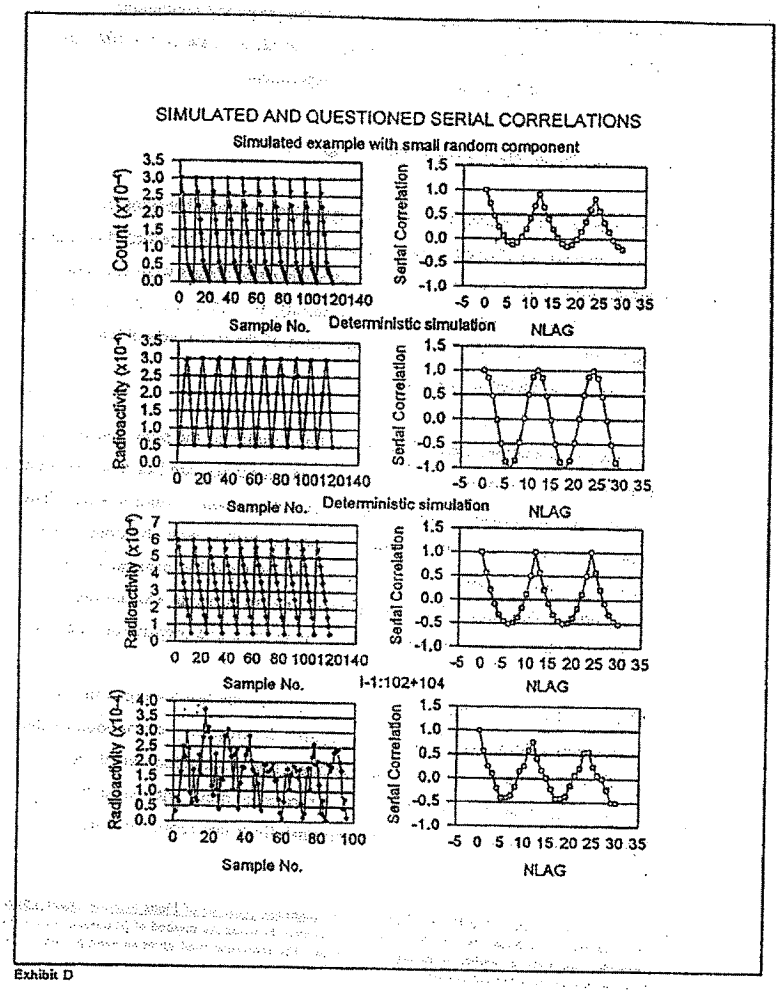



Figure 2–2: Computer simulations showing how certain patterns in data can cause strong cyclical patterns in the observed serial correlations. (Office of Research Intergrity Report, 1994)

In Figure 2-3 below we reproduce the third simulation, which was deterministic, from Figure 2-2.

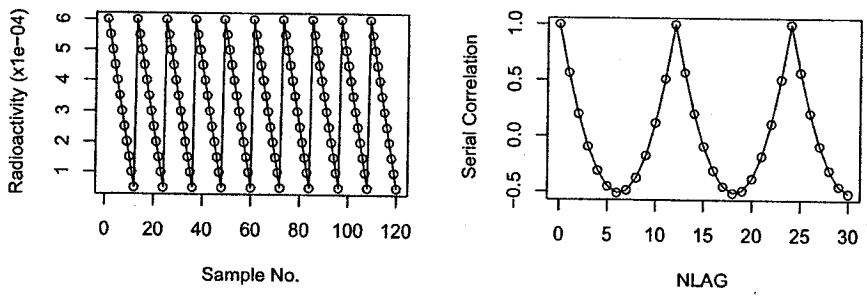



Figure 2–3: One of the deterministic simulations in Figure 2-2 is reproduced here. NLAG refers to the number of lag variables ($j$ in our case).

Reading off the values from the left plot in Figure 2-3, we have 120 samples of radiation count, $w_1, w_2, \ldots, w_{120}$, where

$$\begin{aligned} w_1 &= w_{13} = \ldots = w_{109} = 6 \times 10^{-4} \\ w_2 &= w_{14} = \ldots = w_{110} = 5.5 \times 10^{-4} \\ &\vdots \\ w_{12} &= w_{24} = \ldots = w_{120} = 0.5 \times 10^{-4}. \end{aligned}$$

Then, using (2.4) we obtain a sequence of 30 correlation coefficients, $\hat{\rho}_j$ for $j = 1, \ldots, 30$, which are plotted in the right graph in Figure 2-3.

In its analysis, the ORI assumed that the observations taken from neighboring wells were independent. Now, if the $n$ radiation counts $W_1, \ldots, W_n$ were independent

and identically distributed, then we should have that $\rho_0 = 1$ and $\rho_j = 0$ for all $j = 1, 2, \ldots, 30$. In this case, if $n$ is large, the estimated autocorrelations, $\hat{\rho}_j$ for $j = 0, 1, 2, \ldots, 30$ from the observed radiation counts $w_1, \ldots, w_n$ will be roughly independently and normally distributed with mean 0 and variance $1/n$. (Box et al., 1994)

We next recount the evidence for fraud that the ORI accumulated based on these statistical analyses.

## 2.2 Evidence of Scientific Fraud

Essentially, the ORI carried out two types of comparisons. First, it compared the serial correlations estimated from data on different pages from Imanishi-Kari's notebook only. Second, it compared the serial correlations estimated from Imanishi-Kari's data to the serial correlations estimated from unquestioned data recorded in the notebooks of Reis and Weaver.

In the first type of comparison, the ORI compared the serial correlations estimated from data on pages 102, 103 and 104 in Imanishi-Kari's notebook. The ORI's main observation was that the serial correlations estimated from data on pages 102 and 104 exhibited a strong cyclical pattern with a cycle of length 12 but the relationship differed, it became weaker, when data from page 103 was included in the estimation of the serial correlations. Figure 2-4 below exhibits the graphs of these serial correlation functions.

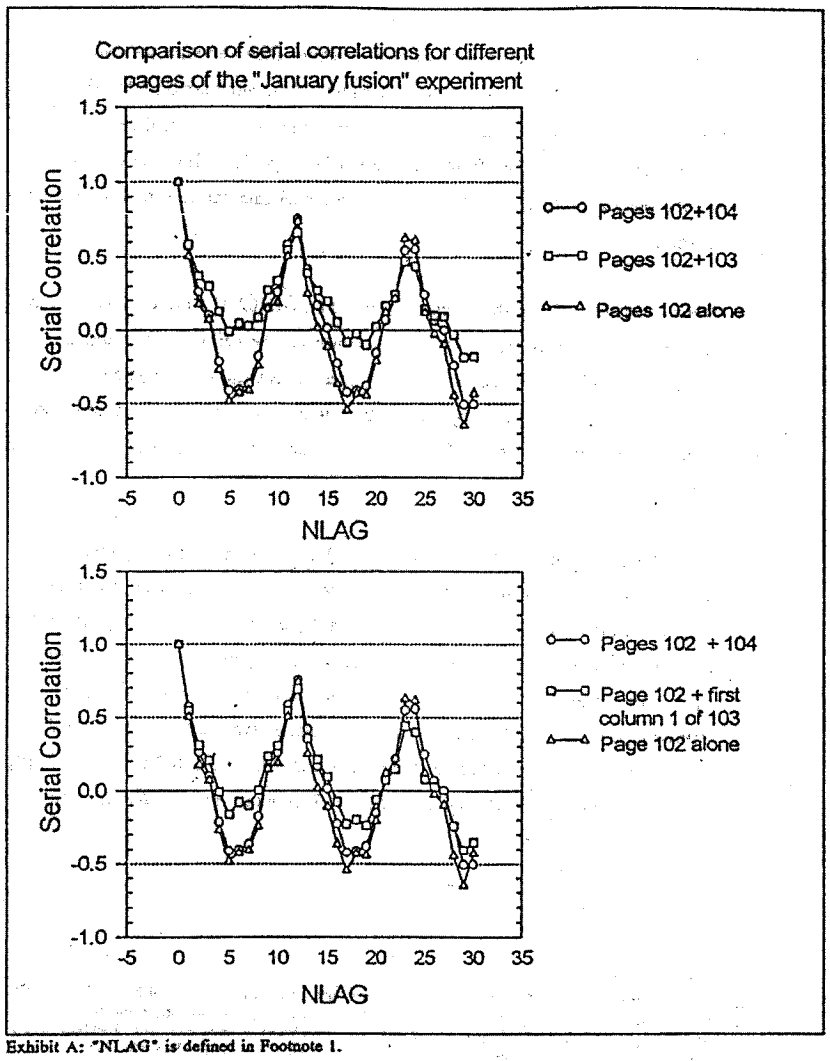



Exhibit A: "NLAG" is defined in Footnote 1.

Figure 2–4: Comparisons of serial correlations estimated from Imanishi-Kari's data alone. (Office of Research Intergrity Report, 1994)

According to the report, there existed independent forensic evidence, which suggested that data on page 103 could not have been part of the continuous experimental record reported on pages 102 and 104. The forensic analysts claimed that the counts were recorded before Imanishi-Kari had obtained the mice! According to the ORI, the statistical analysis only gave more support to this finding. With the serial correlations exhibiting a different pattern depending on whether or not data from page 103 was included, the ORI concluded that data on page 103 had been inserted into the data reported on pages 102 through 104 and, hence, that "the purported continuous experimental record was actually discontinuous." (Office of Research Intergrity Report, 1994)

In the second type of comparison, the ORI found that the serial correlations estimated from Imanishi-Kari's data revealed a strong cyclical pattern. In contrast, when the ORI compared the estimated serial correlations obtained from comparable fusion experiments by Reis and Weaver, it observed no apparent cyclical pattern. Figures 2-5 and 2-6 below illustrate the serial correlations estimated from data from Imanishi-Kari and from Weaver, respectively.

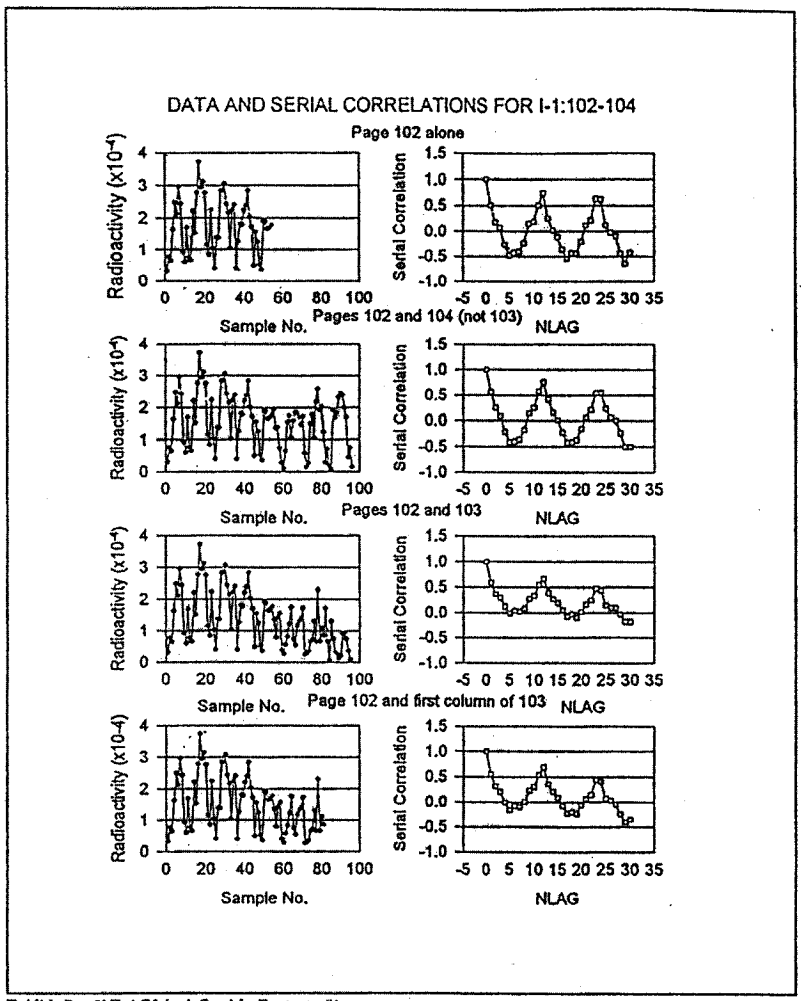



Exhibit B - "NLAG" is defined in Footnote #1.

Figure 2–5: Serial correlations estimated from data from Imanishi-Kari's notebook. (Office of Research Intergrity Report, 1994)

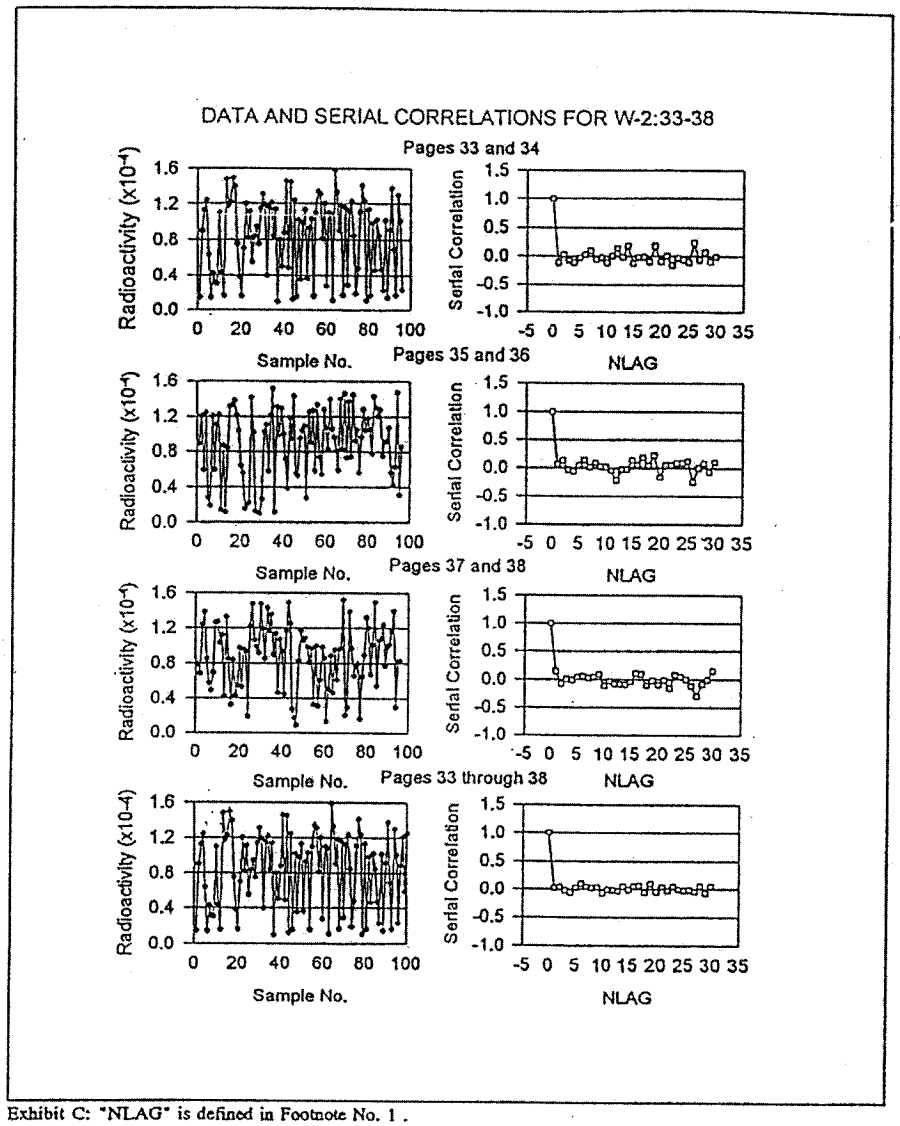



Exhibit C: "NLAG" is defined in Footnote No. 1 .

Figure 2–6: Serial correlations estimated from data from Weaver's notebook. (Office of Research Intergrity Report, 1994)

The ORI claimed that this cyclical pattern was "profoundly different than the pattern seen in the unquestioned data, and appear[ed] to be highly non-random." (Office of Research Intergrity Report, 1994) As the ORI frazed it, Imanishi-Kari's data appeared only "superficially similar" to Weaver's data. The ORI considered this as an additional evidence that the data set did not represent an authentic experimental record.

It is important to point out that, presumably, the ORI combined radiation counts on several pages to increase the sample size. As was mentioned earlier, for a reliable estimation of the correlation function, practice suggests that it is necessary to have the sample size, $n$, greater than 50. Also mentioned earlier was the fact that, if the observations are independent and identically distributed, we would expect the estimated serial correlations to be roughly independently and normally distributed with mean 0 and variance $1/n$. Although no formal justification is presented, the graphs in Figures 2-5 and 2-6 show that Weaver's data are consistent with the null hypothesis of independence of the radiation counts but Imanishi-Kari's data are not.

Since the estimated serial correlations were not consistent with those arising from independent counts, the ORI doubted their authenticity. While examining these well counts, the ORI investigators came across what they believed to be an anomalous sequence of positive wells.

The cells, once pipetted into the wells, can either grow or not grow. Whether the cells grow or not is considered to be a completely random process. Wells with cell growth are referred to as positive wells. The investigators found that of 260 wells, 137 were positive. The last 15 wells, that is, 246-260, were all positive. Mosimann

calculated that the "probability of ending with 15 consecutive positive wells [was] less than 0.0001." (Office of Research Intergrity Report, 1994) Explicitly, given that the data are not fraudulent,

$$Pr(\text{last 15 of 260 all positive}) = \frac{2^{245}}{2^{260}} = \frac{1}{2^{15}} = 0.00003 < 0.0001.$$

Hence, the ORI believed that the low probability of the event added more weight to the case of fraud.

In the next section, we examine the objections raised to these analyses by the Appeals Board based on the testimony by Speed.

### 2.3 The Counterarguments

Speed countered that there were three major flaws with the analysis above. He said that the right question had not been asked, he doubted the underlying assumptions and he questioned the appropriateness of the control data sets.

Speed argued that, when calculating the probability of 15 consecutive positive wells, Mosimann asked the wrong question. Using the analogy of obtaining 20 consecutive heads in 100 tosses of a coin, Speed explained that Mosimann had found the probability of the event of 15 consecutive positive wells on a particular plate when, in fact, he should have found the probability of the event on *any* of Imanish-Kari's many plates. (Kevles, 1998)

Next, the ORI assumed that for all such fusion experiments, the level of antibody in one well is not influenced by the level of antibody obtained in the surrounding wells. (Office of Research Intergrity Report, 1994) Speed pointed out that although

this may be the case in an ideal situation, in an experiment this assumption may not hold.

Finally, Speed argued that Weaver's data set was not comparable to Imanishi-Kari's and, hence, was not a valid control. (Kevles, 1998) In fact, the Appeals Board concluded that there were "problems with relying on the "controls."" (Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel, 1996) This is discussed in more detail in the next chapter on Poisson mixture models.

## 2.4 A Third (My) Opinion

In addition to the objections raised by Speed, there are various other issues with these analyses that need to be addressed.

**Covariance Estimators.** As was mentioned earlier, there seems to be some ambiguity whether (2.2) or (2.3) was used as an estimator of $\gamma_j$ when calculating the serial correlations. Figure 2-7 below compares, graphically, the two estimators.

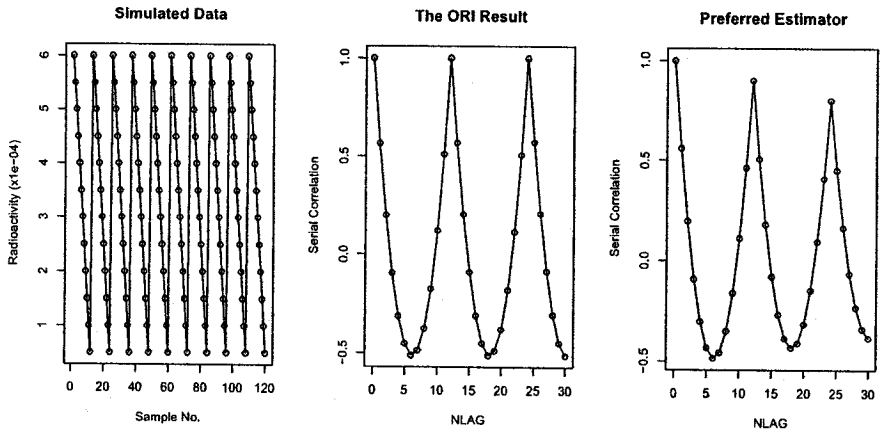



Figure 2–7: Comparing the serial correlations estimated using (2.3) (the ORI result) and (2.2) (the preferred estimator).

Since we do not have the original data sets, we cannot investigate if using (2.2) instead of (2.3) when estimating the serial correlations from Imanishi-Kari's data would have lead to different conclusions by the ORI.

**Post Hoc Analyses.** Another possibly serious flaw in the analyses performed by the ORI is the fact that many of them were carried out post hoc.

For instance, the data generated for the computer simulations followed a cyclical pattern with a cycle of length 12. The choice for the length of the cycle seems to have been influenced by the fact that the ORI observed a similar pattern in the serial correlations estimated from Imanishi-Kari's data. In other words, the choice of cycle period was a post hoc choice.

As another example, Mosimann calculated the probability that the last 15 wells would all be positive because he had observed this outcome in Imanishi-Kari's

records. However, one could argue that all such probabilities should be calculated *conditioning* on the fact that the event has already taken place. This changes the sample space and, therefore, the probability. Hence, Mosimann's calculation could be rendered invalid.

Letting

$$A = \{\text{last 15 consecutive positives in 260 wells}\} \tag{2.5}$$

and

$$B = \{\text{observed event A}\},$$

then, trivially, the conditional probability is given by

$$Pr(A|B) = 1,$$

in contrast to the unconditional probability of 0.00003 calculated by Mosimann.

**Incorporating Available Information in the Probability Calculation.** It is worth pointing out that in the sample Mosimann was considering for the probability calculation, more than half of the wells were positive. Define the event

$$C = \{\text{observed 137 positives in 260 wells}\}$$

and let the event $A$ be defined as in (2.5).

Pipetted cell cultures in a well may or may not grow. De Blas et al. (1981) estimate the probability of growth, $p$, for monoclonal (single clone) cell cultures to

be

$$\hat{p} = \frac{\text{number of wells with growth}}{\text{total number of wells}}.$$

Any particular well may contain one or more clones of cells. The more cells there are, the likelier it is that they will grow. (Newkirk, 2006) This suggests that the probability of cell growth is not identical across the wells.

Imanishi-Kari claimed that most wells contained a single clone and the few that did not, contained two or at most three clones. (Kevles, 1998) However, since the number of cells a well would contain was completely determined by chance and not induced in any systematic way by Imanishi-Kari, the probability of cell growth remains constant across the wells. This explains why, as Kevles (1998) points out, it can be expected that the wells with cell growth are randomly scattered across the plate. Therefore, for a single plate of wells, the assumption that all outcomes are equally likely is valid.

Another important observations to make is that the size of the sample in question implies that the counts came from more than one plate of wells. However, since these counts came from the same set of cell fusions simply spanning several well plates, the assumption of equiprobable outcomes is not affected.

Now, we have a situation where all outcomes have equal probability and in any one well we can observe a 0 (no growth) or a 1 (growth). We can, therefore, use the Fermi-Dirac method to establish the total number of outcomes with a specified number of positive wells. (Feller, 1960) In particular, if we let r be the number of

positive wells and n be the total number of wells, then by the Fermi-Dirac method we have a total of $\binom{n}{r}$ possible outcomes.

Hence, we have that

$$Pr(A|C) = \frac{Pr(A \bigcap C)}{Pr(C)} = \frac{Pr(A)}{Pr(C)} = \frac{\frac{2^{245}}{2^{260}}}{\frac{\binom{260}{137}}{2^{260}}} = \frac{2^{245}}{\binom{260}{137}} \simeq 0.0009.$$

As expected, the $Pr(A|C)$ is larger than the $Pr(A)$. Therefore, it would have been more correct to take into account the fact that 137 of the 260 wells were all positive, albeit this difference seems to be of little practical importance.

**Analysis of the Probability of Fraud.** Mosimann's probability calculation could be analyzed from another point of view. Define the event

$$F = \{\text{falsified or fabricated data}\}$$

and let the event $A$ be as defined in (2.5).

In a recent survey of scientists, Martinson et al. (2005) found that about 0.5% of researchers in their early careers and about 0.2% researchers in their mid-careers falsify research data. Although, the survey period does not coincide with the time when Imanishi-Kari carried her research, we assume that $Pr(F)$ has not changed over time and we use this assumption to carry a simple analysis of the $Pr(F|A)$ for illustrative purposes only.

Now, we found earlier that $Pr(A|F^c) = 0.00003$. The ORI's claim was that the low value of $Pr(A|F^c)$ added weight to the case of fraud. Essentially, the ORI reasoned, under the "null hypothesis" (of no fraud) the observed event, $A$, has very low probability and, therefore, is unlikely to have arisen if no fraud had taken place.

Using the results of the study carried by Martinson et al. (2005), if we let $Pr(F)$ lie between 0.001 and 0.005, then we obtain the following graph for various values of $Pr(A|F)$.

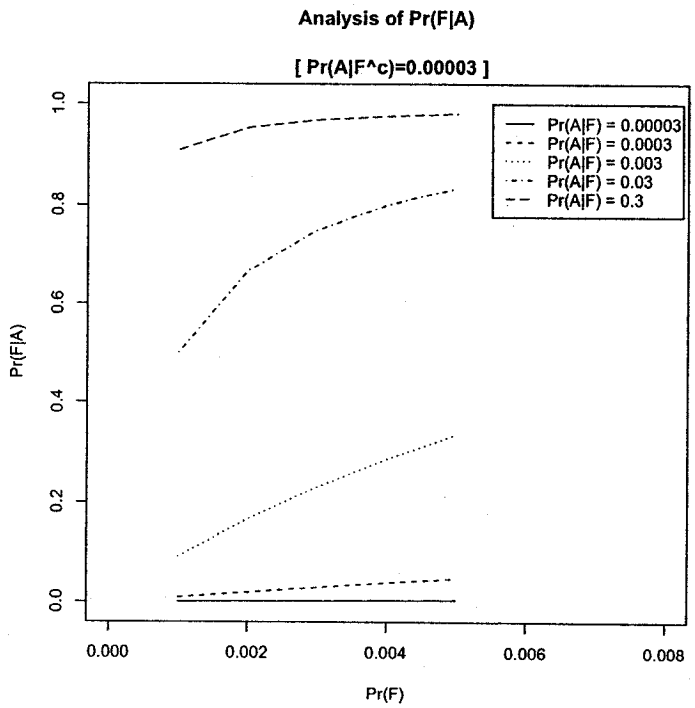



Figure 2–8: Analyzing the behavior of $Pr(F|A)$ over different $Pr(F)$ when $Pr(A|F^c)$ is kept fixed at 0.00003 and $Pr(A|F)$ is allowed to vary from 0.00003 to 0.3.

The solid line in Figure 2-8 illustrates that when $Pr(A|F) = Pr(A|F^c)$, then $Pr(F|A) = Pr(F)$ as is expected since then the events $A$ and $F$ would be independent. The remaining lines in the figure above simply show that no matter what $Pr(A|F)$ is as long as it is larger than $Pr(A|F^c)$, $Pr(F|A)$ greatly exceeds $Pr(F)$, that is, $Pr(F|A) \gg Pr(F)$.

Further, consider the behavior of the odds

$$\frac{Pr(F|A)}{1 - Pr(F|A)},$$

as $Pr(A|F)$ increases, illustrated in the figure below.

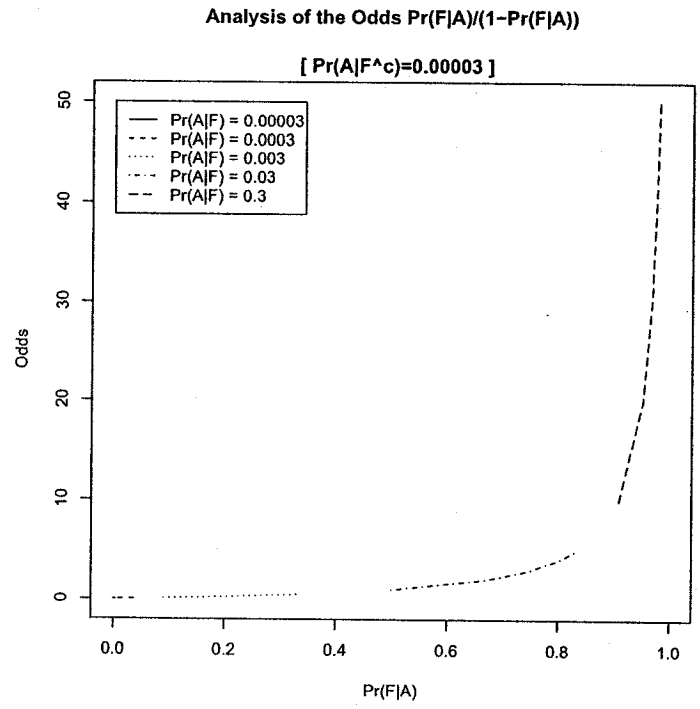



Figure 2–9: Analyzing the behavior of the odds $\frac{Pr(F|A)}{1-Pr(F|A)}$ as $Pr(A|F)$ increases.

We see from Figure 2-9 that as long as $Pr(A|F) \leq 0.3$, the odds of fraud having taken place given that $A$ has been observed are less than 10. We believe that when a reputation of a scientist is at stake, such odds do not represent a sufficiently strong supporting evidence for fraud. Therefore, as this analysis simply aims to show, the

ORI's approach at evaluating such evidence was rather flawed.

In this chapter, we presented the serial correlation analyses carried out by the ORI. First, we defined the estimator, which the ORI seemed to have had used when estimating the serial correlations. Next, we described the data, from which the serial correlations were estimated, as well as the computer simulations the ORI used as supporting evidence for the case of fraud. We then recounted the statistical evidence, which the ORI put forward against Imanishi-Kari, based on these serial correlation analyses. Finally, we outlined several possible flaws with these analyses. We present in the next chapter, in a similar manner, the second set of analyses carried out by the ORI, namely, the Poisson mixture model analyses.

# CHAPTER 3

## Poisson Mixture Models

Mixture distributions have been increasingly studied over the last few decades. With improved computational power they form an important class of statistical models. In its second set of statistical analyses, the ORI used a mixture of Poisson distributions to examine the veracity of the low count data. In this chapter, we present the details of these analyses.

At the time when the ORI published its report there appears to have been only one application in the literature of a mixture of Poisson distributions. (Titterington et al., 1985) The example modeled the number of death notices of women 80 years of age and older, reported in the *Times* newspaper for each day over the 3-year period from 1910 to 1912. Modeling the data with a standard Poisson distribution gave a very poor fit. A mixture of two Poisson distributions gave a much better fit. One possible explanation was that death rates in winter could be expected to be higher than in summer. Clearly, the mixture model, by modeling two different mean death rates, allowed for such heterogeneity in the sample.

A recent search in the literature revealed more examples on the application of Poisson mixture models, including those of order greater than two. The fields of

computer science, marketing and even chemistry have benefitted from the development of these models. (Church and Gale, 1995; Willse and Tyler, 2002; Brijs et al., 2004)

We, therefore, begin by giving some basic notions of Poisson mixture models.

## 3.1 Basics of Poisson Mixture Models

Poisson mixture models can be finite or infinite. We will describe first the general Poisson mixture model and then show how two special cases arise from it, namely, the Negative Binomial model (an infinite mixture of Poisson distributions) and the finite Poisson mixture model. Lastly, we will provide the formal definition of a finite mixture model. In the next section, we will address the ORI's justifications for the use of the Poisson mixture model.

Recall the notation we introduced in Chapter 2, where we let $W_1, \ldots, W_n$ be $n$ random well counts and $w_1, \ldots, w_n$ be the corresponding *observed* well counts. Without loss of generality, we drop the subscripts and we define $W$ as the recorded counts coming from a single well. Let $f(w)$ be the probability mass function of $W$.

If we model $f(w)$ as an infinite mixture of Poisson distributions, it will take the form of

$$f(w) = \int_0^\infty f(w|\lambda) dS(\lambda) \tag{3.1}$$

where $f(w|\lambda)$ is the Poisson probability mass function with parameter $\lambda$, the observed value of the "random" Poisson rate, $\Lambda$, and $S$ is *any* cumulative probability distribution function of $\Lambda$.

Therefore, it is easy to see how differently defined cumulative probability distribution functions, $S(\lambda)$, would give rise to different mixtures of Poisson distributions. This is the motivation behind a special case of the Poisson mixture model, the Negative Binomial model, that the ORI also studied.

The Negative Binomial Model

If $S(\lambda)$ is any absolutely continuous density function, then (3.1) can be written as

$$f(w) = \int_0^\infty f(w|\lambda)dS(\lambda) = \int_0^\infty f(w|\lambda)S'(\lambda)d\lambda. \tag{3.2}$$

In particular, if we let $S'(\lambda)$ be the Gamma probability density function with parameters $\alpha = r$ and $\beta = (1-p)/p$, it can be easily verified that $f(w)$, in (3.2), is the Negative Binomial probability mass function with parameters $r$ and $p$.

Mostly, however, the ORI relied on the finite mixture model, which also arises as a special case of the infinite mixture, given by (3.1).

The Finite Mixture of Poisson Distributions

If the mixture is finite, the number of components, $g$, is fixed. Further, the infinite set of $\lambda$'s in (3.1) is actually a finite set consisting of $\lambda_1, \ldots, \lambda_g$. And, the corresponding component weights become $\pi_1, \ldots, \pi_g$, that is, also a finite set.

Let $\lambda_1^*, \ldots, \lambda_g^*$ be a permutation of $\lambda_1, \ldots, \lambda_g$ such that $\lambda_1^* < \ldots < \lambda_g^*$ and let $\pi_1^*, \ldots, \pi_g^*$ be the corresponding permutation of $\pi_1, \ldots, \pi_g$. We can define the probability mass function of $\Lambda$, the random Poisson rate, as

$$s(\lambda) = \begin{cases} \pi_i^* & \text{if } \Lambda = \lambda_i^*,\ i = 1, \ldots, g; \\ 0 & \text{otherwise.} \end{cases}$$

Let $\lambda_0^* = 0$. Then, it follows that the cumulative probability distribution function of $\Lambda$ is given by

$$S(\lambda) = \begin{cases} 0 & \text{if } \infty < \Lambda \leq \lambda_0^*; \\ \sum_{i=1}^{k} \pi_i^* & \text{if } \lambda_{k-1}^* < \Lambda \leq \lambda_k^*,\ k = 1, \ldots, g; \\ 1 & \text{if } \Lambda > \lambda_g^*. \end{cases}$$

Therefore, (3.1) can be written as a Riemann-Stieltjes integral,

$$f(w) = \int_0^\infty f(w|\lambda) dS(\lambda) \tag{3.3}$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} f(w|\lambda_j) \Delta S(\lambda_j) \tag{3.4}$$

$$= \sum_{i=1}^{g} f(w|\lambda_i^*) s(\lambda_i^*), \tag{3.5}$$

that is, as a finite mixture of Poisson probability mass functions. To visualize this better consider the graph below.

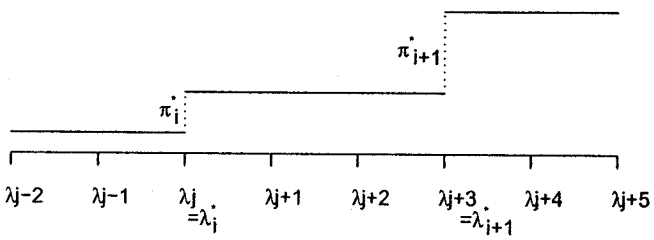



Figure 3–1: A jump of height $\pi_i^*$ occurs whenever $\lambda_j = \lambda_i^*$.

As Figure 3-1 illustrates, in (3.3) we are essentially integrating over a step function with jumps of height $\pi_i^*$ at $\lambda_i^*$ for $i = 1, \ldots, g$.

We now present the formal definition of a finite mixture model given by McLauchlan and Peel (2000).

**Definition 3.1** Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ denote a random sample of size $n$, where $\mathbf{Y}_j$ is a $p$-dimensional vector with probability density function $f(\mathbf{y}_j)$ on $\mathcal{R}^p$. Let $\mathbf{y}_j$ denote the observed value of the random vector $\mathbf{Y}_j$. Then, the probability density function (or the probability mass function (pmf) in the case of discrete random sample) $f(\mathbf{y}_j)$ of $\mathbf{Y}_j$ is said to be a finite component mixture density if,

$$f(\mathbf{y}_j) = \sum_{i=1}^{g} \pi_i f(\mathbf{y}_j|\boldsymbol{\theta}_i) \tag{3.6}$$

where

- the number of components, $g$, is fixed;
- the quantities $\pi_1, \pi_2, \ldots, \pi_g$ are the mixing proportions or weights, with $0 \leq \pi_i \leq 1$ for $i = 1, 2, \ldots, g$ and $\sum_{i=1}^{g} \pi_i = 1$;
- the densities $f(\mathbf{y}_j|\boldsymbol{\theta}_1), f(\mathbf{y}_j|\boldsymbol{\theta}_2), \ldots, f(\mathbf{y}_j|\boldsymbol{\theta}_g)$ are the component densities of the mixture parameterized by $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_g$, respectively.

Any convex linear combination of densities gives a density and, hence, $f(\mathbf{y}_j)$ is a density. Its corresponding distribution function, $F(\mathbf{y}_j)$, is called the $g$-component finite mixture distribution function.

In particular, if we consider a univariate model, where $f(y_j|\theta_i) = f(y_j|\lambda_i)$, $i = 1, 2, \ldots, g$, are Poisson probability mass functions, then $f(y_j)$ is said to be a "$g$-component mixture of Poisson mass functions."

We next examine how the Poisson mixture models came to be used in the accusations against Imanishi-Kari.

### 3.2 Why the ORI Used Poisson Mixture Models

The ORI claimed that the count data from uncontested experiments of the type carried out by Imanishi-Kari should roughly follow a mixture of Poisson distributions. We shall see later in this chapter, however, that this assumption is difficult to justify. We proceed to examine the details of the ORI's arguments in support of its claim.

We recall that the amount of radiation emanating from each well is measured by the gamma counter in counts per minute, which we denoted by $W$. Therefore, define $W(t)$ as the recorded counts coming from a single well over time $t$. That is, $\{W(t), t \geq 0\}$ is a random function of time and, by definition, a stochastic process. It has been experimentally verified that radiation emitted from a radioactive source is

adequately described by a homogeneous Poisson process. In particular, the counting process $\{W(t), t \geq 0\}$, that arises from counts recorded by a gamma counter, fit a Poisson process very well.

A formal definition of a Poisson process follows.

**Definition 3.2** A stochastic process $\{N(t), t \geq 0\}$ is a Poisson process with mean rate $\nu$, if the following two assumptions are satisfied:

1. $\{N(t), t \geq 0\}$ has stationary independent increments;
2. the number of events in the time interval $[t, t+\Delta t]$ follow a Poisson distribution with parameter $\nu\Delta t$, that is,

$$Pr(N(t+\Delta t) - N(t) = k) = \frac{\exp(-\lambda\Delta t)(\lambda\Delta t)^k}{k!} \tag{3.7}$$

and

$$E(N(t+\Delta t) - N(t)) = Var(N(t+\Delta t) - N(t)) = \nu\Delta t. \tag{3.8}$$

Next, we present an alternative definition of a Poisson process that is more often adopted in practice as justification for the use of a Poisson process. (Parzen, 1965)

**Definition 3.3** Let $\nu$ be any positive constant and suppose that a counting process, $\{N(t), t \geq 0\}$, satisfies the following axioms:

*Axiom 0.* Define $N(0) = 0$.

*Axiom 1.* The process $\{N(t), t \geq 0\}$ has stationary independent increments.

*Axiom 2.* $Pr(N(t+\Delta t) - N(t) = 0) = 1 - \nu\Delta t + o(\Delta t)$

*Axiom 3.* $Pr(N(t+\Delta t) - N(t) = 1) = \nu\Delta t + o(\Delta t)$

*Axiom 4.* $Pr(N(t+\Delta t) - N(t) \geq 2) = o(\Delta t)$

Then, $\{N(t), t \geq 0\}$ is a Poisson process as defined by Definition 3.2.

Axiom 4 shows that the events of a Poisson counting process, $\{N(t), t \geq 0\}$, are rare, being highly unlikely that we will observe more than one event over a small interval $[t, t + \Delta t]$. This is the underlying reason for the claim by Willse and Tyler (2002) that the Poisson distribution represents low count data more accurately than other models, and is, presumably, also the reason why the ORI included only counts less than 600 cpm in their Poisson mixture analyses.

All counts in both the questioned and the unquestioned experimental records were counts *per minute*. Therefore, following Definition 3.2, we have,

$$W = W(t+1) - W(t).$$

Thus, if we denote the mean rate of the radiation emanating from a well by $\lambda$, from Definition 3.2 we see that the pmf $f(w)$ of $W$ is given by

$$f(w) = \frac{\exp(-\lambda)(\lambda)^w}{w!}. \tag{3.9}$$

Now, if $W$ were to have a Poisson distribution, then $E(W) = Var(W)$. However, this assumption was most likely violated for the data that arose from all examined experiments, both contested and uncontested, that the ORI analyzed. These data tended to be overdispersed, with $Var(W) \gg E(W)$. To account for this excessive variation in these data, the ORI decided to fit them using a mixture of Poisson distributions. There are at least two different reasons (other than overdispersion, which is an empirical reason) for considering a Poisson mixture model.

Provided the data were authentic, one of the ORI's experts, other than Mosimann, claimed that since different assays had been carried out and different reagents used, a mixture model would be appropriate. That is, different reagents may have given rise to different distributions for the well counts. That is, a particular well count, $W$, could have arisen with a certain probability from one of a number of different Poisson distributions.

On the other hand, Mosimann claimed that a mixture model was appropriate to model authentic count data because the wells had different "cooking" times meaning that the amount of radiation emanating from the wells during a certain time interval varied for each well. As this amount was measured by the gamma counter in counts per minute, the counts recorded for the different wells would not be expected to follow the same Poisson distribution.

Consequently, the ORI's justifications for the use of the mixture model in this setting were not consistent. We will return to the problems with its justification in section 3.6. In summary, the ORI used the Poisson mixture model,

$$f(w) = \sum_{i=1}^{g} \pi_i f(w|\lambda_i) = \sum_{i=1}^{g} \pi_i \frac{\exp(-\lambda_i)\lambda_i^w}{w!}, \tag{3.10}$$

with mixing proportions, $\pi_i$, $i = 1, \ldots, g$, estimated using the method of maximum likelihood. The component intensities, $\lambda_i$, $i = 1, \ldots, g$, were also estimated using the same approach. Furthermore, the number of components, $g$, was unknown and calculated from the data as well (see section 3.4 for details). Due to computer limitations, the ORI was forced to restrict $g \leq 9$.

Next, we present the spikiness index, which Mosimann developed as an additional measure of assessing whether the data were consistent with a mixture of Poisson distributions.

### 3.3 The Spikiness Index

Mosimann reasoned that excessively "spiky" data were not likely to follow a Poisson mixture model. Therefore, he believed that a high spikiness index provided further evidence against the null hypothesis that the data fit the model.

**Definition 3.4** Let $n_i$ denote the number of observations in cell $i$, $i = 1, \ldots, m$, of a frequency histogram of a sample of $N$ observations. Let $\|n_{i-1} - n_i\|$ be the absolute value of the difference between heights of the adjacent cells $i-1$ and $i$. Then, the spikiness index, $SI$, is defined as

$$SI = \sum_{i=1}^{m+1} \|n_{i-1} - n_i\| \tag{3.11}$$

where $n_0$ and $n_{m+1}$ are the number of observations before the first and after the last positive cell frequency, respectively, and are taken to be $n_0 = n_{m+1} = 0$.

The spikiness index is always an even number and it is always less than or equal to twice the sample size, $N$. The former property of the index is not necessary in understanding the analysis to follow and, hence, its derivation is omitted. We will return to the latter property in the next chapter when we discuss the *relative* spikiness index. We present its proof now, which is not provided in the documentation of the Imanishi-Kari case.

**Theorem 3.5** Let $\max\{SI\}$ denote that maximum possible value of the spikiness index. Then, $\max\{SI\} = 2N$.

**Proof** Proof by induction.

*Base Case.* Suppose there is only 1 cell with $n_1 = N > 0$ observations. Then, by definition, the spikiness index, $SI$, is given by

$$SI = \|0 - n_1\| + \|n_1 - 0\| = n_1 + n_1 = 2n_1 = 2N.$$

Hence, we have verified the base case.

*Inductive Hypothesis.* Suppose there are m cells with $n_1, n_2, \ldots, n_m$ observations in each cell, where $\sum_{i=1}^{m} n_i = N$. Without loss of generality, assume that $n_1 > 0, n_2 > 0, \ldots, n_m > 0$. Then, by definition, the spikiness index, $SI$, is given by

$$SI = \|0 - n_1\| + \sum_{i=2}^{m} \|n_{i-1} - n_i\| + \|n_m - 0\| \leq 2\sum_{i=1}^{m} n_i = 2N.$$

*Inductive Step.* Then, for $m+1$ cells with $\sum_{i=1}^{m+1} n_i = N + n_{m+1} = K$, we have

$$\begin{aligned}
SI &= \|0 - n_1\| + \sum_{i=2}^{m+1} \|n_{i-1} - n_i\| + \|n_{m+1} - 0\| \\
&= \|0 - n_1\| + \sum_{i=2}^{m} \|n_{i-1} - n_i\| + \|n_m - n_{m+1}\| + \|n_{m+1} - 0\| \\
&= \|0 - n_1\| + \sum_{i=2}^{m} \|n_{i-1} - n_i\| + \|n_m - 0\| - \|n_m - 0\| + \|n_m - n_{m+1}\| + \|n_{m+1} - 0\| \\
&\leq 2\sum_{i=1}^{m} n_i - \|n_m - 0\| + \|n_m - n_{m+1}\| + \|n_{m+1} - 0\| \\
&= 2\sum_{i=1}^{m} n_i + \|n_{m+1} - 0\| + \|n_m - n_{m+1}\| - \|n_m - 0\| \\
&= 2\sum_{i=1}^{m} n_i + n_{m+1} + \|n_m - n_{m+1}\| - \|n_m - 0\| \\
&= 2\sum_{i=1}^{m} n_i + n_{m+1} + \|n_m - n_{m+1}\| - \|n_m\|
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\sum_{i=1}^{m} n_i + n_{m+1} + \|n_m - n_{m+1} - n_m\| \\
&= 2\sum_{i=1}^{m} n_i + n_{m+1} + \| - n_{m+1}\| \\
&= 2\sum_{i=1}^{m} n_i + n_{m+1} + n_{m+1} \\
&= 2\sum_{i=1}^{m} n_i + 2n_{m+1} \\
&= 2N + 2n_{m+1} \\
&= 2(N + n_{m+1}) \\
&= 2K
\end{aligned}
$$

Q.E.D.

Having introduced Poisson mixture models and the spikiness index, we now describe in detail the data analyses that the ORI carried out using these statistical techniques.

### 3.4 Data Under Scrutiny

In this set of statistical analyses, the ORI focused mainly on what became known as the June subcloning data. Although this data set was not published in the disputed paper, it played a major role throughout Imanishi-Kari's case. She provided the June subcloning data to the National Institutes of Health (NIH) during their first investigation and she also referred to it in a correction letter she submitted to the journal Cell a year after the investigation had started. (Office of Research Intergrity Report, 1994; Imanishi-Kari et al., 1989) At that time, the NIH did not dispute the veracity of this data set.

However, it was during the second investigation that this data set was scrutinized, mainly because forensic analysis of the ink showed that the tapes on which Imanishi-Kari had recorded her counts were created earlier than the date she had indicated. This was a serious problem because the cells producing the antibodies did not exist at the time the forensic analysis showed the tapes were created. Thus, the forensic experts concluded that the data were fabricated and falsified.

Because of the forensic evidence, the ORI doubted the authenticity of the handwritten counts from Imanishi-Kari's experimental records. In order to assess if these data represented true experimental records, it matched the questioned data sets with data sets whose veracity was not contested. It selected, as controls, tape counts that were never in dispute from Imanishi-Kari's and Reis's notebooks. Since the questioned data consisted of handwritten counts, as further controls, the ORI selected the unquestioned handwritten counts from Reis's records.

Overall, Mosimann examined 10 data sets, of which four were the contentious handwritten counts belonging to Imanishi-Kari. Of all the data sets, two belonged to Reis. Lastly, five of the data sets consisted of tape count data and the other five consisted of handwritten count data. We could not directly assess any of these data sets as we did not have access to them. Hence, we shall mainly rely on the ORI report to describe the analyses that it carried out on these data sets.

For each data set, the ORI estimated the number of components, $g$, from the data. In particular, Mosimann tried to fit the data to mixture models of orders ranging from one to nine "successively," which, to our understanding, means that for each value of $g \leq 9$ that he tried, Mosimann carried out a goodness-of-fit test.

He stopped whenever he found a model to fit the particular data set. Strangely, as will become apparent in the next section, for the data sets that did not fit a Poisson mixture model, he reported the results of the goodness-of-fit tests for models of order less than 9. Instead, we would have expected him to have reported that he had attempted to fit models of all orders up to $g = 9$, a limit imposed by computational constraints.

Mosimann estimated the other unknown parameters of the model, namely, the mixture components, $\pi_i$, $i = 1, \ldots, g$ as well as the Poisson rates of the component densities, $\lambda_i$, $i = 1, \ldots, g$, using the method of maximum likelihood. We do not know what algorithm he used and, further, the ORI report does not provide the estimates of these unknown parameters. From the report, we only had at our disposal the histograms of the fitted models.

For example, Figure 3-2 below shows the frequency histogram of a data set comprising of 34 well counts, the smallest of all data sets that Mosimann analyzed, and the fitted Poisson mixture model.

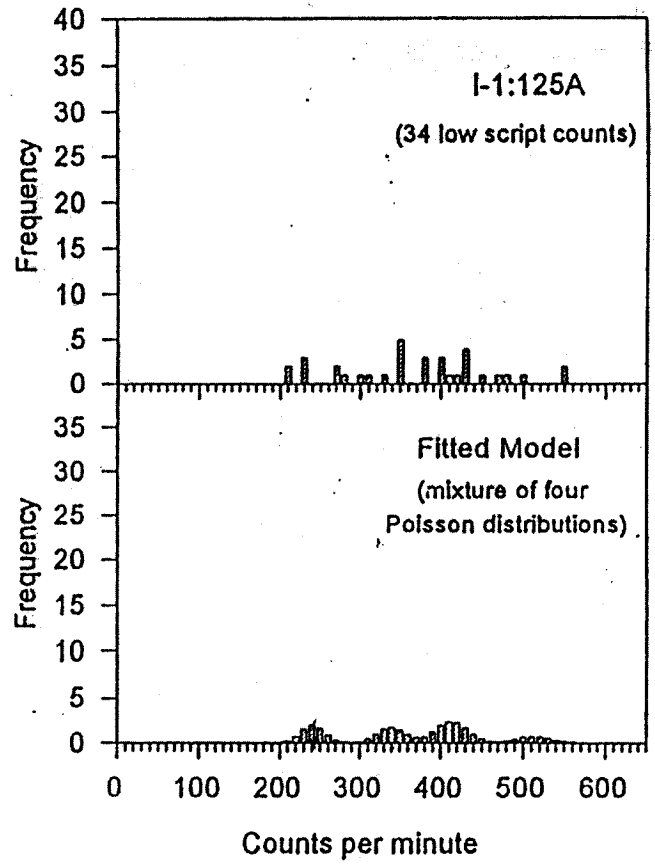



Figure 3–2: A frequency histogram of the well counts on page 125A of Imanishi-Kari's notebook and the corresponding Poisson model fitted to the data. (Office of Research Intergrity Report, 1994)

For the purpose of completeness, in section 3.7 we first derived the maximum likelihood estimators of the mixture model parameters by invoking the EM algorithm.

We then approximated the data set of 34 counts by reading them off the histogram to illustrate Mosimann's putative analysis. We assumed that $g = 4$ as did Mosimann.

We recall that Mosimann tried to fit all of the data sets to the Negative Binomial model as well, to overcome the restriction $g \leq 9$. Since, as we saw in section 3.1, the Negative Binomial distribution arises as an infinite mixture of Poisson distributions, the ORI believed that any data, which fit a Poisson model of any number of mixtures, should fit a Negative Binomial model - an assumption which, as we shall see, came into question.

Mosimann further believed that authentic experimental records from a mixture of Poisson distributions should not exhibit excessive spikiness. He was convinced that Imanishi-Kari's data were unusually spiky. This motivated him to compute the spikiness indices for each and every one of the data sets he examined and compare the observed spikiness indices for the different data sets. He relied on computer simulations to assess the statistical significance of the observed indices, as this was a new statistical procedure.

In the next section, we will present all the statistical evidence that the ORI accumulated as further evidence of fraud based on the Poisson mixture model analyses.

### 3.5 Evidence of Scientific Fraud

All of the control data sets fitted a Poisson mixture model, whereas none of the questioned data sets fitted. Further, Imanishi-Kari's handwritten counts exhibited a spiky behavior in contrast to the counts from the data sets whose veracity was undisputed. We next present the results of the Poisson mixture model analysis and the spikiness analysis.

### 3.5.1 Poisson Mixture Model Analysis of Examined Data

Mosimann fitted the 10 examined data sets that consisted of four questioned and six unquestioned data sets to mixtures of up to nine Poisson distributions. As we see in Table 3-1 below, he also tried to fit all of these data sets to the Negative Binomial distribution.

Exhibit STAT-2
Fits of Poisson Mixture Models to Background Counts

| Notebook:Pages | Mixture Model | Chi Square | Degrees of Freedom | Probability | Model Fits Data? |
|---|---|---|---|---|---|
| I-1:97-99 | N. Binomial[3] | 24.7 | 26 | NS* | Yes |
| | 7-mixture[4] | 20.1 | 15 | NS | Yes |
| I-1:102-104 | N. Binomial | 17.2 | 16 | NS | Yes |
| | 5-mixture | 8.2 | 5 | NS | Yes |
| I-1:12-19 | N. Binomial | 13.0 | 17 | NS | Yes |
| | 6-mixture | 15.7 | 6 | .02 | No |
| I-1:48-58 | N. Binomial | 52.1 | 42 | NS | Yes |
| | 9-mixture | 60.3 | 32 | .002 | No |
| R-1:19-24 | N. Binomial | 22.2 | 24 | NS | Yes |
| | 7-mixture | 20.9 | 14 | NS | Yes |
| I-1:7-8 | N. Binomial | 198.7 | 32 | <.00001 | No |
| | 7-mixture | 172.4 | 22 | <.00001 | No |
| I-1:121 | N. Binomial | 105.8 | 22 | <.00001 | No |
| | 4-mixture | 75.9 | 13 | <.00001 | No |
| I-1:125A | N. Binomial | 30.1 | 14 | .007 | No |
| | 4-mixture | 27.1 | 7 | .0003 | No |
| I-1:124-128A (not 125A) | N. Binomial | 246.7 | 28 | <.00001 | No |
| | 5-mixture | 163.2 | 20 | <.00001 | No |
| R-2:114-124 | N. Binomial | 86.1 | 44 | .0002 | No |
| | 7-mixture | 31.2 | 32 | NS | Yes |

* NS - not significant at the 0.05 probability level. This means that the model "fits" the data at issue.

Table 3–1: The results of fitting the data to Poisson mixture models and to the Negative Binomial model. (Office of Research Intergrity Report, 1994)

The first five rows and the last row of Table 3-1 correspond to the control data sets. The results of the fit to the June subcloning data set are presented in the second to last row.

If any of the data sets fitted neither a finite mixture of Poisson distributions nor the Negative Binomial distribution (that is, an infinite mixture of Poisson distributions) the ORI concluded that the data did not fit *any*-component mixture of Poisson distributions. For all other situations, the ORI concluded that the Poisson mixture model "adequately" described the count data. Table 3-2 below summarizes these conclusions.

Exhibit STAT-1
Notebook Pages from which Background Counts were Studied

| Notebook:Pages | Tape/Script | Number of Low Counts (<600) | Fit by Mixtures of Poissons? |
|---|---|---|---|
| I-1:97-99 | Tape | 59 | Yes |
| I-1:102-104 | Tape | 112 | Yes |
| I-1:12-19 | Tape | 116 | Yes |
| I-1:48-58 | Tape | 574 | Yes |
| R-1:19-24 | Tape | 183 | Yes |
| I-1:7-8 | Script | 94 | No |
| I-1:121 | Script | 72 | No |
| I-1:125A | Script | 34 | No |
| I-1:124-128A (not 125A) | Script | 265 | No |
| R-2:114-124 | Script | 161 | Yes |

Table 3–2: A summary of the ORI's conclusions whether or not the data analyzed are fitted by a mixture of Poisson distributions. (Office of Research Intergrity Report, 1994)

The highlighted handwritten counts in Table 3-2 correspond to the four data sets from Imanishi-Kari's notebook that the ORI examined for fraud. As we can see, none of these data sets fitted a Poisson mixture model, finite or infinite, and, hence,

the ORI deemed these results as support for its speculation that the data had been fabricated and falsified.

Since different reagents were used in the experiments, the ORI analyzed certain sets of count data separately for each different reagent. The results of these analyses reflected what the ORI found for the samples without grouping the counts by the individual reagents. In other words, the handwritten counts from Imanishi-Kari's notebook did not fit a mixture of Poisson distributions, whereas the tape counts from Reis's notebook fitted a mixture of Poisson distribution.

Mosimann believed that Imanishi-Kari's data possibly did not fit a mixture of Poisson distributions because they were spiky. He reasoned that spiky data were unlikely to fit a Poisson mixture model and, therefore, he set out to examine whether that was really the case. In the next section, we briefly summarize the spikiness index analysis that he carried out.

### 3.5.2 The Spikiness Index Analysis of the Examined Data

Further analysis of the handwritten counts from Imanishi-Kari's notebook exhibited a very spiky behavior in contrast to the sets of counts that were used as controls, which exhibited a ""solid base" appearance." To support this claim, Mosimann computed the spikiness indices for each set of counts, both questioned and unquestioned. He found that the four sets of data which did not fit a Poisson mixture model exhibited a statistically significant spiky behavior, evaluated using computer simulations. This was in sharp contrast to the spikiness observed for the control data sets. The ORI made the same observation from the analysis that it carried out for several of the data sets separately by reagent.

It is important to note that the ORI acknowledged in its report that spikiness could have occurred due to rounding of the counts. To account for this fact, Mosimann analyzed counts for spikiness, which were recorded in Imanishi-Kari's notebook and which were known to have been rounded. These counts did not exhibit a spiky behavior.

We summarize the ORI's conclusion based on the results of the Poisson mixture model and the spikiness index analyses.

### 3.5.3 The ORI's Conclusion

The ORI set out to investigate if there was any compelling statistical evidence that, as the forensic evidence had suggested, the June subcloning data as well as other handwritten counts from Imanishi-Kari's notebooks had been fabricated. To objectively assess the veracity of these questioned data, it relied on control data sets whose veracity was never disputed.

Overall, from the Poisson mixture model and the spikiness index analyses that Mosimann carried out, it can be seen that all of the control data sets fitted a Poisson mixture model and did not exhibit spikiness, whereas none of the questioned data sets fitted a Poisson mixture model and, further, exhibited a statistically significant spiky behavior.

Consequently, from these results, the ORI was led to believe that the handwritten counts of the June subcloning data, the most critical of all the questioned data, were "likely" to have been produced by human selection rather than generated from a gamma counter.

In the next section, we discuss in detail the counterarguments summarized in the decision of the Appeals Board, that were based on testimonies by expert scientific witnesses such as Terence Speed. (Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel, 1996)

### 3.6 The Counterarguments

Speed attacked Mosimann's mixture model analyses on several fronts. In particular, he questioned the underlying assumptions and several of the techniques used in these analyses. He also disputed the inferences Mosimann made from these analyses.

Barron, the ORI's other statistical expert, generally agreed with Mosimann's analyses. Nevertheless, he pointed out that, based on them alone, one could not conclude that Imanishi-Kari had fabricated her data. Although he did not clearly indicate whether or not he agreed with the validity of the underlying assumptions in the analyses, one can assume an implicit acceptance of these assumptions.

**Validity of the Poisson Mixture Distribution Model.** Largely based on the testimonies by Speed, the Appeals Board questioned the validity of the Poisson mixture distribution model and concluded that its use was not properly justified. We examine the reasons for Speed's complaints.

First, as we have seen earlier in this section, if the gamma radiation counts come from a constant source (that is, a single well), then it might be reasonable to assume that the counts over a certain time interval would follow a Poisson distribution. However, these experiments consisted of many wells, each of which provided separate counts. The Appeals Board argued that no scientific or empirical evidence existed

to support the validity of the Poisson distribution assumption when the radiation source was not constant.

Moreover, Speed asserted that Mosimann had not justified the use of the mixture models in this experimental setting. We describe below the two common situations when mixture models arise.

Two Interpretations of Mixture Models

Recall that a finite mixture model takes the form

$$f(y) = \sum_{i=1}^{g} \pi_i f(y|\boldsymbol{\theta}_i), \tag{3.12}$$

where $y$ is the observed value of a random variable $Y$, $g$ is the number of components of the mixture, $\pi_i$ is the mixing proportion for component $i$, and $f(y|\boldsymbol{\theta}_i)$ is the density of $Y$ with parameters $\boldsymbol{\theta}_i$ for mixing component $i$.

*Interpretation I*

Suppose a chemist is studying a new chemical compound consisting of three chemical elements and is particularly interested in its flammability. A common measure for that is the amount of heat released as the compound burns. This quantity depends on the specific mixture of the chemical elements forming the compound rather than on the amount of heat released by each chemical element when it burns separately. The usual approach would be to model $Y$, the amount of heat released by the compound, using a mixture model.

In this instance of the use of the mixture model, a random sample from this population would consist of homogeneous observations. That is, each observation would arise from the same underlying mixture distribution, where the mixing proportions,

$\pi_i$ for $i = 1, 2, 3$, represent the actual fractions of the three chemical elements forming the compound, respectively.

*Interpretation II*

On the other hand, consider what appears to be one of the earliest uses of a mixture model attempted by Karl Pearson. The data he analyzed consisted of measurements of a certain body characteristic in a sample of crabs. The scientist, who had asked Pearson's help with the analysis of these data, believed that the crabs were evolving toward two new subspecies. Pearson managed to fit the data to a 2-component Normal mixture model supporting the speculation of the scientist. (McLauchlan and Peel, 2000)

In this second instance of its use, the mixture model is actually a mixture of probability models. A random sample from this population would consist of heterogenous observations each arising from only one of a number of possible subpopulations. In particular, if we let $y$ in (3.12) be the observed measurement for a single crab, we obtain

$$f(y) = \sum_{i=1}^{2} \pi_i f(y|\boldsymbol{\theta}_i) = \pi_1 f(y|\boldsymbol{\theta}_1) + \pi_2 f(y|\boldsymbol{\theta}_2) \tag{3.13}$$

where $f(y|\boldsymbol{\theta}_i)$ is the density of a Normal random variable with parameters $\boldsymbol{\theta}_i = (\mu_i, \sigma_i)$ for $i = 1, 2$. Here, $y$ arose either from the $Normal(\mu_1, \sigma_1)$ distribution with probability $\pi_1$ or from the $Normal(\mu_2, \sigma_2)$ distribution with probability $\pi_2$.

Returning to the count data in Mosimann's analyses, it can be seen that the first interpretation of a mixture model does not apply to them. Rather, it is the second

interpretation, which possibly motivated the use of mixture models in this context. Since counts from a single well can be assumed to follow a Poisson distribution, the important question to ask is whether a mixture of counts, coming from different wells, can be expected to follow a mixture of Poisson distributions - the assumption the ORI made.

We recall once again that Mosimann fitted the count data to Poisson mixture models because of variation from one experimental preparation to another. Now, consider for example, the one set of count data in Imanishi-Kari's notebook that fitted a 5-component mixture of Poisson distributions. Was the fit successful because there were five separate experimental preparations from which these counts arose? Not necessarily. Most likely, the variability of the data simply *happened* to be explained well by a 5-component mixture model. Consider the diagram below.

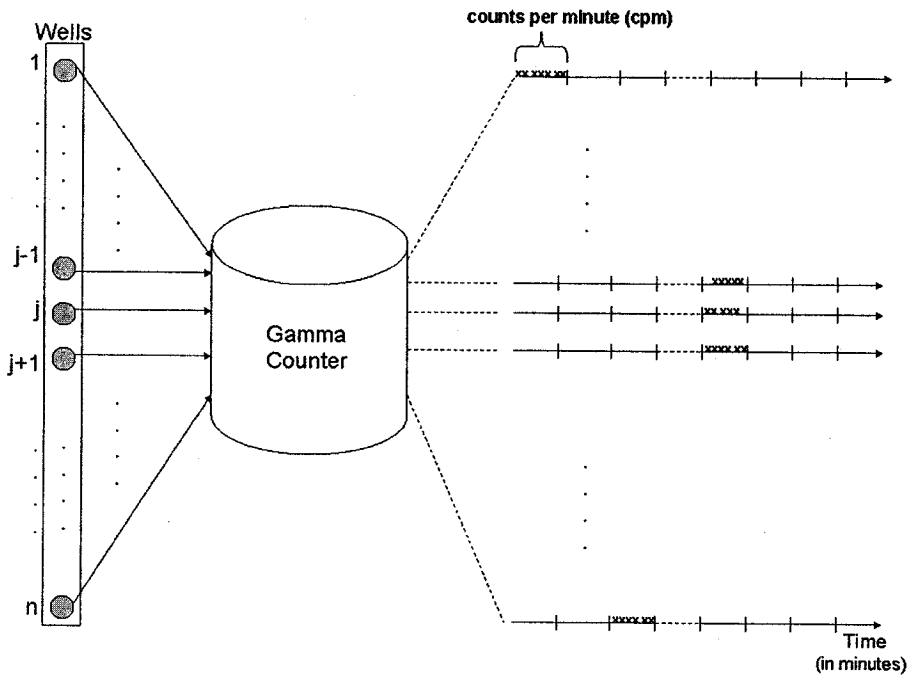



Figure 3–3: How the gamma counter generated the well counts.

From Figure 3-3, we can see that different wells may sometimes give rise to different Poisson distributions. This phenomenon might even arise from wells on the same plate, since wells from the same plate may emanate radiation at different Poisson rates due to the variability of the experimental preparation from well to well.

Mosimann argued that one might consider one of $g$ possible Poisson candidate distributions for producing each well count. Thus the well counts recorded for each of the $n$ wells may be thought of as being selected with probability $\pi_i$ from the Poisson distribution with mean $\lambda_i$. However, there was no concrete example to support this explanation of why a mixture model might be appropriate.

In particular, consider $w_1$, the cpm recorded for the first well. If we *know* for example that $W_1$ arises from the Poisson distribution with parameter $\lambda_3$, say, then the pmf $f(w_1)$ of $W_1$ is simply

$$f(w_1) = \frac{\exp(-\lambda_3)\lambda_3^{w_1}}{w_1!}.$$

If, on the other hand, we do not know from which of $g$ possible Poisson distributions $W_1$ arises but the value of $g$ is known or supported by strong scientific argument, then the pmf $f(w_1)$ of $W_1$ becomes

$$f(w_1) = \sum_{i=1}^{g} \pi_i f(w_1|\lambda_i),$$

that is, a $g$-component mixture of Poisson distributions.

However, there was no prior experiment that indicated that there was a "master set" of $\lambda$'s from which a particular $\lambda_i$ would have been "selected" with probability $\pi_i$, for each well count.

In summary, there was no compelling scientific evidence to support the ORI's claim that the count data should fit a Poisson mixture model. Moreover, as we shall see in the next counterargument, there is a statistical flaw in the ORI's analysis.

**Problems with the Control Data Sets.** The handwritten counts from Imanishi-Kari's notebook were the subject of close scrutiny. The ORI selected, as controls, data sets from Imanishi-Kari's tape counts as well as tape and handwritten counts from Reis's notebook. The Appeals Board disagreed with the analysis of the control data sets for the following reasons.

First, Mosimann believed that the model was valid because it explained the control data sets well. Consider the graph below.

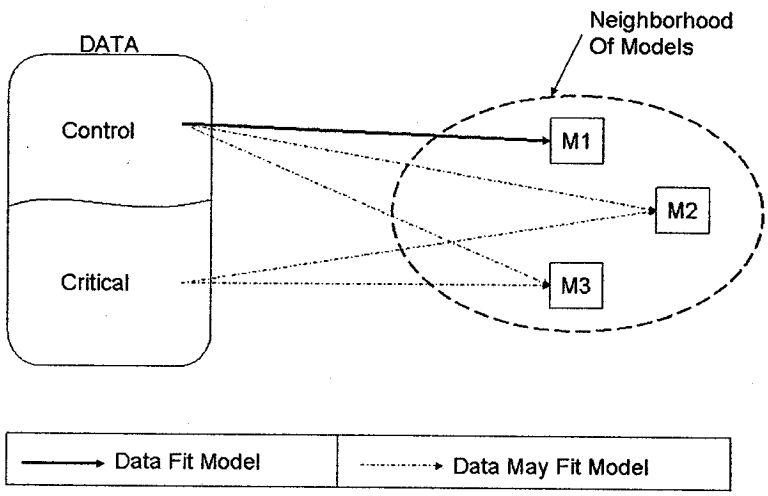



Figure 3–4: Several candidates for the valid (true) model to fit the count data.

Suppose the ORI fitted the control data sets to model M1 in Figure 3-4. Now, consider models M2 and M3, in a neighborhood of model M1. These two models may also fit the control data sets and possibly also fit the critical data sets. The point is that there may be many such models which explain the control data well. Lack of fit of the contentious data to model M1 does not mean that the models M2 and M3 would not have fit these data. Selecting the model M1 as *the* model the questioned data should fit, is poor statistical practice.

Moreover, goodness-of-fit tests in general and, particularly, for Poisson mixture models have low power except for very large samples. Therefore, the argument that because the control data "fit" Poisson mixture models so should the contentious data, was flawed.

Lastly, Mosimann did not justify his choice of control data sets. This was a problem, in particular, because the control data sets were not comparable to the questioned data sets, since they were from experiments performed at a different time from those experiments in question. The tape counts used as controls were the exact (no rounding) tape counts obtained from the counter, in contrast to the questioned data which were heavily rounded. The control data sets were from less complicated experiments and for different reagents.

**Binning Data.** It appears that Imanishi-Kari rounded the low counts (less than 600 cpm) to the nearest 10. To carry out the goodness-of-fit tests, the ORI needed to decide how to bin these count data. Speed complained that Mosimann's choice of binning was subjective.

Because of Imanishi-Kari's rounding procedure, Mosimann used a bin size of 10. That is, he divided the count data into data classes of width 10. Speed argued that he should have followed instead the general rule to bin data, so that there were at least five observations of counts per bin, whereas the rule Mosimann followed resulted in many bins with less than five observations. In addition, the ORI had cited another version to the "5-per-bin" rule, which was to have five observations or more in at least 80% of the bins and at least one observation in all of the bins.

However, Mosimann had violated this variation to the rule as well, with most of the data sets he analyzed.

If the frequency in the bins is not large enough, the chi-square distribution might not serve as a good large sample approximation to the null distribution of the chi-square test statistic. Hence, the p-values obtained based on the observed values of the test statistic might not be valid.

In fact, Speed pointed out that if he rebinned a particular critical data set that he had selected to ensure at least 5 counts per bin, then he found that the contentious data set that had not previously fit the Poisson mixture model, then fitted it. Yet, Mosimann countered by arguing that the rule of minimum 5 values in each bin was not a necessary rule to follow at all times as he believed it hindered the effectiveness of the chi-square test.

Now, Cochran (1952) questions a variant of the 5-per-bin rule, that the smallest *expected* number of counts in any class should be 5 or 10. He suggests, as Mosimann did, that the sensitivity of the chi-square test is likely to be jeopardized if many cells are pooled together at the tails in order to alleviate the "too-few-per-bin" problem. Importantly, it is in the tails, where the differences between the observed and the theoretical distributions are often most apparent and thus easily detected. That is, large bin sizes lead to lack of sensitivity for the test.

In its report, the Appeals Board was concerned about the contradictory views of two adversarial expert witnesses, especially as they had come to different conclusions. This concern supported Speed's case, that there was at least considerate doubt as to the validity of Mosimann's statistical conclusions.

**Problems with the Negative Binomial Model.** The Appeals Board considered Mosimann's Negative Binomial analysis a failed attempt to deal with the computer limitations that the ORI faced, since they were able to fit mixture models of order up to nine only.

We saw earlier in this chapter that the Negative Binomial model arises as an infinite mixture of Poisson distributions. As a result, contrary to what Mosimann claimed, it did not constitute an alternative independent analysis to lend support to the Poisson mixture model analysis already carried out.

Moreover, the ORI asserted that any data which fit a Poisson mixture model should fit the Negative Binomial model. Yet, in the analysis the ORI performed, one of the unquestioned data sets fitted a 7-component Poisson mixture model but did not fit the Negative Binomial model.

In addition, as will be suggested by the simulations we carried out in the next section, a single mixture model can produce data sets quite different in appearance. In fact, each data set could well be described by a number of different mixture models. Further, Speed pointed out that fitting the data to an infinite mixture of Poisson distributions was not equivalent to fitting the data to a Poisson mixture model with a finite number of components.

**Spikiness Index and Computer Simulations.** Since Mosimann worried about certain aspects of the chi-square test, he developed the spikiness test. It is not, however, a widely acceptable statistical procedure and, hence, accusations of fabrication based on the spikiness statistic were questionable. Speed pointed out

that a formal spikiness test was unnecessary since the histogram conveyed the notion of spikiness well enough.

After Mosimann carried out all of these analyses, he came to believe that a 5-component mixture of Poisson distributions explained best the whole set of the June subcloning data. He carried out simulations based on this model. The purpose of these simulations was ostensibly to examine the behavior of a 5-component mixture model so that he could compare this behavior with that of the data from Imanishi-Kari's experiments. In particular, the simulated data enabled him to carry out significance tests for the spikiness index. The Appeals Board thought the simulations had the same problems of subjectivity and invalid assumptions as the actual analysis. We have attempted to reconstruct the analysis and simulations in section 3.7 for a particular data set.

### 3.6.1 Conclusion of the Appeals Board

Overall, the Poisson mixture model analyses carried out by the ORI were seriously flawed, because

1. the underlying assumptions of the model were not justified;
2. the control data sets fitted the model well but this did not prove its validity;
3. the Negative Binomial model did not lend support to these analyses nor the spikiness index served as further indication that fraud had taken place.

Yet, the Appeals Board argued, even if the statistical analyses were not flawed, there were other possible explanations as to why the data did not resemble a random sequence of radiation counts. For instance, it had been already established that

Imanishi-Kari did not intend to be exact when recording the counts and that such behavior was not unusual among scientists who carry out such experiments.

In conclusion, the Appeals Board did not find these statistical analyses convincing and reliable enough to conclude that Imanishi-Kari had intentionally fabricated or falsified experimental data.

### 3.7 A Third (My) Opinion

Mosimann fitted the data by the method of maximum likelihood. Up to this day, no widely distributed software is available which fits count data to a mixture of Poisson distributions. Personal communication with Barron (the ORI's other statistical expert) confirms that Mosimann must have written the software procedures to carry out the tests himself. We do not have the original data nor do we know what statistical software package, if any, Mosimann used. Further, we do not know the algorithm, which he developed to fit the data and estimate the unknown parameters.

However, to enhance our understanding of the analysis, we attempted to fit the 34 counts on page 125A of Imanishi-Kari's notebook by employing the EM algorithm. We read off these counts from the histogram in Figure 3-2. In the next few pages, we provide, for completeness, a detailed description of how the EM algorithm can be applied in this setting.

We established earlier in this chapter that the probability mass function $f(w_j)$ for well count $W_j$, $j = 1, \ldots, n$, is given by

$$f(w_j) = \sum_{i=1}^{g} \pi_i f_i(w_j) = \sum_{i=1}^{g} \pi_i \frac{\exp(-\lambda_i)\lambda_i^{w_j}}{w_j!}.$$

For simplicity, let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_g)^T$ and let $\boldsymbol{\theta} = (\pi_1, \ldots, \pi_{g-1}, \boldsymbol{\lambda})^T$. Then, the likelihood of $n$ independently observed well counts will be given by

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{w}) = f(\mathbf{w}|\boldsymbol{\theta}) = \prod_{j=1}^{n} \left( \sum_{i=1}^{g} \pi_i f_i(w_j) \right) = \prod_{j=1}^{n} \left( \sum_{i=1}^{g} \pi_i \frac{\exp(-\lambda_i)\lambda_i^{w_j}}{w_j!} \right)$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)$.

This likelihood is difficult to maximize and, hence, we resorted to the EM algorithm by formulating the problem as an incomplete-data problem. We defined the random variable $Z_{ij}$ as

$$Z_{ij} = \begin{cases} 1 & \text{if } W_j \text{ arises from group } i; \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, g$ and $j = 1, \ldots, n$. The group assignments of the observations are unknown, which is the reason why our problem became one of a "missing data" problem.

Thus, for a particular observation $W_j$, we have that

$$Pr(W_j = w_j | Z_{1j} = 0, \ldots, Z_{kj} = 1, \ldots, Z_{gj} = 0) = f(w_j|\lambda_k),$$

and also that $Z_{1j}, \ldots, Z_{gj}$ follow the multinomial distribution with parameters $\pi_1, \ldots, \pi_g$ and $n = 1$. Explicitly, we have that

$$\begin{aligned} & Pr(Z_{1j} = 0, \ldots, Z_{kj} = 1, \ldots, Z_{gj} = 0) \\ = & \frac{n!}{z_{1j}! \cdots z_{kj}! \cdots z_{gj}!} \pi_1^{z_{1j}} \cdots \pi_k^{z_{kj}} \cdots \pi_g^{z_{gj}} \\ = & \frac{1!}{0! \cdots 1! \cdots 0!} \pi_1^0 \cdots \pi_k^1 \cdots \pi_g^0 \\ = & \pi_k \end{aligned}$$

Then, the joint density of $W_j$ and $Z_{1j}, \ldots, Z_{gj}$ is given by

$$\begin{aligned}
& Pr(W_j = w_j, Z_{1j} = 0, \ldots, Z_{kj} = 1, \ldots, Z_{gj} = 0) \\
= \; & Pr(W_j = w_j | Z_{1j} = 0, \ldots, Z_{kj} = 1, \ldots, Z_{gj} = 0) \\
& \times \; Pr(Z_{1j} = 0, \ldots, Z_{kj} = 1, \ldots, Z_{gj} = 0) \\
= \; & f(w_j|\lambda_k)\pi_k \\
= \; & [f(w_j|\lambda_1)\pi_1]^0 \times \cdots \times [f(w_j|\lambda_k)\pi_k]^1 \times \cdots \times [f(w_j|\lambda_g)\pi_g]^0 \\
= \; & [f(w_j|\lambda_1)\pi_1]^{z_{1j}} \times \cdots \times [f(w_j|\lambda_k)\pi_k]^{z_{kj}} \times \cdots \times [f(w_j|\lambda_g)\pi_g]^{z_{gj}} \\
= \; & \prod_{i=1}^{g} [f(w_j|\lambda_i)\pi_i]^{z_{ij}},
\end{aligned}$$

where the $z_{ij}$ are the realized values of $Z_{ij}$, for $i = 1, \ldots, g$ and $j = 1, \ldots, n$.

We then let $\boldsymbol{Z_j} = (Z_{1j}, \ldots, Z_{gj})^T$ and let $\boldsymbol{Z} = (\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n})^T$. Since the observations $W_j$, $j = 1, \ldots, n$, are independent, it immediately follows that the random vectors $\boldsymbol{Z_j}$, $j = 1, \ldots, n$, are also independent. Finally, the complete-data likelihood is thus given by

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{z}) = f(\boldsymbol{w}, \boldsymbol{z}|\boldsymbol{\theta}) = \prod_{j=1}^{n} \prod_{i=1}^{g} [\pi_i f(w_j|\lambda_i)]^{z_{ij}},$$

and the complete-data log likelihood is given by

$$\log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{z}) = \sum_{j=1}^{n} \sum_{i=1}^{g} z_{ij} \{\log \pi_i + \log f(w_j|\lambda_i)\},$$

where $\boldsymbol{z} = (\boldsymbol{z_1}, \ldots, \boldsymbol{z_g})$.

At this stage, we applied the EM algorithm. It is an iterative procedure in two steps, the first being the E (Expectation) step and the second being the M (Maximization) step.

E-Step

In this step, we found the expected value of $\log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{z})$ conditional on the observed data $\boldsymbol{w}$. Let $\boldsymbol{\theta}^{(k)}$ be the value of $\boldsymbol{\theta}$ after the $k^{th}$ iteration of the algorithm and let $\boldsymbol{\theta}^{(0)}$ be the initialized value of $\boldsymbol{\theta}$. Then, at the $(k+1)^{th}$ iteration we evaluated the expected value of $\log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{z})$ conditional on the observed data $\boldsymbol{w}$ using the values we estimated for $\boldsymbol{\theta}$ from the $k^{th}$ iteration. That is, we evaluated

$$\begin{aligned}
& \boldsymbol{E}_{\boldsymbol{\theta}^{(k)}}\{\log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{z})|\boldsymbol{W} = \boldsymbol{w}\} \\
= \; & \boldsymbol{E}_{\boldsymbol{\theta}^{(k)}}\left\{\left(\sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\{\log \pi_i + \log f(w_j|\lambda_i)\}\right)|\boldsymbol{W} = \boldsymbol{w}\right\} \\
= \; & \sum_{j=1}^{n}\sum_{i=1}^{g} \boldsymbol{E}_{\boldsymbol{\theta}^{(k)}}(Z_{ij}|\boldsymbol{W} = \boldsymbol{w})\{\log \pi_i + \log f(w_j|\lambda_i)\} \\
= \; & \sum_{j=1}^{n}\sum_{i=1}^{g} \boldsymbol{E}_{\boldsymbol{\theta}^{(k)}}(Z_{ij}|W_j = w_j)\{\log \pi_i + \log f(w_j|\lambda_i)\}
\end{aligned}$$

for $k = 0, 1, \ldots$.

Now,

$$\boldsymbol{E}_{\boldsymbol{\theta}^{(k)}}(Z_{ij}|W_j = w_j) = Pr_{\boldsymbol{\theta}^{(k)}}(Z_{ij} = 1|W_j = w_j)$$

where, in general, for $1 \leq i \leq g$,

$$\begin{aligned}
& Pr(Z_{ij} = 1|W_j = w_j) \\
= \; & Pr(Z_{1j} = 0, \ldots, Z_{ij} = 1, \ldots, Z_{gj} = 0|W_j = w_j)
\end{aligned}$$

$$= \frac{Pr(Z_{1j}=0,\ldots,Z_{ij}=1,\ldots,Z_{gj}=0,W_j=w_j)}{Pr(W_j=w_j)}$$
$$= \frac{\pi_i f(w_j|\lambda_i)}{\sum_{m=1}^{g}\pi_m f(w_j|\lambda_m)}$$

and, therefore,

$$\boldsymbol{E}_{\boldsymbol{\theta}^{(k)}}(Z_{ij}|W_j=w_j) = Pr_{\boldsymbol{\theta}^{(k)}}(Z_{ij}=1|W_j=w_j) \tag{3.14}$$
$$= \frac{\pi_i^{(k)} f(w_j|\lambda_i^{(k)})}{\sum_{m=1}^{g}\pi_m^{(k)} f(w_j|\lambda_m^{(k)})}. \tag{3.15}$$

M-Step

Recall that we obtained earlier the following likelihood

$$\log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w},\boldsymbol{z}) = \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\{\log\pi_i + \log f(w_j|\lambda_i)\}. \tag{3.16}$$

We maximized this likelihood with respect to $\pi_i$ subject to the constraint that $\sum_{i=1}^{n}\pi_i = 1$ using the method of Langrange multipliers as follows

$$\frac{\partial}{\partial\pi_i}\left\{\sum_{j=1}^{n} z_{ij}\{\log\pi_i + \log f(w_j|\lambda_i)\} + \kappa\sum_{m=1}^{g}\pi_m\right\} = 0 \tag{3.17}$$
$$\iff \frac{\partial}{\partial\pi_i}\left\{\sum_{j=1}^{n} z_{ij}\log\pi_i + \kappa\sum_{m=1}^{g}\pi_m\right\} = 0 \tag{3.18}$$
$$\iff \sum_{j=1}^{n} z_{ij}\frac{1}{\pi_i} + \kappa = 0 \tag{3.19}$$
$$\implies \kappa\sum_{i=1}^{g}\pi_i = -\sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij} \tag{3.20}$$
$$\iff \kappa = -\sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij} = -n. \tag{3.21}$$

Thus, from (3.19) it follows that the maximum likelihood estimator of $\pi_i$ is given by

$$\widehat{\pi_i} = \frac{1}{n}\sum_{j=1}^{n} z_{ij}$$

for $i = 1, \ldots, g$.

Similarly, we maximize the likelihood in (3.16) with respect to $\lambda_i$ such that

$$\begin{aligned}
&\frac{\partial}{\partial \lambda_i}\sum_{j=1}^{n} z_{ij}\{\log \pi_i + \log f(w_j|\lambda_i)\} = 0 \\
\iff &\sum_{j=1}^{n} z_{ij}\frac{\partial}{\partial \lambda_i}\log f(w_j|\lambda_i) = 0 \\
\iff &\sum_{j=1}^{n} z_{ij}\frac{\partial}{\partial \lambda_i}\log\left\{\frac{e^{-\lambda_i}\lambda_i^{w_j}}{w_j!}\right\} = 0 \\
\iff &\sum_{j=1}^{n} z_{ij}\frac{\partial}{\partial \lambda_i}\{-\lambda_i + w_j\log\lambda_i - \log w_j!\} = 0 \\
\iff &\sum_{j=1}^{n} -z_{ij} + \frac{z_{ij}w_j}{\lambda_i} = 0.
\end{aligned}$$

Solving for $\lambda_i$ we obtain the maximum likelihood estimator of $\lambda_i$

$$\widehat{\lambda_i} = \frac{\sum_{j=1}^{n} z_{ij}w_j}{\sum_{j=1}^{n} z_{ij}}$$

for $i = 1, \ldots, g$.

Lastly, since we cannot observe $z_{ij}$ we replaced them in both estimators $\widehat{\pi_i}$ and $\widehat{\lambda_i}$ by the conditional expectation of $Z_{ij}$ evaluated in the E-step derived in (3.15).

We stopped the algorithm whenever

$$|\log \mathcal{L}(\boldsymbol{\theta}^{(k+1)}|\boldsymbol{w}, \boldsymbol{z}) - \log \mathcal{L}(\boldsymbol{\theta}^{(k)}|\boldsymbol{w}, \boldsymbol{z})| < \epsilon$$

for some small $\epsilon > 0$.

We used the R statistical software package and we obtained the following fit.

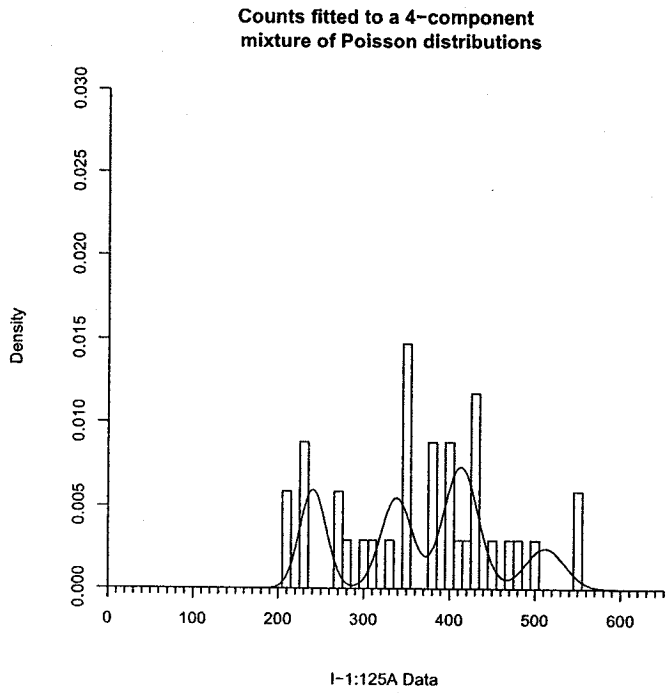



Figure 3–5: Fitting counts from page 125A in Imanishi-Kari's notebook to a mixture of four Poisson distributions using the EM algorithm.

We see that the fit depicted by the solid line in Figure 3-5 is very close to the fitted model in Figure 3-2.

We continued by simulating several samples of 34 counts from the fitted model. We present in the figure below the first 12 samples we generated.

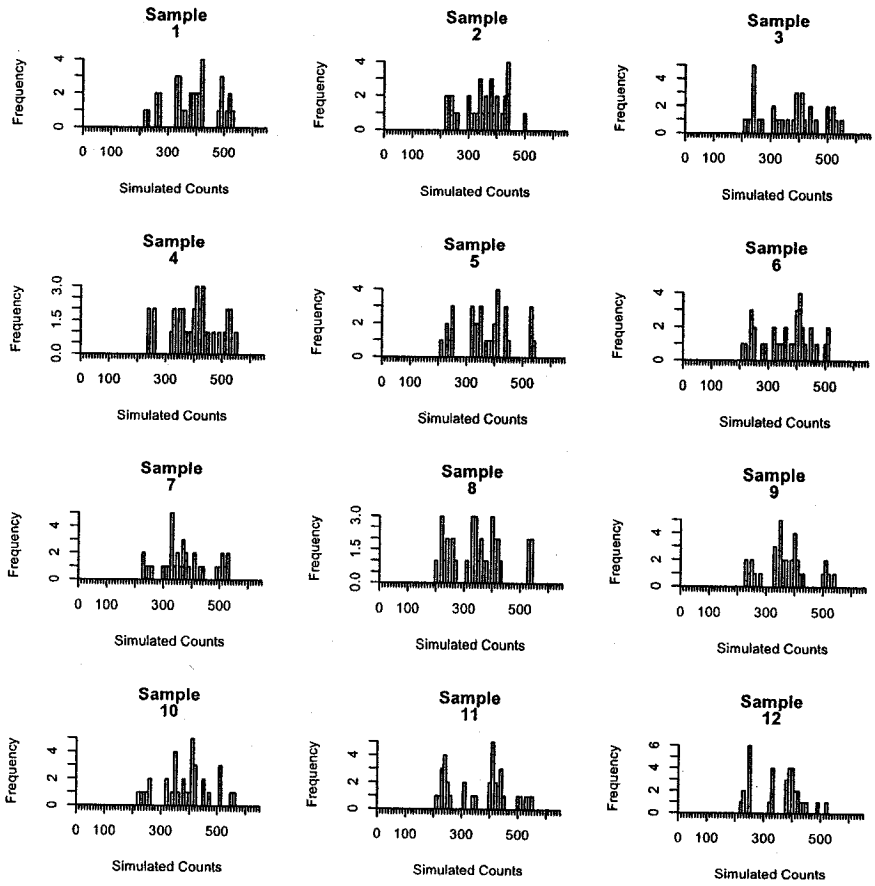



Figure 3–6: Simulating 12 samples of 34 counts from the fitted mixture of four Poisson distributions.

From this figure it becomes immediately apparent that the data sets from the same mixture model are highly variable. We have roughly unimodal data in Sample 1 and multimodal data in Sample 8. Also, the data from Sample 10 appear to be spikier than the data from Sample 4.

This type of variability in the data increases the plausibility of the explanation that the data from Imanishi-Kari's records were simply a "chance" sample of random counts but nevertheless still possible to be observed as opposed to being fabricated or falsified.

We simulated 500 samples of size 34 from the fitted mixture of four Poisson distributions in the same way that we simulated the 12 samples in Figure 3-6 above, and computed the spikiness index for each of these 500 samples. Finally, we found the number of times those indices exceeded the value of 56, the index observed from the original 34 counts. Letting S be the random variable representing the spikiness index, then we found based on this current simulation, that

$$\widehat{Pr}(S > 56) = 0,$$

since out of the 500 spikiness indices computed, none exceeded 56.

Mosimann found a probability of 0.0004. Although, this simulation was solely done for the purpose of illustration, it does seem to support Mosimann's claim that the count data of sample 125A were more spiky than would be expected from a mixture of four Poisson distributions.

In this chapter, we presented the Poisson mixture model analyses carried out by the ORI. First, we developed the background necessary to understand these analyses. Next, we described the data under scrutiny, in particular, the June subcloning data set and the various control data sets. We then recounted the statistical evidence, which the ORI put forward against Imanishi-Kari, based on these mixture model

analyses. Finally, we carried out some simulations of a 4-component Poisson mixture model, where the unknown parameters were estimated using the EM algorithm, and we also presented an example of how the significance of the spikiness index was evaluated. We present in the next chapter, in a similar manner, the third and final set of statistical analyses carried out by the ORI, namely, the analyses on the distribution of digits in the handwritten counts from the June subcloning data set.

# CHAPTER 4

## Distribution of Digits

Suppose you were asked to flip a coin 200 times and record whether you observed a head or a tail. You have two choices. You can either repeatedly flip the coin 200 times and record the outcome each time or just "pretend" you flipped the coin and simply write down a sequence of heads and tails made up by you to appear as if they had arisen as true outcomes of coin tosses. After all, how difficult can cheating on this exercise be if you were to go unnoticed?

Well, it turns out it is not that easy. What people believe to be a random sequence of heads and tails may not produce anything like a true random sequence of heads and tails. In fact, it can be shown, by using a simple Markov chain argument, that a true random sequence of 200 tosses will produce at least one run of six or more heads or six or more tails with probability around .95. However, such runs are rarely observed in the recorded outcomes by people who cheat.

This is a simple instance of the fact that people are rarely able to generate a truly random sequence of numbers because of an inherent preference of certain digits over others, and because randomness includes sequences that humans would be tempted to regard as nonrandom because of runs of repeated outcomes. As a result, this makes faking or fabricating experimental numerical data quite difficult.

Not surprisingly statistical techniques relying on this fact are available to analyze data sets for their veracity. This is precisely the motivation behind this third and last statistical analysis that the ORI carried out.

It is necessary to differentiate between the distribution of leading and non-leading digits. Since Imanishi-Kari had rounded her counts, the ORI was interested in the non-leading digits of the counts it analyzed. In particular, the following algorithm was used to give the sets of digits, whose analysis was to determine whether falsification had taken place.

1. "Starting from the right side of the number and moving left, the first nonzero digit (called x) was found.
2. If there was at least one more digit to the left of digit x, then digit x was selected as the digit from the number; however, if there was no digit further to the left of x, then x was discarded and no digit was selected from that number." (Office of Research Intergrity Report, 1994)

The ORI explained the algorithm with the following two examples. Consider the count 38200 cpm. Then, following the first step, they set x=2. From the second step, they verified that there was a digit to the left of x. Therefore, they selected to use in their analyses the digit 2 from this count. On the other hand, consider the count 800. Here, they set x=8 but since there was no other digit to the left of digit 8, they selected no digit from this count.

The whole point of this algorithm was to obtain a data set of "insignificant" digits ranging from 1 to 9. If the counts were to be randomly generated, then the

ORI asserted that these digits should follow the discrete uniform distribution with the probability of each digit occurring equal to 1/9.

Handwritten and tape counts that the ORI used as the control data sets, which were not rounded, were subjected to a rounding algorithm designed by Mosimann. The design of the algorithm attempted to reflect the way Imanishi-Kari had rounded her data. The rounding rules were as follows:

- "numbers over 10,000 were rounded to the nearest "thousand" (e.g. 12,450 to 12,000, 17,800 to 18000);
- numbers less than 10,000 but over 1,000 were rounded to the nearest "hundred" (e.g. 1,245 to 1,200, 1,780 to 1,800);
- numbers less than 1,000 but over 100 were rounded to the nearest "ten" (e.g. 242 to 240, 585 to 590); and
- numbers less than 100 were not used." (Office of Research Intergrity Report, 1994)

The ORI also computed the "relative spikiness index" for each sample of digits.

**Definition 4.1** The relative spikiness index, $RSI$, is defined as the spikiness index, $SI$, (see Definition 3.4) of a sample of N observations divided by the sample size, $N$. (Office of Research Intergrity Report, 1994) That is, $RSI$ is given by

$$RSI = \frac{SI}{N}.$$

**Theorem 4.2** Let $\max\{RSI\}$ be the maximum value of the relative spikiness index. Then, $\max\{RSI\} = 2$.

**Proof** From Theorem 3.5, we know that $SI \leq 2N$. It immediately follows, therefore, that

$$RSI = \frac{SI}{N} \leq \frac{2N}{N} = 2.$$

Q.E.D.

Now, recall that the rounding rules changed at the points 100, 1000 and 10000 disturbing the uniformity of the data. Hence, departure from uniformity alone could not lead to the conclusion that the data were counterfeit. If, in addition, the data were spiky, then the ORI considered it as evidence of fraud. The ORI relied on the $RSI$ as the measure of spikiness.

### 4.1 Data Under Scrutiny

For the distribution of digits analyses, the ORI focused on the June subcloning data set as well. In fact, all but one of the data sets used in these analyses were the same as those used in the Poisson mixture model analyses. However, for the distribution-of-digits analyses, unlike for the Poisson mixture model analyses, both low counts (less than 600 cpm) as well as high counts (greater than or equal to 600 cpm) were analyzed.

The data sets were essentially divided into two groups, the "critical" data sets and the "neutral" data sets. The critical data sets consisted of the handwritten counts by Imanishi-Kari. The "neutral" data sets consisted of tape and handwritten counts from Reis's notebook as well as tape counts from Imanishi-Kari's notebook. For each data set, where applicable, the distribution of digits from high counts was analyzed separately from the distribution of digits from low counts.

Of particular importance was one of the control data sets from Imanishi-Kari's notebook consisting of high tape counts, for which the corresponding rounded counts by the author were also available. The ORI included part of that data set in its report. We reproduce it below in Table 4-1.

Exhibit STAT-7
Rounding Behavior of the author of Notebook I-1

| Tape CPM | Script CPM, Rounded | Tape CPM minus Rounded CPM | Error (Per Cent) | Tape CPM | Script CPM, Rounded | Tape CPM minus Rounded CPM | Error (Per Cent) |
|---|---|---|---|---|---|---|---|
| 19890 | 19000 | 890 | 4.5 | 4540 | 4500 | 40 | 0.9 |
| 1736 | 1800 | -64 | -3.7 | 4919 | 4900 | 19 | 0.4 |
| 19275 | 19000 | 275 | 1.4 | 16221 | 16200 | 21 | 0.1 |
| 8229 | 8000 | 229 | 2.8 | 34938 | 35000 | -62 | -0.2 |
| 4029 | 4000 | 29 | 0.7 | 31313 | 31000 | 313 | 1.0 |
| 2121 | 2200 | -79 | -3.7 | 5221 | 5200 | 21 | 0.4 |
| 1468 | 1400 | 68 | 4.6 | 8062 | 8000 | 62 | 0.8 |
| 4459 | 4400 | 59 | 1.3 | 1032 | 1000 | 32 | 3.1 |
| 8644 | 8600 | 44 | 0.5 | 3874 | 3800 | 74 | 1.9 |
| 1997 | 2000 | -3 | -0.2 | 2102 | 2000 | 102 | 4.9 |
| 13485 | 13000 | 485 | 3.6 | 5292 | 5300 | -8 | -0.2 |
| 7758 | 7800 | -42 | -0.5 | 2522 | 2500 | 22 | 0.9 |
| 2235 | 2200 | 35 | 1.6 | 1928 | 1900 | 28 | 1.5 |
| 1051 | 1000 | 51 | 4.9 | 1627 | 1600 | 27 | 1.7 |
| 3927 | 3900 | 27 | 0.7 | 792 | 800 | -8 | -1.0 |
| 15766 | 15700 | 66 | 0.4 | 4846 | 4800 | 46 | 0.9 |
| 19327 | 19000 | 327 | 1.7 | 7577 | 7600 | -23 | -0.3 |
| 13467 | 13000 | 467 | 3.5 | 11003 | 11000 | 3 | 0.0 |
| 1089 | 1000 | 89 | 8.2 | 3386 | 3300 | 86 | 2.5 |
| 2248 | 2200 | 48 | 2.1 | 503 | 500 | 3 | 0.6 |
| 17971 | 17000 | 971 | 5.4 | 2476 | 2500 | -24 | -1.0 |
| 3347 | 3000 | 347 | 10.4 | 1301 | 1300 | 1 | 0.1 |
| 6574 | 6600 | -26 | -0.4 | 2948 | 2900 | 48 | 1.6 |
| 13582 | 13500 | 82 | 0.6 | 10441 | 10400 | 41 | 0.4 |
| 19006 | 19000 | 6 | 0.0 | 1975 | 1900 | 75 | 3.8 |
| 13069 | 13000 | 69 | 0.5 | 9010 | 9000 | 10 | 0.1 |
| 3825 | 3800 | 25 | 0.7 | 8334 | 8000 | 334 | 4.0 |
| 1132 | 1100 | 32 | 2.8 | 1412 | 1400 | 12 | 0.8 |
| 17393 | 17000 | 393 | 2.3 | 11712 | 11700 | 12 | 0.1 |
| 3074 | 3000 | 74 | 2.4 | 2329 | 2300 | 29 | 1.2 |
| 1697 | 1700 | -3 | -0.2 | 6629 | 6600 | 29 | 0.4 |
| 4433 | 4000 | 433 | 9.8 | 24395 | 24300 | 95 | 0.4 |
| 8229 | 8000 | 229 | 2.8 | 37107 | 37100 | 7 | 0.0 |

Table 4–1: Original tape counts and the corresponding rounded counts by Imanishi-Kari. (Office of Research Intergrity Report, 1994)

As an illustration, we apply the rounding algorithm and the digit selection algorithm to three counts from Table 4-1.

| Tape Count | Rounded by Imanishi-Kari | Rounded by Algorithm | Selected Digit: Rounded by Algorithm (Imanishi-Kari) |
|---|---|---|---|
| 19890 | 19000 | 20000 | none (9) |
| 1736 | 1800 | 1700 | 7 (8) |
| 792 | 800 | 800 | none (none) |

Table 4–2: Examples illustrating how the ORI applied the rounding rules and the digit selection algorithm to the data in its analyses.

From the few examples in the Table 4-2 above, we see that often the rounded counts from Imanishi-Kari and from the algorithm differ. Although the discrepancy is not big, it inevitably results in a different distribution of digits from the same sample of counts. Figure 4-1 below illustrates this point explicitly.

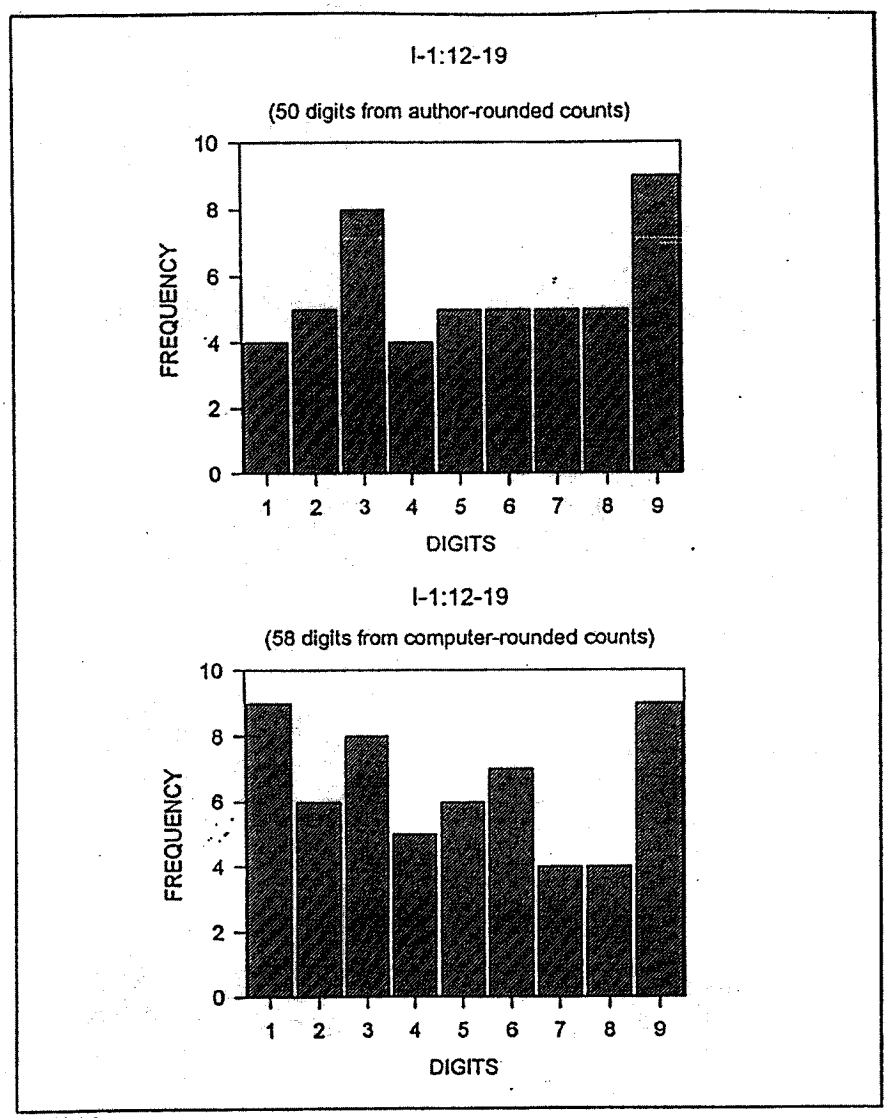



Figure 4–1: A total of 50 digits were selected from the counts rounded by the author and a total of 58 digits were selected from the same counts but rounded using the rounding algorithm. (Office of Research Intergrity Report, 1994)

We see from Figure 4-1 that the counts rounded by Imanishi-Kari led to the selection of 50 digits, whereas the counts rounded by the computer algorithm led to the selection of 58 digits for exactly the same sample of counts. The analysis of both samples, however, led to the same conclusion that the digits were not inconsistent with the null hypothesis of uniformity.

Explicitly, for the various data sets the ORI analyzed, it tested, using the chi-square goodness-of-fit test, the null hypothesis that the digits 1 to 9 followed the discrete uniform distribution with probability 1/9 against the alternative that the digits did not follow the specified distribution. In addition, the ORI calculated the relative spikiness index for each sample of digits.

## 4.2 Evidence of Scientific Fraud

Figure 4-2 below provides the summary of results of the digit distribution analyses the ORI carried out.

Exhibit STAT-8
FREQUENCIES OF DIGITS 1 TO 9

| PAGES, HIGH OR LOW | N | PROBABILITY OF CHISQUARE | RELATIVE SPIKINESS[1] |
|---|---|---|---|
| CRITICAL DATA SETS: Handwritten counts by the author of notebook I-1 | | | |
| High script: | | | |
| I-1:125A | 43 | 0.0451 | 0.698 |
| I-1:121 | 47 | 0.0001 | 0.979 |
| I-1:124-128A (not 125A) | 61 | <.0001 | 0.820 |
| (BET-1 alone) | 26 | 0.0495 | 1.000 |
| (AF-6 alone) | 11 | 0.0424 | 1.273 |
| (Anti-$\mu$ alone) | 24 | 0.1879 | 0.750 |
| I-1:7-8 | 53 | <.0001 | 1.245 |
| I-1:9-10 | 69 | 0.0038 | 0.667 |
| Low script: | | | |
| I-1:125A | 33 | 0.0023 | 1.455 |
| I-1:121 | 68 | <.0001 | 1.412 |
| I-1:124-128A (not 125A) | 260 | <.0001 | 1.123 |
| (BET-1 alone) | 79 | <.0001 | 1.342 |
| (AF-6 alone) | 99 | <.0001 | 1.192 |
| (Anti-$\mu$ alone) | 82 | 0.0001 | 1.049 |
| I-1:7-8 | 63 | <.0001 | 1.492 |
| I-1:9-10 | 26 | <.0001 | 1.462 |
| NEUTRAL DATA SETS: Author's rounding versus computer's rounding, for the same counts | | | |
| High tape: | | | |
| I-1:20-21, Author of I-1 | 50 | 0.8233 | 0.520 |
| I-1:20-21, Computer | 58 | 0.7902 | 0.621 |
| COMPUTER'S ROUNDING OF TAPE COUNTS | | | |
| High tape: | | | |
| I-1:12-19 | 68 | 0.4627 | 0.618 |
| I-1:124-128A | 149 | 0.0049 | 0.483 |
| I-1:48-58 | 254 | 0.1841 | 0.402 |
| I-1:102-104 | 147 | 0.1512 | 0.558 |
| I-1:97-99 | 60 | 0.4838 | 0.533 |
| R-1:19-24 | 95 | 0.4543 | 0.568 |
| Low tape: | | | |
| I-1:12-19 | 100 | 0.3665 | 0.460 |
| I-1:124-128A | 22 | 0.2176 | 0.636 |
| I-1:48-58 | 497 | 0.9931 | 0.266 |
| I-1:102-104 | 100 | 0.9924 | 0.380 |
| I-1:97-99 | 55 | 0.7863 | 0.509 |
| R-1:19-24 | 161 | 0.8399 | 0.360 |
| COMPUTER'S ROUNDING OF HANDWRITTEN COUNTS BY THE AUTHOR OF R-2 | | | |
| Low script: | | | |
| R-2:114-124 | 144 | 0.9769 | 0.347 |

[1] "Relative spikiness" is the spikiness index divided by the number of observations, N.

Figure 4–2: A summary of the results of the digit distribution analysis. (Office of Research Intergrity Report, 1994)

For the 15 "neutral" data sets tested, which consisted mostly of tape counts, we see from Figure 4-2 that almost all were consistent with the uniform distribution. In fact, the high tape counts from only one data set seemed to be inconsistent with the uniformity assumption (p-value = 0.0049). However, the ORI accounted for the fact that 15 simultaneous tests were carried out and, hence, subjected the p-value to a Bonferroni correction. The corrected p-value was no longer significant (corrected p-value = 15*0.0049 = 0.074). The ORI also pointed out that since the relative spikiness index was low at 0.483, the departure from uniformity for that sample of digits was not due to the fact that the digits were spiky. Overall, the ORI concluded that the uniform distributions "adequately describe" the distributions of digits coming from the control data sets. (Office of Research Intergrity Report, 1994)

On the other hand, for the 16 "critical" data sets tested, which consisted mostly of handwritten counts, none except one fitted the uniform distribution. In particular, consider the *low* handwritten counts only. Even after the Bonferroni correction was applied to all the p-values by multiplying each p-value by 16 for the number of simultaneous tests carried out, they still remained significant thus indicating that these samples of digits did not fit the uniform distribution. Furthermore, the ORI pointed out, the relative spikiness indices calculated for these samples were quite high all being greater than 1.

Now, consider the *high* handwritten counts. With the Bonferroni correction, four of these samples of digits appeared to be nonuniform. In addition, these samples also had higher relative spikiness indices than any of the indices calculated for the control data sets.

Overall, for the most critical data set consisting of the handwritten counts from the June subcloning data set found on pages 124-128A of Imanishi-Kari's notebook, the ORI found that the digits were not uniformly distributed and were spiky. With the low handwritten counts, visual examination revealed that the digits 1,3,7 and 8 appeared more frequently than expected and with the high handwritten counts, the digits 7 and 8 appeared more frequently than expected. Figure 4-3 below graphically illustrates this fact.

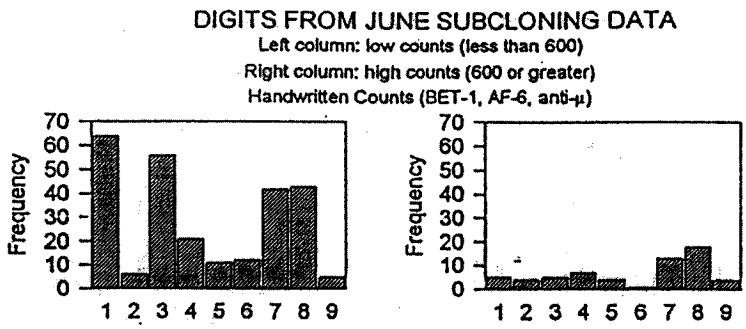



Figure 4–3: Histograms showing the frequency of digits in the June subcloning data for both high and low counts. (Office of Research Intergrity Report, 1994)

Imanishi-Kari's known rounding behavior did not account for the nonuniformity and spikiness of the sample of digits from the critical data sets. This was because the rounded counts in Imanishi-Kari's notebook, for which the original tape counts were also available, were consistent with the assumption of uniformity and were not spiky.

In summary, all the control data sets were consistent with the uniform distribution and were not spiky. In contrast, most questioned data sets and, in particular,

the June subcloning data set, departed from uniformity and had comparatively high relative spikiness indices. Since these results were not explained by Imanishi-Kari's rounding behavior, the ORI concluded that the handwritten counts, including the June subcloning counts, were likely to have been fabricated.

### 4.3 The Counterarguments

There was general consensus that the handwritten counts in Imanishi-Kari's notebook were the result of some sort of human intervention. However, what the statistical analysis aimed at showing was the fact that this human intervention was intentional. The Appeals Board did not find this analysis convincing enough to conclude that Imanishi-Kari had fabricated or falsified her data for the following reasons.

First, the statistical analysis of the distribution of digits of the June subcloning data showed that certain digits occurred more often than others. This reflected Imanishi-Kari's personal preference for certain digits over others. However, this fact did not in itself prove that the counts were fabricated. An equally plausible explanation could be that she had not been meticulous when she was recording the counts.

Second, the assumption of uniformly distributed digits applies to rightmost digits only. The ORI based its analysis on this assumption although it was not the rightmost digits that were being analyzed. The algorithm Mosimann used to select the digits for the analysis described earlier in this chapter gave digits, which were the rightmost digits that were nonzero and that were also not the leftmost digits.

The ORI did not justify that the assumption of uniformity was valid in that case as well.

Third, the ORI claimed it carried out its analysis on "insignificant" digits, that is, the digits which did not contain any information. However, the Appeals Board questioned the way in which digits were judged to be significant or insignificant. In its analysis, the ORI considered, for example, the digit 8 in 800 to be significant but the digit 7 in 27000 not to be significant.

Lastly, Mosimann at first claimed that his rounding algorithm did not aim at mimicking exactly Imanishi-Kari's rounding behavior but in his testimony he said that it did follow her rounding behavior. In actual fact, the Appeals Board found that the rounding algorithm, to which the control data sets were subjected did not really model the rounding behavior of Imanishi-Kari. This observation was supported by a data set from Imanishi-Kari's notebook, for which the tape counts were available in addition to the handwritten counts.

Although Barron deemed valid the statistical methods used by Mosimann in this analysis, he said that he would not conclude that fabrication had occurred based on this analysis alone. Barron asserted that other evidence was necessary to reach such a conclusion.

In summary, the Appeals Board concluded that "the statistical analyses [were] not reliable evidence that the June subcloning data or other questioned data were created by fabrication of falsification as alleged by [the] ORI." (Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications

Panel, 1996)

In this chapter, we presented the last statistical analyses carried out by the ORI, namely, the distribution of digits analyses. First, we provided the necessary background to understand them. Next, we described the data that were examined. We then recounted the statistical evidence, which the ORI put forward against Imanishi-Kari, based on these digit distribution analyses. Finally, we outlined several possible flaws with them. We present in the next chapter our concluding remarks on the Imanishi-Kari's alleged case of fraud.

# CHAPTER 5

## Conclusion

Up until 1986 there was no central body within the United States National Institutes of Health (NIH) that was designated specifically for dealing with cases of scientific misconduct. In 1989, after the NIH had concluded its first investigation into the case of Imanishi-Kari, the United States Public Health Service created the Office of Scientific Integrity (OSI) and the Office of Scientific Integrity Review (OSIR). A month after that, the NIH reopened its investigation into the case within the newly created OSI.

A few months just before, United States Congress had become involved in Imanishi-Kari's case as well. The Congressional Investigations Committee of Representative John Dingell held the hearings. Dingell, a Democrat from Michigan, chaired the House Energy and Commerce Committee, which had control over the NIH. He was interested in scientific fraud particularly because the NIH had a very high operating budget and he believed that his committee could not "afford to divert precious dollars into areas of meaningless or fraudulent work." (Kevles, 1998) Dingell forced Imanishi-Kari to provide all her research records and he sent them to the Secret Service for analysis.

In 1992, the OSI and the OSIR merged to form the currently known Office of Research Integrity (ORI). About two years later, the ORI issued its 231 page report with attachments accusing Imanishi-Kari on 19 charges of scientific misconduct. She appealed the decision to an independent body, the United States Department of Health and Human Services Departmental Appeals Board, whose mission is to provide "prompt, fair, and impartial dispute resolution services to parties in many different kinds of disputes involving components of the Department of Health and Human Services." (Department of Health and Human Services, Departmental Appeals Board) In 1996, an adjudication panel of three members appointed by the Appeals Board concluded that the "ORI did not prove by a preponderance of the evidence that Dr. Imanishi-Kari engaged in scientific misconduct as charged." (Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel, 1996)

In essence, at the time the case was unfolding, the institution then playing the role of the ORI today, was undergoing major reorganization. There were no well established procedures for such cases nor were there people with well defined responsibilities to take care of these procedures. No doubt, this situation further complicated the case.

## 5.1 Dealing with Scientific Misconduct Today

We have used the phrase "scientific misconduct" numerous times throughout this thesis, but what exactly does it mean? It is a challenging task to come up with a comprehensive definition for a term such as this. The currently accepted definition by the ORI is the following:

"Scientific misconduct or misconduct in science means fabrication, falsification, plagiarism, or other practices that seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting, or reporting research. It does not include honest error or honest differences in interpretations or judgements of data." (Office of Research Integrity Model Policy, Revised February 1997)

It is of interest to note that the National Research Council Canada has also adopted this definition of scientific misconduct. (National Research Council Canada)

Now that we have defined the problem, the natural question arises. How are we to solve it? How do institutions such as the ORI currently deal with an alleged case of scientific misconduct?

The ORI has come a long way since the case of Imanishi-Kari. It has established many policies and procedures for handling misconduct ranging from the protection of the whistleblower to the complete review of final reports in the investigations to assure quality and fairness. (Office of Research Integrity Model Policy, Revised February 1997; Office of Research Integrity Model Procedures, Revised February 1997) It has also published numerous studies on or relating to scientific integrity and has various other studies currently in progress.

The ORI's website offers a plethora of information not only on scientific misconduct and all related information in the United States but also on the current state of this issue around the world. The ORI has become an exceptional institution of its kind, the steps of which the newly created UK Research Integrity Office, for example, hopes to follow. (BBC News, 2006a)

Without doubt the major goal of such institutions is at the least discouraging and at the best preventing research misconduct. Clearly, scientific integrity cannot be forced upon the researcher but various aspects of the research practices can be. Scientific journals and research institutions such as the universities have a major role to play. The good news is that these bodies have become increasingly aware of the need to take action. For example, the BMJ (British Medical Journal) has an ethics committee, which is an independent body set up to advise editors and peer-reviewers on research ethics. As another example, by July 1995 the government of Canada had required from all its research universities to establish policies and guidelines for handling scientific fraud. (Science, 1994)

In general, what can countries do to handle scientific misconduct effectively before and after it has happened? Richard Smith, a former editor of BMJ for 25 years, presents the following arguments: Naturally, the first step is to recognize that the problem exists. He points out that a widely accepted definition of scientific misconduct is essential. Next, all issues of or relating to scientific misconduct must be dealt with by a centralized independent body. Also, protection of the whistleblowers as well as a fair system has to be guaranteed as much as possible. Finally, a code of good research practice has to be established and widely promoted and encouraged. (Smith, 2001)

The implementation of these steps would only indicate strong attempts by governments to enforce good research practice in their scientific communities. However, the international scientific community today is so intertwined that measures beyond

the national level are necessary. Unfortunately, the drawback is that too many regulatory or monitoring policies may also create an unproductive and an undesirable research environment.

There is a wide range of reasons why one pursues a career in science the most common being his or her genuine interest in the advancement and understanding of that particular scientific field. Researchers today face enormous pressure to obtain results and to build a reputation through a sound list of publications. There is intense competition and there shall always be circumstances which will trigger unethical research practices. Therefore, the responsibility in preventing those practices from taking place rests in no other but the scientists themselves, young and old alike.

## 5.2 Years Later: Update on the Players

Thereza Imanishi-Kari is currently an associate professor of pathology at Tufts University. She has one PhD student under her supervision. The main focus of their research is the pathogenesis of systemic lupus erythematosus, a systemic autoimmune disease.

David Baltimore is currently the president of the California Institute of Technology and has held that position since 1997. He supervises a laboratory with nine postdoctoral fellows, five staff members and one undergraduate student. Topics being researched in his laboratory include various critical factors affecting HIV genome expression.

The latest records of the whereabouts of Margot O'Toole date to 1998, when she was working for a biotechnology company in Boston. Her husband at the time

and possibly today, Peter Brodeur, currently also holds a position as an associate professor of pathology at Tufts University.

Baltimore, together with three of the other five authors, retracted the disputed paper. (Weaver et al., 1991) Imanishi-Kari did not retract the paper. One of the inconsistencies that O'Toole found in the published data was on the specificity of one of the reagents used in the experiments. Imanishi-Kari replied to this with a letter to the editor of Cell. (Imanishi-Kari et al., 1989) In general, however, the authors claimed that the errors in the paper were scientifically insignificant. Baltimore doubted that if any other scientific paper was so harshly scrutinized it would be found errorless.

Recently, we have ascertained that the study in contention was rather complicated and no one other than Imanishi-Kari has really attempted to repeat these experiments and confirm or refute the results published in the paper. In addition, the theory which this study aimed at supporting, has fallen out of fashion. There seems to be some evidence for its plausibility but there do not seem to be many researchers willing to pursue the topic further.

So, what did really happen? Did Imanishi-Kari fabricate and falsify data or was she a victim of serious injustice? Were O'Toole's strong beliefs in the correctness of science an excuse for the fraud investigators in their endless pursuit of guilt? Was Baltimore an arrogant scientist or a heroic figure who stood behind Imanishi-Kari at all costs?

It is neither in our ability nor was it our intent to objectively answer these questions. Imanishi-Kari was exonerated because there was insufficient evidence to

prove fraud and not because there was strong evidence that she was innocent. But innocent or not, a doubt has been cast on her reputation as a scientist.

It was the goal of this thesis, however, to assess critically the statistical arguments that were put forward against Imanishi-Kari and the counterarguments to them raised before the Appeals Board. As we did not have access to the actual experimental records, this limited to a certain extent our ability to present a more comprehensive study of the case from the statistical point of view. We briefly summarize the three statistical analyses that the ORI carried out to prove fraud and the reasons why they were disputed.

First, the method of serial correlations was used to establish whether or not the data represented a true experimental record of gamma radiation counts. What took place in any particular well was assumed to be independent from what took place in its neighboring wells. Hence, if the counts were authentic, little or no serial correlation was expected to be observed. Although, theoretically, this is a sound statistical approach at evaluating independence in a data set, its application by the ORI had serious flaws. As Speed pointed out, the control data sets were not comparable to Imanishi-Kari's and, therefore, any difference observed between the critical and the control data sets was not very informative and conclusive. Furthermore, he pointed out that the fundamental assumption of independence between the observations was probably invalid in an actual experimental setting such as those being studied.

Second, different "control" data sets were fitted by Poisson mixture distribution models, and it was expected that the contentious data sets should, therefore, also fit Poisson mixture models. Yet, there are various models which may explain equally

well, any particular data set, should "a fit" be declared; there is no reason to require that the contentious data should also fit a Poisson mixture model. As importantly, the physical/biological justification that the data should follow a mixture of Poisson distributions in the first place was seriously put in doubt by Speed, who pointed out that there was neither any compelling scientific nor any empirical evidence to support that claim.

The third and last set of statistical analyses involved the distribution of insignificant digits in the count data. These analyses were based on the well known assumption that leading digits in random numerical experimental data follow the discrete uniform distribution. Unfortunately, the digits analyzed were not the leading, significant, digits in the counts. Moreover, which digits were deemed "insignificant" for the analyses, suggested serious subjectivity on the part of the ORI.

Recently, Nature Medicine devoted a special section on scientific fraud. Imanishi-Kari's case was presented as the "stuff of folklore," and one that will retain its popularity despite the fact that there have been and continue to be increasingly many allegations of scientific misconduct. (Basu, 2006)

Professor Ranjit Kumar Chandra, a prominent Canadian researcher in the field of nutrition and immunology, was recently accused of scientific fraud. He is said to have published at least 10 studies, which are based on fraudulent data and is known to have attracted substantial amounts of research funds to support them. (CBC News, 2006) As another example, Professor Hwang Woo Suk, a South Korean researcher and termed a national hero based on his stem cell research, recently admitted to

faking data. He has also been able to attract millions and millions of dollars in public and private funding to support his research. (BBC News, 2006b)

We bring Imanishi-Kari's story to an end by pointing out that, true or otherwise, with each allegation of scientific fraud there is more at stake than just the protection of the whistleblower and the reputation of the scientist involved. One of the most precious things we hold as scientists, particularly in academia, is our freedom to carry out our research. We will be allowed to keep it so long as the public continues to trust us and the work that we carry. This is exactly the reason why we, as scientists, have the greatest role to play in keeping high ethical standards in scientific research and the greatest interest in doing so if we want to see future generations of scientists follow from where we leave off.

## REFERENCES


B. Abraham and J. Ledolter. *Statistical Methods for Forecasting.* John Wiley & Sons Inc., NY, 1983.

C. Anderson. *NIH: Imanishi-Kari guilty. Nature*, 350:262–263, March 1991.

P. Basu. *Where are they now? Nature Medicine*, 12:492, May 2006.

BBC News. *Drive to cut fraudulent research.* Available from `http://news.bbc.co.uk/1/hi/health/4903184.stm`, April 2006a. Last accessed 8 June 2006.

BBC News. *Profile: Hwang Woo-suk.* Available from `http://news.bbc.co.uk/1/hi/world/asia-pacific/4554704.stm`, May 2006b. Last accessed 8 June 2006.

G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, third edition, 1994.

T. Brijs, D. Karlis, G. Swinnen, K. Vanhoof, G. Wets, and P. Manchanda. *A multivatiate Poisson mixture model for marketing applications. Statistica Neerlandica*, 58:322–348, August 2004.

CBC News. *The Secret Life of Dr. Chandra.* Available from `http://www.cbc.ca/national/news/chandra/index.html`, January 2006. Last accessed 8 June 2006.

K.W. Church and W.A. Gale. *Poisson Mixtures. Journal of Natural Language Engineering*, 1995.

W.G. Cochran. *The $\chi^2$ Test of Goodness of Fit. The Annals of Mathematical Statistics*, 23:315–345, September 1952.

A.L. De Blas, M.V. Ratnaparkhi, and J.E. Mosimann. *Estimation of the number of monoclonal hybridomas in a cell fussion experiment. Effect of post-fusion cell dilution on hybridoma survival. Journal of Immunological Methods*, 45:109–115, September 1981.

Department of Health and Human Services, Departmental Appeals Board. *DAB Website.* Available from `http://www.hhs.gov/dab/`. Last accessed 8 June 2006.

Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel. *Decision.* Available from `http://www.hhs.gov/dab/decisions/dab1582.txt`, June 1996. Last accessed 8 June 2006.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons Inc., NY, second edition, 1960.

G. Grigg. *ORI finds researcher guilty of misconduct.* Available from `http://www.hhs.gov/news/press/pre1995pres/941125.txt`, November 1994. Last accessed 8 June 2006.

T. Imanishi-Kari, D. Weaver, and D. Baltimore. *On the specificity of BET-1 antibody. Cell*, 57:515–516, May 1989.

M. Kendall, A. Stuart, and J.K. Ord. *The Advanced Theory of Statistics*, volume 3. Charles Griffin & Co. Ltd., London, fourth edition, 1983.

D.J. Kevles. *The Baltimore Case: A Trial of Politics, Science, and Character.* W. W. Norton & Company, 1998.

B.C. Martinson, M.S. Anderson, and R. de Vries. *Scientists Behaving Badly. Nature,* 435:737–738, June 2005.

G. McLauchlan and D. Peel. *Finite Mixture Models.* John Wiley & Sons Inc., NY, 2000.

National Research Council Canada. *Human Resource Manual.* Available from `http://hr-rh.nrc-cnrc.gc.ca/Documents/Manuals.nsf`. Last accessed 1 May 2006.

M.N. Newkirk. Associate Professor, Faculty of Medicine, McGill Uiveristy, Monteal, Canada, April 2006. Personal Communication.

Office of Research Integrity Model Policy. *ORI Model Policy for Responding to Allegations of Scientific Misconduct.* Available from `http://ori.dhhs.gov/documents/model_policy_responding_allegations.pdf`, Revised February 1997. Last accessed 8 June 2006.

Office of Research Integrity Model Procedures. *ORI Model Procedures for Responding to Allegations of Scientific Misconduct.* Available from `http://ori.dhhs.gov/documents/model_procedures_responding_allegations.pdf`, Revised February 1997. Last accessed 8 June 2006.

Office of Research Intergrity Report, November 1994. Case 072, Thereza Imanishi-Kari. Courtesy of Terence Speed.

E. Parzen. *Stochastic Processes.* Holden-Day, Inc, California, 1965.

Science. *Canadian Schools: Draft Misconduct Rules or Lose Funds,* December 1994.

R. Smith. *Fraud in medical research.* Available from `http://bmj.bmjjournals.com/talks/`, 2001. Last accessed 8 June 2006.

D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons Inc., NY, 1985.

D. Weaver, C. Albanese, F. Constantini, and D. Baltimore. *Retraction: Altered repertoire of endogenous immunoglobulin gene expression in transgenic mice containing a rearranged Mu heavy chain gene. Cell*, 65:536, May 1991.

D. Weaver, M.H. Reis, C. Albanese, F. Constantini, D. Baltimore, and T. Imanishi-Kari. *Altered Repertoire of Endogenous Immunoglobulin Gene Expression in Transgenic Mice Containing a Rearranged Mu Heavy Chain Gene. Cell*, 45:247–259, April 1986.

A. Willse and B. Tyler. *Poisson and multinomial mixture models for multivaiate SIMS image segmentation. Analytical Chemistry*, 74:6314–6322, December 2002.

COPYRIGHTED MATERIAL WAIVERS

**NATURE PUBLISHING GROUP LICENSE**
**TERMS AND CONDITIONS**

Oct 13, 2006

This is a License Agreement between Nora Bohossian ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

| | |
|---|---|
| License Number | 1484920465197 |
| License date | Jun 09, 2006 |
| Licensed content publisher | Nature Publishing Group |
| Licensed content publication | Nature Medicine |
| Licensed content title | Where are they now? |
| Licensed content author | Paroma Basu |
| Volume number | 12 |
| Issue number | 5 |
| Pages | 492-493 |
| Year of publication | 2006 |
| Portion used | Figures |
| Number of figures | 1 |
| Requestor type | Student |
| Type of Use | Thesis/Dissertation |
| High-res requested | No |
| Total | $0.00 |
| Terms and Conditions | |

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, **and for no other use**, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).

**NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.**

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:

   **Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)**

   For AOP papers, the credit line should read:

   **Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)**

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   **Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)**

7. Translations of up to a whole article do not require NPG approval. The translation should be credited as follows:

   **Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)**

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

---

---

# NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Oct 13, 2006

This is a License Agreement between Nora Bohossian ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

| | |
|---|---|
| License Number | 1487101077929 |
| License date | Jun 13, 2006 |
| Licensed content publisher | Nature Publishing Group |
| Licensed content publication | Nature |
| Licensed content title | Science meets forensic science |
| Licensed content author | Christopher Anderson |
| Volume number | 347 |
| Issue number | 6293 |
| Pages | 507-507 |
| Year of publication | 1990 |
| Portion used | Figures |
| Number of figures | 1 |
| Requestor type | Student |
| Type of Use | Thesis/Dissertation |
| Total | $0.00 |
| Terms and Conditions | |

## Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, **and for no other use**, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

   Where print permission has been granted for a fee, separate permission must be

obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).

**NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.**

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:

   **Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)**

   For AOP papers, the credit line should read:

   **Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)**

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   **Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)**

7. Translations of up to a whole article do not require NPG approval. The translation should be credited as follows:

   **Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)**

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

---

---

From "Permissions" <Permissions@WWNORTON.com>
Subject RE: Permission to use figures
Date Thu, June 22, 2006 3:25 pm
To "Nora Bohossian" <nora@math.mcgill.ca>

---

Dear Nora Bohossian:

Thank you for your request to use our material in your dissertation. Provided the material you wish to use is uncredited in our work to another source, this letter will grant you one time, nonexclusive rights to use the material in your dissertation, and in all copies to meet university requirements including University Microfilms edition, subject to the following conditions:

1. Such material must either be reproduced exactly as it appears in our publication, or if edited to be shown as adapted from our publication;

2. Full acknowledgment of the title, author, copyright and publisher is given;

3. You must reapply for permission if your dissertation is later published.

4. You may reproduce no more than 10% of our book in your dissertation.

Best wishes for the success of your paper.

Sincerely,

Elizabeth Clementson
Permissions Manager
WW Norton & Company, Inc
500 Fifth Avenue
New York, NY 10110

---

From: Nora Bohossian [mailto:nora@math.mcgill.ca]
Sent: Thursday, June 15, 2006 11:57 AM
To: Permissions
Subject: RE: Permission to use figures

Dear Mrs. Clementson,

Here is the list of figures that I would like to use in my thesis:

1) Figure 4-B, page 37

2) Figure 7, "Figure 1 of the Cell Paper", page 52

3) Figure 10, "Among the Seventeen Pages: the Mistyped Mouse", page 70

4) Figure 14, "Statistical Analyses of the Background Count Digits", page 220

Thank you very much for your time!

Nora

---

From: Permissions [mailto:Permissions@WWNORTON.com]
Sent: Thursday, June 15, 2006 10:40 AM
To: Nora Bohossian
Subject: RE: Permission to use figures

Nora - Please cite the figures and page numbers from The Baltimore Case that you are interested in publishing. I will research the rights and let you know if Norton is able to grant permission.

Yours,

Elizabeth Clementson
Permissions Manager
WW Norton & Company, Inc
500 Fifth Avenue
New York, NY 10110

---

From: Nora Bohossian [mailto:nora@math.mcgill.ca]
Sent: Tuesday, June 13, 2006 10:44 AM
To: Permissions
Subject: Permission to use figures

Dear Madam or Sir,

I am writing a master thesis and I would like to use some of the figures

presented in the book by Daniel J. Kevles, "The Baltimore Case."

Since the reproduction is of a scholarly intent, I believe I do not need to request a permission to use these figures. Could you please confirm that with me?

Also, I looked at the online permission request form and I noticed that it is specific to the US only. I am in Canada so would that change anything?

Thank you very much,

Nora