

**1<sup>st</sup> rejection letter from AOAS March 13, 2012**

Dear Helene Z Hill,

We appreciate your submission of "ANALYSIS OF KEY EXPERIMENTS OBTAINED IN DISCOVERY IN A QUI TAM CASE" to The Annals of Applied Statistics, but regret that it has been rejected.

Please see the Editor's commentary and the test below for the review.

I have read through your paper and have determined it is not suitable for AOAS and thus have decided to return it to you, so you can consider other options.

In a nutshell, while you present an interesting case study in what could be called forensic radiobiology, the statistical contents of your manuscript are not of sufficient interest, nor of sufficient originality, to make it successfully through our review process.

Specifically, the statistical techniques used are routine, with the possible exception of the "max prob" analysis. However, if there is originality in the max prob analysis, then it is probabilistic rather than statistical, and furthermore it is subject to technical concerns, as detailed below. The statistical inference part of the "max prob" analysis again is completely routine. The manuscript is written in the style of a life sciences journal, and parts of it, such as pages 8-11 and 17, are inaccessible and to any non-radiobiological audience.

Let me add a few thoughts you might want to consider in pursuing this work:

1. The "max prob" analysis raises concerns, as you do not state the null hypothesis under which you compute the nominal probabilities, nor do you provide theoretical or computational support for a range of claims about these probabilities. A more formal approach is critical here; let me make an attempt in developing it. As I understand, you consider three independent draws from a Poisson distribution with parameter  $\lambda$ . Let  $M_1 \leq M_2 \leq M_3$  denote their order statistics, and let  $\mu$  denote the rounded arithmetic mean of  $M_1$ ,  $M_2$  and  $M_3$ . Then you seem to conjecture that, for all  $\lambda > 0$ , and all nonnegative integers  $m_1$  and  $m_3$ , the conditional probability

$\text{Poisson}_{\lambda}(M_2 = \mu \mid M_1 = m_1, M_3 = m_3)$

is bounded above by  $1.3/(m_3 - m_1 + 1)$  if  $m_3 - m_1$  is even, and is bounded above by  $2.6/(m_3 - m_1 + 1)$  if  $m_3 - m_1$  is odd. I wonder why you cite "simulations" as evidence for these conjectures. Why don't you, instead, compute these conditional probabilities explicitly, given the values of  $m_1$  and  $m_3$  in the data, and using a plausible estimate of  $\lambda$ ? While any such procedure would be computationally intense, it may lead to a much more persuasive test, as it would use actual probabilities, rather than conjectured upper bounds.

2. In the tables, you repeatedly state a p-value of " $< 9.9 \times 10^{-10}$ ". Why not simply " $< 10^{-9}$ "?

While I am sorry to deliver this unwelcome news, I would encourage you to pursue this work along the lines indicated above, and then to submit to a

more specialized journal, most likely in radiobiology, or, if the material on pages 8-11 and 17 is cut short, to a quantitative forensic journal.

Sincerely yours,

Tilman Gneiting  
Editor, The Annals of Applied Statistics

Submission URL: <http://www.e-publications.org/ims/submission/index.php/AOAS/>

Title:

ANALYSIS OF KEY EXPERIMENTS OBTAINED IN DISCOVERY IN A QUI TAM CASE

Authors:

Helene Z Hill, Joel Pitt

Abstract:

Numerical and radiobiological data subpoenaed during Discovery in a qui tam lawsuit were analyzed. Coulter ZM cell counts and survival colony counts that comprised data in 8 publications, a successful R01 grant proposal and its renewal were reviewed. Chi-squared analysis was performed on the terminal digits of the recorded numbers, which were expected to be approximately uniformly distributed. Counts produced by one researcher in the laboratory diverged significantly from this expectation and raised questions about the data. The p-value for the goodness of fit to a uniform distribution applied to the terminal digits of 5,155 Coulter ZM particle counter values from 171 of this individual's experiments was less than  $2.2 \times 10^{-16}$ . The chi-squared p-value for the terminal digits of 3,501 of colony counts in 114 experiments was also less than  $2.2 \times 10^{-16}$ . In data produced by several other members of the same laboratory, the terminal digits in 2,759 cell counts in 99 experiments from the same Coulter ZM particle counter and 1,556 colony counts in 59 experiments were consistent with uniformity: the p-values similarly obtained were 0.12 and 0.57, respectively. In the analyzed data of the questioned researcher, the average or near-average value of triplicate colony counts appears as one of the triples at a frequency that also greatly exceeds expectations. Additionally, results of two key experiments could not be replicated in 22 attempts. Tritiated thymidine survival kinetics in the 22 experiments conform to radiobiological predictions but differ by orders of magnitude from the questioned individual's exponential survivals. This analysis underscores the importance of access to raw data that form the bases of publications, reports and grant applications in order to evaluate the correctness of the conclusions. Methods employed in this study may prove to be useful to others in screening numerical data for anomalous results.

**2<sup>nd</sup> rejection (with permission to resubmit) from AOAS 9/26/2012**

Dear Helene Z Hill and Joel Pitt,

Thank you for submitting your paper "FORENSIC STATISTICS: ANALYSIS OF UNUSUAL PATTERNS IN DATA OBTAINED IN DISCOVERY IN A QUI TAM CASE" for possible publication in Annals of Applied Statistics.

An Associate Editor and a referee have reviewed your paper. While both found promise in the paper, they had major concerns about the suitability of the

paper for fuller review at Annals of Applied Statistics. The major criticism was that it is written in a form that makes the paper difficult to evaluate, and a form that does not give your paper the best chance for success.

I have also reviewed the paper and agree with this opinion. I also found the paper challenging to evaluate initially, but it is notable that persons with specialized expertise in the topic had trouble with this as well. In the report for the paper (available via download from EJMS), there are some specific suggestions about how to make the paper more appropriate for the journal.

Thus, the decision is a 'rejection with resubmission.' This means that the paper is being rejected. However, you are welcome to submit a NEW submission that addresses the issues raised in the review. If you submit a new paper, please submit a detailed response that describes how you change the paper given the critiques, and also note this manuscript's submission number in the cover letter. Similar to any new submission, I cannot predict whether your paper will ultimately be successful at Annals of Applied Statistics; however, following the referee's guidance will improve your chances.

In addition to these comments you may also find a review report posted on EJMS.

Thank you for considering Annals of Applied Statistics as a venue for your work.

I wish you success in finding a suitable outlet for publication.

Sincerely,  
Susan M Paddock  
Editor, The Annals of Applied Statistics

Submission URL: <http://www.e-publications.org/ims/submission/index.php/AOAS/>

Title:

FORENSIC STATISTICS: ANALYSIS OF UNUSUAL PATTERNS IN DATA OBTAINED IN DISCOVERY IN A QUI TAM CASE

Authors:

Joel H Pitt, Helene Z Hill

Abstract:

Radiobiological data subpoenaed during Discovery in a qui tam lawsuit were reviewed employing statistical tests used to identify anomalous data. Conventional statistical tests related to the relative frequencies of insignificant terminal digits as well as newly developed tests for anomalous patterns in the triplicate data commonly produced in radiobiological research were applied to data sets produced by a defendant in the case and to those of 10 other investigators from the same and other laboratories whose data served as controls. When applied to controls the results of the various tests were in all cases consistent with the null hypotheses suggested by our probabilistic models while results for data from the one questioned investigator resulted in rejection of these null hypotheses with p-values well below 0.001. This analysis underscores the importance of access to raw data that form the bases of publications, reports and grant applications in order to evaluate the correctness of the conclusions. Methods employed in

this study may prove to be useful to others in screening numerical data for anomalous results.

Here is what the referee had to say (it is not clear what is his/her native language)

Here's the conclusion of the paper

Bishayee's results challenge radiobiological expectations, and statistical analyses indicate that his numerical results diverge significantly from expectations of uniformity or randomness and are extremely unlikely to have resulted from chance alone. Because the experiments that he performed form the background of 8 publications, we suggest that it would be appropriate for Howell, the Principal Investigator, and Bishayee to retract these papers.

---

In my opinion this paper is not ready to be refereed. There are various inconsistent approaches to bibliography, trivial R functions are described and excel computations are reported, the supporting material is at a plaintiff's personal web site, remarks on reproducible research discipline are dated and anodyne.

They may have submitted this to AOAS in light of some parallels to Baggerly and Coombs. But their title is completely generic, the statistical and data analytic content is elementary, and the tables are mostly very peculiar. I'd suggest that they go back to the journal, look for a brief paper that has some similarity to theirs (e.g., Gutierrez and Quintano Dec. 2011) and use it as a model for structure and content.

They should consider the problem less generally than "unusual patterns of data ... in qui tam" which has almost no specific meaning, and more generally than the Bishayee anomaly -- to contribute to AOAS, some procedure with general applicability has to be identified.

They have some formalism on gaps in count data, appearance of mean values in sequences of realizations, and digit preferences, but it is scattered and presented fairly informally. They should find an AOAS paper to emulate, and use it produce a much briefer account of the statistical problem and solution than we have here.

We asked for further guidance and Dr Paddock responded

10/12/12

Dear Helene,

I apologize for the delay in responding. Let me provide you with some guidance that I give to many authors, and also address the major specific issue in the review of your paper. While I would be glad to provide advice via conversation, I should caution you that I rely very heavily on the Associate Editor's recommendation, so my advice would be to instead stick to addressing the comments provided by the referees' report and also consider some of the more general advice I provide below.

The key issue at this point is that the paper was not clearly written – some specific comments are provided in the referee’s report. Being more clear and organized in the writing and presentation should take your paper a long way toward being acceptable for peer review by AOAS. For that, I urge you to heed the referee’s advice to examine several AoAS papers (such as the Baggerly/Coombes paper, the Gutierrez and Quintano Dec. 2011 paper mentioned by the referee, but also looking at a few others would be valuable) to get a feel for the journal’s style and find some examples to follow in terms of how you might organize your paper.

More information in terms of style and content, here is the general guidance I provide authors:

“For AOAS, we expect a paper to begin with a substantive question arising in a practical context and then introduce the data that can be used to address the question. Only then should the paper introduce new or existing methodology to help provide the answer. The bulk of an AOAS paper should be devoted not to the methodology, but rather to the analysis of the data, including comparisons with alternative methods and the method’s performance is best judged in that context and, most importantly, to the extent that it better answers the original substantive question than existing methods or approaches.”

Also, please note the four criteria provided on the journal’s website:

<http://imstat.org/aoas/referee.html>

After your paper is written in a style that more closely adheres to the journal’s format, the peer reviewers will be in a much better position to assess the scientific merit and overall fit of the paper for the journal.

I hope this helps. Please let me know if you have other questions.

Sincerely,

Susan Paddock

Editor

### **Our responses following “reject with resubmission”**

Responses to critiques of the referee:

1. Inconsistent approaches to bibliography: *we believe that the bibliography is now more relevant*
2. Trivial R functions: *while some of the R functions may be “trivial” they are a propos for answering the key question and, to our knowledge, ppoibin has not been used in the context in which we employ it here*

3. Excel computations reported: *all of our computations are now performed using R. However, as noted at the end of the paper, we can provide an Excel spreadsheet to approximate our key analyses that can be used by researchers who are unfamiliar with R*
4. Supporting material on a website: *This was done for purposes of review only. It was not intended for final accessibility. There are over 300 PDF files which would be quite a challenge to download for the review*
5. Remarks on reproducible research discipline are dated and anodyne: *we no longer discuss the lack of reproducibility and only mention it as support for our belief that the data have been fabricated*
6. Title is completely generic: *the title has been changed and we hope is now more specific*
7. Statistics and data analytic content are elementary: *true for some but the new model is not elementary. In any case, we argue that we are applying statistics – both new and old - in a context that is of increasing concern in scientific performance and publication in the world today – the problem of detecting data falsification and fabrication*
8. Tables are mostly very peculiar: *we are not exactly sure what is meant by this but, in any case, we have deleted Tables 4 and 5. There are now only 4 Tables and the former Table 6 has become Table 4. In addition, we eliminated Figure 4, the bar graphs, as it presents the same data that are in what is now Table 4. We still have 4 figures as we compare all of Bishayee's mid-ratio data to that of other investigators in what is now Figure 2. The new Figure 3 shows Bishayee's mid-ratios broken down by isotope. Figure 1 now depicts a different experiment by Bishayee that is not as blatant as the experiment shown in the earlier Figure 1*
9. Use Gutierrez & Quintano as model: *We followed their outline as closely as we could*
10. "unusual patterns of data...in qui tam" has almost no specific meaning: *This phrase appears in the title which has now been changed to **A FORENSIC APPROACH TO ANALYSIS OF DATA IN CELLULAR AND RADIATION BIOLOGY** which we believe is more specific*
11. Some procedure with general applicability has to be identified: *Our new model should be generally applicable to the types of data generated in cell survival studies such as are produced in radiobiology and drug studies. We provide a means for laboratory chiefs to affirm that data being produced by their laboratories are reliable as noted at the end referring to the availability of our spreadsheet*
12. Formalism on gaps in count data, appearances of mean values in sequences of realizations and digit preferences is scattered and presented informally: *We believe that the reorganization of the paper should take care of this*
13. Should find an AoAS paper to emulate and use it to produce a much briefer account: *We have used both Gutierrez and Baggerly as models. We estimate that Gutierrez is about 56,000 characters and Baggerly is about 86,000. Our paper falls in between at 72,000 characters*

Responses to critiques made by Susan Paddock:

1. Not clearly written: *as noted, the paper is reorganized to conform more to the organization of Gutierrez and of Baggerly*

2. Examine several AoAS papers e.g. Baggerly/Coombes and Gutierrez/Quintano and a few others for organization: *This we have done and hope that our organization is now compatible with the journal*

3. Style and content

a. begin with substantive question arising in a practical context: *We state the key question in section 1.1: **Can the unexpected patterns in Bishayee's data be due to chance alone?***

b. introduce data to address the question: *This we do in section 3: the data sets*

c. Then introduce new and/or existing methodology to provide answer: *the new model provides for new methodology i.e. the detection of the overuse of the mean or the near mean while existing methodology provides for the analysis of terminal digits and duplicate terminal digits*

d. Bulk of paper devote not to methodology but to analysis of data: *We have moved the bulk of the development of the model to an Appendix*

e. Include comparisons with alternative methods. Method's performance best judged in "that context" and to extent it better answers original substantive question than existing methods or approaches *We did approach the problem of the averages in the triples using t-tests. However, by using an explicit probability model we can calculate estimates for the probability distribution of the number of mean containing triples in a collection (and the number of triples with mid-ratios in given intervals). This allows us to attach probability bounds on the chance occurrence of unusually high numbers of mean containing triples or mid-ratios concentrated at the mean. We could not do this with a t-test.*

### **Final Rejection:**

Dear Helene Z Hill,

We appreciate your submission of "A FORENSIC APPROACH TO ANALYSIS OF DATA IN CELLULAR AND RADIATION BIOLOGY" to Annals of Applied Statistics.

An Associate Editor and a referee have carefully read your paper. While we appreciate your efforts at making the paper more accessible for the audience of Annals of Applied Statistics (AOAS), the comments from the Associate Editor and the referee are still negative. They felt that the statistical content is standard and simply not appropriate for AOAS. I have reviewed the paper as well and agree with their assessment.

This is the third time AOAS has reviewed this paper (I was not made aware of the fact that another editor handled your paper previously until just a week ago) and the assessments have all concluded that the paper is not appropriate for publication in AOAS. Given this, I must reject your paper without the possibility of resubmission. While this may be disappointing, it will give you an opportunity to target your paper to a more appropriate journal, either in the field of biology or statistics. The Associate Editor suggested *The American Statistician* as a target for your work. *Chance* or *Significance* are other publications that might be interested in this type of a paper (real case studies that provide an interesting and compelling story but that do not necessarily require the same level of methodological novelty that appears in AOAS). I only offer these as suggestions and cannot comment on the chances of success you might have with these outlets, but they might be worth considering.

However, for either journal I would recommend further tightening the writing style of the manuscript, since both reviewers commented on that once again despite the improvements thus far you have made to it.

I wish you success in finding a suitable publication outlet for your work.

Sincerely yours,  
Susan Paddock  
Editor

Submission URL: <http://www.e-publications.org/ims/submission/index.php/AOAS/>

Title:

A FORENSIC APPROACH TO ANALYSIS OF DATA IN CELLULAR AND RADIATION BIOLOGY

Authors:

Joel Pitt, Helene Z Hill

Abstract:

Raw data from drug and radiation survival studies are often in the form of cell and colony counts. Although modern techniques have been developed to automate counting processes, the fact remains that numbers that underlie results can still be manipulated. In the light of the rising awareness that data falsification and fabrication occur, and the ever increasing rate of retractions, there is a pressing need to verify numerical output. During the discovery period in a qui tam law suit, we obtained access to raw data from several hundred experiments in a single radiobiological laboratory. In processing this data we identified several distinct anomalous patterns in data sets from one researcher in that laboratory, none of which were present in data sets from ten other investigators in the same and other laboratories. Application of two conventional statistical tests and a newly developed test for anomalous patterns in the triplicate data produced in such research to various data sets reported by this researcher resulted in repeated rejection of the associated hypotheses (often at p-levels well below 0.001) that these anomalous patterns might have occurred by chance. This analysis underscores the importance of access to raw data that form the bases of publications, reports and grant applications in order to evaluate the correctness of the conclusions, as well as the utility of methods for detecting anomalous, and fabricated,



numerical results.

I wrote the following to Dr Efron, the Managing Editor

Saturday, December 8, 2012

Dr Bradley Efron  
Stanford University

Dear Dr Efron,

Our paper entitled “A forensic approach to analysis of data in cellular and radiation biology (AOAS1211-0210)” has been rejected by the Annals of Applied Statistics. This was our 3<sup>rd</sup> submission.

We chose to submit our work to your journal because of your title “Annals of *Applied Statistics*”, since our analysis applies statistics to experimental data generated in a radiation research laboratory, published in 8 articles in peer reviewed journals and designed to improve radiation dosages and exposures used in nuclear medicine. We deemed the results that we analyzed to be highly suspicious, quite likely fraudulent and we fear that radionuclide doses calculated therefrom could seriously harm patients.

Not only did the title of your journal appeal to us, it was your journal that published the landmark paper by Baggerly and Coombes entitled “Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology”. As you know, this paper deals with highly suspicious results from chemotherapy trials. We thus concluded that your journal was not afraid to venture into an area of increasing concern: that of data fabrication and scientific misconduct.

The first version of our paper, entitled “Analysis of key experiments obtained in discovery in a qui tam case (AOAS1203-020)” was submitted to AoAS in March of 2012. It was rejected. However, the Associate Editor, Tilman Gneiting, in encouraging us to pursue this work, made some very cogent suggestions, especially about the statistical derivations, which we took very seriously. Dr Pitt spent many hours developing the model for the analysis of triplicate samples. We decided to resubmit our new version to your journal. It was entitled “Forensic statistics: Analysis of unusual patterns in data obtained in discovery in a qui tam case (AOAS1209-019)” and was assigned to Dr Susan Paddock who at the time was apparently unaware that there was an earlier version of the paper. One may wonder why it was not assigned to Dr Gneiting again, but that, I suppose, is neither here nor there. This time it was found to be promising and was rejected “with resubmission”. I have appended our responses to the comments by the referee and by Dr Paddock. As you should see, we again took the critiques very seriously. Dr Paddock had noted that the major problem with the paper was the format. We revised the form of the paper using two suggested publications in your journal as guides.

We believe we have been treated very unfairly. We were encouraged by the permission to resubmit. We rewrote and reorganized the paper as instructed. The statistics are vastly more detailed and complex when compared to the first submission in March. We argue that our statistical analysis is original. We challenge you to find any other paper that puts forward a model to analyze triplicate colony samples that is in any way comparable to ours. One could hardly call this statistical content “standard”. We do use standard statistics (Mosimann analysis and equal digit analysis) as additional support.

We were (mis)led to believe that our paper failed the second time due primarily to organization. We were advised to use the paper by Gutierrez and Quintano as a guide to reorganizing our paper, and we followed this advice. Our paper was also said to be too long. We therefore shortened it by moving some of the statistical derivations to an Appendix.

. We have no way to refute any further arguments made after the third submission as they have not been made available to us.

The statistics were not changed between the second and third submissions, only presented more clearly. If the “statistical content is standard and simply not appropriate for AOAS” why then were we encouraged to resubmit? Why have your Associate Editor and the referee wasted our time?

The message in our paper is an important one. If you follow the scientific news, you well know that scientific fraud and misconduct are very much front and center these days. The Office of Research Integrity of the US Public Health Service is coming out almost weekly with new accusations of research misconduct. If you look carefully, you will see that most of these involve image manipulation, a technique that is easy to spot. Very few, if any, have to do with statistical detection of fraud. Our paper should stand out along with the Mosimann publications in this regard. Our paper is unique in that it provides any entirely new method for detecting fraud in triple colony forming replicates that are standard practice in survival studies used in cell biology to determine the efficacy of drugs and radiation. Our paper should serve as a guide for the development of statistical models when one encounters suspicious data that may not fit any of the usual molds.

We entreat you to reconsider and to accept our paper for publication in the Annals of Applied Statistics.

Sincerely yours,

Helene Z Hill, PhD

Professor of Radiology

NJ Medical School

Here is Dr Efron’s response:

Dear Dr Hill,

First of all let me plead guilty to the change in Editors. Somehow I missed that the 2nd submission was a resubmission. My main role is to send the papers to the appropriate area, I don't interfere in Editor's judgements.

That being said, I don't see much purpose in sending the paper out once again. Dr Paddock is not likely to change her previous opinion and there would just be further delay in getting your message out. I agree that you have something valuable to say to the reseach community. It doesn't seem that the Annals of Applied Statistics is going to be the vehicle. There are various reasons here, having to do with form and writing. The readers wanted a much shorter, crisper, more polished presentation, and three tries wasn't enough to reach agreement.

From my own experience as an author I can vouch that editorial review is an uncertain exercise, and that errors of both kinds are frequent. You need to try again at a new venue. "Statistics in Medicine" would be an obvious choice.

Again my apologies for what has been an arduous and frustrating process from your viewpoint.

Sincerely yours,

Bradley Efron