

Report on the Statistical Analyses in Gilman v. Schwarzenegger

Richard Berk
Department of Statistics
Department of Criminology
University of Pennsylvania

11/28/2011

1 Introduction

I was asked by Judge Lawrence Karlton to provide neutral, technical assistance to him on the statistical issues in the case of Gilman v. Schwarzenegger. In order to do that, I reviewed the following materials.

1. Plaintiff Exhibits submitted 04/06/2011
2. (Corrected Version) Plaintiff Exhibits submitted 04/06/2011
3. Plaintiffs' Pre-Hearing Brief, ECF No. 331
4. Defendants' Pre-Hearing Brief, ECF No. 332
5. Transcript of April 6, 2011 Hearing, ECF No. 347
6. Plaintiffs' Post-Hearing Brief, ECF No. 358
7. Defendants' Post-Hearing Brief, ECF No. 357
8. Ninth Circuit Opinion, ECF No. 322

The translation from legal reasoning to statistical reasoning can be difficult. Even when key concepts are broadly similar, important details will usually differ. It is difficult, therefore, to determine from the written documents alone exactly what statistical points the plaintiffs are trying to make and what statistical points the defendants are trying to refute. As a result, I have had to make some inferences that may differ from what both sides intend. I am to happy to consider revising my report if such is the case. I am also happy to expand the discussion should that be helpful. I have tried to get quickly to the primary issues.

2 Summary of Conclusions

As I explain more completely below, the plaintiffs appear to be making three related claims. The first is that if inmates with parole hearings after Proposition 9 had their hearings before Proposition 9, the “more burdensome ... punishment” caused by deferrals and their associated procedural changes would have not existed. There would have been on the average no increased “risk of prolonging [a prisoner’s] incarceration” but for Proposition 9. I conclude that although many of the changes in the hearing procedures brought about by Proposition 9 appear to be consistent with the Plaintiffs’ claims, *the empirical case from a statistical point of view is at least incomplete.*

A second claim implicitly made is that differences in the pre Proposition 9 and post Proposition 9 deferral consequences are not the result of “chance.” *This matter is not directly addressed by the plaintiffs.*

A third claim, also made implicitly, is that whatever the implications of Proposition 9 for deferrals found for the inmates studied, they apply as well to the current experiences of California inmates. *This claim too is not directly addressed by the plaintiffs.*

In summary, from the materials I reviewed a reasonable statistical evaluation on balance is “can’t tell.” Several important statistical matters are insufficiently addressed.

3 The Implicit Research Design And Causal Effects

The plaintiffs are making several causal claims. The essential claim seems to be that but for the parole hearing provisions changed by Proposition 9, the time served after a parole board deferral by the cohort of inmates with hearings between January 2009 and December 2010 would on the average have been substantially shorter. The individuals for whom causal claims are being made are those in this post Proposition 9 cohort. It is at least these individuals for whom allegations of disadvantage are being made.

In statistics, a causal effect is *defined* as a comparison of hypotheticals. Here, that means defining a causal effect as what would happen if an inmate had a parole hearing deferral before Proposition 9 compared to what would happen if that inmate had a parole hearing deferral after Proposition 9. Note that data are irrelevant. How a causal effect is defined is a conceptual matter. Such definitions are important as a guide for any attempt to *estimate* a causal effect.

For empirical estimates of a causal effect, data must be brought to bear. Now, data matter. The practical obstacle, known as “the fundamental problem of causal inference,” is that such a comparison cannot be directly undertaken. In this case, one only gets to observe the post Proposition 9 experience of the post Proposition 9 inmates. One cannot observe what would have happened to *these* inmates under the pre Proposition 9 parole hearing regime. In statistical terms, the pre Proposition 9 experience is a “counterfactual” because it cannot be observed.

The plaintiffs implicitly recognize this problem. In order to estimate the causal effect of Proposition 9, they provide surrogate sets of inmates as a stand in for the post Proposition 9 cohort. The idea is to compare the experience of the post Proposition 9 cohort to the experiences of *other* inmates who are alike the post Proposition 9 inmates except for the impact of Proposition 9. Therefore, everything depends on how alike they really are. If they differ in important ways, what one takes as the causal effect of Proposition 9 could actually be the casual effect of some other factors.

One very important surrogate group is the cohort of inmates with hearings between January 2007 and December 2008. Their outcomes are compared to those of the post Proposition 9 cohort. The plaintiffs apparently assume that the pre and post cohorts are similarly situated except for the impact of

Proposition 9. Yet, I can find no real evidence supporting that assumption — it may or may not be a reasonable claim. For example, according to the plaintiffs, the pre Proposition 9 cohort had parole granted in 6.4% of the hearings. The post Proposition 9 cohort had parole granted in 17% of hearings. One might wonder if the greater number of paroles granted for the post Proposition 9 cohort left behind more inmates whose cases were problematic. Again, this may or may not be true, but it is the kind of matter to which data could have been directed.

A second surrogate group is the 442 inmates who had the Proposition 9 deferral periods modified to the original law deferral periods. A third surrogate group is the set of cases brought to the governor between 2007 and 2010. How comparable are they to the post Proposition 9 cohort studied? No information is provided.

What the plaintiffs provide is a detailed discussion of how the parole deferral process operated before and after Proposition 9. I gather that the intent is to make a plausible argument that one should expect more time served with the changes brought about by Proposition 9. In effect, they are proposing a set of causal mechanisms by which the causal effects of Proposition 9 are manifested. This is not a purely statistical matter although it goes to the credibility of any causal claims.

I stress that I am asserting nothing one way or the other about “but for” comparability. The key issue is whether the comparability is sufficient so that the differences in outcomes cannot reasonably be explained by other factors such as the mix of inmates or others changes in the way parole cases have been handled.

4 The Role of Chance and “Statistical Significance”

There is also the matter of “chance” differences. There are several ways one can formulate the role of change in this setting. Probably the most appropriate is based the following “thought experiment.” Imagine that the set of inmates in both the pre and post Proposition 9 cohorts were assigned at random either to the “before” condition or the “after” condition. It is as if nature conducted a randomized clinical trial.

It would then be possible by chance alone for the inmates in the after

condition to have different parole board outcomes compared to those in the before condition even if Proposition 9 had no impact. Perhaps the longer time served after a deferral is simply the result of a chance shuffling placing more problematic parole cases in the post Proposition 9 group.

Of course, the two groups were not literally assigned at random. But one might use this formulation to test the hypothesis that Proposition 9 had no impact. One can proceed *as if* the inmates wound up in either the before or after condition by the equivalent of random assignment. Then, under assumption that Proposition 9 had no effect, one can compute the probability that pre-post differences as large or larger than those observed could have have been the result of the random assignment alone. If that probability were sufficiently small, one could reject the null hypothesis of no effect.

A test result in which the null hypothesis is rejected is sometimes called “statistically significant.” It does not necessarily mean that the result is substantively or legally important. A test result in which the null hypothesis is not rejected is usually taken as evidence that one can proceed as if the null hypothesis is true (although that is technically not a proper statistical interpretation). Similar reasoning might be used with the other surrogate groups.

I do not have the information needed to conduct such tests and statisticians will differ on whether such tests makes sense in this setting. I happen to be among the skeptics, but these are matters on which respected statisticians can disagree. In any case, there was no real effort to address the possible role of chance by either the plaintiffs or the defendants.

5 What is the Relevant Population of California Inmates?

The plaintiffs claim that because of certain provisions of Proposition 9, the time served by certain California inmates has increased. A key question, therefore, is which inmates? Clearly, the inmates who had hearings between January 2009 and December 2010 are included. But what about inmates who had hearings after December 2010 and beyond? This matters because if the claims being made are only for the subset actually studied, one can treat the January 2009 to December 2010 inmates the relevant population.

However, it is also likely that the plaintiffs see the difficulties they describe as carrying forward in time. Therefore, the post proposition hearings studied are meant to also represent the experiences of inmates after December 2010 as well. Now, the inmates studied are a sample from which inferences are being drawn to a population.

Ideally, when inferences are to be drawn from a sample to a population, the sample is selected at random. Such data sets are called probability samples defined so that every member of the population has a known probability of selection.

In this setting, the post Proposition 9 data set cannot be selected at random. As a fallback position, it is necessary to make the case that at least in the near future the consequences of parole deferrals for time served will be essentially the same as the consequences alleged from the post Proposition 9 data set. No such case was made by the plaintiffs perhaps because they may have thought generalizations to the near future were obvious. Such generalizations may be a plausible hypothesis, but ultimately the plausibility must be empirically demonstrated.

6 Conclusions

In settings such as these, statistical concepts and analyses should be seen as providing information that can help inform decisions of fact. The statistical concepts and analyses should not be seen as determining the decisions of fact. I have tried to provide a context in which those concepts and analyses can be understood. I am taking no position with respect to any questions of fact and certainly not with respect to any questions of law. With respect to the statistical analyses, several important statistical matters were insufficiently addressed for firm conclusions to be reached.