Quality of Operations at Private and Public Prisons:
Using Trends in Inmate Misconduct to Compare Prisons*

Scott D. Camp scamp@bop.gov

Dawn M. Daggett ddagget@bop.gov

Office of Research and Evaluation
Federal Bureau of Prisons
400 Building, Room 3002
320 First Street, NW
Washington, DC 20534

June 27, 2005

Scott D. Camp is a Senior Social Science Research Analyst with the Office of Research at the Federal Bureau of Prisons. He joined the office in 1992 after completing his Ph.D. in Sociology at The Pennsylvania State University. His current research focuses on performance measurement of prisons, inmate misconduct, staff issues in corrections, evaluating prison programs, mental health issues of inmates, and prison privatization. He is also a coauthor of the recent book *Measuring Prison Performance: Government Privatization and Accountability* (2004).

Dawn M. Daggett, M.S., Northeastern University (1995), came to the Office of Research at the Bureau of Prisons as a Research Analyst in 2000 from the Boston Police Department, where she worked for the Domestic Violence Unit. She is currently a Ph.D. student in the Criminology and Criminal Justice Department at the University of Maryland. Her primary research focuses on the mental health status of inmates, identifying and tracking skills for inmate reentry, and the medical classification of inmates. She has published on competency to stand trial from extensive prior research with inmates in forensic hospitals.

Quality of Operations at Private and Public Prisons:
Using Trends in Inmate Misconduct to Compare Prisons

ABSTRACT

A model-based approach was used to develop performance measures from inmate misconduct data to compare public and private prisons. The performance measures indicated the impact of different prisons upon raising or lowering the probability of inmate misconduct. Data for all misconduct and two categories of misconduct, violent and drug, were generated for the 36 month period between January of 1999 and December of 2001 for all prisons within the Federal Bureau of Prisons (BOP) and one low-security private prison under contract to the BOP. The private prison performed within the lower range of performance for low-security prisons within the BOP.

INTRODUCTION

The Federal Bureau of Prisons (BOP) uses private contractors to hold over 20,000 inmates in secure beds for adult offenders. Given this extensive use of private contractors, the question is whether the private prisons operate as efficiently and as safely as the BOP's own facilities. This study compares the quality of operations at one of the private prisons in comparison to similar BOP prisons to partially answer the question about safety. Healthy competition is often cited as one of the key benefits of privatization (Harding 2001), with the assumption being that competition with the private sector places pressure on the public sector. The comparisons presented here provide some feedback on the nature of that competition.

In this paper, we examine how the performance at the private prison compared to operations in the BOP, especially operations at three facilities. The three comparison facilities and the private prison were constructed and activated at approximately the same time, were built upon the same architectural design, and were designed for low-security inmates. The outcome measures used to assess quality of operations are based on official records of inmate misconduct. Public safety and institutional safety are among the highest priorities of any correctional system, and misconduct data are the most direct expressions of these safety concerns. The methods reported here demonstrate the proper types of techniques that are necessary to insure fair comparisons of prison operations. The study is unique in that it combines methodologically sophisticated analysis of operational data from an entire correctional system with the opportunity to observe the operations of four prisons that were very similar in design and mission. Other studies have compared public and private operation of otherwise similar prisons, but the analyses were not conducted with the same comprehensive methodological plan as the current study.

REVIEW OF THE LITERATURE

There is continuing disagreement about the political desirability of private prisons (Segal and Moore 2002; Shichor 1995), and much of the literature on private prisons presents a polemic for or against. Proponents of prison privatization have long argued that private prisons both operate more cheaply and with higher levels of performance than public prisons (Crants 1991). Privatization proponents are vague, however, on identifying the mechanisms that push private prisons to be cheaper and better. They refer to innovations and market pressures, but without identifying how these factors are specifically translated into differences in operations (Camp and Gaes 2001; Gaes, Camp, Nelson, and Saylor 2004). The lack of specificity is interesting because Hart, Schleifer, and Vishny (1997) note that under the theory of ownership of residual rights of control, ambiguity in a contract provides the vendor with greater opportunities to manipulate quality of service to maximize profits. This study focuses on the respective performance of private and public prisons in achieving order and safety behind bars.

The review of the literature starts with a brief examination of cost issues. Cost issues are intertwined with quality, but the main reason to examine cost issues is to set the stage for an investigation of prison quality. A more detailed investigation of the research literature on prison quality follows. The shortcomings in the research design of existing research are examined with a few examples. The methodological work of Camp and his colleagues is presented to demonstrate how previous pitfalls can be avoided (see the review of the literature in Gaes, Camp, Nelson, and Saylor 2004: Chapters 4 and 5). Finally, a method is proposed for taking the results from cross-sectional analyses, such as those used by Camp and his colleagues, to analyze trends over time.

Studies comparing the costs of public and private prisons have generated controversy. Proponents of prison privatization promise cost savings of 5 to 15 percent, and some research has found that these savings have been achieved (Segal and Moore 2002). Likewise, a study financed by the Association for Private Correctional and Treatment Organizations that was not covered in the Segal and Moore (2002) review found that states with private prisons spent almost 9 percent less on corrections than states without private prisons (Blumstein and Cohen 2003). Pratt and Maahs (1999) did a meta-analysis of 33 different cost evaluations and found that private prisons were no more cost-effective than public prisons. Instead, they found other factors to be more predictive of costs, factors such as the age of the facility, the economy of scale realized, and security level.

Some researchers claim that the private sector has largely missed the opportunity to realize significant cost savings in the United States because of their inability to penetrate the markets in the Midwest and Northeast where labor costs and the presence of unions are highest (Austin and Coventry 2003). More to the point, researchers who examined the existing cost studies noted that the results of cost comparisons were specific to the prisons examined and did not generalize to prison privatization in general (General Accounting Office 1996; McDonald, Fournier, Russell-Einhorn, and Crawford 1998; Perrone and Pratt 2003). In addition, there is controversy over the methods used in the studies (Perrone and Pratt 2003; Thomas, Gookin, Keating, Whitener, Williams, Crane, and Broom 1996), especially the handling of certain costs in the comparisons, most particularly overhead costs (Nelson 1998). The most general problem is that overhead costs of the contracting agencies are not typically separated into those costs that are avoided by contracting with the private sector and those overhead costs that were incurred (unavoidable costs) regardless of whether the private sector operated the prison or not.

5

Just as there is uncertainty regarding the existing cost studies, researchers have reviewed evaluative studies of the quality of prison performance at public and private prisons and arrived at different conclusions despite examining essentially the same studies. The present discussion focuses upon evaluations of prison quality in the United States. Most private prison beds are located in the U.S., and a discussion of experiences in the U.S. is most appropriate for the private prison under examination. On the one hand, Moore concluded that private prisons typically do as well as public prisons and often better on quality dimensions such as public safety, protection of staff and inmates, compliance with professional standards, quality of programs, and quality of health care and food (Moore 1998). On the other hand, researchers at the General Accounting Office and Abt Associates Inc. concluded that there was insufficient scientific evidence to determine whether private and public prisons differed in terms of their general operations (General Accounting Office 1996; McDonald, Fournier, Russell-Einhorn, and Crawford 1998). In a less systematic fashion, Austin and Coventry (2003) also reported that the evidence on the comparative quality of public and private prisons is mixed and inconclusive.

Both the GAO and Abt studies recognized that there were two limitations to existing research. First, despite the ample opportunities for investigating the performance of private prisons, there are few actual evaluations. Second, the prisons that were compared were not randomly selected from a population of prisons nor did they encompass the population of prisons. The present study is atypical in the sense that the results for quality of operations are based upon all BOP prisons, not a subset of BOP prisons. Other studies of prison quality completed to date do not use all prisons in a given correctional system nor do they randomly sample prisons, so the existing studies do not support generalizations concerning how well private prisons operate in comparison to entire correctional agencies. The present study does not

generalize to all types of prisons, but it does generalize to the performance of one private prison

in comparison to all BOP prisons.[1]

A third limitation of existing studies of quality at private prisons is the lack of

appropriate control variables in the statistical analyses. The clearest review to date of

methodological problems with existing studies was written as an appendix to a 1998 Abt report.

The researchers concluded that the research designs used in existing studies did not allow for a

determination of whether observed differences between prisons were due to differences in

management practices, differences in the types of inmates held, or differences in other factors

not related to management performance – such as the age of the prison or the architectural design

(Gaes, Camp, and Saylor 1998).

Gaes et al. critiqued an evaluation of prison operations in Louisiana (Archambeault and

Deis 1996), and the limitations of this study are common in the literature. Archambeault and

Deis evaluated three prisons that were built at the same time but were operated by different

entities. One of the prisons was operated by the state of Louisiana and the other two were

operated by private prison firms, one by Corrections Corporation of America and the other by

Wackenhut Corrections Corporation. (Wackenhut Corrections Corporation is now an

independent corporation and operates as the Geo Group.) Even though the authors recognized

that the three prisons had demonstrable differences in factors unrelated to performance,

Archambeault and Deis did not adjust for these differences in their analyses. Instead, their

research design treated the prisons as though they differed in only random ways when this was

---

[1] Despite having over 20,000 inmates in private beds, only the private prison examined here is comparable to BOP prisons with general population inmates. The other private prisons that hold inmates for the BOP hold special population inmates.

clearly not the case. In essence, then, the statistical model used by Archambeault and Deis was misspecified.

While the Archambeault and Deis study is singled out here for its importance, the criticisms apply equally well to other major evaluations comparing the quality of operations at public and private prisons. Charles Logan, who pioneered the components to examine in quality evaluations (Logan 1990), faced a similar methodological issue when he conducted an actual comparison of a private female prison in New Mexico, the old female prison formerly operated by the state, and a female prison operated by the Federal Bureau of Prisons (Logan 1992). Even though Logan documented that the characteristics of staff and inmates differed across the prisons, he did not adjust for those differences when using staff and inmate evaluation to compare prison operations. The comparisons were based upon the number of favorable evaluations for over 200 items. Since research has consistently demonstrated the importance of individual characteristics when evaluating prison conditions (Britton 1997; Saylor and Wright 1992), the validity of Logan's results were limited by his inability to statistically adjust the comparisons.

The methodological work of Camp, Gaes and their colleagues have demonstrated how survey as well as operational data can be statistically adjusted to create prison performance measures (Camp 1999; Camp and Gaes 2002; Camp, Gaes, Klein-Saffran, Daggett, and Saylor 2002; Camp, Gaes, Langan, and Saylor 2003; Camp, Gaes, and Saylor 2002; Camp, Saylor, and Harer 1997; Camp, Saylor, and Wright 1999; Gaes, Camp, Nelson, and Saylor 2004; Gaes, Camp, and Saylor 1998). Armstrong and MacKenzie (2003) used similar methods to assess differences in environmental quality at public and private juvenile correctional facilities. Camp and his colleagues claim that three key methodological points are important. First, does it make

8

sense to summarize the data about individuals into an institutional indicator? Second, do the measures need to be adjusted to account for the characteristics of the units being summarized? And, finally, are the actions or attitudes being measured influenced by the clustering of individuals within the respective organizational units, in this case prisons?

Concerning the first point, it is important to note that there is not always a direct connection between measurements collected on individuals and organizational summaries or aggregates that can be computed from the individual measurements. Consider a case where there is a connection: race. Race is a characteristic of individuals, but when summarized over an organizational unit such as a prison, it can also be an organizational measure that reflects the proportion or percentage of individuals of a given race. Even when the connection exists, as in the case of race, the meaning may be quite different for the two levels of measurement. Where race measures identification to a socially defined group for individuals, it may serve as a measure of segregation when examining racial composition of prisons or schools.

Not all individual measures have a direct connection to organizational processes. Lincoln and Zeitz (1980) examined job satisfaction in their now classic article on aggregating individual-level measures. Job satisfaction clearly measures a trait of individuals, how happy or satisfied they are with the work they do. However, Lincoln and Zeitz argued that it is not clear what it means to take the average level of job satisfaction as a measure of an organizational property. Organizations do not have job satisfaction, even if organizations may contribute to job satisfaction.

Taking the argument the next level, even if there is a theoretical or logical connection between individual measures and aggregations of those measures, it still remains an empirical question as to whether the aggregated measures express anything about the organization. The

9

question is whether the organization generates variability in the measures collected from individuals. In lay terms, it is a question of whether the same individual would provide different evaluations of the measure in question if they were located in different organizational units. For example, with inmate misconduct, the question is whether the level of inmate misconduct is pushed up or down by being located at different prisons. As we discuss below, there are statistical procedures known collectively as multilevel models that provide assessment of the organizational properties of aggregated measures, but almost no systematic evaluations of private and public prisons have employed the techniques with the exception of the work done by Camp and colleagues as noted previously.

The second point has to do with adjusting aggregate measures for exogenous, or outside, factors that influence the measure in question. Take the case of the relationship between age and inmate misconduct. Most studies of inmates find that more young inmates are likely to engage in institutional misconduct than older inmates. Assume that we are going to compare two prisons on the percentage of inmates involved in misconduct over the past six months. Unless the institutions are perfectly matched on the distribution of age in the respective inmate populations, the percentage of inmate misconduct as a measure of prison performance is confounded with differences between the prisons. The institution having a greater percentage of younger inmates is subject to an unfair disadvantage. The goal is to develop a measure that indicates the likelihood that inmates similar to one another become involved in misconduct at different prisons. Again, multilevel models permit this type of adjustment for differences in individuals as well as organizational differences, such as differences in architectural design or the proportion of inmates who are non-citizens.

The final point is that aggregate measures created from individual-level data must account for what statisticians call interdependence among observations. Observations (in our case behavioral data provided by inmates) are interdependent when a measure for one individual is affected by another individual in the study. Traditional statistical techniques assume that observations are independent of one another. For example, it is extremely unlikely that any two individuals selected for a national survey know one another. Therefore, the information provided by each individual is completely independent of the attitudes of other respondents. In the case of staff attitudes about prison or inmate misconduct, it is extremely likely that the individuals providing data *do* know each other, and there is a non-ignorable chance that their responses or behaviors are conditioned by this proximity and knowledge. Again, the point is that a statistical adjustment is needed to deal with the interdependence of observations, and while multilevel models do this, these models have not been employed. Interdependence of observations causes a loss in statistical power in determining the statistical significance of observed relationships. This increases the likelihood of rejecting the null hypothesis and concluding that an effect exists for a given variable when in reality there is no effect (Type I error). The actual number of respondents providing independent information is smaller than the sample of individuals participating in the study, usually by an unknown amount.

Multilevel models have been described very clearly in the literature (Goldstein 1995; Kreft and De Leeuw 1998; Raudenbush and Bryk 2002). Multilevel models, also known as hierarchical linear models or mixed models, allow us to simultaneously handle data on individuals and the prisons at which the study participants work or are incarcerated. As such, multilevel models can be employed to address the three issues associated with aggregated measures: are the measures affected by organizational differences, do the measures need to be

adjusted for differences among individuals providing the data, and are the measures impacted by interdependence among observations? If the three conditions exist, then we can aggregate the variables with multilevel techniques and construct appropriate measures on which to compare prisons.

Camp, Saylor and Wright (1999) presented a practical discussion of the consequences of not adjusting measures of organizational behavior with techniques such as multilevel models. For a measure of the organizational contribution to institutional commitment, some prisons that produced slightly lower evaluations of institutional commitment than would have been expected on the adjusted scores had very high scores of institutional commitment on the unadjusted measure. Had an evaluation of organizational performance been based upon the unadjusted measures, the evaluation would have concluded that one particular prison had the highest level of performance in terms of increasing workers' institutional commitment. In reality, after adjusting for the characteristics of the staff and prison that produce favorable evaluations of institutional commitment, the prison actually produced lower evaluations than would have been expected.

Camp, Gaes, Langan and Saylor (2003) used multilevel models to examine misconduct for a one month time period at the Federal Bureau of Prisons. The analysis demonstrated that it is feasible to use multilevel models to rank institutions on the amount of variance added to measures of inmate misconduct, suggesting that some factor(s), presumably management performance, contributed to the likelihood that inmates would become involved in misconduct. Empirical Bayes estimates of institution effect were used to measure the extent to which the likelihood of inmate misconduct was raised or lowered by housing inmates at particular prisons.

The findings presented by Camp et al. (2003) suggested that the performance at a private prison was not equal to that of three BOP comparison prisons. A limitation of the study, acknowledged by the authors, was that it assessed performance for only one month. In this study, we adopt the same basic methodological approach used by Camp et al (2003), e.g., analyzing inmate misconduct with multilevel models. However, instead of analyzing one month's worth of data, we examine data for 36 months, from January 1999 through December of 2001. As described below, we analyze each month's data with multilevel models, generate Empirical Bayes estimates of institutional performance for each month, and then examine the plots of the residuals over time. In this manner, we examine trends in misconduct at the respective prisons for the 36 month period.

DATA AND METHODS

Data were analyzed on all sentenced inmates incarcerated in the Bureau of Prisons between January of 1999 and December of 2001. In January of 1999, this included 87,465 inmates incarcerated at 141 different facilities, and by December of 2001 it involved 116,268 inmates at 160 facilities. The count of facilities is for separate physical facilities, although the different facilities can be part of the same prison compound. For example, penitentiaries often have prison camps associated with them, but the prison camps were analyzed separately from the main facility. Since operational data were used, the data were complete.

Performance measures were constructed from inmate misconduct data using multilevel methods. As noted, multilevel techniques allow for the simultaneous examination of individual-level and organizational-level data, and in most cases this technique is needed to provide properly adjusted measures (Heinrich and Lynn 2000). Misconduct was categorized in three different ways. First, all types of misconduct were grouped together. The multilevel models

examined whether inmates had an occurrence of any type of misconduct during each of the 36 months examined. Second, violent forms of misconduct were analyzed separately as were drug forms of misconduct. The latter types of misconduct tend to be of special concern for prison administrators. The BOP misconduct codes and the respective categories used for analysis are available upon request from the authors. The form of the model was as follows:

Level 1:
$$\log\left[\frac{prob(Y_{ij}=1)}{1-prob(Y_{ij}=1)}\right] = \beta_{0j} + \sum \beta X_{ij}$$

Level 2:
$$\beta_{0j} = \gamma_{00} + \sum \gamma W + u_{oj}$$

Since β_{0j} was the only coefficient allowed to randomly vary and modeled at the institutional level, the model is a random intercepts model. The unique value for each institution (u_{0j}), estimated with an Empirical Bayes algorithm, is the value of institutional performance and measures the deviation of the institution from the overall estimate of the intercept. In other words, the measure of institutional performance captured the unique contribution of the institution's environment on the likelihood of inmate misconduct.

The variables used at the individual level, represented by the term $\Sigma \beta X_{ij}$ in the equation, included controls for criminal history and socio-demographic characteristics of the inmates. For criminal history, there were variables indicating the classification score of the inmate, the count of prior misconduct (of the type of misconduct under examination), and the period of time the inmate was at risk of a new offense.[2] The socio-demographic variables included age, citizenship (coded as dummy variables to represent U.S., Mexican, Cuban, and other citizenship), sex,

---

[2]The classification score measures the risk that inmates pose to institutional and public safety and has been validated against prison misconduct. The first score assigned after incarceration was used in the analyses.

ethnicity (Hispanic or not), and race (coded as dummy variables to indicate black, white, and other).

At the institutional level, the intercept was modeled with variables for the average age of the inmates at the prison, the proportion of inmates who were black, the proportion of inmates who were of other race, crowding at the prison, and the security level of the prison (dummy variables for minimum security, low security, medium security, and administrative security with high security as the reference category). Other variables were available, but Camp et al. (2003) identified the variables used here as statistically significant or of theoretical concern, especially crowding. In particular, they found that characteristics of the prison staff (percent female, black, or new) were not statistically significant. It had been theorized that these characteristics would impact upon the likelihood of charging inmates with misconduct. At least as represented in analyses of misconduct convictions, there is no evidence for this position either in prior studies or the current study. In addition to the variables excluded by Camp et al. (2003), we examined and excluded from consideration aggregate variables for region of the country, admission rate, disciplinary rate, average inmate custody score, percentage of inmates with drug offenses, median sentence length, and percentage of inmates with other citizenship than U.S.

For each of the 36 months examined, models for the categorizations of misconduct were specified in the HLM software (Raudenbush, Bryk, and Congdon 2001). After estimating the equations, residual files containing the u_{0j} terms were generated to provide monthly performance measures for each type of misconduct. The results for each type of misconduct, each analyzed over the 36 months, were joined together in computer files, and the graphs for each respective type of misconduct were produced using PROC GPLOT in SAS 8.2 (SAS International 2002). In addition to plotting the actual Empirical Bayes performance measures for each month, the SAS

software was used to estimate and plot a regression line to represent the trend in misconduct for the different prisons over the entire time span. The regression model chosen to represent the trends in misconduct included a square and cube term that allowed the lines to bend and change direction. The equation is $Y_i=\beta_0+\beta_1x+\beta_2x^2+\beta_3x^3+e_i$ where i indexes each institution, x represents month, and e is the normal error term. The output of the regression model for each institution provided information about how well the regression line fit the observed data points.

Two different views of the data are presented for each type of misconduct. While somewhat redundant, the views do emphasize information about different aspects of trends in misconduct. The first view of the data presents the regression lines for the private prison and the three BOP comparison prisons. While the models were based upon all BOP prisons, presenting more than a handful of prisons in any one graph makes the results difficult to discern, and these four prisons are of most direct interest for this discussion. It is straightforward to compare the relative performance of the private prison and the three BOP comparison prisons. The second representation of the data again presents the regression line through the monthly observations which are not shown, but only for the private prison. In addition, a 95 percent confidence interval around the estimated regression is presented. This allows for an inspection of whether the different parts of the estimated trend included zero. A value of zero, which is represented by a horizontal reference line in the different graphs of misconduct, indicates the expected value for each institution. Values above the reference line indicate that the institution contributed to inmate misconduct whereas values below the reference line mean that the institution detracted from the probability that an inmate would be involved in misconduct. In other words, if the confidence interval around the regression included zero, it was not possible to conclude that the

16

private prison performed differently than would be expected given the characteristics of the inmates and the institution itself.

RESULTS

Descriptive statistics for June 2000 are presented in Table 1. Since the HLM models were conducted on data for thirty-six months in this analysis, it was impractical to present descriptive statistics for all months. Nevertheless, it is helpful to provide descriptive statistics to provide some feel for the quantitative analysis summarized in the graphs below. June 2000 was chosen as a representative month as it falls midway in the series.

The mean values in Table 1 that are less than 1 are for variables that were dummy coded. That is, a value of 1 was used to represent the characteristic named in the variable label and a zero for the reference group. So, the values represent proportions. For example, the mean value of 0.03 for the any misconduct variable meant that in June of 2000, the proportion of inmates who were convicted of any misconduct was 0.03. With some simple math, we see that 3 percent of the inmates in June 2000 had a conviction for any type of misconduct. For both violent and drug misconduct, only about 1 percent of the inmates were convicted for these types of misconduct. Most of the rest of the values in Table 1 are self-explanatory, except perhaps for the variables measured at both the individual and institutional levels. The mean age of inmates measured from their individual-level data was 36.83. However, the average of the mean ages for the 146 prisons was 37.39. The latter estimate of average age occurs because the average age computed from institutional data did not weight the prisons by the numbers of inmates incarcerated there. This can be confusing for individuals not familiar with multilevel analyses.

The first graphs generated from the multilevel models are for all types of misconduct considered together. As noted in previous research (Camp, Gaes, Langan, and Saylor 2003), all

misconduct is the most reliable form of misconduct for generating performance measures to compare prisons. The first view of all misconduct is presented in Figure 1.[3] As can be seen there, the private prison contributed to a higher probability that inmates would be involved in overall misconduct for much of the time period than the BOP comparison prisons. The private prison, indicated as P_D in the graph, started off with very favorable negative values for the first six months of the observation period, but after that initial period all of the observations are above the expected level although there is a trend downward toward the end of 2001. Prior to the downward trend, though, the private prison was one of the worst performing low-security prisons.

{Insert Figure 1 about here.}

The three BOP comparison prisons showed different performance than the private prison. The probability of misconduct at public prison C (P_C) was pushed to the lowest levels for any of the BOP comparison prisons. For public prison A (P_A) and public prison B (P_B), the performance was more modest as they performed somewhere in between. Nonetheless, as the values at these institutions hovered around zero, the institutions contributed to inmate misconduct in a manner not very different than would have been expected.

The trend noted for the private prison noted from Figure 1 is reinforced with the information contained in Figure 2. The confidence interval presented for the regression line for the private prison in Figure 2 bolsters the idea that performance at the private prison was less than desirable for most of the observation period. Performance was favorable early on as noted previously, but the regression line moved above the zero reference line early on and remained

---

3 The results presented here for the four comparison prison fall within the range of results for all BOP low-security prisons. This point holds for all of the types of misconduct examined in this report.

18

there. The confidence interval shows that until the very end of 2001, the performance was significantly worse than expected for most of the observed time period. Clearly, though, performance was moving in a favorable direction (toward 0 in the graph) by the end of 2001.

It is also possible to use the results of the regression models to assess the fit of the regression lines to the observed data. For the private prison, the regression of the monthly performance measures against time as independent variables (with squared and cubed terms for the time unit to allow the curve to change direction) is statistically significant (F=11.12 with 2 and 32 degrees of freedom, p<.0001), as suggested by the fairly tight confidence intervals around the line. The adjusted R^2 is 0.46 (see Table 2). For the private prison, at least, the fit measures suggest that we can trust the regression line for displaying the trend. This was not the case for public prisons A and B. For these two prisons, the F value for the model was not statistically significant (see Table 2). This suggests that there was neither a trend to the data or a statistically significant departure from the expected value of zero for the intercept of the model. In other words, it was not possible to conclude that these two prisons changed inmate misconduct in any fashion over the observed time period. For public prison C, a significant departure was noted. As already discussed, inmates at public prison C were less likely to be involved in misconduct. A graph with the confidence interval would show that, but because of space limitations such a graph is not presented here.

{Insert Figure 2 about here.}

Information on violent forms of misconduct is presented in Figures 3 and 4. The information provided in these graphs suggests that whatever the problems at the private prison that contributed to negative performance ratings for all types of misconduct considered together, the problems were not related to violent forms of misconduct. As can be observed in Figure 3,

19

being an inmate at the private prison seemed to reduce the probability of violent misconduct during the observation period. In fact, the private prison was the best performing prison for most of this period. The performance at public prison B (P_B) and public prison C (P_C) was also favorable for most of the observation period, as inmates at these institutions had lower probabilities of violent misconduct, although public prison B started the observation period with higher contributions to violent misconduct. Performance at public prison A, though, was in marked contrast. For almost the entire observation period, inmates at public prison A were more likely to be involved in violent misconduct than expected, although this negative performance trailed off at the end of 2001.

{Insert Figure 3 about here.}

Figure 4 provides more information about the performance of private prison D. The confidence intervals demonstrate that there was greater month to month dispersion of measures of violent misconduct as the intervals are wider, indicating less precise fit of the model. Where the information in Figure 3 suggested that inmates at the private prison were less likely to be involved in misconduct, the information in Figure 4 shows that statistically the performance did not differ from the expected value as zero falls within the upper confidence interval for almost every month. In fact, although the needed graph is not presented here, the performance at the private prison and public prison A only differed in a statistically significant manner for a couple of the months during the observation period when the confidence intervals around the estimated regression lines are considered for both institutions. This emphasizes both the need to have reliable measures and to consider reliability when making comparisons. Otherwise, erroneous conclusions can be reached.

{Insert Figure 4 about here.}

The fit statistics for monthly performance measures of violent misconduct at the private prison confirm that the measures did not track changes in performance over time (see Table 2). The overall fit of the model suggested that the trend was basically a flat line that could not be differentiated from zero in a statistically significant fashion (F=0.29 with 2 and 32 degrees of freedom, p=.8296). The explained variance was correspondingly non-significant as well and low (R^2=.02). In short, even though visual inspection of the regression line drawn through the monthly points made it look as though the likelihood of violent misconduct at the private prison was reduced, the values could not be statistically distinguished from the private prison having no impact upon the likelihood of violent misconduct. It is important to remember that the private prison is a low-security prison, and violent misconduct is relatively uncommon among low-security prisoners. The fit statistics in Table 2 also demonstrate that there was no statistically significant pattern for the violence measures at the public prisons.

Information about drug misconduct at the respective prisons is presented in Figures 5 and 6. Part of the problem experienced at the private prison with all forms of misconduct considered together was related to the impact of the private prison upon inmate drug misconduct. As can be seen in Figure 5, the private prison had the highest contribution to the probability of drug misconduct of all low-security prisons examined for most of the time period. Almost as unfavorable in terms of performance was the contribution of public prison B to drug misconduct, although the poor performance of both public prison B and the private prison did tail off dramatically at the end of 2001.[4]

---

[4] It is important to note that the downward movement for private prison D at the end of the time period is entirely due to the performance measure calculated for the final month of the observation period. If that measure is treated as an outlier and removed from the analysis, then the graph of performance for private prison D remains at levels that are much higher (worse) than the BOP comparison prisons. However, since it was not possible to discover a legitimate reason for excluding this observation, it was included in this analysis. Clearly, future analyses are needed

{Insert Figure 5 about here.}

The information in Figure 6 shows that the confidence interval around the projected regression line was again larger than observed for all forms of misconduct considered together. Again, this is a function of the lower reliability of basing performance upon only drug misconduct. Nonetheless, performance at the private prison pushed the probability of inmate drug misconduct so far above zero that for most of the time period the performance at the private prison was worse than expected. If not for the lower than expected contribution to drug misconduct in the final month of the observation period, private prison D would have remained much higher than the comparison prisons for the entire time period.

{Insert Figure 6 about here.}

The fit statistics for the model of drug misconduct at the private prison confirm that the regression line presents a valid depiction of the trend at the private prison. The overall fit statistics for the model are statistically significant (F=5.25 with 2 and 32 degrees of freedom, p=.0046). The adjusted R^2 is 0.27, suggesting a moderate fit of the line to the observed data. More to the point, the intercept of the model is positive and significant, confirming that the trend tends to stay above zero in a significant fashion. For the comparison prisons, the fit statistics suggest that there was a significant trend for all of the prisons with the exception of public prison A which tended to hover around the expected value of zero.

Analyses and graphs were produced for other forms of misconduct–accountability, security, property, and other–in the manner described by Camp et al. (2003). These forms of misconduct are no less important than those considered so far to prison administrators, but they

---

to determine if this represented a change in direction or a one-time deviation. We thank one of the anonymous reviewers for this suggestion.

22

are not as often the focus of scrutiny by those outside of corrections. As such, the results for the other categorizations of misconduct were not presented here to conserve space, but the graphs are available upon request from the authors. Generally speaking, the results presented in these graphs demonstrate that the private prison did worse than the comparison prisons for these types of misconduct, with the exception of security misconduct where the performance of the private prison was very good. For accountability, property, and other types of misconduct, the performance of the private prison was worse than expected for most of the observation period.

In sum, the performance at the private prison as measured by the multilevel models of individual-level data was generally less favorable than the performance of the BOP comparison prisons. The private prison did perform well on two categories of misconduct, violent misconduct and security misconduct, but this favorable performance was not sufficient to offset the poor performance in other areas of misconduct, as captured by the overall category of misconduct.

Note on Interpretation of Residuals

The substantive interpretation of the Empirical Bayes residuals was not emphasized in the discussion other than to note how the residuals from multilevel models could be used to discuss relative performance. The concern with relative performance springs from the main methodological goals of this study, which were to develop methods for assessing the relative performance of prisons and to then look at one specific prison, a private prison, in comparison to three public prisons that were similar in ways not captured in the statistical models. Nonetheless, the monthly performance measures, the Empirical Bayes residuals, do have a substantive meaning as well. For example, in Figure 3, many of the monthly observations for the private prison hovered around a value of 0.35 for many of the middle months. By working with the

exponential of any value of the Empirical Bayes residuals, it is possible to assess the impact

upon the log odds of the event occurring as follows. $\Delta odds = 100 * (e^{u_{oj}} - 1)$

So for example, if a monthly value was exactly 0.35 at the private prison, then the odds

of inmate misconduct at the private prison were almost 42 percent higher than the odds for other

inmates for any form of misconduct, all other things being held constant.

CONCLUSIONS

This study was designed to demonstrate the viability of an untested method for

comparing the relative performance of prisons. The specific context of this study was to compare

the performance of a private prison to similar public prisons, but the method is valid for any type

of prison comparison. No attempt was made to establish baseline levels (or benchmarks) of

acceptable performance. This is an entirely separate issue, and it generally falls under the

purview of agency-conducted program reviews or prison audits under the auspices of

professional accrediting bodies.[5]

The results presented for all misconduct, violent misconduct, and drug misconduct come

from sophisticated statistical models that involve the estimation of Empirical Bayes residuals as

measures of prison performance. Most prison administrators and policy-makers dealing with

correctional issues lack the patience or even the need to parse through this last sentence, much

less the actual methods implied. For whatever historical reasons, prison administrators are often

reluctant in trusting the technology behind statistical measures, whereas they are more accepting

of the technology behind wireless computer networks or infrared motion detectors. Nonetheless,

there is a real need to produce rigorous performance measures, whether in response to internal

government initiatives such as the National Partnership for Reinventing Government under the previous presidential administration or the need to produce rigorous performance measures for public and private entities that can withstand the scrutiny of all parties with competing vested interest.

The technique of presenting complex results in graphical form screens the end users from the technical details used to generate the results, much in the manner that a speedometer in a car alleviates the need to record the number of wheel rotations per unit of time and then calculate speed. While the graphs used here were for the Empirical Bayes estimates, the graphs could be easily adapted to present changes in odds of misconduct for the more discerning end users. The comparative conclusions are completely unchanged with either approach as they are simply a re-expression of the same information. In the sense of developing a new method for making relative comparisons, the method presented here takes monthly performance measures from multilevel models and then smooths the monthly fluctuations to reveal trends. The end product is readily understood by non-technical audiences after some coaching.

The second objective of the current study was to present the new methods within the context of comparing the relative performance of one private prison with the performance of three public prisons. This was done by extracting the results derived for these four prisons from the models for all BOP prisons. This objective raises the larger issue of how the methods presented here are used in actual practice. First, it is important to recognize that results on misconduct are only one piece of information from a larger palate of data. The results are important as they are derived from rigorous and replicable procedures, but they do not present an entire representation of the performance of the respective prisons. For example, no information

---

[5] The four prisons examined here, for example, are accredited by the American Correctional Association.

has been provided here on such crucial areas as cost, provision of health care, or programs targeting inmate re-entry. Similar methods that adjust outcomes for factors not related to performance need to be developed for these areas as well. With the exception of costs, the methods would probably greatly resemble those used for this analysis of misconduct. The advantage of using misconduct as an outcome is that almost all prison systems keep meticulous records on it. Likewise, information on events too rare for normal statistical analysis but substantively important were not presented. For example, one of the prisons examined here had two escapes from inside the secure perimeter and a disturbance involving 1,000 inmates. Clearly, this information is also necessary for informed decision making.

The results demonstrated that the private prison did not perform as well as the three comparison prisons in the pubic sector, on the whole. For certain measures, the performance of the private prison was exemplary, as was noted for violent misconduct and security-related misconduct. For the other forms of misconduct, the results were less favorable for the private prison. This use of the results is relatively straightforward and non-controversial. It is simply a summarization of the trends noted in the analyses.

There are much harder, and correspondingly more important, questions that can be addressed with the results presented here. Did the results presented rise to the level where action was needed on the part of public-sector administrators overseeing the contract for the private prison, or did the results suggest that private-sector administrators running the facility needed to evaluate their procedures? This study is ill-equipped to answer these comparative types of questions. Likewise, the study is not prepared to deal with issues internal to the operation of one prison such as the substantive significance of changes where poor performance in earlier periods gets superceded by good performance in later periods. Do prison administrators in such cases get

rewarded for the good performance or chastised for the poor? All of these issues are operational

matters, and they are generally made at the highest levels of the agency. Generally speaking,

applied researchers are not involved directly in operations. Instead, applied researchers work

within a predefined framework to present the information to those administrators and policy-

makers who make judgment calls about operational issues. This is not to excuse researchers

entirely as they are certainly responsible for insuring that administrators and policy-makers

arrive at conclusions that are consistent with the data. For the purposes of this paper, decisions

about success or failure of whether administrative intervention is required are not directly

relevant to the demonstration of the methods within the context of measuring performance at

three pubic prisons and one private prison.

There are always caveats presented about the limits to the usefulness of the results

presented in an analysis. Studies generally provide a list of potential shortcomings that is similar

to the list on side-effects that are distributed with prescription medications. The current study is

no exception. First, the results presented here do not generalize to prisons outside of the BOP,

although the methods do generalize. Second, the four prisons that provided the primary base of

comparison were very similar to one another with respect to architectural design, size, age, and

other factors not captured in the models. The comparison prisons were even similar in terms of

the money spent per inmate, and this insured that the federal prisons did not have an advantage

in terms of spending higher amounts on factors related to controlling inmate misconduct, e.g.,

higher staffing levels. This advantage of comparability on unmeasured factors is probably not

often achieved. Third, the comparisons were done within a prison system that is national in

nature. Most prior evaluations of private prisons in the United States have focused upon prisons

in single states in the southern part of the country, although some of the evaluations have

covered states in the western regions. This means that the results of the analyses may have been influenced by regional factors not captured in the models. Regional differences have not shown up in other analyses of BOP misconduct data (Camp et al., 2003), but the effects may be more subtle than are captured in models with dummy variables for region of the country.

As a final word, the results for the analyses presented here are useful in their own right, but they are more important in demonstrating the proper methods for comparing the relative performance of prisons. With tools such as those employed here, better policy decisions are possible because the graphical display of results make the presentation of complex findings more accessible to policy makers and prison administrators and moves them out of the rarified world of academics and methodologically sophisticated applied researchers.

Figure 1

All Misconduct at Private (P_D) and BOP Comparison Prisons (P_A, P_B & P_C)
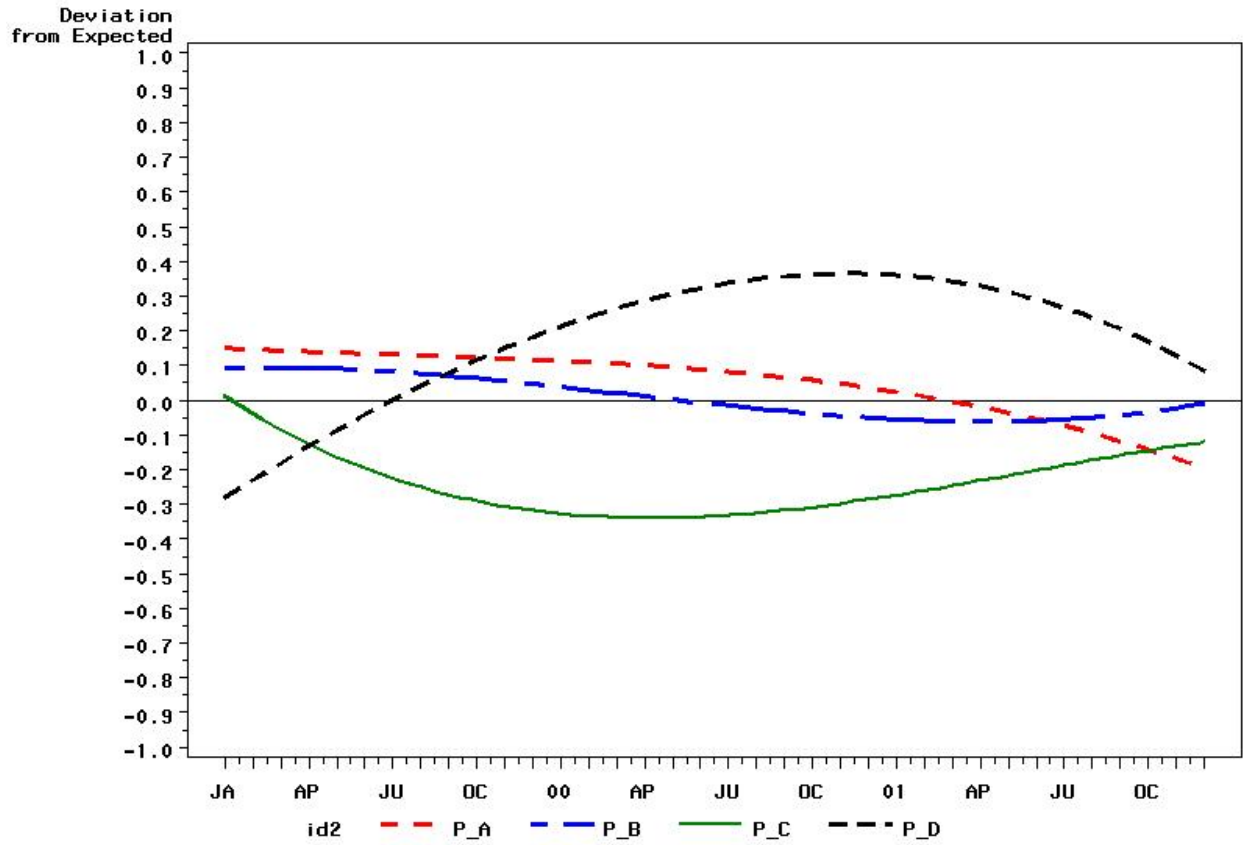Multilevel Models, January 1999 to December 2001

Figure 2

All Misconduct at the Private Prison (P_D), with 95% Confidence Interval
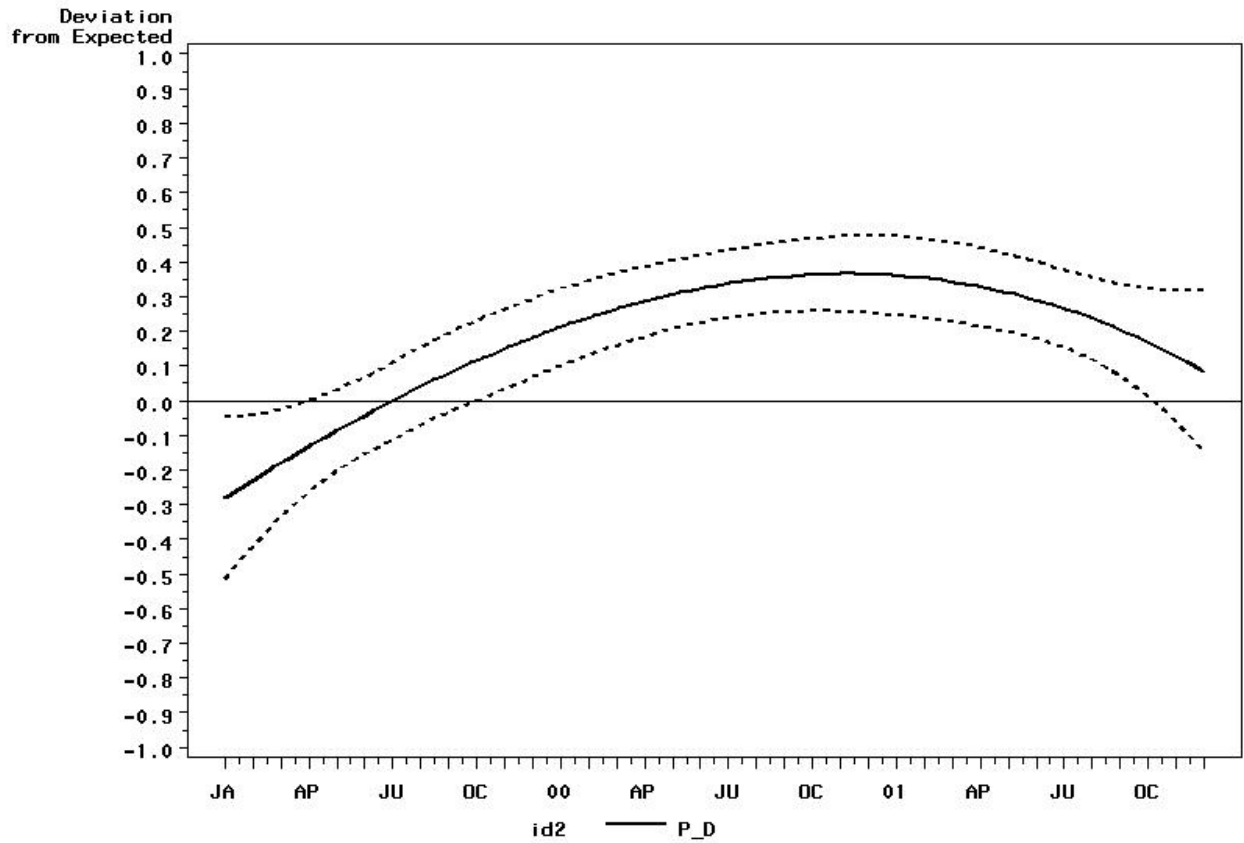
Multilevel Models, January 1999 to December 2001

30

Figure 3

Violent Misconduct at Private (P_D) and BOP Comparison Prisons (P_A, P_B & P_C)
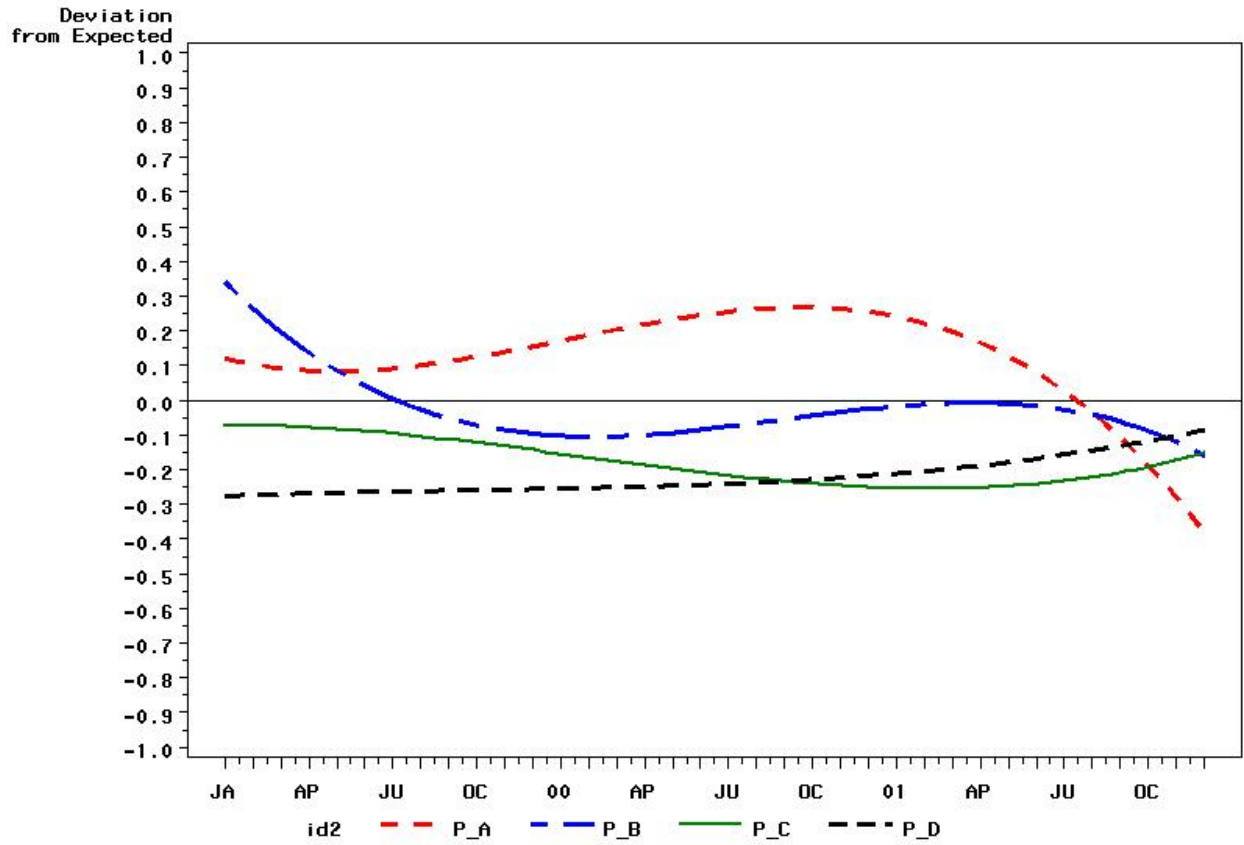Multilevel Models, January 1999 to December 2001

Figure 4

Violent Misconduct at the Private Prison (P_D), with 95% Confidence Interval
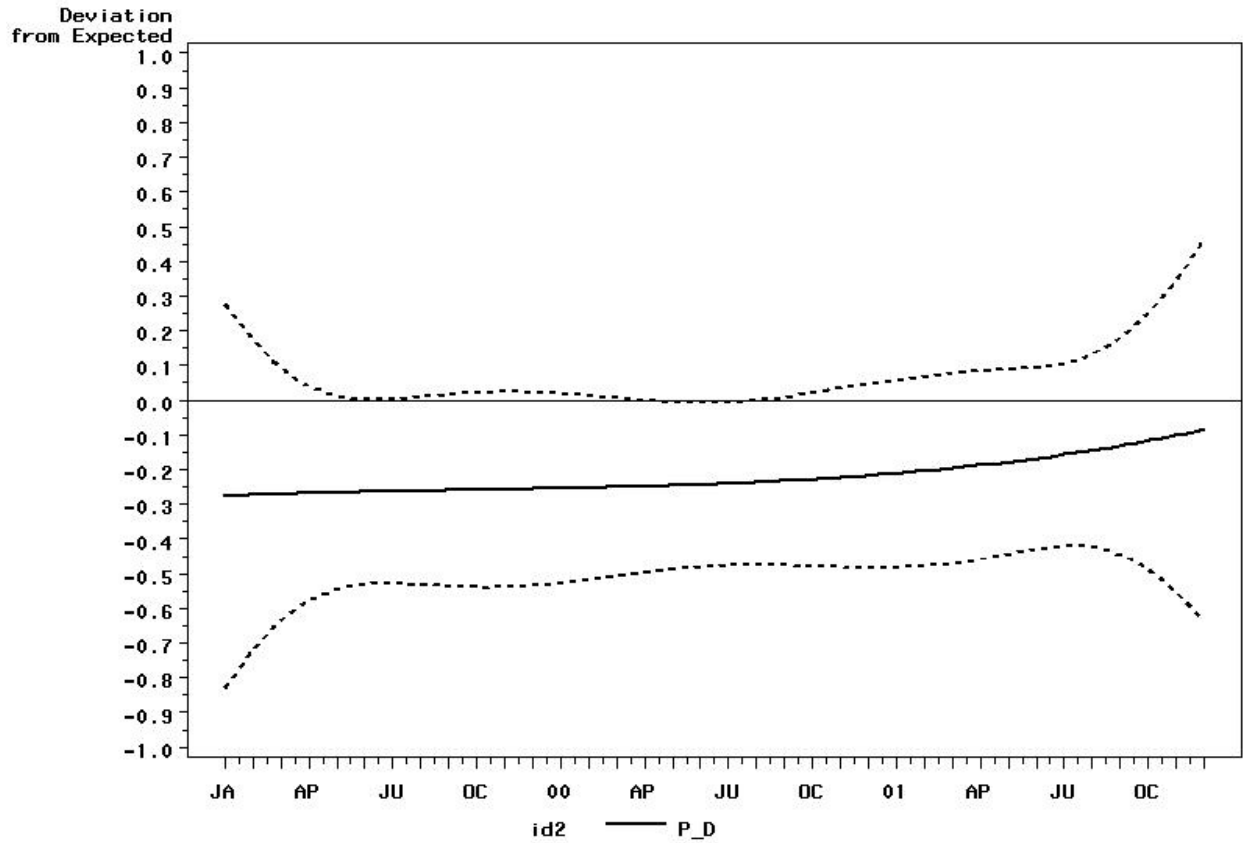Multilevel Models, January 1999 to December 2001

Figure 5

Drug Misconduct at Private (P_D) and BOP Comparison Prisons (P_A, P_B & P_C)
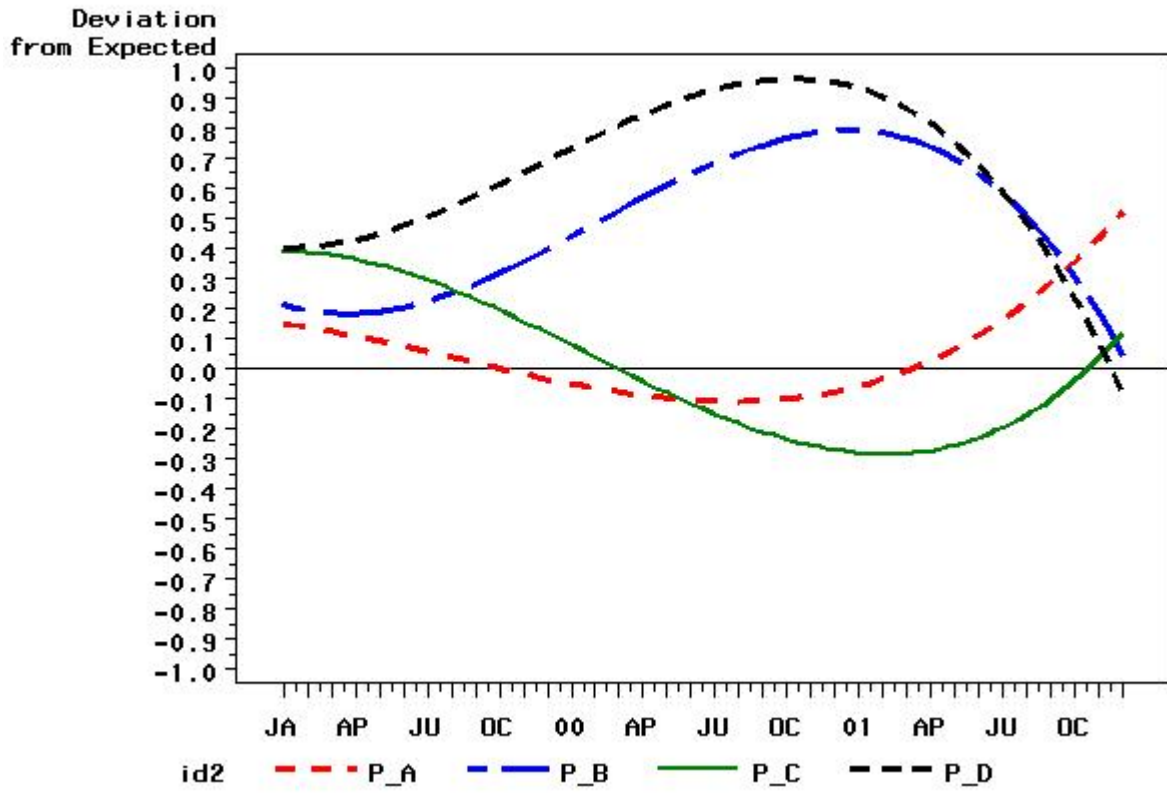Multilevel Models, January 1999 to December 2001

Figure 6

Drug Misconduct at the Private Prison (P_D), with 95% Confidence Interval
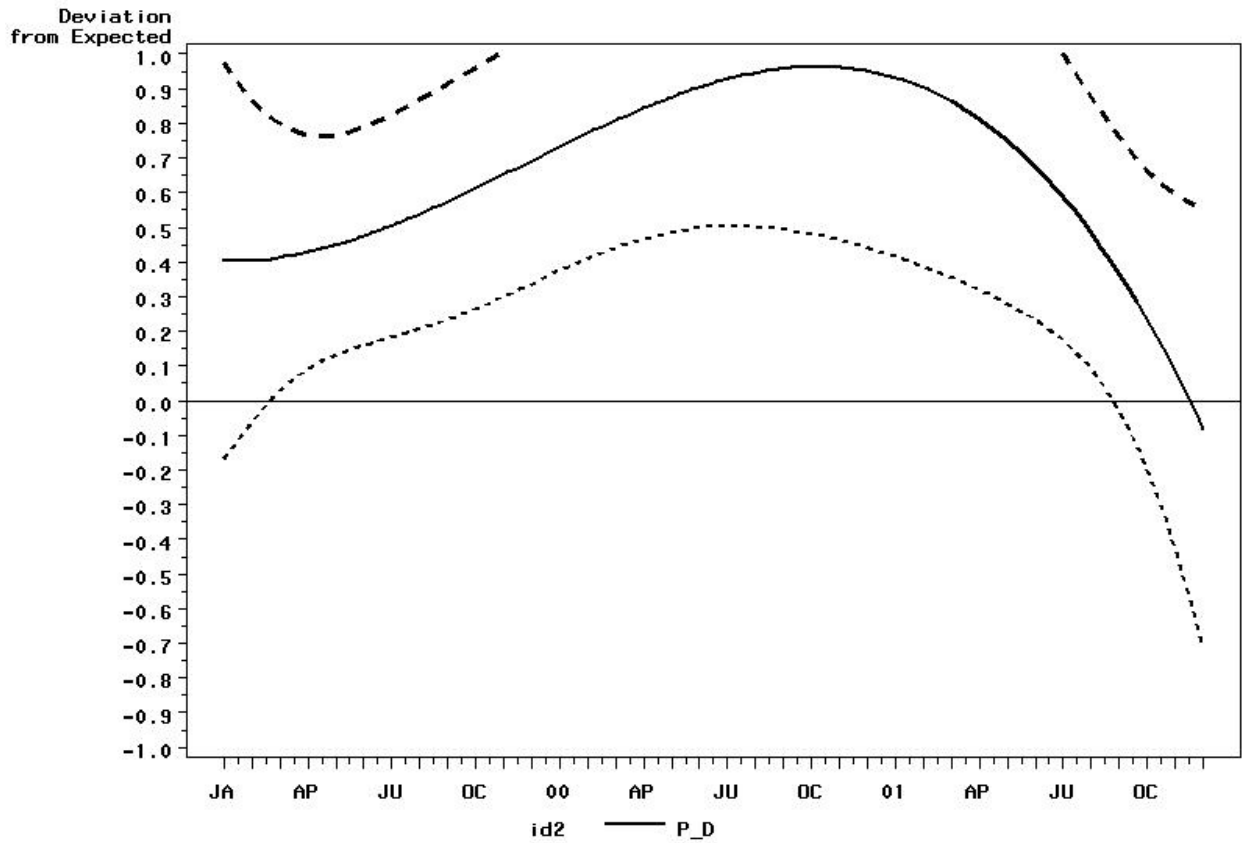Multilevel Models, January 1999 to December 2001

Table 1
Selected Descriptive Statistics for June 2000

Variable Name	N	Mean	Standard Deviation
Individual-Level Variables			
Any Misconduct	114,835	0.03	0.18
Violent Misconduct	114,835	0.01	0.07
Drug Misconduct	114,835	0.01	0.07
Age	114,835	36.83	10.44
Female	114,835	0.08	0.27
Black	114,835	0.40	0.49
Other	114,835	0.03	0.17
Hispanic	114,835	0.30	0.46
Mexican Citizen	114,835	0.14	0.34
Cuban Citizen	114,835	0.02	0.15
Other Citizen	114,835	0.11	0.32
Institutional Variables			
Average Age	146	37.39	2.24
Percent Black	146	39.81	16.58
Percent Other	146	3.10	3.66
Percent Beds Used	146	128.64	43.48
Minimum-Security	146	0.42	0.49
Low-Security	146	0.19	0.40
Medium-Security	146	0.21	0.41
Administrative-Security	146	0.10	0.30

Table 2

Fit Statistics for the Regression Models Used in Figures 1, 2, and 3

Prison	R^2	Adjusted R^2	F	Probability > F
Figure 1 – All Misconduct				
P_A (Public A)	0.1313	0.0500	1.61	0.2054
P_B (Public B)	0.0893	0.0039	1.05	0.3856
P_C (Public C)	0.2989	0.2332	4.55	0.0092
P_D (Private D)	0.5105	0.4646	11.12	0.0000*
Figure 1 – Violent Misconduct				
P_A (Public A)	0.1744	0.0970	2.25	0.1011
P_B (Public B)	0.1233	0.0411	1.50	0.2331
P_C (Public C)	0.0442	0.0000	0.49	0.6891
P_D (Private D)	0.0268	0.0000	0.29	0.8296
Figure 1 – Drug Misconduct				
P_A (Public A)	0.0586	0.0000	0.66	0.5803
P_B (Public B)	0.2695	0.2010	3.94	0.0169
P_C (Public C)	0.2801	0.2126	4.15	0.0136
P_D (Private D)	0.3300	0.2672	5.25	0.0046

* 0.0000 indicates a probability less than 0.0001, the limit of the statistical software used.

REFERENCES

Archambeault, William G. and Donald R. Deis. 1996. "Cost Effectiveness Comparisons of
Private Versus Public Prisons in Louisiana: A Comprehensive Analysis of Allen,
Avoyelles, and Winn Correctional Centers." Louisiana State University, Baton Rouge.

Armstrong, Gaylene Styve and Doris Layton MacKenzie. 2003. "Private versus Public Juvenile
Correctional Facilities: Do Differences in Environmental Quality Exist." *Crime &
Delinquency* 49:542-563.

Austin, James and Garry Coventry. 2003. "A Second Look at the Private Prison Debate." *The
Criminologist* 28:1, 3-9.

Blumstein, James F. and Mark A. Cohen. 2003. "The Interrelationship Between Public and
Private Prisons: Does the Existence of Prisoners Under Private Management Affect the
Rate of Growth in Expenditures on Prisoners Under Public Management." Corrections
Corporation of American and Association for Private Correctional and Treatment
Organizations, Alexandria, VA.

Britton, Dana M. 1997. "Perceptions of the Work Environment Among Correctional Officers: Do
Race and Sex Matter?" *Criminology* 35:85.

Camp, Scott D. 1999. "Do Inmate Survey Data Reflect Prison Conditions? Using Surveys to
Assess Prison Conditions of Confinement." *Prison Journal* 79:250-268.

Camp, Scott D. and Gerald G. Gaes. 2001. "Private Adult Prisons: What Do We Really Know
and Why Don't We Know More?" Pp. 283-298 in *Privatization of Criminal Justice: Past,
Present and Future*, edited by D. Shichor and M. Gilbert. Cincinnati, OH: Anderson.

—. 2002. "Growth and Quality of U.S. Private Prisons: Evidence from a National Survey."
Criminology and Public Policy 1:427-449.

Camp, Scott D., Gerald G. Gaes, Jody Klein-Saffran, Dawn M. Daggett, and William G. Saylor. 2002. "Using Inmate Survey Data in Assessing Prison Performance: A Case Study Comparing Private and Public Prisons." *Criminal Justice Review* 27:26-51.

Camp, Scott D., Gerald G. Gaes, Neal P. Langan, and William G. Saylor. 2003. "The Influence of Prisons on Inmate Misconduct: A Multilevel Investigation." *Justice Quarterly* 20:501-533.

Camp, Scott D., Gerald G. Gaes, and William G. Saylor. 2002. "Quality of Prison Operations in the Federal Sector: A Comparison with a Private Prison." *Punishment & Society* 4:27-53.

Camp, Scott D., William G. Saylor, and Miles D. Harer. 1997. "Aggregating Individual-Level Evaluations of the Organizational Social Climate: A Multilevel Investigation of the Work Environment at the Federal Bureau of Prisons." *Justice Quarterly* 14:739-761.

Camp, Scott D., William G. Saylor, and Kevin N. Wright. 1999. "Creating Performance Measures from Survey Data: A Practical Discussion." *Corrections Management Quarterly* 3:71-80.

Crants, Doctor R., III. 1991. "Private Prison Management: A Study in Economic Efficiency." *Journal of Contemporary Criminal Justice* 7:49-59.

Gaes, Gerald G., Scott D. Camp, Julianne Nelson, and William G. Saylor. 2004. *Measuring Prison Performance: Government Privatization and Accountability*. Lanham, MD: Alta Mira Press.

Gaes, Gerald G., Scott D. Camp, and William G. Saylor. 1998. "Appendix 2: Comparing the Quality of Publicly and Privately Operated Prisons: A Review." Pp. 1-38 in *Private Prisons in the United States: An Assessment of Current Practice*, edited by D. McDonald, E. Fournier, M. Russell-Einhorn, and S. Crawford. Boston, MA: Abt Associates Inc.

General Accounting Office. 1996. *Private and Public Prisons: Studies Comparing Operational Costs and/or Quality of Service*. Washington, D.C.: U.S. General Accounting Office.

Goldstein, Harry. 1995. *Multilevel Statistical Models*. New York: John Wiley.

Harding, Richard W. 2001. "Private Prisons." Pp. 265-346 in *Crime and Justice: A Review of Research*, vol. 28, edited by M. Tonry. Chicago: University of Chicago Press.

Hart, O., A. Schleifer, and R. W. Vishny. 1997. "The Proper Scope of Government: Theory and an Application to Prisons." *The Quarterly Journal of Economics* 112:1127-1161.

Heinrich, Carolyn J. and Laurence E. Lynn. 2000. "Means and Ends: A Comparative Study of Empirical Methods for Investigating Governance and Performance." *Journal of Public Administration and Theory* 11:109-138.

Kreft, Ita and Jan De Leeuw. 1998. *Introducing Multilevel Models*. Thousand Oaks: Sage Publications.

Lincoln, James R. and Gerald Zeitz. 1980. "Organizational Properties from Aggregate Data: Separating Individual and Structural Effects." *American Sociological Review* 45:391-408.

Logan, Charles H. 1990. *Private Prisons: Cons and Pros*. New York: Oxford University Press.

—. 1992. "Well Kept: Comparing Quality of Confinement in Private and Public Prisons." *The Journal of Criminal Law and Criminology* 83:577-613.

McDonald, Douglas C., Elizabeth Fournier, Malcolm Russell-Einhorn, and Stephen Crawford. 1998. *Private Prisons in the United States: An Assessment of Current Practice*. Boston, MA: Abt Associates Inc.

Moore, Adrian T. 1998. "Private Prisons: Quality Corrections at Lower Cost." Reason Public Policy Institute, Los Angeles.

Nelson, Julianne. 1998. "Appendix 1: Comparing Public and Private Prison Costs." Pp. 1-34 in *Private Prisons in the United States: An Assessment of Current Practice*, edited by D. McDonald, E. Fournier, M. Russell-Einhorn, and S. Crawford. Boston, MA: Abt Associates Inc.

Perrone, Dina and Travis C. Pratt. 2003. "Comparing the Quality of Confinement and Cost-Effectiveness of Public versus Private Prisons: What We Know, Whey We Do Not Know More, and Where to Go from Here." *The Prison Journal* 83:301-322.

Pratt, Travis C. and Jeff Maahs. 1999. "Are Private Prisons More Cost-Effective Than Public Prisons? A Meta-Analysis of Evaluation Research Studies." *Crime & Delinquency* 45:358-371.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

Raudenbush, Stephen W., Anthony S. Bryk, and Richard T. Congdon. 2001. "HLM for Windows 5.0: Hierarchical Linear and Nonlinear Modeling." Lincolnwood, IL: Scientific Software International.

SAS International. 2002. "SAS 8.02." Cary, NC: SAS International.

Saylor, William G. and Kevin N. Wright. 1992. "Status, Longevity, and Perceptions of the Work Environment among Federal Prison Employees." *Journal of Offender Rehabilitation* 17:133-160.

Segal, Geoffrey F. and Adrian T. Moore. 2002. "Weighing the Watchmen: Evaluating the Costs and Benefits of Outsourcing Correctional Services." Reason Public Policy Institute, Los Angeles.

Shichor, David. 1995. *Punishment for Profit: Private Prisons/Public Concerns*. Thousand Oaks, CA: Sage.

Thomas, Robert C., Kathy Gookin, Beth Keating, Valerie Whitener, Robert M. Williams, Richard Crane, and Cheryle Broom. 1996. "Department of Corrections Privatization Feasibility Study." Legislative Budget Committee, State of Washington.