Do Inmate Survey Data Reflect Prison Conditions?
Using Surveys to Assess Prison Conditions of Confinement*

Scott D. Camp, Ph.D.
Social Science Research Analyst

August 21, 2000

Office of Research and Evaluation
Federal Bureau of Prisons
400 Building, Room 3025
320 First Street, NW
Washington, D.C. 20523
202-307-3871
202-307-5888 (fax)
scamp@bop.gov

*I would like to thank Dr. Christopher A. Innes, Chief of the Statistical Reporting Section in the Office of Research and Evaluation at the Federal Bureau of Prisons, for providing the data analyzed here. I would like to thank William G. Saylor, Assistant Director in the Office of Research and Evaluation at the Federal Bureau of Prisons, for his helpful comments on an earlier version of the paper. This paper appears in *The Prison Journal* 79: 2: 250-268.

Do Inmate Survey Data Reflect Prison Conditions?
Using Surveys to Assess Prison Conditions of Confinement

Abstract

Survey data of opinions are often collected to answer questions about individuals and processes affecting individuals. This is an appropriate use of survey data as the unit of analysis matches the unit of measurement. Sometimes, though, survey data (individual-level data) are used to create group-level measures. For example, the percentage of inmates at one prison who responded that safety is a problem is compared to the respective percentage of inmates at another prison. This latter use has definitely shown up in evaluations of prison privatization. In the latter case, the unit of analysis is no longer the individual, it is the group, the respective prisons. While this shift in the unit of analysis may be justified, the theoretical and methodological properties of group-level measures created from individual-level data are rarely examined, and the shift in the unit of analysis is not always justified.

This report examines whether survey data collected from inmates can be used to create group-level measures of prison conditions. Inmates often carry a stigma that they are never to be trusted. A subset of a national survey of inmates was used to examine how inmates incarcerated in prisons operated by the Federal Bureau of Prisons answered questions about safety, noise, and job assignments at their prisons. In particular, this report demonstrates that inmate answers to the questions vary in a systematic fashion that lends credence to using survey data from inmates to obtain information about the prisons in which they are incarcerated. However, proper techniques for using survey data have not been practiced in existing evaluation studies comparing public and private prisons.

Do Inmate Survey Data Reflect Prison Conditions?

Correctional experts claim that there are clear differences between prisons that are well and poorly run. Even when there is agreement about what constitutes a well-run prison, and often there is no such agreement (for contrasting opinions see DiIulio, 1987; Wright, 1994), capturing these differences in a systematic and defensible fashion has proven to be challenging for both inside and outside corrections experts. Comparing the relative performance of prisons *within* a given prison system has always been an interest of public-sector prison administrators, but a new interest is comparing prisons under the legal authority of one prison system that are run by different entities. This, of course, is an allusion to the legal stipulations that often exist when an existing public prison system is charged with contracting the operations of a prison to the private sector. There is often language that the private-sector contract can be awarded or renewed only if costs are lower and quality of services are comparable at the private prison or alternatively if costs are comparable and services are superior at the private prison. But how do researchers or policymakers evaluate the quality of services at the respective public and private prisons? (Others can examine the cost issues.)

One well-established approach to comparing prisons (at least in the United States) is to conduct prison audits. This was the approach used in the evaluation of a private prison in Tennessee (Tennessee Select Oversight Committee on Corrections, 1995). Audits, though, tend to have several shortcomings in identifying the well from the poorly functioning prisons even when the prisons are governed by identical policy. First, audit procedures are predisposed toward being paper exercises that document adherence to policy. While it is true that a properly

functioning prison adheres to good policy, it is probably the case that adherence to policy is a necessary but not sufficient condition of operating a well-run prison. The American Correctional Association (ACA), for example, makes no claim that ACA accreditation is necessarily an indication of superior performance. Second, when the auditors attempt to generalize their findings into an overall rating, there is much room for subjectivity, and this may inhibit auditors from making critical remarks that cannot be directly supported by the procedural evidence. This may encourage "grade inflation" in the evaluations of prisons. Third, audits tend to be costly, especially in the use of human resources. This last point about cost is especially problematic when the comparison prisons are operated under different policy, which is the case at least in the federal sector where performance-based contracting is being used.[1]

A complementary approach to using audits is to survey the opinions of those most intimately involved with prison operations. While this can involve gathering feedback from local prison administrators, it can also involve soliciting the opinions of line staff and inmates. While line staff and inmates generally do not have detailed knowledge about certain aspects of prison operations, matters such as the prison budget, personnel administration, or technical details about the physical plant, there are many relevant aspects of prison operations about which staff and inmates are informed by their normal day-to-day activities. It seems reasonable to assume that line staff and inmate evaluations of these aspects, factors such as inmate safety, staff safety, quality of programs, inmate idleness, the accessibility and quality of medical care, and the quality of food operations, are partly influenced by specific practices and resources at the prison.

Managers, though, are suspicious of relying upon the evaluations of those working under them to provide feedback about their effectiveness. Wardens are no exception to this pattern, and

if wardens are suspicious of anything more than staff evaluations, it is inmate evaluations. After all, everyone wants evaluations (whether of themselves or the work setting) to be produced by trustworthy evaluators. But are staff and inmates trustworthy in providing evaluations of those things about which they have direct knowledge? Surprisingly, little research has been done on this topic, especially for data provided by inmates. The gut instinct of wardens seems to be that evaluations of prison operations provided by staff, and certainly by inmates, can be viewed as little more than individual and collective "whining." That is, a warden does not want to receive a poor overall evaluation simply because she or he had malcontents providing the evaluations or had an antagonistic union leader who influenced the evaluations provided by line staff.[2]

Of course, having an antagonistic union relationship typically says something about prison operations, but what is really needed to determine whether staff and inmate data can be used to produce prison-level measures is empirical demonstration of two points: first, that the average responses provided by staff and inmates at prisons actually differ, and, second, that the differences are independent of the individual characteristics of the inmates and staff providing the evaluations. This demonstration, believe it or not, is not so different from the problem of evaluating the ability of school managers and teachers to influence the test performance of students. When comparing average test scores for students at different schools, you want to know how much of the average test score reflects knowledge added by the school, not just whether or not the school had good (or bad) students to start with. As such, education researchers have been very active in popularizing multilevel methods for untangling the effect that schools produce on student achievement (Bryk & Raudenbush, 1992; Raudenbush & Willms, 1991), although certainly the methods can be traced in sociology to suggestions by Lincoln and Zeitz (1980) and the exposition

by Mason, Wong and Entwisle (1983). Multilevel models are becoming more common in criminological research to separate out individual and community effects on crime (Horney, Osgood, & Marshall, 1995; Rountree, Land, & Miethe, 1994).

Camp and colleagues (Camp, Saylor, & Harer, 1997; Camp, Saylor, & Wright, 1999) have researched the use of staff survey data to create prison-level measures employing multilevel techniques. Using data collected at the Federal Bureau of Prisons, they have demonstrated that the prison-level "average" for some measures (such as a job satisfaction scale) cannot be properly used because the measures are not sensitive to differences between prisons. They report other measures (such as an organizational commitment scale) where prison-level measures are appropriate, but where substantive differences arise for prison-level scores computed by different methods. In the most extreme case of incongruence, an institution that ranked 59th out of 80 federal prisons on the "naive" average level of commitment actually ranked 8th when appropriate controls were introduced.

The research by Camp and his colleagues suggests that staff survey data can be used to create some types of prison-level measures. This paper tries to extend that research by looking at whether survey data collected from inmates also demonstrate prison-level sources of variation. The data analyzed here are from the 1997 Survey of Inmates of Federal Correctional Facilities (referred to hereafter as the 1997 Inmate Survey). This survey focused primarily upon the criminal history of inmates confined in federal prisons, but there is a section of the survey on the conditions of confinement. In that section, there are several candidate questions for creating prison-level measures, questions about prison safety, noise conditions in living units, and job opportunities. Unfortunately, there were no appropriate questions on other topics where inmates could provide

information, areas such as food quality, access to medical care, etc. Nonetheless, with the items that do exist, it is possible to address the larger issue of whether inmates provide trustworthy evaluations of prison conditions, or whether inmates simply give negative evaluations regardless of the actual conditions of the prison. While it is never adequate to rely upon a single source of information, if inmate evaluations are trustworthy, then inmate evaluations are a further piece of information to use in assessing the functioning of prisons. The other pieces of information that have been used in the past to evaluate prison performance are operational data, especially official records of misconduct, audit data, and survey data collected from staff. The General Accounting Office encourages the use of as many types of data as feasible, or data triangulation, when conducting evaluations of private and public prisons (General Accounting Office, 1996).

Given the consequences of poor prison management, the more reliable information that can be obtained at reasonable cost, the better. Triangulating indicators of the quality of prison services is especially important in evaluating public and private prisons. The enormous financial stakes involved make prison privatization research especially contentious. This inquiry focuses on the viability of one select source of information, survey data provided by inmates. Survey data has been used in other evaluations of prison privatization (Hatry, Brounstein, & Levinson, 1994; Urban Institute, 1989), including survey data from both inmates and staff (Logan, 1991; Sechrest & Shichor, 1994). Logan, though, when faced with discrepancies between the information provided by staff and inmates chose to ignore the data from inmates, data which provided information contrary to the conclusions he reached. Part of the problem with Logan's use of survey data was methodological. As demonstrated below, there are three analytical steps that must be followed to use survey data appropriately(whether from inmates or staff). None of the

researchers who have produced existing evaluation studies advanced methodologically beyond the first step.

Data and Methods

The data analyzed here are for inmates at federal prisons included in the 1997 Survey of Inmates. The Bureau of Justice Statistics (BJS) coordinated inmates' interviews in the federal prison system and the state systems in 1991 and 1997. In 1997, BJS and the Federal Bureau of Prisons split the costs of surveying federal inmates, and 4,041 federal inmates agreed to be interviewed, 3,167 of whom were males (Bureau of Justice Statistics, 1997). This analysis is restricted to male inmates simply because there are more of them, and hence, more prisons that house males. Since attention here focuses on deriving prison-level measures from survey data, it makes sense to select the more numerous groups of prisons. Also, there does not seem to be any valid reason to include female prisons with male prisons. The 3,167 males included in this analysis were housed in 32 different prisons. Two of the prisons are administrative prisons, four are high-security prisons, eleven are medium-security prisons, eight are low-security prisons, and seven were minimum-security prisons.

Analyses were conducted on the data after weights were applied. The weights were necessary because of the manner in which respondents were sampled. Because of the large number of drug offenders in federal prisons, one of every three drug offenders otherwise identified by the sampling procedures was included in the sample (Waite, 1997). This procedure insured that more resources could be devoted to inmates serving sentences for offenses other than drug convictions.

The analysis examines whether male inmates, housed at the 32 separate prisons, evaluated their conditions of confinement differently and whether any of the differences were due to differences between the prisons. While there were many items in the conditions of confinement section of the Inmate Survey, most of the questions asked inmates about conditions they have faced since admission to the prison system. A much smaller subset of questions asked inmates about their experiences at or evaluations of the specific prison in which they were housed at the time of the interview. Fortunately, those few questions are some of the more interesting questions which probe inmate evaluations of prison conditions. In particular, there are two relevant questions about inmate safety, one question about noise during sleeping hours, and two questions about work opportunities.

The questions about safety are: 1) SAFE, "How safe do you feel from being hit, punched, or assaulted by other inmates?" and 2) STREET, "How safe is it here compared to the streets around where you lived?" The five possible choices for the SAFE question are (presented to an inmate in either ascending or descending order) Safe, Somewhat safe, Neither safe nor unsafe, Somewhat unsafe, and Unsafe. There were three choices for the STREET item, Here is safer, Streets are safer, and About the same. Alternatively, half of the inmates were presented the categories in a different order, About the same, Streets are safer, and Here is safer. For this analysis, all responses have been reordered to the same ordering: Streets are safer, About the same, and Here is safer. The NOISY question asked inmates, "How noisy has it been in your (appropriate type of living unit inserted on computer screen as determined by previous question) during sleeping hours?" The five choices (presented to the inmates in either ascending or descending order) were Very noisy, More than moderately noisy, Moderately noisy, Slightly

noisy, and Not at all noisy. The JOB item, Yes-No, was created from two similar questions in the survey "Do you have a work assignment OUTSIDE this prison facility for which you leave the prison grounds?" and "Do you have a work assignment HERE, either inside this facility or on the prison grounds?" An answer of yes to either question was sufficient to code the JOB item as yes. For those people with a job, follow-up questions were used to determine the number of hours (HOURS) that were worked in the last week on the job(s).

There are three steps that were necessary to determine whether inmate responses to the above questions were influenced by conditions at their respective prisons. The first analytic step was to determine whether there were any 'average' differences in the responses given by inmates at the different prisons. At this first step, it was necessary to compare only those prisons with the same overall security level designation.[3] While overall security level is only a general indicator of the security features of the prisons, the different functions of the prisons, and the security needs of the inmates housed at the prisons, security level is generally recognized as a factor differentiating how inmates are managed in prison.[4] Where there were meaningful differences between prisons in terms of the responses provided by inmates, a second step was taken.

The second step in the analysis was to determine whether any of the differences in average inmate response across prisons remained after controlling for relevant characteristics of the inmates. Socio-demographic controls were entered into the models for age of the inmate in years rounded to a tenth of a year at the time of the interview (AGE), race of the inmate coded as white, black, or other (RRACE), Hispanic ethnicity (V15), marital status coded as married, widowed, divorced, separated, never married (V27), U.S. citizenship (CITIZEN), and highest level of schooling completed recorded in years from 0 to 20 (ED_ATT). There were also some rudimentary

controls for criminal history: type of offense determining longest sentence currently being served recorded as violent offense, property offense, drug offense, public order offense, and other offense or do not know (V2023), previous incarceration (V2025), time served during current incarceration (TIMESRVD), illegal drug use (v2027) and alcohol use (v2026). There was also a control for whether the inmate reported a disability that limited the ability to work (V1316) and a control for security level of the facility in which the inmate was housed (V2043). The last control is the only control variable used in this analysis that was measured at the level of the prison instead of at the level of the inmate.[5]

The models used in the second analytic step were multilevel models, or what Bryk and Raudenbush (1992) called two-level hierarchical models. The structure of the data reflected the two-stage sampling design. Responses provided by randomly selected inmates were nested within prisons that were randomly selected from stratum defined by the region of the country in which the prisons were located and the overall security designations of the prisons. As such, it was necessary to control for this nested structure if there was reason to believe that features of the prison may affect the responses provided by inmates. As it happens, this is the very question of interest. In essence, the multilevel models can be thought of as a modified analysis of covariance that deals with the nested nature of the data and the random selection of only some prisons. As such, the variance in the responses was partitioned into components due to individual-level variability and group-level (prison) variability. In the analysis here, interest was given to determining whether group-level variability remained after controlling for potential influences on individual-level variability. Usually, attention is given to examining the factors that influence

responses at the individual level, but in this analysis these factors were treated primarily as controls.

The PROC MIXED procedure of SAS System 6.12 was used to solve the multilevel models. A good introductory discussion of using SAS for multilevel modeling is provided by Littell, Milliken, Stroup, and Wolfinger (1996). Since four of the five conditions of confinement variables were not continuous and normally distributed, the GLIMMIX macro which permits for the analysis of dichotomous dependent variables, as discussed by Littell et al., was used for these four variables: safe from attack (SAFE), streets or prison safer (STREET), noise level when sleeping (NOISE), and currently working (JOB). The four variables were transformed into dichotomies for this analysis as discussed below. Only the number of hours worked in the past week (HOURS) was analyzed as being a continuous variable.

The third step of the analysis was to demonstrate how to use the corresponding results to examine differences between prisons. While other uses are possible, the most appropriate use of the results is to identify prisons that were performing significantly better or worse in raising or lowering inmates evaluations than would be expected given the organizational mission (security level) and inmate composition of the prisons.

Results

First Step: Do Prisons Differ?

All of the conditions of confinement variables appeared to be good candidates for prison-level measures, at least for medium, low, and minimum security prisons (see Table 1). For medium, low, and minimum security prisons, statistically significant differences between prisons

were noted for average inmate response on all five of the conditions of confinement variables. This pattern did not hold for administrative and high-security prisons. However, there are several points worth noting about the administrative and high-security prisons.

First, there were only two administrative prisons included in the 1997 Inmate Survey, and these two prisons served different functions in the Federal Bureau of Prisons. One of the prisons was a metropolitan detention center (MDC), holding pre-trial inmates, and the other was a medical center (FMC) housing inmates with serious or chronic medical conditions. While these two very different prison missions did produce different evaluations among the respective inmates, for example, 74% of inmates at the FMC rated themselves as being safe from being hit, punched or assaulted (SAFE) as compared to only 59% of the inmates at the MDC (complete results not reported here), the differences in evaluations generally were not large enough to negate the disadvantage of having only two comparison points.

Second, regarding the employment variables (JOB and HOURS), given the special functions of the two administrative prisons, a larger number of inmates were not assigned to normal prison jobs (although even in these two prisons about 60% of the inmates held jobs). Third, there were only four high-security prisons in the 1997 Inmate Survey, and this may have created a problem of not having sufficient statistical power to identify differences in the population of high-security prisons for the two safety measures (SAFE and STREET). Alternatively, there may have been more agreement among inmates at all high-security prisons about their safety than was the case for inmates at the lower security levels. Nonetheless, the significant differences for the other three conditions of confinement measures (NOISY, JOB, and HOURS), suggested that the second step of analysis should be undertaken with all conditions of confinement variables for the four security

levels of high, medium, low, and minimum. Administrative prisons were a special case, and there were not enough of them in this sample to analyze properly.

Second Step: Do Prisons Differ after Introducing Relevant Controls

The results for the second step of analysis begin with a discussion of hours worked. Hours worked was the only variable that could be reasonably treated as a continuous variable and analyzed with the standard features of PROC MIXED in SAS. The other variables were analyzed with the features built into the GLIMMIX macro to handle binomial dependent variables. The variable indicating whether the inmate respondent had a prison job (JOB) is a true dichotomy and thus appropriate for GLIMMIX. The variables for safety from assault (SAFE) and noise in prison (NOISY) have five ordinal responses. To make the variables dichotomies, SAFE was recoded into those who felt that they were "safe" from assault and those who were not as sure (who responded "somewhat safe," "neither safe nor unsafe," "somewhat unsafe," and "unsafe"), and NOISY was recoded into those who reported the prison was "not at all noisy" as compared to those who reported some problems with noise ("slightly noisy," "moderately noisy," "more than moderately noisy," and "very noisy"). This recoding strategy had intuitive appeal as it compared those who rated the prison most favorably (safest and quietest) against those who did not, it divided the respondents into approximately equal groups for both variables, and it made the variables appropriate for analysis with GLIMMIX. The variable asking inmates to compare prison safety to their safety on the street had three possible categories. It was recoded to compare those who claimed the prison was more safe than the streets to those who said the prison was either more

dangerous than the streets or about the same. Recoding the variable this way had similar advantages as the manner in which SAFE and NOISY were recoded.

All three variables with three or more categories (SAFE, NOISY, and STREET) were also analyzed as though they were continuous variables with PROC MIXED. While this analysis is not defensible in a formal statistical sense, many analysts have shown that ordinal variables with five or more categories behave fairly well when analyzed as continous variables. The additional analysis for these conditions of confinement variables was provided to bolster confidence in the analyses based on the recoded variables and to provide some idea about the values of the intraclass correlations for these variables. As logit models (the link function used with GLIMMIX for the models presented here) do not have residual variance at the individual level (which is necessary to calculate the "within" group component of total variance), the logit models for the dichotomized variables by definition could not be used to generate intraclass correlations (see discussion below of the intraclass correlation).

Hours of employment. The results presented in Table 2 are for the multilevel model of hours worked in the past week for those inmates who held a job. The results are for a model that included all of the control variables discussed previously, including time served (TIMESRVD). The time served variable was problematic in all of the models considered here because it was missing for a large number of cases. However, clearly the amount of time served in prison is intuitively important. Therefore, the analysis for HOURS (and the other conditions of confinement variables) was also performed without the time served variable to recover the cases deleted because of missing values on time served. In terms of the prison-level properties of inmate reports of hours worked, the results did not change.

After controlling for the individual-level characteristics of the inmates and the security level of the prison, there was a portion of the variability in hours worked that is associated with differences between prisons. As seen in Table 2, the intraclass correlation (ICC) is .056. One way to interpret the ICC is to view it as the percentage of between prison variance to total variance. In this case, the ICC was equal to the ratio of between prison variance (11.856) divided by total variance (sum of 11.856 and 198.186). This ratio is the ICC reported in Table 2, .056. To make the ratio a percentage, all that is required is to multiply it by 100. In other words, 5.6% of the variance in the number of hours worked in the past week was due to differences between prisons. Conversely, 94.4% of the variance was due to differences between individual inmates. While it is desirable to have an ICC larger than .056, this compares favorably to the largest ICC value (ICC=.058 for a scale of institutional operations) reported by Camp, Saylor and Wright (1997) for prison-level properties of three scales created from staff survey data.

Prison Job. Parameters for the model of prison job (and the other dichotomous dependent variables) produced by GLIMMIX were estimated with a binomial error distribution and a logit link function, the most typical choices for binomial data. The results presented in Table 3 for JOB suggest that at least some of the prisons differ in the proportion of inmates holding jobs after controlling for security level and the individual characteristics of inmates identified here. The actual value of the variance estimate does not make much sense in that it is in the unit of the link function (the logit), but it was significant, and a method for using the model results is presented in the *Third Step* section.

Safety of Prison. The results for between-prison variance in the SAFE variable were not as unequivocal as the findings for the other variables considered to this point. As seen in Table 3, the

between-prison variance for SAFE was not statistically significant if there is a strong preference for a cutoff level of p = .05. A firm cutoff of .05 seems a bit too strict, but the results presented in Table 4 where SAFE was treated as a continuous variable also raise some doubts. As seen there, the between-prison variance was still marginal, and the amount of variance accounted for by between prison differences was not even quite 2%. This variable is not a strong contender for capturing differences between prisons.

Prison Compared to Street Safety. It is possible to see in the results produced by GLIMMIX that prisons differed in terms of the proportions of inmates evaluating the prison as safer than the streets from which they came (STREET), even after controlling for security level and the individual socio-demographic and criminal history characteristics of the inmates (see Table 3). When STREET is treated as a continuous variable, it is obvious that the intraclass correlation was low for this safety variable as it was for the safety variable probing inmates about assaults and fights (SAFE). For the STREET variable, the percentage of the variance occurring between prisons is less than 2%.

Noise when Sleeping. Inmates reported that prisons clearly differed in the noise levels that existed while they were trying to sleep. The results for those inmates who reported that noise was not a problem at all in comparison to inmates who reported at least some problem with noise (NOISY) are clearly significant in both Tables Y and Z, and the intraclass correlation for NOISY considered as a continuous variable is .068. According to this result, almost 7% of the difference in inmate evaluations of noise levels resulted from inmates being housed at different prisons. This is the largest intraclass correlation reported in this study.

Third Step : Adjusted Measures of Prison Differences

The multilevel model results can be used to produce a check of prison-level effects. The output can be used to identify prisons that were doing either significantly better or worse than would be expected in affecting inmates self-reports. Since the main interest of this analysis does not lie with the substantive interpretation of each condition of confinement variable, only one condition of confinement variable, hours worked (HOURS), was examined with respect to this use of the multilevel output.

The models specified in this analysis produced three parameter estimates (not including the error term) that are associated with the group, or prison, level of analysis. First, an overall intercept was produced for the model, and this is a fixed effect in the models. Second, there was a fixed effect specified for the security level designation of the prison. This parameter modified (either up or down depending upon the sign of the coefficient) the intercept to reflect an adjusted intercept for each security level. Finally, there was a random parameter introduced into the models to account for the effect of how each prison differed from the intercept. In essence, the parameter can be interpreted as the amount the specific prison was above or below (again depending upon the sign of the coefficient) the intercept for all prisons after controlling for security level and the individual characteristics of the inmate raters. The effect was specified as random as not all federal prisons, only a random sample, were analyzed.

It is the last parameter, the random coefficient for prison (SITEID in this analysis), that is of most interest for identifying prisons that had the most negative or positive impact upon inmate evaluations. The coefficients are the predicted values of the random variable SITEID with an

overall expected value for SITEID of 0. This means that values indicate the extent to which the prison under question moves the typical inmate response above or below the intercept (which is adjusted for the group-level effect of security level in this analysis and the individual characteristics of the inmates). The values of SITEID were created for data collection purposes, and they do not allow for actual identification of the actual federal prisons surveyed without a lookup sheet. For present purposes, it was not necessary to actually identify the prisons.

From the output presented in Table 5, it is possible to identify prisons that differ significantly in the number of hours worked on average, as reported in the inmate surveys. In this example, the values indicate the number of additional hours (fewer if the value is negative) that the inmates reported working after controlling for other variables included in the model. Using a typical cutoff value of p # .05 (with a corresponding absolute value of t $1.96), inmates at prisons 28021, 28041, and 32041 worked significantly more hours than inmates at other prisons. On the other hand, several prisons had inmates reporting that they worked fewer hours than would be otherwise expected, prisons 23041, 25011, 29011, 29031, 29041, and 29071.

In the example used here, the coefficient for SITEID is directly interpretable. For the prison identified as 28041 in Table 5, those inmates who have a prison job worked on average 5.1 more hours in the week prior to the survey as result of being at prison 28041. For the other models presented in Table 3 where the dependent variables are treated as dichotomies, the SITEID coefficients are not as directly interpretable since the effects were defined on the link function, the logit in these models. It would be a simple matter, though, to transform the results to make them interpretable. The transformed results would provide the differences between the prisons in the proportion of inmates who chose the “event” category and those that did not. In the models

presented here, the variables were coded so that "event" meant respondents had a job (JOB), respondents said the prison was safe (SAFE), respondents reported that the prison was safer than the street (STREET), and respondents claimed that noise was not a problem (NOISY).

Discussion

Inmate survey data indicated systematic prison-level differences in this analysis. Inmates did not provide uniform responses to the survey questions examined here irrespective of the prison where they were housed. For the number of hours worked, for whether or not the inmate held a job in prison, and for reports of noise in the prison when sleeping, inmates at some prisons reported on average—as a group—differently than inmates at other prisons. The presumption is that the prisons actually differed on these measures, although it is not possible to rule out alternate explanations. For the two safety measures, the evidence was not compelling, but it suggested that prison-level measures of safety could be developed. Note that single survey questions were analyzed. A scale created from multiple measures of inmate safety might perform better. Unfortunately, there were not enough safety items in the 1997 Inmate Survey to perform appropriate analyses to create the scales. It is also worth noting that the results for noise were particularly encouraging. Anyone who has ever visited different prisons recognizes that prisons definitely differ in noise level.

What are the implications of these findings for using inmate data to evaluate the performance of public and private prisons? As noted previously, it probably would not be a good idea to bet the farm on any one piece of information, certainly not survey data. However, it does seem clear that survey data, both staff survey data as demonstrated by Camp and colleagues

(Camp et al., 1997; Camp et al., 1999) and inmate survey data as demonstrated here, can have a role in identifying differences between prisons in evaluations. In fact, as noted previously, inmate and staff survey data have been included in past evaluations of private prisons (Hatry et al., 1994; Logan, 1991; Sechrest & Shichor, 1994; Urban Institute, 1989). The questions is, does the analysis presented here inform the manner in which the data were used in the previous studies?

The answer to this latter question is a resounding yes. Survey data have not been used properly in previous evaluations of private and public prisons. Admittedly, most of the substance of the findings presented here is rather dry reading. Some prisons are more noisy than other, but that is not news. But, it was important to demonstrate that inmate data reflected that some prisons are more noisy than others as this finding legitimates the use of this measure (and any other measures that pass the same analytic steps used here) to distinguish prisons. Plus, if prisons are to be compared on the numbers (or proportions) of inmates or staff who responded to a survey question in a given manner, then the analysis presented here also suggests that it is necessary to control for differences between prisons in the types of inmates they house.

Conclusion

The collection of survey data is relatively inexpensive. Nonetheless, methods proposed here raise the stakes for using survey data in evaluations of prison operations. The stakes are raised both in terms of using appropriate statistical methods and in terms of collecting more data. The existing evaluations of private prisons (that have used survey data) were limited to data from only a handful of prisons, often three or fewer prisons. The models employed in this analysis to identify prison-level differences require data on more prisons than three. In this analysis, inmates

in 30 different male prisons provided data. Do appropriate analyses require data from 30 or more prisons? The answer to that question is extremely complex, but the safest conclusion is that the more data you have from different prisons, then the more statistical power at the disposal of the analyst to identify prison differences.

The only alternative to collecting data on multiple prisons is to use the data in less appropriate ways, which has been the case in all existing studies that used survey data in evaluating private and public prisons. Even with data from three or fewer prisons, there are simple statistical techniques (such as ordinary least squares or logistic regression where the prisons are treated with dummy variables) that can be employed to control for differences between inmates at the respective prisons.[6] Even these simpler methods were not employed in the existing evaluations that used survey data.

The task that remains is to continue research to identify the most appropriate survey questions that yield desirable information about differences between prisons. A natural extension of the research reported here is whether or not differences exist between different prison systems (the state systems as well as the federal) in terms of the condition of confinement variables reported on by inmates. Such an examination would take place with a three-level hierarchical model, with the two levels examined here nested within the state prison system. Also, at the higher aggregate level of comparing state/federal prison systems, the other condition of confinement variables included in the Inmate Survey that ask about experiences since entering the prison system could be analyzed with two-level hierarchical models, with the two levels being the individual and the state/federal levels.

Endnotes

1. Unlike many states where contractors are required to follow all state policy in operating private prisons, the federal sector, at least as represented by the Federal Bureau of Prisons, uses performance-based contracting. With this type of contracting, the contract specifies certain performance targets that the contractor must meet, but the contractor is given latitude in implementing policy to achieve these objectives. As much as possible, Bureau of Prisons' policy is not imposed on the private contractor, except where necessary to maintain official inmate records or compliance with federal law or regulations.

2. William G. Saylor pointed out the following. "Many staff would say that using inmate opinions to evaluate them puts them in an untenable position. If they do their job well they will be evaluated poorly. Many managers might take the same position with respect to evaluations by their subordinates. For some areas this may have some merit. However, one can pose some areas, e.g., safety, where doing a good job in an organizational context would also result in a favorable evaluation by inmates or subordinate staff. This also suggests another means of ensuring a fair (meaningful) evaluation. The organizational elements of the questionnaire data should demonstrate an expected correspondence with relevant objective (official agency) measures."

3. Controlling for security level in itself implies that inmate responses are conditioned by the type of prison in which they are housed. Formally, the conditions of confinement measures should have been examined for variance across the different prisons in step one without controlling for security level. Given prior knowledge that security level usually conditions survey responses, the decision

was made to impose the more rigorous criteria of looking for variance across prisons after controlling for security level. See also the following note.

4. In a crosstabulation of security level with the conditions of confinement variables considered here, inmates at the different security levels answered the questions in ways that most corrections practitioners would expect. For example, inmates at higher security-level prisons were less likely to rate themselves as being safe from physical attack than prisoners at lower security-level prisons. Inmates at higher security-level prisons were more likely to view the streets as being safer than the prison where they were serving time. The one item that demonstrated a relationship with security level that may be somewhat counterintuitive upon first examination is the noise level when trying to sleep. Inmates at medium and high security prisons were less likely to report that noise is a problem than inmates at low and minimum security level prisons. However, this finding is not as surprising if you consider that inmates at high and medium level federal prisons are housed in cells, where inmates at lows and minimums are typically housed in dormitory or pod types of arrangements. Also, a oneway analysis of variance was conducted on hours of employment with security level. The results indicated differences across at least some of the security levels. The crosstabulations and oneway analysis of variance for the conditions of confinement variables by security level are not reported here, but they are available upon request from the author.

5. In reviewing the paper, William Saylor noted that inmate evaluations may be confounded by factors related to inmate and institution security level. For example, inmates sometimes work their

way “down” in security level. Inmates who start their sentences in a medium-security prison can work their way down in some instances to lower security-level prisons if their in-prison behavior is exemplary. Likewise, poor adjustment and misconduct can move inmates up in security level. In both cases, the inmate evaluations might be systematically influenced by the mismatch between the prison they are currently in and the other prison(s) in which they were incarcerated which had different security features. Unfortunately, it is not possible to test for these effects as the survey data do not permit for identification of inmates who have been housed at prisons of differing security levels.

6. Technically, the simpler techniques are less adequate than the methods outlined here. Using dummy variables, and thereby statistically treating the prison effect as fixed, is appropriate when all “levels” or “doses” of the prison variable are represented. In most prison systems, though, all levels are not represented unless every prison in the prison system is included in the model.

Table 1

Summary of Prison Differences for Conditions of Confinement Items within Each Security Level

	SAFE†	STREET†	NOISY†	JOB†	HOURS‡
Administrative	no*	no	yes	no	no*
High	no	no	yes	yes	yes
Medium	yes	yes	yes	yes	yes
Low	yes	yes	yes	yes	yes
Minimum	yes	yes	yes	yes	yes

†Difference between prisons based on a Pearson P^2 test performed on the crosstabulation of the condition of confinement variable with the respective prisons. Complete results are available from the author.

‡Difference based on a one-way analysis of variance of hours by the respective prisons within each security level. Complete results are available from the author.

*.05 # p # .10.

Table 2

Measurement Properties of the Aggregate (Prison) Measure of Hours Worked (HOURS)

	Hours Worked	p-value
Within Prison Variance	11.856	0.0049
Between Prison Variance	198.186	0.0001
Intraclass Correlation	0.056	

Table 3

Measurement Properties of the Aggregate (Prison) Measure of Prison Job, Safety of Prison, Prison Safety Compared to Street Safety, and Noise Level when Sleeping

Results Produced with GLIMMIX

Between Prison Variance for:	Estimate	p-value
Prison Job (JOB)	0.539	0.015
Safety of Prison (SAFE)	0.093	0.059
Prison Compared to Street Safety (STREET)	0.106	0.048
Noise when Sleeping (NOISY)	0.474	0.002

Table 4

Measurement Properties of the Aggregate (Prison) Measure of Prison Job, Safety of Prison, Prison Safety Compared to Street Safety, and Noise Level when Sleeping

Results Produced with PROC MIXED

Between Prison Variance for:	Estimate	p-value	Intraclass Correlation
Prison Job (JOB)	--	--	--
Safety of Prison (SAFE)	0.019	0.067	0.017
Prison Compared to Street Safety (STREET)	0.011	0.051	0.018
Noise when Sleeping (NOISY)	0.126	0.003	0.068

Table 5

Effect of Prison Site (SITEID) on Reported Hours Worked (HOURS)

Site ID	Estimate	Standard Error	t	Prob. > |t|
23011	1.30930586	1.86180633	0.70	0.4820
23021	0.52726029	2.01283119	0.26	0.7934
23031	0.18874880	1.73583625	0.11	0.9134
23041	-5.03106102	2.29886774	-2.19	0.0288
23051	-2.43006091	1.97764723	-1.23	0.2193
24011	- 0.59720886	2.00121822	-0.30	0.7654
24311	0.55992838	1.94068143	0.29	0.7730
25011	-4.37100803	1.89719065	-2.30	0.0213
25021	-0.27499434	2.32672337	-0.12	0.9059
26011	2.92196671	1.89016221	1.55	0.1223
26031	2.12859525	1.83996810	1.16	0.2475
27031	-1.12234970	2.07772095	-0.54	0.5891
28021	3.97828326	1.91498380	2.08	0.0379
28031	0.85129439	1.89684079	0.45	0.6536
28041	5.13269597	1.82842216	2.81	0.0051
29011	-4.45132873	1.83734129	-2.42	0.0155
29021	2.32651314	1.89221649	1.23	0.2190
29031	- 4.27472823	1.76500142	-2.42	0.0155
29041	-4.12774096	1.84879390	-2.23	0.0257
29061	3.40592481	2.29467874	1.48	0.1379
29071	-4.12308866	1.94396691	-2.12	0.0341
30011	3.21552463	2.00623743	1.60	0.1092
30021	0.02530417	1.84009532	0.01	0.9890
30031	-1.50009162	2.01642747	-0.74	0.4570
30041	-1.13665284	1.85874238	-0.61	0.5409
31011	0.00874445	1.88245874	0.00	0.9963
31021	0.57426914	1.93931507	0.30	0.7672
32021	1.90013055	2.30221590	0.83	0.4093
32031	0.66350823	1.79422609	0.37	0.7116
32041	4.77683644	1.94566232	2.46	0.0142

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.

Bureau of Justice Statistics. (1997). *Survey of Inmates of Federal Correctional Facilities* (Washington, D.C.). Interviews conducted by U.S. Department of Commerce, Bureau of Census.

Camp, S. D., Saylor, W. G., & Harer, M. D. (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*(4), 739-761.

Camp, S. D., Saylor, W. G., & Wright, K. N. (1999). Creating Performance Measures from Survey Data: A Practical Discussion. *Corrections Management Quarterly, 3*(1), 71-80.

DiIulio, J. J. (1987). *Governing Prisons: A Comparative Study of Correctional Management*. New York: The Free Press A Division of Macmillan, 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.

Hatry, H. P., Brounstein, P. J., & Levinson, R. B. (1994). Comparison of Privately and Publically Operated Corrections Facilities in Kentucky and Massachusetts. In G. W. Bowman, S. Hakim, & P. Seidenstat (Eds.), *Privatizing Correctional Institutions* (pp. 193-212). New Brunswick, N.J.: Transaction.

Horney, J., Osgood, D. W., & Marshall, I. H. (1995). Criminal Careers in the Short-Term: Intra-Individual Variability in Crime and Its Relation to Local Life Circumstances. *American Sociological Review, 60*(5), 655-673.

Lincoln, J. R., & Zeitz, G. (1980). Organizational Properties from Aggregate Data: Separating Individual and Structural Effects. *American Sociological Review, 45*, 391-408.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, NC: SAS Institute Inc.

Logan, C. H. (1991). *Well Kept: Comparing Quality of Confinement in a Public and a Private Prison* . Washington, D.C.: National Institute of Justice.

Mason, W. M., Wong, G. Y., & Entwisle, B. (1983). Contextual Analysis through the Multilevel Linear Model. In S. Leinhardt (Ed.), *Sociological Methodology 1983-1984* (pp. 72-103). San Francisco: Jossey Bass.

Raudenbush, S. W., & Willms, J. D. (Eds.). (1991). *Schools, Classrooms, and Pupils: International Studies of Schooling from a Multilevel Perspective*. San Diego: Academic Press.

Rountree, P. W., Land, K. C., & Miethe, T. D. (1994). Macro-Micro Integration in the Study of Victimization: A Hierarchical Logistic Model Analysis across Seattle Neighborhoods. *Criminology, 32*(3), 387-414.

Sechrest, D. K., & Shichor, D. (1994). *Final Report: Exploratory Study of California's Community Correctional Facilities* : Parole and Community Services Division, California Department of Corrections.

Tennessee Select Oversight Committee on Corrections. (1995). *Comparative Evaluation of Privately-Managed Corrections Corporation of America Prison (South Central Correctional Center) and State-Managed Prototypical Prisons (Northeast Correctional Center, Northwest Correctional Center)* : Tennessee Select Oversight Committee on Corrections. Nashville, TN.

Urban Institute. (1989). *Comparison of Privately and Publicly Operated Correctional Facilities in Kentucky and Massachusetts*. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Waite, P. J. (1997). 1997 SIFCF: Inmate Sample Selection : U.S. Department of Commerce, Bureau of the Census.

Wright, K. N. (1994). *Effective Prison Leadership*. New York: William Neil Publishing.