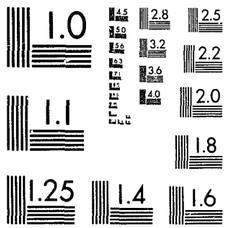

CUTBACK MANAGEMENT AND SERVICEABILITY IN CRIMINAL ADJUDICATION SYSTEMS

by

Joan E. Jacoby
Executive Director
Jefferson Institute for Justice Studies
Washington, D. C.

Paper presented at the

American Society for Public Administration (ASPA)
42nd Annual Conference

April 12 - 15, 1981
Detroit, Michigan

This paper was prepared and supported as part of NIJ grants numbers 78-NI-AX-0091 and 80-IJ-CX-0032. The views and opinions of the author are not necessarily those of the National Institute of Justice or the U. S. Department of Justice.

CUTBACK MANAGEMENT AND SERVICEABILITY IN CRIMINAL ADJUDICATION SYSTEMS

Joan E. Jacoby
Executive Director
Jefferson Institute for Justice Studies
Washington, D. C.

Abstract

The Washington Post on Sunday, February 15, 1981 headlined a feature about Detroit that said curing its economic problems would involve one of two treatments; the first would inflict agony, the second, extreme pain. When the choices facing local governments are not where to cut budgets but how much, then certain issues that rarely rise to importance in more affluent times have to be considered.

These issues involve public agencies and their ability to provide mandated services, defined narrowly as those that are provided for by law or specified in the state (or Federal) constitution. The questions posed by many who have an interest in, or who have suffered from, cutbacks revolve around the basic issue of serviceability -- its definition, measurement, and interpretation.

This paper examines the problem of measuring the adequacy of services as it applies specifically to the public agencies of prosecution and defense. It presents various approaches to defining levels of service and the problems involved in measuring them. It discusses two dimensions of service, quantity and quality, that should be considered in deciding when an agency is no longer able to provide services and outlines techniques useful in assessing this state.

The quantification of the concept of serviceability is based on the author's research in developing performance measures for prosecution

and public defense. agencies. [1] Currently in its second phase, the long term nature of this research reflects the priority being given by NIJ to determine how performance of the criminal justice system and its component members should be measured and for what purposes. Not the least of these is the subject addressed here.

Defining Serviceability

The problems associated with insufficient funds, financial embarrassments and incurred liabilities eventually give focus to the issue of funding discretionary or mandated services and as part of this issue, the need for some objective way to measure the adequacy of services delivered to the public. In response to reductions in revenues and after other funding sources are exhausted, [2] local jurisdictions generally follow a pattern of first, reducing appropriations to discretionary services and then ordering reductions in mandated ones. The definitional problems of what are discretionary services as opposed to mandated ones; what legitimacy they have in making claims against the budget; and what mix of discretionary and mandated services should be maintained are critical. Indeed, these and other related criminal justice issues have been posed and examined by the auditor in Multnomah County (Portland) Oregon (Lansing: 1977) and contested in the courts by the Prosecuting Attorney in Wayne County (Detroit) Michigan. [3]

1/ Supported by the National Institute of Justice grants numbers 78-NI-AX-0091 and 80-IJ-CX-0032. The views of the author are not necessarily those of the National Institute of Justice or the U. S. Department of Justice.

2/ For a comprehensive discussion of this impact and various responses, see the special issue of Public Administration Review, Vol. 41 Jan. 1981.

3/ Wayne County Prosecutor vs. Wayne County Board of Commissioners, 93 Mich. App. 114 (1979).

These issues and questions are too broad to be addressed here, but ultimately they ask a common question. Namely, after reductions in resources, how does one know when an agency can no longer sustain further reductions because it would be unable to provide services?

There is little controversy about the fact that sufficient funds should be made available to the prosecutor or public defender so they can provide services (whether mandated or discretionary). Controversy arises when attempts are made to define the level of these services and most importantly the minimum level below which one can say that services are not provided. In its decision reversing the trial court's finding that the Wayne County Board of Commissioners acted legally in ordering a fifteen percent reduction in personnel costs, the court described a serviceable level of funding as "the minimum budgetary appropriation at which statutorily mandated functions can be fulfilled." (93 Mich. App. at 124). It further stated that a serviceable level was not met when the failure to provide funds either (1) eliminated the function or (2) created an emergency that immediately threatened the existence of the function. It declared that a function funded at a serviceable level would be carrier out in a "barely adequate manner" but it would be carried out. A function funded below this budgetary appropriation, by definition, would not be fulfilled as required by statute.

The court did not go beyond these broad statements to specify how these levels were to be measured or what criteria were to be established. As a result its omission leaves room for exploration and research; but its phasing hints at the directions it envisions will be followed in examining the concept of serviceability. The questions that need answers include:

(1) What constitutes serviceability? Can it be defined? Can its component parts be identified? (2) How should serviceability be measured? What are the measures that should be used? (3) How do we find out when a function is not being carried out?

Measuring Serviceability in Quantitative Terms

The answers may be sought from two perspectives, one using quantitative measures, the other, attempting to introduce a consideration of quality. Quantitatively, several approaches can be taken to address these questions. A "relevant constituency" approach (Connolly and Deutsch, 1978) would propose that standards of serviceability can be defined and measured by the degree of satisfaction expressed by the people who receive the service or are affected by it. Based on this type of approach, a preponderance of negative responses could be defined as violating the "barely adequate" level of service. There are two weaknesses in this approach. The first lies in trying to define what constitutes a relevant constituency; the second is defining what level of negative responses constitutes dissatisfaction.

Does one measure satisfaction as expressed by police agencies, defendants, attorneys, the courts, corrections or the general public? Since there is no closure to the sets within this universe of relevant constituents, the number of groups that can be formed and measured approach infinity. Similarly, arbitrarily setting a level of negative responses and defining that as dissatisfaction is at best easily and endlessly contested. This is even more certain when the conflicting interests and goals of the groups surveyed are considered. What is satisfactory to the public defender (or even the prosecutor) may be unsatisfactory to the public, for example.

A more practical approach would be to use cost analysis to measure levels of service. One could compute the difference between the resources available to a prosecutor's or public defender's office (including dollars, personnel, space and equipment) and the time and costs needed to perform functions. If these latter figures were derived from independent estimations, the difference between the actual and the estimated could be defined to indicate levels of serviceability.

However, merely knowing differences between actual and estimated costs would not by itself, establish minimum or "below adequate" levels. This is because there is no assurance that the estimates derived for the prosecution or defense measure what is "barely adequate" or what is optimum. In fact, it is more likely that they reflect some unknown level of service depending upon the resources available to the office or the procedures adopted. Despite these difficulties, this approach is less ambiguous than the relevant constituency one because a beginning and an end can be specified. It also lends itself to a measurement of identifiable units like the dollar costs and man hours required to perform certain tasks.

Levels of service could also be determined using system simulation techniques. A set of models could be developed that would predict when a function was eliminated or when an emergency was created. One such model would define this state as occurring when the equation between input and output was not in balance. This means that a prosecutor's or public defender's agency would be defined as functioning when the case input (I) plus the pending cases (P) is not larger than the output of the agency (O). This can be stated as $I + P \leq O$. When the sum of new and pending cases exceeds the output of the organization, then one could define the

organization as being unable to carry out its functions (namely, dispose of its caseload). In this case, $I + P > 0$.

The tilt produced when the pending caseload exceeds the system's capacity to dispose of it could be rectified by either decreasing intake, $I \leq 0-P$, (thereby letting the system "catchup") or by an increasing output capacity. The latter could be achieved in a number of ways such as increasing court capacity (and sometimes with it prosecutorial resources) or changing the mix of dispositions (using more plea negotiation, for example) or by adding new dispositional outlets to the system such as diversion, mediation, community service or other alternative treatment programs.

The difficulty with this particular model is that, analogous to the cost analysis approach, specifying the maximum output capacity of the adjudication system may be impossible. Unless this difficulty can be overcome, solutions to the tilt will first call for increasing productivity as a means of increasing capacity. Certainly, this is a valid point. It was affirmed in Orleans Parish in 1974 when the output capacity of the court system was in effect doubled because the judges started working a full day, and it was reaffirmed in our recent evaluation of Brooklyn's felony nighttime jury trial project showed that productivity in the night court was one and one half times h gher than that estimated for the day time courts. Part of this could be attributed to the simple fact that more hours were worked in the evening session than the day.

Nevertheless, there are limits to increasing productivity. When these are reached, then priority decisions have to be made such as when the Wayne County Prosecutor, among other actions, shifted the attorneys from the juvenile section to Recorders Court (the adult felony court) to meet the

demands of the volume of work. In reality, prosecutors and public defenders continually adjust their resources to the shifting areas of workload or priority. This makes identification of maximum efficiency and productivity very difficult.

Even though the maximum capacity of the system is probably impossible to define, another approach used by prosecutors and public defenders is to use national or comparative standards to measure (1) the amount of difference from the norm that their office displays; or (2) the probability of this occurring based on the experience of other jurisdictions. Since the resources and expenditures of an office remains fairly stable over time, comparisons and standards remain relatively constant as well. These data describe the current operating levels of adjudication from which means and standard deviations can be computed so that probability statements can be made about the likelihood of deviations occurring too far from the average.

For example, a 1980 survey of prosecutors, (Jacoby, Mellon, Smith, 1981) found that the average number of jury trials that can be disposed of by a single judge was 24 per year. In Brooklyn, it was recently reported as 16. Although these averages reflect only operating norms, they do exist within a universe of bounded rationality (Simon, 1957). For example, it would be highly unlikely that trial judges could triple or quadruple the average. Limits can be applied to the norms.

Ratios also can serve as standards. Based on a survey of over 700 jurisdictions, the ratio of the number of prosecutor's assistants in an office to the number of felony cases annually processed by the office for prosecution has, since 1972, covered about 1 to 100. In 1972 (Jacoby, 1972)

the ratio was 1 in 99. In 1980 (Jacoby, Mellon, Smith, 1981) the ratio was 1 in 96. This means that about one assistant should be on staff for every 100 felony cases processed by the office. The public defender is not immune to these standards either: one assistant public defender is recommended for every 150 felony defendants per year. (NLADA, 29; NAC, Courts, 19).

Comparing output capacity to national averages and standards is helpful in assessing the ability of an agency to provide services. The standards reflect normal operating conditions and thus, give perspective to the requests of an individual prosecutor or public defender for more (or the same) amount of resources. But, these comparisons are not sufficient to determine when services are inadequate.

Introducing Quality as a Factor

One reason is because in addition to volume, time and cost, the quality of the services being provided must be taken into consideration. Quality adds another set of standards that can be used to assess the adequacy of service. These standards would not condone allowing a rapist to plead guilty to a disorderly conduct misdemeanor and pay a fine; or incarcerating a youthful, first offender for joy riding. Although these are extreme examples (and sometimes unfortunately they occur), they suggest that there exists a range of dispositions that can be labelled as acceptable for different types of criminal cases. For our purposes here, quality can be defined as the ability to achieve minimally acceptable dispositions, or better.

Acceptable dispositions are essentially those that match the sanction with the seriousness of the offense and the criminality of the defendant. They can be thought of as a band of dispositions; some preferred, some acceptable and some only barely acceptable. Below this barely acceptable level are dispositions that can be called "not acceptable". Above the preferred band is another "not acceptable" set. Violating the lower band constitutes injustice to the public, violating the upper band does injustice to the defendant. With this rule, the prosecutor's and public defender's interests are made clear. In this paper our focus is on the ability of the prosecutor's or public defender's agency to obtain minimally acceptable dispositions. These are defined to include either the lowest band or the preferred. For semantic simplicity, both concepts will be referred to as minimally acceptable levels, below or above which justice is denied.

When the minimal acceptable levels of dispositions cannot be obtained, then one can claim that service is not adequate. This implies that the quality of dispositions is a function of two variables: an expected disposition (plea, acquittal, dismissal) weighted by an expected sanction. The two are not independent of one another. Charging decisions, plea negotiations and trial tactics all take into consideration the potential sanctions that one can expect to be imposed on the defendant. These expected levels and sanctions will vary by defense and prosecution functions; but, because the relationship between these two agencies is reinforced by daily contact and communication, wide discrepancies in expectations are not expected to exist.

If it were possible to identify and measure the least acceptable dispositions, then one could argue with strong justification that an agency which consistently violates the minimally acceptable standards is not providing adequate services to the public or its clientele. Acceptable dispositions are not an unknown phenomenon. They are articulated daily by prosecutors and public defenders as they "bottom line" cases, decide the "best offer" or "hang tight and go all the way". The win or loss of a case in these agencies is not simply the difference between acquittal or conviction but is measured more often by the sanctions imposed upon the defendant. Thus, whether a case falls below a minimum acceptable level is possible to define and should be quantifiable. The primary reason why such a task appears to be feasible, is because the dynamics of the adjudication system are rational and consistent, ordering its caseload and work by a set of priorities. A testing of 855 prosecutors in 15 jurisdictions showed that the priority of cases for prosecution had a significant effect on the probability of cases:

- being accepted for prosecution -- the higher the priority, the more likely acceptance
- disposed of by a trial, plea or other means -- the higher the priority, the less likely a plea or non trial disposition, the more likely a trial
- disposed of by a plea to a reduced charge -- more likely for lower priority cases
- ending with a sentence of incarceration -- the higher the priority, the more likely imprisonment. (Jacoby, Mellon, Ratledge, Turner, 1980)

These priorities as expressed by a single number include all three dimensions of a criminal case -- the seriousness of the offense, the criminality of the defendant, and the legal-evidentiary strength of the case. They describe relationships to dispositions. The most serious crimes receive the highest priority: likewise the criminality of the defendant. Priority is affected by the legal and evidentiary factors of the case. If constitutional issues are involved, for example, a poor search and seizure or the defendant's Miranda rights were not read, the priority of the case is decreased. In contrast, priority is increased if there is corroboration by two or more police and/or civilian witnesses, the defendant admits to an involvement in the crime, the defendant is known to the victim and there is a gun involved. Based on these factors, it is possible to rank cases by their dispositional priority and the severity of their expected sanctions.

Furthermore, the priority groupings reflect both dispositions and sanctions. On a scale of 1 to 7 (with 1 representing the lowest priority, 4 an average, and 7 top priority) workload in an office follows a predictable trend. Cases falling into the lowest priority group (1 and 2) tend to be declined at intake and not even accepted for prosecution. Cases in the middle, 3-5 range tend to be disposed of by pleas, with pleas to a reduced charge occurring at the lower three end and pleas to an original charge more likely to occur at the 5 level. Priority cases, the 6's and 7's tend to be disposed of by a trial.

When this rational ordering is combined with the results of a survey of almost 100 prosecutors offices (Jacoby, Mellon, Smith, 1981) it was possible to attach disposition rates to these same priority groupings. At one end, there is an almost immutable 9% of the caseload that are disposed of by trials regardless of volume and trials are reserved for the top priority cases (6, 7). At the other, for those offices who review charges (and 15% of those surveyed, do not) from 10% to 45% of cases referred by the police are declined at intake (the mean is 20%) and these are found in the lowest priority group (1 and 2). Of the remainder (the 3-5 group) 49% are disposed of by pleas, some reduced, some to the original. Figure 1 displays the relationship between priority, type of disposition, and the average percents of dispositions. [4/]

FIGURE 1

Relationship between Case Priority Type of Disposition and Average Disposition Rates

	Priority Cases		
	1 - 2	3 - 5	6 - 7
Type of Disposition	Decline	Plea	Trial
Average Rate.....	20%	49%	9%

4/ Most of the difference between these average disposition rates and 100% is due to dismissals.

This combination shows how the adjudicatory disposition system aligns its resources to meet priorities. Now, assuming that this system is fairly stable and that cutbacks have been imposed upon the agency, we can speculate as to some of its reactions. If the prosecutor or the public defender, for example, refuses to change his procedures, (for example, if he has a strong trial emphasis and a no plea bargaining stance), then with a lack of resources, a backlog should develop. Input should exceed the output and cases will be lost either through dismissals because of speedy trials rules being violated, or because the evidentiary strength of the case decays over time. When that begins to occur, a change has to be made. In this example, the most likely one is to relax the no plea bargaining stance and dispose of more cases by pleas. However, when the changes are too great, when the burglar who has been arrested for the third time is allowed to plead to a misdemeanor or when the dismissal rate continues to increase, the agency begins to violate the acceptable standards of justice.

Figure 1 indicates how accommodations to cutbacks can occur. If the present declination rate in the office is low, it may be increased. If the office, for example, is operating at a 10% declination rate, it may be possible to increase this to 20% or even 30% with more intensive screening. This means that the system would be rid of all cases in the 1-2 priority category and start to cut into the cases that are labelled 3.

In addition, the number of cases going to trial might be decreased if the percent disposed by trial is above the rather constant 9%. For

example, if the trial rate is 20%, an expanded use of negotiated pleas could be encouraged to reduce this resource consuming level to 10%. Graphically, this would mean moving some of the 5 or 6 priority cases into the plea range and out of the trial range. Finally, of course, the use of plea bargaining in the middle range (3-5) can be made more efficient and timely by negotiating pleas earlier in the system.

At first, all of this can probably be done without conflict because it merely means changing policy or lowering standards without violating the quality of the prosecution or the defense. However, as the agencies are put under more intensive pressure to cutback, they are more and more prone to violate the boundaries of acceptable dispositions.

Conclusions

To return to our original question: how does one know when a prosecutor's or public defender's office is unable to provide services? As we have seen here, adequate levels of service can be defined to exist when two conditions are satisfied: (1) the agency can dispose of its input; that is, intake and pending cases do not equal or exceed dispositions and (2) the dispositions are reasonable; that is, they are at least minimally acceptable by the chief prosecutor or the public defender. Conversely, there are two ways to define a prosecutor or a defender agency as providing inadequate services. First, it violates the quantitative conditions such that its caseload cannot be disposed of; and second, it violates the qualitative conditions so that the distribution of its dispositions does not correspond to the urgency of the cases for prosecution or defense.

Ideally, to define what constitutes a barely adequate level of service, one should be able to specify both the quantitative and qualitative dimensions. This means that: (1) output capacity and resources have to be measured to determine whether an agency can dispose of its caseload; and (2) the actual dispositions need to be tested against minimally acceptable dispositional standards. Unfortunately, based on our current levels of research and knowledge, we are not able to do this.

However, there are some options that a prosecutor or public defender may substitute as reasonable alternatives to this ideal state. To satisfy the quantitative rule, the following procedures could be undertaken: (1) Within the office, measure the volume of caseload being disposed, by type of dispositions and the volume of cases referred to the office in addition to the number of cases currently pending. Parenthetically, this should be done in all areas serviced by the prosecutor or public defender because, as we noted, the resources in one area may be transferred to another, thereby creating inadequacies in the area not under primary attention. (2) These measures of caseload and associated resources can be compared to other jurisdictions who operate in similar environments or they can be compared to some nationwide norm (such as we noted with the number of assistants per felony intake). The extent of deviation from these operational norms may lend support to the argument that the services are being denigrated. (3) Finally, in some instances there are standards against which the gap between the reality of one office and the goals as expressed by professionals in the field can be also utilized. One of

the most commonly used standards is, of course, the length of time from arrest through disposition and as articulated in speedy trial rules.[5/]

The use of the quantitative rule does not consider quality as a factor. Indeed, some prosecutors or public defenders would argue that if quality were changed, then, this by itself would constitute a disservice to the public or the client. Leaving this point, however, the evaluation of serviceability based on standards of quality also can be approached in a number of ways. Most detailed is to review each disposition and evaluate whether it is acceptable to the office. This can be subjectively by senior policymakers or objectively, based on some rules or criteria. In its simplest form a count of those dispositions not meeting acceptable levels and a computation of their proportion to all cases disposed will provide indicators of the extent to which quality is being changed (or denigrated).

Another way to obtain this indicator is to count certain classes of dispositions, especially dismissals and/or nolles for speedy trial violations, witness no shows, evidence missing, government or people not ready, or complaining witness or arresting police officer not present. These and other reasons similar to them reflect the problems with case management associated with delay and implicitly cutbacks in resources.

5/ A note should be made here that one should not use as an indicator of serviceability the number of cases that are very old (Church, 1978) or fall in the fourth quartile. The difficulty with this indicator is that it is not a true measure of the quality of services since many of these cases are old simply because they should never have been in the system, have been through an evidentiary decaying process or are so complicated or complex and of such low priority that they should simply be dismissed.

Another technique that can be used to assess the dynamics of quality and to identify decay is to periodically assess the dispositions with respect to their priority. If the caseload, spread along the continuum illustrated here, has shifted drastically over time, the quality may be deteriorating. For example, if a higher proportion of the low priority 1's and 2's are being accepted for prosecution or if plea cases in the 4-5 range are moving into a trial status, then concern should be given to the balance in the system. By examining dispositions based on their priority ranking periodically, the agency head should be able to track the consistency between priority and resources.

This technique begs the question of acceptable dispositions. That calls for the development of a system specifying a range of dispositions and sanctions acceptable for certain types of cases. Much like sentencing guidelines, deviations from the acceptable band would be allowed with justification, but the frequency could be noted and reported as part of the quality assessment.

All of this, of course, points up the complexity that is involved in this issue called serviceability. Given the present state of research and data collection there is too little knowledge or data to advance beyond these techniques and clearly, there is need for a long range, systematic development and study of this complex issue. Serviceability is an issue

that will remain important to contemporary society for a long time. As more and more remedies for inflation are sought through the use of cutback management, the development of measures and techniques for evaluating the functions of the prosecutor and the public defender along with other public service agencies is becoming increasingly urgent.

NOTES

Church, Thomas, Jr. , Justice Delayed: The Pace of Litigation in Urban Trial Courts, Williamsburg, VA, National Center for State Courts, 1978

Connolly, Terence, and Stuart J. Deutsch, "Performance Measurement: Approaches to the Development of Measures" (Atlanta, GA: Georgia Institute of Technology, Unpublished, 1978) also "Performance Measurement: Some Conceptual Issues."

Jacoby, Joan E., First Annual Report of the National Center for Prosecution Management, Washington, D. C.: NCPM, 1972)

Jacoby, Joan, Mellon, Leonard and Smith, Walter: Policy and Prosecution, Washington, D. C.: BSSR, 1981)

Lansing, Jewel A., Staffing Analysis of the District Attorney's Office, Department of Justices Services, Multnomah County: Multnomah County, Oregon, 1977)

National Advisory Commission on Criminal Justice Standards and Goals: Courts, Washington, D. C., U. S. Government Printing Office, 1975)

National Legal Aid and Defender Association, The Other Face of Justice, Washington, D. C. NLADA, 1973

Simon, Herbert A., Administrative Behavior (2nd Ed.), New York: The McMillan Company, 1957

END