

How do we choose proper performance measures for prisons?

By [Sasha Volokh](#)

February 27, 2014

This is the fourth post in a series about my new article, *Prison Accountability and Performance Measures*, which is in the current issue of the *Emory Law Journal*. In [Monday's post](#), I introduced the issue and advocated greater use of performance measures, which I'll come back to later this week. In [Tuesday's post](#), I discussed why we don't know much about the comparative cost or quality of public vs. private prisons. In [Wednesday's post](#), I talked about why introducing performance measures would be a good idea.

Today, I'll discuss some of the normative issues involved in what measures to choose. I'll also start on some potential concerns about, and critiques of, implementing performance measures — a discussion I'll finish up tomorrow.

* * *

What Measures to Choose

The earlier discussion of how to define recidivism shows that a lot rides on choosing the outcome measures judiciously. This applies across the board, not just to recidivism. This section considers two distinct aspects of performance measures. The first is that wherever *outcome* measures have been used, *output* measures haven't been abandoned. The second is that what outcomes to measure—and even whether something counts as an output or outcome measure—is inevitably a value-laden question, which must be resolved for a performance-based compensation scheme to go forward. The inevitable incompleteness of outcome measures—and therefore the need to supplement outcomes with outputs—can give rise to undesirable strategic behavior, which I discuss in a later section.

Adopting specific outcomes to measure is equivalent to adopting what [John DiIulio](#) calls an “operational” goal—“an image of a desired future state of affairs that can be compared unambiguously to an actual or existing state of affairs.” “Improving the quality of public education in America’ is a nonoperational goal; ‘Increasing the average verbal and math SAT scores of public school students by 20% between the year 1992 and the year 2000’ is an operational goal.” Similarly, “[r]eforming criminals” is nonoperational, while “[d]oubling

the rate of inmate participation in prison industry programs” is operational. That last goal was output-based, but there’s no reason we can’t, as in the education example, adopt an outcome-based goal—we could just agree on a convenient if arbitrary measure of how well criminals are reformed, such as the two-year reconviction rate. Moreover, there’s no reason to adopt a numerical target as the goal (which would be binary); the goal might merely be (thinking more continuously) to reduce the rate as far as possible. And there’s no reason to adopt a unique goal: multiple operational goals can be implemented in one part of an overall index that determines compensation.

A useful way to explore this question is to examine some existing prison performance measures. Perhaps one of the oldest formal approaches to measuring prison performance is the [Correctional Institutions Environment Scale](#) developed by Rudolph Moos in the late 1960s and often used in the 1970s. The Moos scale contains several subscales:

“Involvement,” “Support,” “Expressiveness,” “Autonomy,” “Practical Orientation,” “Personal Problem Orientation,” “Order and Organization,” “Clarity,” and “Staff Control.” These elements generally aren’t true performance measures, and it’s immediately apparent from their definitions that some are highly impressionistic. The “Involvement” variable “[m]easures how active and energetic residents are”; the “Support” variable “[m]easures the extent to which residents are encouraged to be helpful and supportive”; and so on, with an emphasis on measuring the extent of supportiveness and encouragement. The scale was criticized because it wasn’t clear what the difference between some of the elements was and to what extent they were correlated, and even to what extent they described a real phenomenon. Some critics wrote that “when the CIES is administered and the individual scores are tallied and averaged, we really have no idea what the scores on the nine subscales indicate.” Ultimately, the scale was “determined not to possess acceptable validity.”

A later approach, described in 1980 in a report by [Martha Burt](#), uses five types of measures: “Measures of Security,” including the escape rate and escape seriousness; “Measures of Living and Safety Conditions,” such as victimization, overcrowding, and sanitation; “Measures of Inmate Health” (both physical and mental); “Intermediate Products of Programs and Services” like improvements in basic skills and vocational education completed; and “Measures of Post-Release Success,” including employment success and recidivism. Only the fourth category is explicitly labeled “Intermediate Products,” but some of the other measures are also outputs, not outcomes—see, for instance, the use of hospitalizations and sick days in the measures of inmate health.

The mixing of output and outcome measures is fairly typical; John DiIulio criticizes the BOP’s [Key Indicators/Strategic Support System](#) for also “indiscriminate[ly] mixing . . . process [i.e., input or output] and performance [i.e., outcome] measures.” But [DiIulio](#)

himself has measured prison quality in terms of “order (rates of individual and collective violence and other forms of misconduct), amenity (availability of clean cells, decent food, etc.), and service (availability of work opportunities, educational programs, etc.)”: note the output measures in the inclusion of the availability (not the effectiveness) of programming.

The [MTC Institute](#), the research arm of the private prison firm [Management & Training Corp.](#) (MTC), likewise calls for holding prisons accountable for “outcomes”; but these “outcomes” include not only assaults, escapes, recidivism, overcrowding, and the like, but also outputs like “[s]ubstance abuse education/treatment completions” and “[p]roportion of inmates participating in spiritual development program(s).”

The American Correctional Association’s [performance-based standards](#) for correctional health care raise the same issue. Some of these are true outcomes, like “the rate of positive tuberculin skin tests” or the suicide rate, though others are process measures or expected practices, like whether an offender “is informed about access to health systems and the grievance procedure.” The [Prison Social Climate Survey](#), which is based on inmate and staff surveys, likewise mixes outcomes (such as crowding or safety) with outputs (such as whether the prison is a pleasant place to work for staff).

It is clear, then, that outcomes and output measures tend to go together; no doubt this is because not all outcomes are well measurable. Moreover, the choice of measures, and even the basic question of whether to classify a measure as an output or an outcome, is inevitably value-laden. We can see this clearly by examining [Charles Logan](#)’s “quality of confinement” index, one of the more highly regarded prison performance measures. Logan’s performance indicators focus on eight broad categories:

1. “Security (‘keep them in’).”
2. “Safety (‘keep them safe’).”
3. “Order (‘keep them in line’).”
4. “Care (‘keep them healthy’).”
5. “Activity (‘keep them busy’).”
6. “Justice (‘do it with fairness’).”
7. “Conditions (‘without undue suffering’).”
8. “Management (‘as efficiently as possible’).”

Each of these categories contains a number of subdimensions: for instance, the “security” category contains the subdimensions of security procedures, drug use, significant incidents, community exposure, freedom of movement, and staffing adequacy. The “safety” category contains safety of inmates, safety of staff, dangerousness of inmates, safety of environment, and (again) staffing adequacy.

And, finally, Logan decomposes these subdimensions into specific numerical measures: number of escapes, proportion of staff who have observed staff ignoring inmate misconduct, ratio of resident population to security staff, drug-related incidents, and so on. In all—over all eight dimensions—there are a few hundred measures. Logan used this index [to evaluate three women’s prisons](#) in New Mexico and West Virginia.

None of Logan’s measures involve how many inmates get rehabilitated. But this is also intentional. First, actual rehabilitation is out of the direct control of prisons. Logan has a preference for measuring things that are within prisons’ “direct sphere of influence”; what we measure “ought to be achievable and measurable mostly within the prison itself.” Second, including rehabilitation endorses the rehabilitative model of criminal punishment, and Logan makes it clear that his model is retributive, not rehabilitative. Prisons, in his view, shouldn’t “*add to* (any more than . . . avoid or . . . compensate for) the pain and suffering inherent in being forcibly separated from civil society[;] . . . coercive confinement carries with it an obligation to meet the basic needs of prisoners at a reasonable standard of decency.”

Logan’s concern for focusing on what a prison can control and focusing on the retributive goal merge in the following statement: “a prison does not have to justify itself as a tool of rehabilitation or crime control or any other instrumental purpose at which an army of critics will forever claim it to be a failure.” (Of course “[i]t would be very nice if the prison programs [counted in the ‘activity’ dimension] had rehabilitative effects,” and perhaps they do, but whether they do or don’t doesn’t enter into the index.)

Fair enough. What this illustrates is that you can’t judge particular measures to be desirable unless you have a normative theory that proclaims certain goals to be desirable, and such a political discussion is necessary before one can commit oneself to a particular form of performance measures. “[W]ithout declared goals, we cannot hold a jurisdiction accountable, and performance measurement is meaningless.”

This normative issue arises wherever performance measurements are used. John DiIulio describes how [John Chubb and Terry Moe](#) “measure school performance strictly in terms of pupils’ achievements on a battery of standardized tests, accepting the schools’ value as instruments of socialization and civics training as important but secondary.” On the relative value of test scores vs. socialization, your mileage may vary.

Likewise, for the correctional system, there is a great variety of available goals; prisons should punish, rehabilitate, deter, incapacitate, and reintegrate—all, says John DiIulio, “without violating the public conscience (humane treatment), jeopardizing the public law (constitutional rights), emptying the public purse (cost containment), or weakening the

tradition of State and local public administration (federalism).” So we need to have a political discussion about what the appropriate goals are.

One’s normative theory also affects whether a particular measure is an output or an outcome; this classification, which I’ve been using casually so far as if it were value-neutral, is in fact anything but. If we didn’t care about inmates but only cared about the outside world, perhaps only recidivism would be relevant. The quality of living conditions or inmate literacy would merely be outputs, which we would care about only to the extent that they affected recidivism; they wouldn’t need to independently enter the compensation function as long as we already counted recidivism. But we might independently care about inmates’ living conditions for many reasons; if we do, living conditions become an actual outcome of the system.

Thus, some of Logan’s dimensions, like “activity,” which I’m inclined to call an output measure, might be an outcome measure given Logan’s normative perspective. The same goes for variables like prison employees’ job satisfaction (which I consider an output measure because it’s only instrumentally relevant to prison quality, but which others who care about labor conditions might treat differently) or whether inmates have difficulty concentrating (which—unlike, say, overcrowding or physical safety—many may not consider an appropriate dimension for prison evaluation).

Some of the measures, though, for instance the number of urinalysis tests conducted based on suspicion, are output measures under any definition, and these have the problem that it’s ambiguous whether they’re good or bad. Do we want more or fewer urinalysis tests based on suspicion? More tests could mean that drug use has gone up; or it could mean that prison authorities are getting more serious about controlling drug use. Even worse, prison authorities’ stringency is something prison authorities themselves can control; this is a serious problem, which I discuss below.

As a final note, I’ll mention that while it’s vitally important to have good cost measures that are adequate for comparing public and private prisons, it’s not necessary to include cost in the private contractor’s compensation. If we couldn’t measure quality, perhaps there would be a role for rate-of-return regulation, which might at least limit some of the private sector’s harmful cost-cutting tendencies. But if we’re going to engage in quality measurement, we might as well enforce quality directly by getting the rewards or penalties “right”; let the private firms worry about their own costs.

* * *

Despite the advantages discussed in the previous section, the use of performance measures has its pitfalls.

One concern, so obvious as not to merit its own section heading, is the issue of administrative costs. Recidivism-based contracts require one to track released prisoners adequately. Perhaps there would be substantial startup costs—though current probation and parole systems already track releases and monitor employment, recidivism, and other relevant outcomes, so at least some of these costs are already sunk. Moreover, if performance-based contracting is beneficial at all, its benefits are probably great enough that these startup costs are worthwhile.

This Part focuses on other concerns and critiques. First, there is the concern that one can't set the proper prices in a theoretically defensible way. Second, there's the concern that performance-based compensation will affect market structure, either by driving out the public-interested or by driving out the risk-averse. Third — I'm saving this for tomorrow's post — there's the concern that performance-based compensation will lead to undesirable strategic behavior, for instance via manipulation of the choice of performance goals, by distorting effort across various dimensions of performance, by distorting effort across various types of inmate, and by encouraging outright falsification.

What Prices to Set

The focus on performance measures might seem grating to those who criticize the turn toward efficiency analysis and comparative effectiveness and stress moral considerations. But one can support performance measures without endorsing efficiency in any way—in fact, as a better way of achieving particular moral goals.

I myself have been critical of a focus on efficiency in the context of regulatory cost–benefit analysis, another example of hard-numbers-based accountability. To restate the problems of cost–benefit analysis in the prison context: What's the social value of having less recidivism? To ask this in an economic context, we'd have to know either the maximum amount people would be willing to pay to reduce crime, or the minimum amount people would accept to acquiesce in an increase in crime. These are in general different amounts, and the choice between them is value-laden. Suppose we choose one of these numbers to measure; we may find that, when surveyed, some people—who reject the very notion of paying or being paid for reductions or increases in crime—give answers of zero or infinity for their willingness to pay or accept; the number we're seeking may just not exist for these people. Some people may have true willingness to pay or accept, but they don't even know what these numbers are: we only come to know such numbers because of our experience paying for and consuming goods and services in the real world, but increases and decreases

in crime generally aren't traded in markets. So the very act of asking for the number may bring some number into being, but there's no reason to suppose it's accurate. Or, people may know the number, but there's no incentive for them to truthfully reveal it in surveys.

Even if we use non-survey-based estimation methods—How much higher are house prices in lower-crime areas? How much do people pay to avoid crime?—econometric analysis isn't good enough to give us the correct number. The political process is also likely to manipulate the numbers. Moreover, concerns that are hard to quantify can be systematically slighted.

In short, “[w]hile cost–benefit analysis may look like rationality, perhaps it's merely rationalism.” And these are just the problems for people who *accept* the utilitarian basis of cost–benefit analysis. The problems for those who reject utilitarianism as a moral philosophy are even greater. Surely corrections policy, of all things, should be decided with respect to morality and human values rather than numbers?

These are real problems with cost–benefit analysis, and they potentially infect performance-based contracting as well. Setting the incentives in a performance-based contract means either setting the relative weights of every component of performance, or (equivalently) setting the separate rewards or penalties for every component of performance. Getting the prices “right,” in an efficiency sense, requires knowing the social value of the different components of performance; if that social value doesn't exist or can't be measured, it's an impossible task.

I agree and disagree with this critique.

As to the moral objection, even though moral values have an extremely important place in criminal law and policy, I have no essential problem with using economic incentives to improve outcomes in the process. I've argued elsewhere that the valid arguments for or against private prisons generally are essentially empirical; measuring performance is an essential part of that debate, even though the choice of outcomes to measure is a value-laden enterprise; and attaching incentives to those performance measures is eminently justifiable if the result is a morally more just correctional system.

As to the theoretical incoherence objection, I'm sympathetic. But the enterprise can still be salvaged if we adopt a humble attitude. Rather than trying to achieve incentives that are correct in some abstract sense, we can just try to muddle through and ameliorate the problems of the current system by attaching some weight to factors that traditionally haven't been rewarded. None of this requires buying into the efficiency norm. Maybe the weights will be wrong, but “[t]he basic question . . . is whether the dangers of providing improper incentives through imperfect models outweigh the benefits of providing program direction

and accountability.” Is adding this element of imperfect, numbers-based accountability better than not? The remaining sections in this Part address this question.

Effects on Market Structure

This section discusses how performance-based compensation can change the composition of providers. First, it will attract providers who respond better to market incentives, which might affect the overall public-interestedness of the industry. Second, because performance-based compensation is riskier than flat-rate compensation, it will discourage the more risk-averse providers.

1. Public-Interestedness

[Todd Henderson](#) and [Fred Tung](#) address this concern in the context of performance-based compensation for regulators. If regulators are currently public-interested, introducing market incentives might change the culture within the agency. “Once diligence has been priced, perhaps some regulators will slack.”

This form of compensation will also affect the mix of people who choose to be regulators. “Public service motives might be displaced by financial motivations among new hires Eventually, the composition of the regulatory agency could change for the worse.”

Henderson and Tung conclude, citing the crowding out literature, that this is possible, though not necessary: “public spiritedness and financial reward [might not be] mutually exclusive, at least up to a point.” Moreover, changing the mix of individuals “could be a good,” given the failures of the current crop of people.

The same arguments can be applied to performance-based compensation for prison providers. I would add that, to the extent we’re considering performance-based compensation for private firms rather than public servants, we don’t need to worry about making providers any more mercenary than they already are: if there’s one thing advocates and opponents of private prisons agree on, it’s that private prison providers are a profit-oriented bunch. Not that the profit motive is inconsistent with public-interestedness: public servants “profit” from their employment too without being accused of thereby necessarily becoming mercenaries; moreover, corrections professionals move between the public and private sectors and presumably take their professionalism with them. Finally, as I discuss further below, performance-based compensation, combined with social impact bonds, allows nonprofits to raise money from private investors, so to this extent, introducing the profit motive may turn out to be a great boon for charitable and public-interested providers.

2. Risk and Capital Requirements

a. The Risk Is in the Slope

We've seen, in the discussion of Charles Logan's approach above, the concern that performance measures be based on factors that the relevant actor can actually control. Such concerns crop up frequently; [James Q. Wilson](#) even says, in the context of police departments, that public order and safety aren't "real" measures of overall success because whatever about them is measurable "can only partially, if at all, be affected by police behavior." When he does favor a "micro-level measure of success" of whether the neighborhood is becoming safer and more orderly, he still limits it to cases where the level of danger and disorder is "amenable . . . to improvement by a given, feasible level of police and public action." The concern in the literature over controlling for baselines is similarly motivated.

This seems mistaken: overall public order and safety *are* measures of the success of police departments, and (given that prison programs and conditions affect recidivism to some extent) lower recidivism *is* a measure of the success of prisons. It's true that these measures come with a lot of noise attached—that is, with a lot of omitted variables reflecting the contribution of other people's efforts, as well as environmental variables. But that doesn't mean it's wrong to use them for purposes of accountability, or even to tie compensation to them.

There are two concerns about using these noisy measures: first, that the level of the unobserved variables at the beginning of the contract might establish a high-recidivism baseline, for which the contractor will have to be compensated very highly, or a low-recidivism baseline, for which the contractor will collect more than it deserves; and second, that variation in the unobserved variables might create a lot of risk for the contractor.

As to the first concern, recall the earlier discussion about whether to control for baselines. Whether or not we adjust the contract price to take into account the baseline expected level of performance should have little effect on government expenditures: a high baseline translates into less quality being attributed to the contractor and thus to lower payments, and so the contractor will demand more money at the bidding stage, and vice versa.

The same reasoning addresses the second concern: because controlling for baselines doesn't affect the contractor's payout—it basically amounts to adding or subtracting a constant, which is subtracted or added right back at the bidding stage—it also doesn't necessarily affect risk.

What definitely affects risk is not the *level* of compensation, but its *slope*. A contract that compensates the contractor based on the portion of performance he was able to control isn't

necessarily less risky than one that doesn't, but a contract where the per-quality-unit price is lower is less risky. Thus, in the numerical example discussed earlier, a contract with a \$1 reward per quality unit (regardless of the fixed component of the contract) is riskier than a contract with a \$0.50 reward per quality unit; an even less risky contract is one with a \$0 reward per quality unit, that is, a fixed-price contract, which is close to the norm; and the least risky possible contract is the cost-plus contract typical of rate-of-return regulation. Compensation based on a continuous quality measure is less risky than compensation based on a discrete quality measure (as long as the provider has some chance of being on either side of the cutoff); thus, "\$1 for each quality unit" is less risky than "\$5 but only if you get five quality units."

Do we care? Perhaps large corporations like CCA or The GEO Group, which are publicly traded and diversified across many contracts, can handle the risk; and they cover three-quarters of the industry. Smaller, privately held companies like MTC may be more sensitive to risk. Various potential entrants, especially nonprofits, must be even more sensitive. Adopting high-powered (i.e., high-slope) contracts may scare away the most risk-sensitive potential bidders, leaving the field to a few large corporations. (And it isn't just a matter of risk: if the fixed part of the contract is paid up front while the reward is paid later, possibly a few years later once recidivism statistics come in, this might disadvantage small companies or nonprofits with limited access to capital markets.) This has potential implications for the competitiveness of the industry, possibilities for innovation, and the political influence that drives changes in criminal law.

But the contract doesn't have to be especially high-stakes. The optimal level of risk transfer is probably less than 100%. Rewarding the contractor for increases in quality with a price equal to the social value of quality gives the contractor great incentives but also (since the per-unit reward will be high) subjects him to high risk. Flat-fee contracts are relatively low risk but also low incentive. Some moderate level of risk transfer will optimally balance incentives with risk. Thus, the incentive-based portion of the contract is only 10% of the contract price in U.K.'s Doncaster prison, and was only 5% in the Federal Bureau of Prisons' Taft demonstration project. Recall that in Britain's Job Deal program, 30% of the payment is conditional, and only a third of that is related to "hard outcomes," and even some of those outcomes are slightly "soft."

For the cash-flow issue noted above, one can also "change the timing of payments to providers," for instance by making "a payment every six months for each offender who has not been reconvicted."

b. Financing Nonprofits: Social Impact Bonds

The need to encourage the nonprofit sector calls for innovative funding mechanisms. Nonprofit prisons have been suggested though never implemented. But in light of the widespread concern that private prison firms will cut quality to save money, the nonprofit form seems like an obvious alternative.

[Ed Glaeser and Andrei Shleifer](#) discuss the value of nonprofit status: by weakening the provider's incentives to maximize profits, nonprofit status can be a valuable signal of quality when quality itself is nonverifiable. (Even using performance measures, it's reasonable to suppose that some aspects of quality will remain nonverifiable; the value of nonprofit status depends on how important these remaining nonverifiable components are.) Moreover, altruistic entrepreneurs will tend to be attracted to the nonprofit form.

And [Timothy Besley and Maitreesh Ghatak](#) show that, when both a provider and the government can make productive investments in a project, and when the provider is altruistic, then the provider should own the project if it values it more than the government does. Privatization can thus be more beneficial in the presence of altruistic providers.

But banks or private equity houses are unlikely to finance such nonprofits, especially when the nonprofits don't have much of a track record.

[Social impact bonds](#) have been proposed as a funding mechanism for nonprofits. Rather than contracting directly with a provider, the government contracts with a middleman. This middleman, a "social impact bond-issuing organization," has two functions. First, it hires the staff to provide the service. Second, it sells bonds to investors, particularly philanthropic ones; these bonds are essentially claims to a portion of the performance-based compensation. If the service provider fulfills the performance-based goals and receives its reward from the government, the investors make money; otherwise they don't. At the Peterborough prison in the U.K., the government doesn't pay anything unless recidivism is 7.5% less than in a comparison group, and payments are capped when the difference reaches 13%. The provider's employees may well be paid something like a flat wage, so their monetary incentives aren't great; but the bond-issuing organization and the philanthropic investors (whose money is on the line) are probably better at monitoring the staff than the government would be. It remains to be seen, though, whether the philanthropic sector will provide enough funds for nonprofit prison providers to be a viable alternative to for-profit corporations.

In tomorrow's post, I'll finish up, after discussing how performance measures might lead to undesirable strategic behavior.

🗨️ 2 Comments

Sasha Volokh

Sasha Volokh lives in Atlanta with his wife and three kids, and is an associate professor at Emory Law School. He has written numerous articles and commentaries on law and economics, privatization, antitrust, prisons, constitutional law, regulation, torts, and legal history.

The Washington Post

The story must be told.

Your subscription supports journalism that matters.

Try 1 month for \$1

Podcasts

What one man's death says about the asylum court system

Apple's new heart-healthy tech might be better for the anxious and not the ailing. Also, a dead man's children seek asylum in the same court that denied him.

▶ **Listen** 21:21

2 days ago



Latest episode

What one man's death says about the asylum court system

▶ **Listen** 21:21

Unparalleled reporting. Expert insight. Clear analysis. Everything you've come to expect from the newsroom of The Post – for your ears.

**Our journalism keeps watch on Washington
and the world.**

Try 1 month for ~~\$10~~ \$1

Send me this offer