

PRISON ACCOUNTABILITY AND PERFORMANCE MEASURES

Alexander Volokh*

ABSTRACT

A few decades of comparative studies of public vs. private prison performance have failed to give a strong edge to either sector in terms of quality. That supposed market incentives haven't delivered spectacular results is unsurprising, since, by and large, market incentives haven't been allowed to work: outcomes are rarely measured and are even more rarely made the basis of compensation, and prison providers are rarely given substantial flexibility to experiment with alternative models.

This Article argues that performance measures should be implemented more widely in evaluating prisons. Implementing performance measures would advance our knowledge of which sector does a better job, facilitate a regime of competitive neutrality between the public and private sectors, promote greater clarity about the goals of prisons, and, perhaps most importantly, allow the use of performance-based contracts.

Performance measures and performance-based contracts have their critiques, for example: (1) the theoretical impossibility of knowing the proper prices, (2) the ways they would change the composition of the industry, for instance, by reducing public-interestedness or discouraging risk-averse providers, and (3) the potentially undesirable strategic behavior that would result, such as manipulation in the choice of goals, distortion of effort away from hard-to-measure dimensions or away from hard-to-serve inmates, or outright falsification of the numbers. I argue that these concerns are serious but aren't so serious as to preclude substantial further experimentation.

* Associate Professor, Emory Law School, avolokh@emory.edu. I am grateful to Michael J. Broyde, Russell C. Gabriel, Leonard Gilroy, Linda Hardyman, Erica J. Hashimoto, Peter H. Kyle, Christina Mulligan, Carl Nink, Usha Rodrigues, Joanna E. Saul, Sarah M. Shalf, Vladimir Volokh, and the participants at the Emory/UGA joint faculty colloquium for their input and assistance. I am also grateful to Kedar Bhatia and Julia Hueckel for their able research assistance, and to the law librarians at Emory Law School. Thanks also to the organizers and panelists of the Randolph W. Thrower Symposium, *Privatization: Managing Liability and Reassessing Practices in Local and International Contexts*, on February 7, 2013. A previous version of this Article was presented as the keynote address at the Vermont Law Review's symposium, *Prison Privatization: Optimizing Our Use of a Privatized Resource*, on March 23, 2013.

INTRODUCTION	342
I. THE FAILURE OF COMPARATIVE EFFECTIVENESS STUDIES	347
A. Which Sector Costs Less?	347
1. Difficulties in Calculating Costs	347
2. Competing Cost Estimates	350
a. Texas	350
b. Florida	351
c. Taft	351
B. Which Sector Provides Higher Quality?	353
1. Difficulties in Figuring Out Quality	353
2. Which Sector Leads to Less Recidivism?	357
C. The Limits of Comparative Effectiveness	361
II. WHY USE PERFORMANCE MEASURES?	364
A. The Puzzle of Prisons?	364
B. Accountability, Neutrality, and Goal Setting	369
1. To Know What Works	369
2. To Implement Competitive Neutrality	370
3. To Express What We Want	373
C. For Performance-Based Contracting	373
1. Limited Current Efforts	373
2. The Range of Possible Contracts	375
a. General Considerations	375
b. Rewards or Penalties	377
c. Controlling for Baselines	378
d. Discrete vs. Continuous Measures	382
3. The Feasibility of Merit Pay in the Public Sector	384
D. What Measures to Choose	385
III. CONCERNS AND CRITIQUES	393
A. What Prices to Set	393
B. Effects on Market Structure	396
1. Public-Interestedness	396
2. Risk and Capital Requirements	397
a. The Risk Is in the Slope	397
b. Financing Nonprofits: Social Impact Bonds	401
C. Undesirable Strategic Behavior	403
1. Manipulating the Goals	403
2. Distortion Across Dimensions of Performance	406
3. Distortion Across Types of Inmates	410
4. Falsifying Performance Measures	412

CONCLUSION	415
------------------	-----

“Here arises a feature of the Circumlocution Office, not previously mentioned in the present record. When that admirable Department got into trouble, and was, by some infuriated member of Parliament . . . attacked on the merits . . . as an Institution wholly abominable and Bedlamite; then the noble or right honourable [member] who represented it in the House, would smite that member and cleave him asunder, with a statement of the quantity of business (for the prevention of business) done by the Circumlocution Office. Then would that noble or right honourable [member] hold in his hand a paper containing a few figures, to which, with the permission of the House, he would entreat its attention. . . . Then would the noble or right honourable [member] perceive, sir, from this little document, which he thought might carry conviction even to the perversest mind . . . , that within the short compass of the last financial half-year, this much-maligned Department . . . had written and received fifteen thousand letters . . . , had made twenty-four thousand minutes . . . , and thirty-two thousand five hundred and seventeen memoranda [T]he sheets of foolscap paper it had devoted to the public service would pave the footways on both sides of Oxford Street from end to end, and leave nearly a quarter of a mile to spare for the park . . . ; while of tape—red tape—it had used enough to stretch, in graceful festoons, from Hyde Park Corner to the General Post Office. . . . No one . . . would [then] have the hardihood to hint that the more the Circumlocution Office did, the less was done, and that the greatest blessing it could confer on an unhappy public would be to do nothing.”

—Charles Dickens, *Little Dorrit*¹

“The results obtained from ENRD’s civil and criminal cases in fiscal year 2012 alone were outstanding. We secured over \$397 million in civil and stipulated penalties, cost recoveries, natural resource damages, and other civil monetary relief, including almost \$133 million recovered for the Superfund. We obtained over \$6.9 billion in corrective measures through court orders and settlements, which will go a long way toward protecting our air, water and other natural resources. We concluded 47 criminal cases against 83 defendants, obtaining nearly 21 years in confinement and over \$38 million in criminal fines, restitution, community service funds and special assessments.”

—DOJ’s *Environment & Natural Resources Division Annual Report, 2012*²

¹ CHARLES DICKENS, *LITTLE DORRIT* 517–18 (Oxford Univ. Press 1989) (1857).

² ENV’T & NATURAL RES. DIV., U.S. DEP’T OF JUSTICE, *ENRD ACCOMPLISHMENTS REPORT FISCAL YEAR 2012*, at 8 (2013).

INTRODUCTION

“Isn’t everything to be said on [private prisons] already in print?” asks Sharon Dolovich.³ She means the question to be merely rhetorical; and so do I.⁴ The comparative effectiveness debate, to the extent it’s relevant⁵—and I think it is⁶—has stalled, simply because the empirical literature, exhaustive as it is, is so bad. “The current weight of the evidence on prison privatization in the United States is so light that it defies interpretation,” write prison researcher Gerald Gaes and his coauthors.⁷ (The theory isn’t much better: the same authors characterize prison performance as a “theoretically bereft domain.”)⁸ To intelligently choose between public and private provision, we should at least know which sector costs less, but we don’t; and we should at least know which sector provides higher quality, but we don’t have a great sense of that either.⁹

This seems puzzling: readers of the voluminous debate on private prisons can be forgiven for thinking that market incentives should make private prison firms either (1) cut wasteful expenditures and produce innovative services¹⁰ or (2) cut corners on essential inmate care and security and lead to a humanitarian

³ Sharon Dolovich, *How Privatization Thinks: The Case of Prisons*, in GOVERNMENT BY CONTRACT: OUTSOURCING AND AMERICAN DEMOCRACY 128, 129 (Jody Freeman & Martha Minow eds., 2009).

⁴ Not that her perspective is the same as mine, but we both agree that there’s still something left to say on the subject.

⁵ Dolovich herself is wary of premature engagement with the comparative effectiveness debate without having sorted through the necessary normative issues beforehand. See Dolovich, *supra* note 3, at 128–29; Sharon Dolovich, *State Punishment and Private Prisons*, 55 DUKE L.J. 437, 447 n.20 (2005).

⁶ See generally Alexander Volokh, *Privatization and the Elusive Employee–Contractor Distinction*, 46 U.C. DAVIS L. REV. 133 (2012).

⁷ GERALD G. GAES ET AL., MEASURING PRISON PERFORMANCE: GOVERNMENT PRIVATIZATION AND ACCOUNTABILITY 184 (2004).

⁸ *Id.* at 123.

⁹ These aren’t the only things we should know. For instance, we can also care about where accountability is greater, which sector might be more likely to push the substantive criminal law in a more incarceration direction, and the like. See, e.g., Alexander Volokh, *Privatization and the Law and Economics of Political Advocacy*, 60 STAN. L. REV. 1197, 1199–1205 (2008); *Developments in the Law—The Law of Prisons*, 115 HARV. L. REV. 1838, 1868–91 (2002).

¹⁰ See, e.g., GEOFFREY F. SEGAL & ADRIAN T. MOORE, WEIGHING THE WATCHMEN: EVALUATING THE COSTS AND BENEFITS OF OUTSOURCING CORRECTIONAL SERVICES, PART II: REVIEWING THE LITERATURE ON COST AND QUALITY COMPARISONS 15 (Reason Pub. Policy Inst., Policy Study No. 290, 2002); Samuel Jan Brakel & Kimberly Ingersoll Gaylord, *Prison Privatization and Public Policy*, in CHANGING THE GUARD: PRIVATE PRISONS AND THE CONTROL OF CRIME 125, 134–43 (Alexander Tabarrok ed., 2003).

disaster.¹¹ Let's focus on the positive claims for private prisons: if the private sector is so clearly superior, shouldn't the difference hit us between the eyes?¹²

On second thought, this isn't so puzzling after all. The advantages of market provision are often said to be that, what with the rigidities and low-incentive structure of government agencies, private firms have greater incentive and greater flexibility to figure out how to achieve any desired level of quality. But this assumes that (1) particular levels of quality are desired or encouraged, and (2) private firms are given the flexibility to achieve these levels. It turns out that both of these assumptions are wrong.

Let's take the quality problem first. Why not tally up the quality at a public prison, do the same at a comparable private prison, and compare the two quality measures? The trouble here is that—despite the scores of studies that have been produced purporting to measure quality differences—good performance measures are rarely used. As I document in Part I, this means that comparative quality studies are hard to interpret if one wants to know which sector is better. (This hasn't prevented both partisans and detractors of private prisons from producing loosely reasoned pieces that oversell the findings of their favorite studies.)

It doesn't have to be that way. Criminologists have produced no shortage of performance measures that are appropriate for evaluating prisons, using variables like in-prison violence, the quality of prison health care, the degree of crowding, and—which I think is immensely important—recidivism.¹³ The most important thing about a performance measure is that it measure *performance*, that is, *outcomes*. *Inputs* like money spent, guards hired, or programs offered are of quite limited value, since the whole point is to see whether the money spent is worthwhile, whether the guards hired are necessary, and whether the programs are effective. *Outputs* like the number of doctor visits or the number of graduates of rehabilitative programs—like the number of memos written by Dickens's Circumlocution Office¹⁴ or the number

¹¹ See, e.g., Dolovich, *supra* note 5, at 474–80.

¹² See, e.g., Philippe C. Schmitter, *The “Organizational Development” of International Organizations*, 25 INT'L ORG. 917, 932 (1971) (calling this the “interocular impact test”).

¹³ I first (briefly) advocated performance measures for prison accountability in my student note. See *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1887–88; see also Francis T. Cullen et al., *The Accountable Prison*, 28 J. CONTEMP. CRIM. JUST. 77, 83 (2012) (“The core goal of the accountable prison is to reduce inmates' recidivism.” (italics omitted)).

¹⁴ See *supra* text accompanying note 1.

of years of prison resulting from DOJ prosecutions¹⁵—are also of limited value. Doctor visits might just be make-work; the rehabilitative programs may not actually be rehabilitative. (The Circumlocution Office, whose function is to prevent things from being done,¹⁶ has a zero or negative contribution to performance; and the prosecutions that maximize prison time aren't necessarily the same as those that most improve the environment.) What we care about—prisoner health, decent conditions, actual rehabilitation—are the *outcomes* that we should actually measure, to the extent possible.¹⁷

Why should we use performance measures? There are several reasons, which I canvass in Part II.

First, it's good just to know whether the public or private sector has higher quality, for instance in evaluating whether one's state should outsource or insource a particular project, or whether it should be one of the nineteen states that don't use private prisons at all.¹⁸ Naturally, many factors determine performance other than the quality of the management and the facilities: for instance, a prison can have better performance numbers because it was sent a better crop of people. But certainly having performance measures is better than useless.

Second, using performance measures would help to implement a regime of competitive neutrality, where the public and private sectors could bid against each other and individual projects could shuffle from one sector to another. Competitive neutrality might be better than an all-public or all-private regime, but to implement it properly, the auctions should be evenhanded, which means that proposed costs and proposed quality targets should be fairly comparable. Performance measures would allow a winning contractor to commit to deliver a particular level of performance, and would allow governments to levy the appropriate contractual fine if this level isn't achieved (or grant the appropriate reward if the level is exceeded).

¹⁵ See *supra* text accompanying note 2.

¹⁶ DICKENS, *supra* note 1, at 104–23.

¹⁷ See BERYL A. RADIN, CHALLENGING THE PERFORMANCE MOVEMENT: ACCOUNTABILITY, COMPLEXITY, AND DEMOCRATIC VALUES 15–16 (2006) (defining “input,” “output,” “outcome,” and other terms).

¹⁸ See E. ANN CARSON & WILLIAM J. SABOL, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, BULL. No. NCJ 239808, PRISONERS IN 2011, at 32 tbl.15 (2012) (listing Arkansas, Delaware, Illinois, Iowa, Maine, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, Nevada, New Hampshire, New York, North Dakota, Oregon, Rhode Island, Utah, Washington, and West Virginia as states with no inmates in private prisons in 2011).

Third, it would help policymakers express what's desirable in prisons. One would think that this had been done already; but prison contracts are written in input and output terms because this is largely how the industry works and thinks. Performance measures have been a byproduct of the debate over prison privatization: the different sides in the debate needed them to argue in favor of or against privatization; and the development of these measures has in turn spurred serious thinking about what prisons should accomplish, which has had accountability benefits for the public sector as well.

Perhaps most importantly, the use of performance measures would allow the spread of performance-based contracting, where—instead of levying a fine for not delivering a particular level of performance—one varies the contract fee continuously with the level of performance delivered. Once accountability is tied to actual performance—as is actually being done in the U.K.—giving prison providers the flexibility to choose how to do their job becomes more attractive.

Part III discusses critiques of using performance measures as part of a compensation scheme.

One concern is that the true social benefits of various aspects of performance are unknowable, either in principle or in practice, so that determining the proper prices will inevitably fail. Where a service is closely bound up with justice concerns, a focus on efficiency pricing may be inappropriate: it might demean the service or give insufficient weight to non-efficiency goals.

A second problem is that the use of performance measures will alter the composition of providers in the industry, in ways that are perhaps undesirable. One way this might happen is that, in the presence of monetary incentives, public-interested people may be less attracted to corrections. A different way performance measures can alter the composition of the industry is by increasing risk for providers. Providers can only control inputs, and the connection between inputs and outcomes is highly variable, because it depends on a great many variables, many of which are beyond the prison's control—such as general social conditions or the underlying quality of the inmates. The relationship between any of these variables and outcomes is not very well known. One might care about the fairness of rewarding or penalizing providers based on factors beyond their control, though in an auction system, such windfalls will be canceled out by competitive bidding. More seriously, the riskiness might bias the set of available providers in favor of the largest and

best-capitalized firms, and perhaps discourage experimentation with risky but promising techniques. This means that the sensitivity of price to outcomes might have to be limited, which might also limit the incentive effects.

A third problem is that providers may engage in undesirable strategic behavior. They might manipulate the performance goals so they are easy to meet. They might focus their effort on the measurable dimensions of performance and slight the unmeasurable ones. (For example, what are the true outcomes of the justice system? Some outcomes, like case backlogs, are measurable, but other important outcomes, like accuracy of adjudication, aren't—and measuring one runs the risk of distorting the agency's effort away from the unmeasured outcomes.)¹⁹ Similarly, providers will want to choose the easiest-to-treat populations (“creaming” or “cherry-picking”), and (given a population) fail to treat the hardest-to-treat members (“parking”). And, of course, any system based on particular numbers comes with the risk that someone might try to falsify the numbers.

The good news is that, for prisons, there's hope that these concerns can be fairly addressed. At the very least, these concerns don't seem so serious as to preclude far more experimentation than has been happening so far. We actually have access to reasonably good performance measures that reasonably cover the important dimensions of prison quality, none of which have to be limited to efficiency-based measures. These measures should be set by corrections departments, not by contractors. Riskiness can be addressed, at least in part, by only making part of the payment depend on performance. Social impact bonds have some promise in encouraging nonprofit-sector financing; in any event, the prison market is already highly concentrated, so there is currently no vast population of nonprofits and small companies to lose. Cherry-picking can be addressed by giving contractors no say in what inmates they're given, and parking can be addressed, at least in part, by making monetary rewards depend on observable characteristics of the inmate (if, indeed, it's a problem at all). Outright falsification of performance measures is a serious problem, which requires seriously investing in monitoring and ensuring robust disclosure regimes.

¹⁹ One might think that the reversal rate is a measure of accuracy of adjudication. But this isn't true because (1) the cases selected for appeal aren't random (in the absence of some special process to verify accuracy), and (2) given deferential standards of review, judges can work to insulate their decisions from appellate review if they're so inclined—for instance, by making them more intensely fact-based.

None of these are perfect fixes, but we don't need perfection; we just need an improvement over the status quo.

I. THE FAILURE OF COMPARATIVE EFFECTIVENESS STUDIES

Somewhat surprisingly, for all the ink spilled on private prisons over the last thirty years, we have precious little good information on what are surely some of the most important questions: when it comes to cost or quality, are private prisons better or worse than public prisons?

It's safe to say that, so far at least, the political process hasn't encouraged rigorous comparative evaluations of public and private prisons. Some states allow privatization without requiring cost and quality evaluations at all.²⁰ The nineteen states that don't privatize²¹ might, for all I know, be right to do so, but of course their stance doesn't promote comparative evaluation.

When studies are done, they're usually so inadequate from a methodological perspective that we can't reach any firm comparative conclusions. Section A below discusses the problems with cost comparison studies, and section B discusses the problems with quality comparison studies. Section C takes a broader view and notes that even well-done comparative effectiveness studies don't answer all our questions.

A. Which Sector Costs Less?

1. Difficulties in Calculating Costs

How do we determine whether the private sector costs more or less than the public sector? Ideally, we could work off of a large database of public and private prisons and run a regression in which we controlled for jurisdiction, demographic factors, size, and the like. In practice, this large database doesn't exist, and so the typical study chooses a small set of public and private prisons that are supposedly comparable.

²⁰ See Alexis M. Durham III, *Evaluating Privatized Correctional Institutions: Obstacles to Effective Assessment*, FED. PROBATION, June 1988, at 65, 67; *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1873–74.

²¹ See *supra* note 18 and accompanying text.

Unfortunately, this comparability tends to be elusive; the public and private facilities compared often “differ in ways that confound comparison of costs.”²² Sometimes no comparable facilities exist.²³ Even where there are two prisons in the jurisdiction housing inmates of the same sex and security classification, they generally differ in size, age, level of crowding, inmate age mix, inmate health mix, and facility design.²⁴ In particular, adjusting facilities to take into account different numbers of inmates is problematic, since facilities with more inmates, other things equal, benefit from economies of scale.²⁵

The GAO explained recently that “[i]t is not currently feasible to conduct a methodologically sound cost comparison of BOP [Bureau of Prisons] and private low and minimum security facilities because these facilities differ in several characteristics and BOP does not collect comparable data to determine the impact of these differences on cost.”²⁶ The data problem mostly comes from the private side: information collected by the BOP from private facilities isn’t necessarily reported the same way that public data are reported, and the reliability of the data is uncertain.²⁷ Moreover, “[w]hile private contractors . . . maintain some data for their records, these officials said that the data are not readily available or in a format that would enable a methodologically sound cost comparison at this time.”²⁸

Not only do federal regulations not require that these data be collected,²⁹ but also, and more troublingly, at the time of the GAO study in 2007, the BOP

²² DOUGLAS McDONALD ET AL., PRIVATE PRISONS IN THE UNITED STATES: AN ASSESSMENT OF CURRENT PRACTICE 33 (1998).

²³ See *id.* at 45 (making this claim about the Arizona facilities compared in CHARLES W. THOMAS, ARIZ. DEP’T. OF CORR., COMPARING THE COST AND PERFORMANCE OF PUBLIC AND PRIVATE PRISONS IN ARIZONA (1997)); see also SCOTT D. CAMP & GERALD G. GAES, FED. BUREAU OF PRISONS, PRIVATE PRISONS IN THE UNITED STATES, 1999: AN ASSESSMENT OF GROWTH, PERFORMANCE, CUSTODY STANDARDS, AND TRAINING REQUIREMENTS 15 (2000).

²⁴ See McDONALD ET AL., *supra* note 22, at 34–35; see also Robert B. Levinson, *Okeechobee: An Evaluation of Privatization in Corrections*, PRISON J., Oct. 1985, at 75, 77.

²⁵ Gerry Gaes, *Cost, Performance Studies Look at Prison Privatization*, NAT’L INST. JUST. J., Mar. 2008, at 32, 34; Douglas C. McDonald, *The Costs of Operating Public and Private Correctional Facilities, in PRIVATE PRISONS AND THE PUBLIC INTEREST* 86, 101 (Douglas C. McDonald ed., 1990).

²⁶ U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-08-6, COST OF PRISONS: BUREAU OF PRISONS NEEDS BETTER DATA TO ASSESS ALTERNATIVES FOR ACQUIRING LOW AND MINIMUM SECURITY FACILITIES 4 (2007).

²⁷ *Id.* at 12–13.

²⁸ *Id.* at 5.

²⁹ *Id.* at 13.

didn't believe there was value in developing the data collection methods that would make valid public-private cost comparison methods possible.³⁰

Probably more seriously, public and private prisons have accounting procedures that "make the very identification of comparable costs difficult."³¹

First, public systems, unlike private ones, don't spread the costs of capital assets over the life of the assets, which overstates public costs when the assets are acquired and understates them in all other years.³²

Second, various public expenditures, including employee benefits and medical care, utilities, legal work, insurance, supplies and equipment, and various contracted services, are often borne by various other agencies in government, which might understate public costs by 30%–40%.³³ One of the often-ignored costs in the public sector is the cost of borrowing capital.³⁴ Conversely, governments bear some of the costs of private firms, for instance, in various cases, contract monitoring, inspection and licensing, personnel training, inmate transportation, case management, and maintaining emergency response teams.³⁵

And third, when public or private prisons incur overhead expenditures, there's no obvious way of allocating overhead to particular facilities—Gerald Gaes gives a specific numerical example involving Oklahoma, a high-privatization state, where a difference in overhead accounting can alter the estimate of the cost of privatization by 7.4%.³⁶

³⁰ *Id.* at 7, 19, 30. The BOP's view seems to have been chiefly based on the fact that it used private contractors to run facilities for criminal aliens and wasn't expecting to receive funding to run its own. *Id.* The BOP also believed that the Taft cost study, *see infra* text accompanying notes 56–59, was already a sufficient cost study. U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 26, at 7, 19, 21, 30.

³¹ McDONALD ET AL., *supra* note 22, at 33; *accord* McDonald, *supra* note 25, at 88–89, 97–100.

³² McDONALD ET AL., *supra* note 22, at 35.

³³ *Id.* at 36.

³⁴ *See* McDonald, *supra* note 25, at 106.

³⁵ McDONALD ET AL., *supra* note 22, at 36–37.

³⁶ *See* Gerald G. Gaes, The Current Status of Prison Privatization Research on American Prisons 17–18 (Feb. 2012) (unpublished manuscript), http://works.bepress.com/gerald_gaes/1 ("Other complications arise from the appropriate treatment of property, sales, or income taxes paid by private contractors, as well as profits from inmate phone calls and commissary accounts."); *see also* McDONALD ET AL., *supra* note 22, at 37. Private companies are also loath to divulge their own financial details. *See* McDonald, *supra* note 25, at 89; *see also* OFFICE OF PROGRAM POLICY ANALYSIS & GOV'T ACCOUNTABILITY, FLA. LEGISLATURE, REPORT NO. 95-48, PERFORMANCE AUDIT OF THE GADSDEN CORRECTIONAL INSTITUTION 2 (1996); PUB. ACCOUNTS COMM., LEGISLATIVE ASSEMBLY, N.S.W. PARLIAMENT, REPORT NO. 13/53 (NO. 156), VALUE FOR MONEY FROM NSW CORRECTIONAL CENTRES 23 (2005).

As a bottom-line matter, McDonald says “the uncounted costs of public operation are probably larger than of private operation”;³⁷ I tend to agree, but it’s hard to say for sure.

2. *Competing Cost Estimates*

The best way to see the importance of various assumptions is to look at a handful of cases where different people tried to estimate the same cost. Without committing myself to which way is correct, I’ll provide three examples: from Texas in 1987, from Florida in the late 1990s, and from the federal Taft facility in 1999–2002.

a. *Texas*

In Texas, private prisons were authorized in 1987 with the passage of Senate Bill 251,³⁸ which required that private prisons show a 10% savings to the state compared to public prisons.³⁹ Calculating the per-diem cost of public incarceration in Texas thus became important, since the maximum contract price for private providers would be 90% of that cost.

The Texas Department of Corrections⁴⁰ came up with an estimate of \$27.62 per prisoner per day.⁴¹ The Legislative Budget Board, however, proposed a number of additions to this cost, to better take into account the costs of complying with *Ruiz v. Estelle*,⁴² building costs, the state’s cost to provide additional programs that private firms would be required to provide, and the like.⁴³ All these adjustments raised the estimated per-diem cost by about 50%—to \$41.67.⁴⁴ In the end, contracts were awarded within a range of \$28.72 to \$33.80—between the two estimates, though closer to the first one.⁴⁵

³⁷ McDonald, *supra* note 25, at 100.

³⁸ C. Elaine Cummins, *Private Prisons in Texas, 1987–2000*, at 15 (2000) (unpublished Ph.D. dissertation, American University) (on file with author).

³⁹ *Id.* at 42; *see also* TEX. GOV’T CODE ANN. § 495.003(c)(4) (West 2012).

⁴⁰ Now absorbed into the Texas Department of Criminal Justice. *See* TEX. DEP’T OF CRIMINAL JUSTICE, AGENCY STRATEGIC PLAN: FISCAL YEARS 2013–17, at 2 (2012).

⁴¹ Cummins, *supra* note 38, at 155.

⁴² 503 F. Supp. 1265 (S.D. Tex. 1980) (requiring the Texas Department of Corrections to alleviate overcrowding, increase the number of guards and support staff, and provide adequate health services), *aff’d in part, rev’d in part*, 679 F.2d 1115 (5th Cir. 1982), *amended in part, vacated in part*, 688 F.2d 266 (5th Cir. 1982).

⁴³ *See* Cummins, *supra* note 38, at 156–57.

⁴⁴ *Id.* at 156 tbl.9.

⁴⁵ *Id.* at 158; *see also* GAES ET AL., *supra* note 7, at 87–88. One facility received an extra \$7.41 for an “intensive substance abuse treatment program.” Cummins, *supra* note 38, at 158.

b. Florida

In Florida, the Office of Program Policy Analysis and Government Accountability (OPPAGA) compared two private facilities, Bay Correctional Facility and Moore Haven Correctional Facility, with a public facility, Lawtey Correctional Institution.⁴⁶ After various adjustments, OPPAGA calculated that the per-diem operating cost was \$46.08 at Bay and \$44.18 at Moore Haven, versus \$45.98 at Lawtey; that is, Bay was 0.2% more expensive and Moore Haven 3.9% cheaper than the public facility.⁴⁷

The Florida Department of Corrections had come up with its own numbers: \$45.04 at Bay and \$46.32 at Moore Haven, versus \$45.37 at Lawtey⁴⁸: Bay was 0.7% cheaper and Moore Haven 2.1% more expensive.

The Corrections Corporation of America (CCA), which operated Bay, submitted comments to the OPPAGA report, disputing its analysis.⁴⁹ It disagreed that Lawtey was comparable,⁵⁰ and suggested its own adjustments to OPPAGA's numbers for all three facilities. Under CCA's analysis, Bay cost \$45.16 and Moore Haven cost \$46.32, versus \$49.30 for Lawtey, which comes out to cost savings of 8.4% for Bay and 6.0% for Moore Haven.⁵¹ (OPPAGA, understandably, disputed CCA's modifications.)⁵²

c. Taft

Perhaps the best example of competing, side-by-side cost studies comes from the evaluation of the federal facility in Taft, California, operated by The GEO Group.

A Bureau of Prisons cost study by Julianne Nelson compared the costs of Taft in fiscal years 1999 through 2002 to those of three federal public facilities:

⁴⁶ OFFICE OF PROGRAM POLICY ANALYSIS & GOV'T ACCOUNTABILITY, FLA. LEGISLATURE, REPORT NO. 97-68, REVIEW OF BAY CORRECTIONAL FACILITY AND MOORE HAVEN CORRECTIONAL FACILITY 9 (1998) [hereinafter OPPAGA].

⁴⁷ *Id.*

⁴⁸ FLA. DEP'T OF CORR., 1996-97 ANNUAL REPORT (1997), available at <http://www.dc.state.fl.us/pub/annual/9697/budget.html>. These estimates were analyzed in FLA. DEP'T OF CORR., PRIVATIZATION IN THE FLORIDA DEPARTMENT OF CORRECTIONS (1998). See GAES ET AL., *supra* note 7, at 191 n.4.

⁴⁹ OPPAGA, *supra* note 46, at 55-61 (providing CCA's comments with OPPAGA's comments interspersed throughout).

⁵⁰ *Id.* at 57.

⁵¹ *Id.* at 61.

⁵² See *id.* at 59.

Elkton, Forrest City, and Yazoo City.⁵³ The Taft costs ranged from \$33.21 to \$38.62; the costs of the three public facilities ranged from \$34.84 to \$40.71.⁵⁴ Taft was cheaper than all comparison facilities and in all years, by up to \$2.42 (about 6.6%)—*except* in fiscal year 2001, when the Taft facility was more expensive than the public Elkton facility by \$0.25 (about 0.7%).⁵⁵ Stoppily averaging over all years and all comparison institutions, the savings was about 2.8%.

A National Institute of Justice study by Douglas McDonald and Kenneth Carlson⁵⁶ found much higher cost savings. They calculated Taft costs ranging from \$33.25 to \$38.37, and public facility costs ranging from \$39.46 to \$46.38.⁵⁷ Private-sector savings ranged from 9.0% to 18.4%. Again averaging over all years and all comparison institutions, the savings was about 15.0%: the two cost studies differ in their estimates of private-sector savings by a factor of about five.

Why such a difference? First, the Nelson study (but not the McDonald and Carlson study) adjusted expenditures to iron out Taft's economies of scale from handling about 300 more inmates each year than the public facilities.⁵⁸ Second, the studies differed in what they included in overhead costs, with the Nelson study allocating a far higher overhead rate.⁵⁹

These examples should be enough to give a sense of the complications in cost comparisons; given these difficulties, it's not surprising that most studies have fallen short.

⁵³ JULIANNE NELSON, *THE CNA CORP., COMPETITION IN CORRECTIONS: COMPARING PUBLIC AND PRIVATE SECTOR OPERATIONS* 10, 39 fig.4, 42 fig.5 (2005).

⁵⁴ *Id.* at 42 fig.5.

⁵⁵ *See id.* The study also compared actual GEO costs to hypothetical costs if Taft had been kept in-house. This comparison gave the edge to the public sector, *id.* at 25–26, but I don't stress this result because it's based on a comparison with a hypothetical public institution, not on actual public-sector costs.

⁵⁶ Douglas C. McDonald & Kenneth Carlson, *Contracting for Imprisonment in the Federal Prison System: Cost and Performance of the Privately Operated Taft Correctional Institution* (Oct. 1, 2005) (unpublished manuscript), <https://www.ncjrs.gov/pdffiles1/nij/grants/211990.pdf>.

⁵⁷ *Id.* at 48 tbl.2.18.

⁵⁸ Gaes, *supra* note 25, at 34.

⁵⁹ *Id.* at 34–35; Gaes, *supra* note 36, at 20.

B. Which Sector Provides Higher Quality?

1. Difficulties in Figuring Out Quality

Moving on to quality comparisons, the picture is similarly grim. As with cost comparisons, sometimes no comparable facility exists in the same jurisdiction.⁶⁰ Some studies solve *that* problem by looking at prisons in different jurisdictions, an approach that has its own problems.⁶¹ (If one had a large database with several prisons in each jurisdiction, one could control for the jurisdiction, but this approach is of course unavailable when comparing two prisons, each in its own jurisdiction.) Many studies just don't control for clearly relevant variables in determining whether a facility is truly comparable.⁶²

Often, the comparability problem boils down to differences in inmate populations; one prison may have a more difficult population than the other, even if they have the same security level. Usually prisons have different populations because of the luck of the draw,⁶³ but sometimes it's by design, as

⁶⁰ See MCDONALD ET AL., *supra* note 22, at 54–55 (discussing Arizona facilities compared in THOMAS, *supra* note 23); see also Gerald G. Gaes et al., *The Performance of Privately Operated Prisons: A Review of Research*, in MCDONALD ET AL., *supra* note 22, app. 2, at 12 (discussing Arizona facilities compared in THOMAS, *supra* note 23).

⁶¹ See MCDONALD ET AL., *supra* note 22, at 55 (discussing Charles H. Logan, *Well Kept: Comparing Quality of Confinement in Private and Public Prisons*, 83 J. CRIM. L. & CRIMINOLOGY 577 (1992)); see also CHARLES H. LOGAN, *WELL KEPT: COMPARING QUALITY OF CONFINEMENT IN A PUBLIC AND A PRIVATE PRISON* (1991).

⁶² See, e.g., Gaes et al., *supra* note 60, at 5 (criticizing the use of univariate methods in the comparison of Kentucky facilities in URBAN INST., *COMPARISON OF PRIVATELY AND PUBLICLY OPERATED CORRECTIONAL FACILITIES IN KENTUCKY AND MASSACHUSETTS* (1989)); *id.* at 18 (discussing the lack of information on characteristics of inmate populations in WILLIAM G. ARCHAMBEAULT & DONALD R. DEIS, JR., *COST EFFECTIVENESS COMPARISONS OF PRIVATE VERSUS PUBLIC PRISONS IN LOUISIANA: A COMPREHENSIVE ANALYSIS OF ALLEN, AVOYELLES, AND WINN CORRECTIONAL CENTERS* (1996)); *id.* at 19 (discussing the lack of controls for differences in number of inmates at some comparison prisons in ARCHAMBEAULT & DEIS, *supra*); see also, e.g., GAES ET AL., *supra* note 7, at 51–53 (discussing ARCHAMBEAULT & DEIS, *supra*); Scott D. Camp & Gerald G. Gaes, *Private Adult Prisons: What Do We Really Know and Why Don't We Know More?*, in *PRIVATIZATION IN CRIMINAL JUSTICE: PAST, PRESENT, AND FUTURE* 283, 287 (David Shichor & Michael J. Gilbert eds., 2001) (critiquing ARCHAMBEAULT & DEIS, *supra*, and THOMAS, *supra* note 23).

⁶³ See Gaes et al., *supra* note 60, at 4 (discussing the comparison of Kentucky facilities in URBAN INST., *supra* note 62, where the public sector had a more difficult adult population while the private sector had a more difficult juvenile population); *id.* at 9 (discussing TENN. SELECT OVERSIGHT COMM. ON CORR., *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)* (1995)); *id.* at 11 (discussing ROBERT C. THOMAS ET AL., *LEGISLATIVE BUDGET COMM., STATE OF WASH., DEPARTMENT OF CORRECTIONS PRIVATIZATION FEASIBILITY STUDY* (1996)); *id.* at 20 (criticizing the use of the Angola facility as a

happened in Arizona, when the Department of Corrections chose “to refrain from assigning prisoners to [a particular private prison] if they [had] serious or chronic medical problems, serious psychiatric problems, or [were] deemed to be unlikely to benefit from the substance abuse program that is provided at the facility.”⁶⁴ It’s actually quite common to not send certain inmates to private prisons; the most common restriction in contracts is on inmates with special medical needs.⁶⁵ Not that all prisons must have totally random assignment; it can be rational to tailor prisoner assignment to, say, the programming available at a prison. But such practices do have “the unintended effect of undermining cost comparisons.”⁶⁶ Another practice that undermines cost comparisons is contractual terms limiting the private contractor’s medical costs,⁶⁷ though nowadays it’s increasingly common for contracts to transfer all medical costs to the contractor.⁶⁸

Some performance studies rely on surveys administered to a nonrandom sample of inmates⁶⁹ or potentially biased staff surveys,⁷⁰ or generally to populations of inmates or staff that aren’t randomly assigned to public and private prisons.⁷¹ Survey data aren’t useless, but they’re rarely used with the

comparison facility in ARCHAMBEAULT & DEIS, *supra* note 62); *id.* at 20–21 (discussing that low urinalysis hit rates in ARCHAMBEAULT & DEIS, *supra* note 62, could indicate a population less inclined to use drugs, and low medical risk scores could indicate a population less in need of serious medical care).

⁶⁴ THOMAS, *supra* note 23, at 73.

⁶⁵ CAMP & GAES, *supra* note 23, at 21–22 (noting some restrictions in effect in 62.5% of the contracts surveyed; special medical needs restrictions in 50% of contracts; and other restrictions, including those for high-publicity inmates and gang members).

⁶⁶ THOMAS, *supra* note 23, at 73.

⁶⁷ See, e.g., Contract between the State of Tennessee and Corrections Corporation of America, RFS No.329.44-00408 § A.4.g.13(a) (July 1, 2007), <http://www.capitol.tn.gov/joint/committees/fiscal-review/archives/106ga/contracts/RFS%20329.44-00408%20Correction%20%28CCA%20-%20amendment%201%29.pdf> [hereinafter Tennessee CCA 2007 contract] (“If the inmate is hospitalized, the Contractor shall not be responsible for Inpatient-Hospital Costs which exceed \$4,000.00 per Inmate per admission.”); *id.* § A.4.g.13(b) (“The Contractor shall not be responsible for the cost of providing anti-retroviral medications therapeutically indicated for the treatment of Inmates with AIDS or HIV infection.”). By its terms, this contract covers services at the South Central Correctional Center, *id.* § A.1.j, and runs from 2007 to 2010, *id.* § B.1.

⁶⁸ See, e.g., Notice of Request for Proposal, Ariz. Dep’t of Corr., 5000 Minimum/Medium Security Prison Beds, Solicitation No. 110054DC § 2.14.1 (Jan. 24, 2011) (on file with author) (“There is no medical cap per inmate, per year. The Contractor shall assume all health care related costs.”).

⁶⁹ Gaes et al., *supra* note 60, at 6 (discussing DALE K. SECHREST & DAVID SHICHOR, PAROLE & CMTY. SERVS. DIV., CAL. DEP’T OF CORR., FINAL REPORT: EXPLORATORY STUDY OF CALIFORNIA’S COMMUNITY CORRECTIONAL FACILITIES (1994)).

⁷⁰ Gaes et al., *supra* note 60, at 24 (discussing staff surveys in LOGAN, *supra* note 61; Logan, *supra* note 61).

⁷¹ See GAES ET AL., *supra* note 7, at 74–76 (critiquing Judith Greene, *Comparing Private and Public Prison Services and Programs in Minnesota: Findings from Prisoner Interviews*, 11 CURRENT ISSUES CRIM.

appropriate sensitivity to its limitations.⁷² The higher-quality survey-based studies don't give the edge to either sector.⁷³

Most damningly, many studies don't rely on actual performance measures,⁷⁴ relying instead on facility audits that are largely process-based.⁷⁵ Some supposed performance measures don't necessarily indicate good performance,⁷⁶ especially when the prisons are compared based on a "laundry list" of available data items (for instance, staff satisfaction) whose relevance to good performance hasn't been theoretically established.⁷⁷

Gerald Gaes and his coauthors conclude that most studies are "fundamentally flawed," and agree with the GAO's conclusion that there is "little information that is widely applicable to various correctional settings."⁷⁸

I would add that accountability mechanisms vary widely—the standard U.S. model, the Florida model, and the U.K. model are different,⁷⁹ and these in turn differ from the French model⁸⁰ or the model proposed for prison privatization in Israel before the Israeli Supreme Court invalidated the

JUST. 202 (1999); Judith Greene, *Lack of Correctional Services*, in CAPITALIST PUNISHMENT: PRISON PRIVATIZATION & HUMAN RIGHTS (Andrew Coyle et al. eds., 2003); LOGAN, *supra* note 61; Logan, *supra* note 61); *see also* Scott D. Camp et al., *Quality of Prison Operations in the US Federal Sector: A Comparison with a Private Prison*, 4 PUNISHMENT & SOC'Y 27, 32–34 (2002) [hereinafter Camp et al., *Quality of Prison Operations*]. For a general discussion of methods, *see* Scott D. Camp et al., *Creating Performance Measures from Survey Data: A Practical Discussion*, CORRECTIONS MGMT. Q., Winter 1999, at 71.

⁷² *See* RICHARD W. HARDING, PRIVATE PRISONS AND PUBLIC ACCOUNTABILITY 115–19 (1997).

⁷³ *See* Camp et al., *Quality of Prison Operations*, *supra* note 71, at 49–50; Scott D. Camp et al., *Using Inmate Survey Data in Assessing Prison Performance: A Case Study Comparing Private and Public Prisons*, 27 CRIM. JUST. REV. 26, 31 (2002); *see also* GAES ET AL., *supra* note 7, at 83.

⁷⁴ Gaes et al., *supra* note 60, at 9 (discussing TENN. SELECT OVERSIGHT COMM. ON CORR., *supra* note 63).

⁷⁵ Not that prison audits are useless; Gerald Gaes, in fact, who is a big booster of performance measurement, discusses how audits could be improved to be made more useful. GAES ET AL., *supra* note 7, at 31–37.

⁷⁶ *See* Gaes et al., *supra* note 60, at 20 (discussing, in the context of ARCHAMBEAULT & DEIS, *supra* note 62, how a low count of disciplinary actions could indicate either good or bad performance); *id.* at 25–27 (discussing similar difficulties in interpreting items in LOGAN, *supra* note 61; Logan, *supra* note 61).

⁷⁷ Camp & Gaes, *supra* note 62, at 286 (internal quotation marks omitted).

⁷⁸ Gaes et al., *supra* note 60, at 31 (internal quotation marks omitted) (citing U.S. GEN. ACCOUNTING OFFICE, GAO/GGD-96-158, PRIVATE AND PUBLIC PRISONS: STUDIES COMPARING OPERATIONAL COSTS AND/OR QUALITY OF SERVICE 11 (1996)).

⁷⁹ *See, e.g.*, HARDING, *supra* note 72, at 158–65 (describing the "basic model" of accountability, the U.K. model, and the Florida model, and proposing a new model); David E. Pozen, *Managing a Correctional Marketplace: Prison Privatization in the United States and the United Kingdom*, 19 J.L. & POL. 253, 276–81 (2003) (comparing American and British accountability systems).

⁸⁰ *See* JON VAGG, PRISON SYSTEMS: A COMPARATIVE STUDY OF ACCOUNTABILITY IN ENGLAND, FRANCE, GERMANY, AND THE NETHERLANDS 305–07 (1994).

experiment.⁸¹ When a prison study finds some result about comparative quality, that tells us something about comparative quality *within that accountability structure*; if a private prison performed inadequately under one accountability structure, it might do better under a better one.⁸²

As an example of the problems with current quality metrics, consider the performance evaluations of the private federal Taft facility. As with the cost studies discussed above,⁸³ we have two competing studies, the National Institute of Justice one by McDonald and Carlson⁸⁴ and a Bureau of Prisons study by Scott Camp and Dawn Daggett⁸⁵—the companion paper to Julianne Nelson’s cost paper.⁸⁶

The Bureau of Prisons has evaluated public prisons by the Key Indicators/Strategic Support System since 1989.⁸⁷ Taft, alas, didn’t use that system, but instead used the system designed in the contract for awarding performance-related bonuses.⁸⁸ Therefore, McDonald and Carlson could only compare Taft’s performance with that of the public comparison prisons on a limited number of dimensions,⁸⁹ and many of these dimensions—like accreditation of the facility, staffing levels, or frequency of seeing a doctor⁹⁰—aren’t even outcomes. Taft had lower assault rates than the average of its comparison institutions, though they were within the range of observed assault rates.⁹¹ No inmates or staff were killed.⁹² There were two escapes, which was higher than at public prisons.⁹³ Drug use was also higher at Taft, as was the

⁸¹ See HCJ 2605/05 Academic Ctr. of Law & Bus., Human Rights Div. v. Minister of Fin. ¶18 [2009] (Isr.), available at http://elyon1.court.gov.il/files_eng/05/050/026/n39/05026050.n39.pdf; Volokh, *supra* note 6, at 180–85, 198–99 (discussing this opinion).

⁸² Gaes, *supra* note 36, at 30, also calls for more study of different accountability structures.

⁸³ See *supra* text accompanying notes 53–57.

⁸⁴ McDonald & Carlson, *supra* note 56.

⁸⁵ SCOTT D. CAMP & DAWN M. DAGGETT, U.S. DEP’T OF JUSTICE, FED. BUREAU OF PRISONS, EVALUATION OF THE TAFT DEMONSTRATION PROJECT: PERFORMANCE OF A PRIVATE-SECTOR PRISON AND THE BOP (2005), http://149.101.37.70/news/research_projects/published_reports/pub_vs_priv/orelappin2005.pdf.

⁸⁶ NELSON, *supra* note 53.

⁸⁷ McDonald & Carlson, *supra* note 56, at 119; see also *infra* text accompanying notes 305–06.

⁸⁸ Gaes, *supra* note 25, at 35; *infra* text accompanying note 172.

⁸⁹ McDonald & Carlson, *supra* note 56, at 119.

⁹⁰ *Id.* at 143.

⁹¹ *Id.* at 126, 127 fig.4.2. To focus on the three comparison prisons from the cost analyses, Elkton’s assault rate was similar to what would have been expected, while Taft, like Forrest City and Yazoo City, had lower rates than what would have been expected. Gaes, *supra* note 25, at 36. Yazoo City’s was the lowest. *Id.*

⁹² McDonald & Carlson, *supra* note 56, at 128.

⁹³ *Id.*

frequency of submitting grievances.⁹⁴ On this very limited analysis, Taft seems neither clearly better nor clearly worse than its public counterparts.

The Camp and Daggett study, on the other hand, created performance measures from inmate misconduct data,⁹⁵ and concluded not only that Taft “had higher counts than expected for most forms of misconduct, including all types of misconduct considered together,” but also that Taft “had the largest deviation of observed from expected values for most of the time period examined.”⁹⁶ Camp and Daggett’s performance assessment was thus more pessimistic than McDonald and Carlson’s.⁹⁷

According to Gerald Gaes, the strongest studies include one from Tennessee, which shows essentially no difference, one from Washington, which shows somewhat positive results,⁹⁸ and three more recent studies of federal prisons by himself and coauthors, which found public prisons to be equivalent to private prisons on some measures, higher on others, and lower on yet others.⁹⁹

2. Which Sector Leads to Less Recidivism?

Recidivism reduction is really just one dimension of prison quality, though it’s a particularly relevant one that deserves its own section.

If we found that inmates at private prisons were less likely to reoffend than comparable inmates at public prisons, this would be an important factor in any comparison of public and private prisons. Unfortunately, recidivism comparisons haven’t been very good either.

A study from the late 1990s by Lon Lanza-Kaduce and coauthors reported that inmates released from private prisons were less likely to reoffend than a matched sample of inmates released from public prisons, and they had less serious offenses if they did reoffend.¹⁰⁰ But this study has been critiqued on

⁹⁴ *Id.* at 143.

⁹⁵ CAMP & DAGGETT, *supra* note 85, at 35.

⁹⁶ *Id.* at 59–60.

⁹⁷ *But see infra* text accompanying notes 469–78 (discussing how misconduct rates can be misleading since they depend on accurate and unbiased reporting by prison staff).

⁹⁸ Gaes et al., *supra* note 60, at 31.

⁹⁹ Gaes, *supra* note 36, at 25–26 (citing Camp et al., *Quality of Prison Operations*, *supra* note 71; Scott D. Camp et al., *The Influence of Prisons on Inmate Misconduct: A Multilevel Investigation*, 20 JUST. Q. 501 (2003); Camp et al., *supra* note 73).

¹⁰⁰ See Lon Lanza-Kaduce et al., *A Comparative Recidivism Analysis of Releasees from Private and Public Prisons*, 45 CRIME & DELINQ. 28, 36–37 (1999) [hereinafter Lanza-Kaduce et al., *A Comparative*

various grounds.¹⁰¹ First, not all the recidivism measures are significant: while various reoffense-related rates were found to be significantly lower in the private sector,¹⁰² and while the seriousness of reoffending was found to be significantly lower in the private sector,¹⁰³ a time-to-failure analysis found that there was no significant difference in the “length of time that a releasee ‘survived’ without an arrest during the 12-month follow-up period.”¹⁰⁴ Second, the public inmates seem to not really have been well matched to the private inmates; they only seemed so when their descriptive variables were described at a high level of generality (e.g., custody level vs. “the underlying continuous score measuring custody level,” whether inmates had two or more incarcerations vs. the actual number of incarcerations, etc.).¹⁰⁵ Third, the authors seem to have made the questionable decision to assign an inmate to the sector he was released from, even if he had spent time in several sectors: thus, an inmate who spent years in public prison and was transferred to private prison shortly before his release was classified as a private prison releasee.¹⁰⁶ Fourth, a private releasee who reoffended could take longer to be entered in the system than a public releasee,¹⁰⁷ so the truly comparable number of private recidivists may well have been larger than reported.

A later study by David Farabee and Kevin Knight¹⁰⁸ that “corrected for some of these deficiencies”¹⁰⁹ found no comparative difference in the reoffense or reincarceration rates of males or juveniles over a three-year post-release

Recidivism Analysis]; see also Lonn Lanza-Kaduce et al., *The Devil in the Details: The Case Against the Case Study of Private Prisons, Criminological Research, and Conflict of Interest*, 46 CRIME & DELINQ. 92, 96–97 (2000).

¹⁰¹ The critiques are discussed in GAES ET AL., *supra* note 7, at 24–26. Gaes et al. argue, *see id.* at 27, that several of the critiques continue to apply to a later paper with a longer follow-up period, L. Lanza-Kaduce & S. Maggard, *The Long-Term Recidivism of Public and Private Prisoners* (2001) (unpublished manuscript) (paper presented at the National Conference of the Bureau of Justice Statistics and Justice Research and Statistics Association, New Orleans, 2001).

¹⁰² See Lanza-Kaduce et al., *A Comparative Recidivism Analysis*, *supra* note 100. The difference in rearrest rates is significant at the 1% level and the difference in resentencing rates is significant at the 5% level, but the differences in reincarceration rates and for any indication of recidivism are only significant at the 10% level. *Id.* at 37.

¹⁰³ *Id.* at 37–38.

¹⁰⁴ *Id.* at 38–41.

¹⁰⁵ GAES ET AL., *supra* note 7, at 25 (citing FLA. DEP’T OF CORR., BUR. OF RES. & DATA ANALYSIS, PRELIMINARY ASSESSMENT OF A STUDY ENTITLED “A COMPARATIVE RECIDIVISM ANALYSIS OF RELEASEES FROM PRIVATE AND PUBLIC PRISONS IN FLORIDA” (1998)).

¹⁰⁶ *Id.* at 26.

¹⁰⁷ *Id.*

¹⁰⁸ DAVID FARABEE & KEVIN KNIGHT, *A COMPARISON OF PUBLIC AND PRIVATE PRISONS IN FLORIDA: DURING- AND POST-PRISON PERFORMANCE INDICATORS* (2002).

¹⁰⁹ GAES ET AL., *supra* note 7, at 27.

period, though women had lower recidivism in the private sector.¹¹⁰ However, this study may still suffer from the problem of the attribution of inmates who spent some time in each sector, as well as possible selection bias to the extent that private prisons got a different type of inmate than public prisons did.¹¹¹

Another study by William Bales and coauthors,¹¹² even more rigorous,¹¹³ likewise found no statistically significant difference between public-inmate and private-inmate recidivism.¹¹⁴

A more recent study, by Andrew Spivak and Susan Sharp, reported that private prisons were (statistically) significantly worse in six out of eight models tested.¹¹⁵ But the authors noted that some skepticism was in order before concluding that public prisons necessarily did better on recidivism.¹¹⁶ Populations aren't randomly assigned to public and private prisons: that private prisons engage in "cream-skimming" is a persistent complaint.¹¹⁷ Recall the case in Arizona, where the Department of Corrections made "an effort to refrain from assigning prisoners to [the private Marana Community Correctional Facility] if they [had] serious or chronic medical problems,

¹¹⁰ FARABEE & KNIGHT, *supra* note 108, at ii–iii, 20–25.

¹¹¹ GAES ET AL., *supra* note 7, at 28.

¹¹² William D. Bales et al., *Recidivism of Public and Private State Prison Inmates in Florida*, 4 CRIMINOLOGY & PUB. POL'Y 57 (2005).

¹¹³ See Gaes, *supra* note 36, at 9.

¹¹⁴ Bales et al., *supra* note 112, at 69, 72, 74.

¹¹⁵ Andrew L. Spivak & Susan F. Sharp, *Inmate Recidivism as a Measure of Private Prison Performance*, 54 CRIME & DELINQ. 482, 500 tbl.5, 501 (2008).

¹¹⁶ See *id.* at 503.

¹¹⁷ See, e.g., ARIZ. DEP'T OF CORR., REVISED FY 2009 OPERATING PER CAPITA COST REPORT 2, 4 (2010) (discussing inmates "returned to state prisons due to an increase of their medical scores that exceeds contractual exclusions"); *id.* at 10 (explaining that "[m]edical, dental and mental health treatment is provided but to a healthier inmate population based upon contractual criteria resulting in lower overall medical costs"); *id.* at 12–16 (discussing medical, mental health, and other restrictions on inmates that can be sent to particular private prisons); ARIZ. OFFICE OF THE AUDITOR GEN., REPORT NO. 10-08, DEPARTMENT OF CORRECTIONS—PRISON POPULATION GROWTH 20 (2010) ("[P]rivate prisons do not accept inmates in need of more serious medical care . . ."); GAES ET AL., *supra* note 7, at 28; John J. DiIulio, Jr., *The Duty to Govern: A Critical Perspective on the Private Management of Prisons and Jails*, in PRIVATE PRISONS AND THE PUBLIC INTEREST, *supra* note 25, at 155, 166–67 (stating that private firms "engage in correctional creaming when they bid," meaning that they avoid bidding on facilities that they expect will "bring negative media attention, legislative inquiries, staff unrest, lawsuits, and judicial intervention"—that is, "the Atticas and Rikers Islands of the country"); Dolovich, *supra* note 5, at 505; Richard A. Oppel, Jr., *Private Prisons Found to Offer Little in Savings*, N.Y. TIMES, May 19, 2011, at A1 (discussing Arizona Department of Corrections study stating that private prisons "often house only relatively healthy inmates" and quoting State Representative Chad Campbell calling this practice "cherry-picking"). But see Gaes et al., *supra* note 60, at 34–35 (stressing that the federal Taft facility, the subject of the comparative study reported *supra* text accompanying notes 53–59, 83–94, will house inmates equivalent to those at the comparison facilities).

serious psychiatric problems, or [were] deemed to be unlikely to benefit from the substance abuse program that [was] provided at the facility.”¹¹⁸ But the phenomenon can also run the other way. One of the authors of the recidivism study, Andrew Spivak, writes that while he was “a case manager at a medium-security public prison in Oklahoma in 1998, he noted an inclination for case management staff (himself included) to use transfer requests to private prisons as a method for removing more troublesome inmates from case loads.”¹¹⁹

Moreover, recidivism data is itself often flawed.¹²⁰ Recidivism has to be not only proved (which requires good databases) but also defined.¹²¹ Recidivism isn’t self-defining—it could include arrest; reconviction; incarceration; or parole violation, suspension, or revocation; and it could give different weights to different offenses depending on their seriousness.¹²² Which definition one uses makes a difference in one’s conclusions about correctional effectiveness,¹²³ as well as affecting the scope of innovation.¹²⁴ The choice of how long to monitor obviously matters as well: “[m]ost severe offences occur in the second and third year after release.”¹²⁵ Recidivism measures might also vary because of variations in, say, enforcement of parole conditions, independent of the true recidivism of the underlying population.¹²⁶

The study of the comparative recidivism of the public and private sector could thus use a lot of improvement.¹²⁷

¹¹⁸ THOMAS, *supra* note 23, at 73; *see supra* text accompanying note 64.

¹¹⁹ Spivak & Sharp, *supra* note 115, at 503–04.

¹²⁰ *See* U.S. GEN. ACCOUNTING OFFICE, *supra* note 78, at 29–31 (discussing SECHREST & SHICHOR, *supra* note 69) (“Sufficient data were not available to adequately complete the analysis comparing the inmates released from the community correctional facilities to inmates released from other correctional institutions in the state.”); MICHAEL D. MALTZ, *RECIDIVISM* 58–60 (1984); Gaes et al., *supra* note 60, at 7.

¹²¹ *See* Brakel & Gaylord, *supra* note 10, at 154.

¹²² MALTZ, *supra* note 120, at 62; *see* ASS’N OF STATE CORR. ADM’RS, *ASCA PERFORMANCE-BASED MEASURES SYSTEM COUNTING RULES 15–24* (2013), available at http://www.asca.net/system/assets/attachments/5685/PBMS%20KeyIndicators%204_3_13.pdf.

¹²³ MALTZ, *supra* note 120, at 63; *see also* JAMES DICKER, 2020 PUB. SERVS. TRUST AT THE RSA, *CASE STUDY 2, PAYMENT-BY-OUTCOME IN OFFENDER MANAGEMENT* 16 (2011) (“[N]either reconviction nor re-imprisonment rates capture all re-offending behaviour, as only about 45% of offenders who are reconvicted are incarcerated and it is possible to be recalled to prison for breaching license conditions without being reconvicted.”).

¹²⁴ *See* DICKER, *supra* note 123, at 18.

¹²⁵ *Id.* at 16–17.

¹²⁶ *See* MALTZ, *supra* note 120, at 66–67.

¹²⁷ *See* Gaes, *supra* note 36, at 9–11 (discussing these studies).

C. *The Limits of Comparative Effectiveness*

After having read the foregoing, one should be fairly dismayed at the state of comparative public-private prison research.¹²⁸ In fact, it gets worse. An overarching problem is that most studies don't simultaneously compare *both* cost *and* quality. It is hard to draw strong conclusions from such studies, even if they are state-of-the-art at what they are examining.¹²⁹

If we find that a private prison costs less, how do we know that it did not achieve that result by cutting quality? (This is the standard critique of private prisons.)¹³⁰ If we find that a private prison costs more, how do we know that it did not cost more because of the fancy and expensive educational or rehabilitative programs it implemented?¹³¹ (According to Douglas McDonald, this was exactly the problem with the cost comparison of the Silverdale Detention Center in Hamilton County, Tennessee.)¹³²

Our goal should be to determine the production function for public and private prisons; this is the only way we will find out whether privatization moves us to a higher production possibilities frontier or merely shifts us to a

¹²⁸ Some studies are actually meta-analyses. See Gaes, *supra* note 36, at 3–6 (discussing meta-analyses and literature reviews). Two recent meta-analyses showed little difference between the public and private sectors. One, only analyzing costs, found no statistical difference between the public and private sectors. See Travis C. Pratt & Jeff Maahs, *Are Private Prisons More Cost-Effective Than Public Prisons? A Meta-Analysis of Evaluation Research Studies*, 45 CRIME & DELINQ. 358, 365, 366 tbl.2 (1999). Another, looking at both cost and quality, found that the private sector was both slightly cheaper and slightly worse; but with such small effects, the authors concluded that “prison privatization provides neither a clear advantage nor disadvantage.” See Brad W. Lundahl et al., *Prison Privatization: A Meta-Analysis of Cost and Quality of Confinement Indicators*, 19 RES. ON SOC. WORK PRAC. 383, 392 (2009). A third—more a literature review than a meta-analysis—reported that the comparison was “inconclusive,” Dina Perrone & Travis C. Pratt, *Comparing the Quality of Confinement and Cost-Effectiveness of Public Versus Private Prisons: What We Know, Why We Do Not Know More, and Where to Go from Here*, 83 PRISON J. 301, 301 (2003); and in any event there was no formal attempt to control for differences between the public and private prisons compared. See *id.* at 306.

Given that many of the underlying studies are flawed in various ways, it's not clear how you do better by aggregating them. When studies done in vastly different ways and subject to different sources of bias are aggregated in a meta-analysis, the results are “garbage in, garbage out.”

¹²⁹ See U.S. GEN. ACCOUNTING OFFICE, *supra* note 78, at 13; see also Simon Hakim & Erwin A. Blackstone, *Cost Analysis of Public and Contractor-Operated Prisons* 4, 11 (Apr. 29, 2013) (unpublished working paper) (finding long-run cost savings between 12% and 59% but devoting scarcely any attention to quality).

¹³⁰ See, e.g., Dolovich, *supra* note 5, at 474–80 (discussing the economic incentives inherent to private prison management for saving money by reducing overall quality of service).

¹³¹ See McDONALD ET AL., *supra* note 22, at 34–35; *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1875–78.

¹³² See McDonald, *supra* note 25, at 91.

different cost-quality combination on the existing frontier.¹³³ Realizing this allows us to throw out a lot of studies from the outset.

At least people are taking more seriously the need to develop valid comparisons. Governments need to mandate, by regulation or by contract, that the information necessary to do valid comparisons become available, even if collecting these extra data would add to private facilities' cost.¹³⁴ Until we get a better handle on what works, public and private prisons should be required to live up to the same standards to facilitate comparisons. Private prisons should get the same types of inmates as public prisons—neither better nor worse¹³⁵—and they should be restricted in whom they can transfer out.¹³⁶

Having spent so long bemoaning the paucity of good comparative effectiveness studies, I should note that there's more to life than comparative effectiveness. Even ignoring any differences between the public and private sectors, privatization can have systemic effects, altering how the public sector works.¹³⁷

For one thing, privatization can, for better or worse, change the public sector as well. Suppose private prisons are better than public prisons but competitive pressures lead public prisons to improve as well.¹³⁸ A comparative study may not be able to find any difference between the two sectors, and yet one can still say that privatization was a success.¹³⁹ (Indeed, one study does

¹³³ Cf. Caroline M. Hoxby, *School Choice and School Competition: Evidence from the United States*, SWEDISH ECON. POL'Y REV., no. 2, 2003, at 9, 42 ("If school choice is to be public policy, and not merely an experiment, then the question we need to answer is whether students' achievement would rise if they attended voucher or charter schools that had resources like those available to them in regular public schools. In other words, we should ask the achievement question, holding resources constant (as well as holding students' ability, motivation, and other characteristics constant).").

¹³⁴ See U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 26, at 5, 13–14, 17, 19–20, 30.

¹³⁵ See *supra* text accompanying notes 117–19.

¹³⁶ See OFFICE OF PROGRAM POLICY ANALYSIS & GOV'T ACCOUNTABILITY, STATE OF FLA., REPORT NO. 95-12, REVIEW OF CORRECTIONAL PRIVATIZATION 4 (1995) (recommending restrictions on transfers out of private prisons).

¹³⁷ Cf. Hoxby, *supra* note 133, at 19 (noting that "[school] choice can affect productivity through a variety of long-term, general equilibrium mechanisms that are not immediately available to an administrator," like bidding up the wages of successful teachers and altering the mix of people who choose teaching as a career, making parents into more informed consumers by encouraging the spread of information about schools, altering what curricula are adopted, and the like).

¹³⁸ See Charles W. Thomas, *Correctional Privatization in America: An Assessment of Its Historical Origins, Present Status, and Future Prospects*, in *CHANGING THE GUARD*, *supra* note 10, at 57, 59; see also *infra* Part II.A (discussing how privatization can improve accountability of the public sector).

¹³⁹ Cf. Hoxby, *supra* note 133, at 43 (suggesting that concentrating on the effect on student achievement of private schooling vs. public schooling is wrongheaded in the school choice debate because school choice

suggest that for prisons, privatization might drive public agencies to be more efficient,¹⁴⁰ though the statistical significance of this effect seems highly sensitive to the precise specification,¹⁴¹ and selection bias is a confounding issue.¹⁴²) Similarly, if private prisons really do cost less, and therefore allow for greater increases in capacity, thus relieving overcrowding across the board, that effect will not show up in a comparative study.¹⁴³ Likewise if best practices migrate from one sector to another through a process of cross-fertilization¹⁴⁴: Richard Harding calls this “the paradox of successful cross-fertilization—that regimes progressively become more similar than dissimilar to each other.”¹⁴⁵

Alternatively, what if privatization leads to a race to the bottom? If private prison cost-cutting is harmful, and if public prisons have to cut costs to stay competitive, we may have lower quality, including higher recidivism, across the board.¹⁴⁶

can be a success if, through competition, it leads to improvements in the public sector, so that there never emerges any difference between public and private school outcomes).

¹⁴⁰ See James F. Blumstein et al., *Do Government Agencies Respond to Market Pressures? Evidence from Private Prisons*, 15 VA. J. SOC. POL'Y & L. 446, 454 (2008); see also JAMES F. BLUMSTEIN & MARK A. COHEN, THE INTERRELATIONSHIP BETWEEN PUBLIC AND PRIVATE PRISONS: DOES THE EXISTENCE OF PRISONERS UNDER PRIVATE MANAGEMENT AFFECT THE RATE OF GROWTH IN EXPENDITURES ON PRISONERS UNDER PUBLIC MANAGEMENT? 1 (2003) (concluding that “states that have some of their prisoners in privately owned or operated prisons experience lower growth in the cost of housing their public prisoners”).

¹⁴¹ See Blumstein et al., *supra* note 140, at 465 (finding an insignificant effect with two different specifications but a significant effect with a third).

¹⁴² The authors estimate the effect using a two-stage regression where the first stage represents the probability of privatizing, but this method doesn't always take care of selection effects. See Alexander Volokh, *Do Faith-Based Prisons Work?*, 63 ALA. L. REV. 43, 67–73 (2011). Gaes also critiques the study. See Gaes, *supra* note 36, at 12–14. I have discussed or critiqued selection bias in many places. See Alexander Volokh, *Choosing Interpretive Methods: A Positive Theory of Judges and Everyone Else*, 83 N.Y.U. L. REV. 769, 803–19 (2008); Volokh, *supra* note 9, at 1245–47; Alexander Volokh, *Privatization, Free Riding, and Industry-Expanding Lobbying*, 30 INT'L REV. L. & ECON. 62, 68 (2010) [hereinafter Volokh, *Privatization, Free Riding*]; Alexander Volokh, *The Effect of Privatization on Public and Private Prison Lobbies*, in 3 PRISON PRIVATIZATION: THE MANY FACETS OF A CONTROVERSIAL INDUSTRY 7, 24–26 (Byron Eugene Price & John Charles Morris eds., 2012) [hereinafter Volokh, *The Effect of Privatization*].

¹⁴³ *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1875.

¹⁴⁴ I discuss cross-fertilization at greater length below. See *infra* text accompanying note 194.

¹⁴⁵ Richard Harding, *Private Prisons*, in 28 CRIME AND JUSTICE: A REVIEW OF RESEARCH 265, 334 (2001). But see Tony Ward, Book Review, 3 THEORETICAL CRIMINOLOGY 125, 126 (1999) (reviewing HARDING, *supra* note 72) (conceding that Harding's cross-fertilization argument is valid but noting that “[t]here seems to be a ‘heads I win, tails you lose’ quality to [Harding's cross-fertilization] argument (if public prisons turn out to be better than private ones, that just proves that competition is good for them!)”).

¹⁴⁶ See GAES ET AL., *supra* note 7, at 108; HARDING, *supra* note 72, at 138 (noting that reductions in public prisons' staffing levels in response to competition could be alternatively characterized as “cross-fertilization” or “industrial blackmail” (internal quotation marks omitted)); Gerald G. Gaes, Reaction Essay,

In either of these two cases, good empirical evaluations are necessary, though detecting such dynamic, systemwide effects will require before-and-after studies, not comparative snapshots.

Finally, to step back a bit from the privatization debate, regardless of what comparative effectiveness analysis shows, both sectors may fall short of the ideal, so this exercise should not blind us to the continuing need to reform the whole system.¹⁴⁷ I will add that, even if the public and private sectors are equivalent, one can argue against privatization on the grounds that—assuming it costs less—it enables greater expansion of the prison system and therefore may increase incarceration and hinder the search for alternative penal policies.¹⁴⁸

II. WHY USE PERFORMANCE MEASURES?

A. *The Puzzle of Prisons?*

The moral so far is that the whole empirical literature on public and private prisons is inconclusive.¹⁴⁹ As I noted in the Introduction, this should be somewhat of a puzzle for activists on both sides who claim that privatization should turn prisons into either humanitarian disaster zones or models of quality and efficiency.¹⁵⁰

Of course, that the empirical literature is inconclusive doesn't mean the sectors are equivalent; it means that current methods haven't been good enough to detect the difference. A methodologically deficient literature could hide evidence of either good or bad quality. But if the differences are great enough, you'd think they might show through even with bad methods.¹⁵¹

The tentative conclusion I draw from the literature, though, is that there may be modest, but not huge, quality differences between the sectors; the public sector is better on some dimensions and worse on others, and there's no good evidence that either sector does better at reducing recidivism. And while

Prison Privatization in Florida: Promise, Premise, and Performance, 4 CRIMINOLOGY & PUB. POL'Y 83, 87 (2005).

¹⁴⁷ See Dolovich, *supra* note 5, at 442.

¹⁴⁸ See Volokh, *supra* note 6, at 142–43 & n.30 (collecting sources making this argument).

¹⁴⁹ See also Alexander Volokh, *The Modest Effect of Minneci v. Pollard on Inmate Litigants*, 46 AKRON L. REV. 287, 324 (2013).

¹⁵⁰ See *supra* text accompanying notes 10–11.

¹⁵¹ See *supra* note 12.

the private sector is probably cheaper, it remains to be seen whether the cost savings is on the order of 15% (respectable) or on the order of 3% (somewhat negligible).¹⁵²

But this puzzle largely disappears when we consider the institutional environment of private prisons. In many areas, the private sector has been good at delivering better results at a lower cost. This is because private producers are accountable to customers who care about the quality of the end product, and because they have the flexibility to change how they do things in response to problems they may encounter. Neither of these conditions is true for private prisons—not even slightly, not even as a first approximation.

I have noted above that there is limited evidence of private firm innovation.¹⁵³ But this is because private prisons are highly constrained in how they operate. Private prison contracts essentially “‘governmentalize’ the private sector,”¹⁵⁴ reproducing public prison regulations in the private contract. Privatization can come to resemble an exercise in who can better pretend to be a public prison.¹⁵⁵

For instance, back in 1985, Robert Levinson complained of a contract with the Eckerd Foundation for the management of the Okeechobee School for Boys in which “[v]irtually every” contract item

concerned input activities and pertained to administrative/operational functions. Thus, Eckerd could have been in total compliance with all contractual provisions even if every released client committed a new offense on the first day in the community. Moreover, at no point in

¹⁵² See *supra* text accompanying notes 53–57.

¹⁵³ See, e.g., Camp & Gaes, *supra* note 62, at 287 (“In most of the literature in favor of privatization, . . . little or no detail is offered as to how . . . market pressures actually translate into real differences between public and private prisons . . .”); Scott D. Camp, *Editorial Introduction to Colloquy, Private Prisons & Recidivism*, 4 CRIMINOLOGY & PUB. POL’Y 55, 55 (2005) (stating that “little specific information is provided about why” private prison releasees should have lower recidivism); Dolovich, *supra* note 5, at 476 (noting that “[t]here is . . . little evidence of cost-saving innovation in private-sector prisons”).

¹⁵⁴ Thomas, *supra* note 138, at 64; see also *id.* at 82, 100–02, 116 n.15.

¹⁵⁵ See MCDONALD ET AL., *supra* note 22, at 49; Durham, *supra* note 20, at 67; Gaes et al., *supra* note 60, at 12 (“Generally speaking, the contract [discussed in THOMAS, *supra* note 23] stipulates that [the private provider] run the . . . facility in a manner similar to that in which the state would have operated the prison.”); *id.* at 17 (“Basically, the State of Arizona has taken the position that a private contractor should be given the opportunity to demonstrate it can [outperform] the state in running an Arizona prison according to Arizona Department of Corrections policy.”); Harding, *supra* note 145, at 303; Douglas McDonald & Carl Patten, Jr., *Governments’ Management of Private Prisons* 18 (Sept. 15, 2003) (unpublished manuscript), <https://www.ncjrs.gov/pdffiles1/nij/grants/203968.pdf>.

the contract were the criteria for noncompliance stated nor its consequences specified.¹⁵⁶

More recently, in Arizona, an auditor general report stated,

The Department requires that private prisons mirror state-operated facilities, and performs extensive oversight activities to ensure that its contractors meet its requirements. In order to maintain uniform standards for state and private prisons, the Department requires contractors to follow Department Orders, Director's Instructions, Technical Manuals, Institution Orders, and Post Orders. These requirements extend to specific details, such as following the same daily menus as state-operated facilities. Contractors may request waivers from the Department for policies that are not applicable to private prisons, such as state fiscal management practices, employee evaluations, and employee benefits.¹⁵⁷

The *same daily menus!* In Tennessee, "it even appears that private sector innovation was deliberately thwarted by making the private sector provider . . . abide by [state Department of Corrections] policy" in running the facility.¹⁵⁸

Subjecting private contractors to public regulations is actually quite common;¹⁵⁹ one exception to this trend is Florida, where public and private prisons are controlled by different agencies,¹⁶⁰ and the agency that regulates private prisons tries to balance "setting policy and encouraging innovation."¹⁶¹ More generally, input specification in private-prison contracts is routine, though of course the level of inputs specified can (and should) be "output-driven" in the sense that it's "related to output objectives."¹⁶² For instance, one

¹⁵⁶ Levinson, *supra* note 24, at 87; *see also id.* at 88 (noting that "close, coordinated monitoring of the contract by the state" may be precluded by "vague or nonexistent contract goals").

¹⁵⁷ DEBRA K. DAVENPORT, ARIZ. OFFICE OF THE AUDITOR GEN., REPORT NO. 01-13, ARIZONA DEPARTMENT OF CORRECTIONS: PRIVATE PRISONS 9 (2001); *see also* Thomas, *supra* note 138, at 101.

¹⁵⁸ Gaes et al., *supra* note 60, at 10.

¹⁵⁹ CAMP & GAES, *supra* note 23, at vii ("[P]rivate contractors were typically obligated to use the training standards and policies of the public agencies."); *see also id.* at 28. *But see id.* at ix ("[T]he private sector, even when there is no contractual obligation, has adopted the standards and policies of their public sector counterparts."); *see also id.* at 32.

¹⁶⁰ *Id.* at x, 32–33; *see also* HARDING, *supra* note 72, at 161.

¹⁶¹ CAMP & GAES, *supra* note 23, at x; *see also* Harding, *supra* note 145, at 303–04 (noting a similar situation in Western Australia).

¹⁶² HARDING, *supra* note 72, at 67–68 (internal quotation marks omitted); *see also* Peter H. Kyle, Note, *Contracting for Performance: Restructuring the Private Prison Market*, 54 WM. & MARY L. REV. 2087, 2111 (2013) ("[S]ome states have started to require the provision of vocational services . . ."). Harding does not distinguish between outputs and outcomes, *see supra* text accompanying note 17, so when he refers to outputs

can find liquidated damages provisions for certain input-based breaches like not complying with the state's policies or not filling certain required positions.¹⁶³

If inputs and procedures are highly regulated, it's not surprising that the evidence for private-sector improvements isn't overwhelming. The market is a discovery process; one shouldn't expect different methods to emerge unless innovation is permitted.

And not only permitted: one shouldn't expect different methods to emerge unless the incentives favor it. If the premise of privatization is that incentives work, particularly given the greater flexibility of private industry, micromanaging inputs and failing to incorporate the full range of desirable outcomes into the contract price means giving up on much of the possible benefit of privatization.

But the efforts to measure performance in various areas of government from the Job Training Partnership Act of 1982¹⁶⁴ and the Government Performance and Results Act of 1993¹⁶⁵—and the limited efforts to make funding contingent on those performance measures¹⁶⁶—have largely passed prisons by.

Outcome measures aren't *totally* absent. Contracts do include a limited range of outcome measures—for instance, limited penalties for escapes.¹⁶⁷ But

here, he means something like outcomes. Harding also suggests “intermediate outputs” as a synonym for “output-driven inputs,” HARDING, *supra* note 72, at 67–68 (internal quotation marks omitted); perhaps this concept is close to what I refer to as simply “outputs.” See RADIN, *supra* note 17, at 15 (defining “output” and “intermediate outcome” differently).

¹⁶³ See Leonard Gilroy, *Innovators in Action 2012: Creating a Culture of Competition to Improve Corrections*, REASON FOUND. (May 31, 2012), <http://reason.org/news/show/1012923.html>.

¹⁶⁴ Pub. L. No. 97-300, § 106(b)(1), 96 Stat. 1322, 1333, *repealed by* Workforce Investment Act of 1998, Pub. L. No. 105-220, § 199(b)(2), 112 Stat. 936, 1059 (providing that permissible performance measures for job-training organizations include “(A) placement in unsubsidized employment, (B) retention in unsubsidized employment, (C) the increase in earnings, including hourly wages, and (D) reduction in the number of individuals and families receiving cash welfare payments and the amounts of such payments”).

¹⁶⁵ Pub. L. No. 103-62, 107 Stat. 285 (codified as amended in scattered sections of 5 and 31 U.S.C.); see Laurence E. Lynn, Jr., *Requiring Bureaucracies to Perform: What Have We Learned from the U.S. Government Performance and Results Act (GPRA)?*, POLITIQUES ET MGMT. PUB., June 1999, at 1, 3; Matthew S. Schoen, Note, *Good Enough for Government Work?: The Government Performance Results Act of 1993 and Its Impact on Federal Agencies*, 32 SETON HALL LEGIS. J. 455, 467 (2008).

¹⁶⁶ See *infra* Part III.C.1.

¹⁶⁷ See, e.g., Tennessee CCA 2007 contract, *supra* note 67, § A.4.x.2 (“In the event of an escape resulting in whole or part from Contractor’s failure to perform pursuant to the provisions of this Contract, the State may seek damages in a court of competent jurisdiction.”). Note that there’s no provision for paying for escapes not

by and large, outcome-based compensation is rare.¹⁶⁸ And to the extent there are outcome-based rewards or penalties, Charles Thomas argues, “the amounts involved commonly have little or no correlation with the true magnitude of what independent contractors accomplished or failed to accomplish,” and “the dollar value of the reward or sanction is often too trivial to encourage superior performance or to deter defective performance.”¹⁶⁹ (Of course this isn’t always true: the state of Ohio recently fined CCA nearly \$500,000 for contract violations found during audits, and many of these violations were performance-relevant.)¹⁷⁰ Even developing outcome measures hasn’t been a high priority.¹⁷¹

In 1998—not that long ago—Douglas McDonald and his coauthors identified two exceptional cases of performance-based compensation: the “Bureau of Prisons’ contract with Wackenhut for the operation of the Taft Correctional Institution in California,” which allowed for “an award-fee incentive worth up to 5 percent of paid invoices,” and a District of Columbia contract with CCA for the Correctional Treatment Facility, “which permit[ted] financial rewards for meeting targets based on performance indicators.”¹⁷²

Florida recently would have taken a good step in this direction, if the bill in question¹⁷³ hadn’t been defeated. The bill would have required that private prison contracts make provision for measuring a number of dimensions of performance (though note that some of these are output measures): number of batteries, number of major disciplinary reports, percentage of negative random drug tests, number of escapes, percentage of inmates in “a facility that provides at least one of the inmate’s primary program needs,” and so on.¹⁷⁴ The number of escapes also showed up in a more specific way: the contractor would have

stemming from nonperformance—the contractor need only “exercise its best efforts to prevent escapes.” *See id.* § A.4.x.1.

¹⁶⁸ *See* Kenneth L. Avio, *The Economics of Prisons*, 6 EUR. J.L. & ECON. 143, 150 (1998); Pozen, *supra* note 79, at 282–83; Thomas, *supra* note 138, at 107 (“[I]f there are contracts that include product-oriented requirements that go beyond mere evidence of participation, then they are contracts I have never read.”).

¹⁶⁹ Thomas, *supra* note 138, at 109.

¹⁷⁰ *See Unique Private Prison Deal Leads to Backlash*, CORRECTIONAL NEWS (Mar. 6, 2013), <http://www.correctionalnews.com/articles/2013/03/6/unique-private-prison-deal-leads-backlash>.

¹⁷¹ *See* Durham, *supra* note 20, at 67.

¹⁷² *See* MCDONALD ET AL., *supra* note 22, at 52. “Wackenhut Corrections Corp. changed its name to The GEO Group in November 2003 under the terms of a share purchase agreement with another company.” Volokh, *supra* note 9, at 1229 n.131.

¹⁷³ S.B. 2038, 2012 Leg., Reg. Sess. (Fla. 2012), available at <http://flsenate.gov/Session/Bill/2012/2038/BillText/e1/PDF>.

¹⁷⁴ *Id.* sec. 1, § 944.7115(8)(f)(1)(a)–(r).

been required to reimburse the state for the costs of escapes.¹⁷⁵ The Florida bill also listed required various performance measures for work release centers.¹⁷⁶ (I discuss various other performance measures below.)¹⁷⁷

The following sections develop these themes and discuss two distinct benefits of using performance measures. The first set of advantages of using performance measures, discussed in section B, is a pure accountability advantage: we, as citizens and policymakers, would know how well our prisons are doing; we'd be better informed in deciding which sector to choose, either systemwide or on discrete projects; and we could think more clearly about what prisons should be doing. The second type of advantage, discussed in section C, goes more to harnessing incentives to improve the system over time: incorporating performance measures into contracts, and tying providers' compensation to how well they do, would give providers a reason to care about quality and simultaneously let us grant them greater flexibility. Section D discusses the normative issues involved in choosing the actual measures.

B. Accountability, Neutrality, and Goal Setting

1. To Know What Works

We all want to improve prisons. But forget about that for a moment. Even before any of these improvements were possible, performance measures would have the obvious effect of allowing us to measure performance. This would be a great step forward in researchers' ability to conduct quality studies. We would have a better sense of which sector provides better quality; combine that with better cost studies that take into account the pitfalls described above,¹⁷⁸ and we'd be better able to decide whether to be one of the nineteen states that (as of 2011) don't have private prisons.¹⁷⁹ If we do decide not to use private prisons, performance measures would help us determine which public prisons performed badly and where to look for improvement.¹⁸⁰

¹⁷⁵ *Id.* sec. 1, § 944.7115(11).

¹⁷⁶ These were "(a) The percent of employment of supervised individuals; (b) The illegal substance use by supervised individuals; (c) The victim restitution paid by supervised individuals; (d) Compliance by supervised individuals with no-contact orders; (e) The number of serious incidents occurring at the facility; and (f) The number of absconders." *Id.* sec. 1, § 944.7115(8)(f)(2)(a)–(f).

¹⁷⁷ See *infra* Part II.D.

¹⁷⁸ See *supra* Part I.A.

¹⁷⁹ See *supra* note 18 and accompanying text.

¹⁸⁰ See Aloysius Bavon, *Innovations in Performance Measurement Systems: A Comparative Perspective*, 18 INT'L J. PUB. ADMIN. 491, 493, 502 (1995) (discussing how performance measurement arose as a result of

2. To Implement Competitive Neutrality

Suppose we decide not to use private prisons. Should we then contract out the entire prison system? Probably not: someone has to be able to run a facility if the current contractor has fallen down on the job or gone bankrupt,¹⁸¹ and given how concentrated the private prison industry currently is,¹⁸² it may not always be realistic to count on being able to easily bring in a competitor when this happens.

How much of the system, then, should we privatize? The standard way to proceed is to choose particular prisons to privatize and put them up to bid to private firms, or to contract with private firms to use their own prisons. A more beneficial approach, though, would be to have a regime of “competitive neutrality,” where the public and private sector compete on the same projects.¹⁸³ The best system may be one of mixed public and private management, where private programs “complement existing public programs rather than replace them.”¹⁸⁴ (Health care reformers’ advocacy of the “public option” in health insurance was premised on a similar idea: that public participation can make competition more fair by disciplining private providers more than they would discipline each other.)¹⁸⁵

For instance, Gary Mohr, director of the Ohio Department of Rehabilitation and Correction, has talked about creating a “culture of competition” in corrections.¹⁸⁶ Ohio has pursued a combination of outsourcing and insourcing: some public prisons have been sold or their management has been contracted

the perceived inefficiency of the public sector); Marc Holzer & Arie Halachmi, *Measurement as a Means of Accountability*, 19 INT’L J. PUB. ADMIN. 1921, 1922 (1996) (arguing that measurement improves accountability of the public sector).

¹⁸¹ See HARDING, *supra* note 72, at 158 (“The state must in the last resort be able to reclaim private prisons.”); Michael J. Gilbert, *How Much Is Too Much Privatization in Criminal Justice?*, in PRIVATIZATION IN CRIMINAL JUSTICE, *supra* note 62, at 41, 76–77.

¹⁸² Volokh, *supra* note 9, at 1237–38.

¹⁸³ See WILLIAM D. EGGERS, REASON PUB. POL’Y INST., HOW-TO GUIDE NO. 18, COMPETITIVE NEUTRALITY: ENSURING A LEVEL PLAYING FIELD IN MANAGED COMPETITIONS 6 (1998); Gaes, *supra* note 36, at 24.

¹⁸⁴ Patrick Anderson et al., *Private Corrections: Feast or Fiasco?*, PRISON J., Oct. 1985, at 32, 38.

¹⁸⁵ See JACOB S. HACKER, THE CASE FOR PUBLIC PLAN CHOICE IN NATIONAL HEALTH REFORM: KEY TO COST CONTROL AND QUALITY COVERAGE 1–2 (2008), <http://ourfuture.org/report/case-public-plan-choice-national-health-reform>; see also WILLIAM A. NISKANEN, JR., BUREAUCRACY AND REPRESENTATIVE GOVERNMENT 217 (1971) (“In the 1930’s, the primary case for the creation of public power authorities was to provide a ‘yardstick’ with which to evaluate private electric utility monopolies.”).

¹⁸⁶ Gilroy, *supra* note 163.

out to the private sector, while one private prison has been taken in-house.¹⁸⁷ The result, according to Mohr, is that one can “ratchet[] up the best practices that can be created from both the public sector and multiple private vendors.”¹⁸⁸

But for this sort of system to work, we have to be able to fairly compare private-sector and public-sector bids before the fact. The cross-fertilization that’s supposed to result from competitive neutrality depends on flexibility, otherwise both sectors will try to do the same thing. But, without performance measures, flexibility undermines the ability to do the comparative analysis of bids that’s necessary to successfully implement cross-fertilization; the most straightforward way of making efficiency comparisons without performance measures is to mandate that the private sector replicate every public-sector procedure, down to the tiniest detail. And indeed, this is what Mohr did when contracting out the management of the North Central Correctional Complex facility to the private sector or when selling the Lake Erie Correctional Institution.¹⁸⁹

But with performance measures—and with an understanding of how proposed programs and methods translate into performance—he would have been able to take different proposals, translate them into expected performance, and thus have a basis for comparison, even if the proposals were radically dissimilar.¹⁹⁰ (The beliefs about expected performance would then have to be verified by evaluating the winning contractor’s performance after the fact.)

In particular, recall the problems involved in figuring out the public sector’s true costs¹⁹¹: the same problems can make for unfair competitions if public providers’ bids don’t include the costs they bear that are paid for by

¹⁸⁷ *Id.*

¹⁸⁸ *Id.*

¹⁸⁹ *See id.* (“[I]n the [request for proposals], . . . we replicated the post assignments and the staffing pattern and the policies and the food requirements. We basically said, ‘you must identify a minimum of a 5 percent savings’ from exactly the cost of what it has cost us to operate North Central.”); *see also id.* (noting that “it was the same process” with Lake Erie Correctional Institution).

¹⁹⁰ Ohio actually has performance metrics, which are a combination of output and outcome measures, covering “everything from violence indicators, to use of force indicators, to program completion indicators (GED, etc.), to recidivism data.” *Id.* But they apparently weren’t used in the way described above.

¹⁹¹ *See supra* Part I.A.1.

other departments, their different tax treatment, and the like.¹⁹² So it's not surprising that such a regime is rare in the United States.¹⁹³

One of the advantages of competitive neutrality is that—as in Ohio—prisons can be both outsourced and insourced at different times, depending on who wins the contract, so particular prisons can “churn” between the public and private sectors. The result, according to Richard Harding, would be a “process of positive cross-fertilization,”¹⁹⁴ where best practices migrate from one sector to another.¹⁹⁵ “[T]he opening up of the private sector,” Harding writes, “may heighten awareness of how sloppy public accountability has often been in the past, leading to the creation of innovative mechanisms applicable to both the private and the public sectors.”¹⁹⁶ In fact, Harding argues, systemic improvement has been one of the best consequences of privatization,¹⁹⁷ so narrowly focusing on which sector is better in a static sense is almost beside the point.¹⁹⁸

¹⁹² See EGGERS, *supra* note 183, at 1, 8–11.

¹⁹³ See Thomas, *supra* note 138, at 81, 86 (“I am aware of no example in the United States that reveals fair competition between public and private providers of correctional services. Until both of those policy failures are corrected, achieving many of the potential benefits of privatization will be impossible.”); cf. Harding, *supra* note 145, at 334 (explaining that such competition is also rare in Australia and the U.K.).

¹⁹⁴ HARDING, *supra* note 72, at 115; *accord id.* at 162; *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1890–91; Gilroy, *supra* note 163.

¹⁹⁵ Joanna Saul, Executive Director of Ohio’s Correction Institution Inspection Committee, takes a different view, stating that churning

would actually be horrific in practical application. The transition of prisons from public to private has been very difficult, with negative effects up to a year or more later. Private personnel have been very confused (or just ignorant) about the implementation of Ohio policies, which resulted in the really bad audit that they had at Lake Erie in Nov[ember]/Dec[ember] 2012. Further, the displacement of the public employees to other prisons caused a negative ripple effect across the system that will continue far into the future [because] people lost their seniority [and] the positions they had worked for, they came from different prison cultures, etc.

Comments on a draft of this Article from Joanna Saul, Exec. Dir., Ohio Corr. Inst. Inspection Comm., to Alexander Volokh, Assoc. Professor, Emory Law School (Aug. 31, 2013). Perhaps Harding’s pro-churning and Saul’s anti-churning views can be reconciled: too much churning may be more trouble than it’s worth, since one will always incur transition costs when transferring a prison to new management. But the knowledge that churning *might* happen—that is, that a company could lose the contract to operate a prison—can exert beneficial competitive pressure.

¹⁹⁶ HARDING, *supra* note 72, at 22–23.

¹⁹⁷ See Harding, *supra* note 145, at 272–73, 331–36.

¹⁹⁸ There remains the fear that, instead of systemwide improvement through cross-fertilization, we’ll get a race to the bottom, as Gaes worries. See *supra* text accompanying note 146. But good performance measures help avoid that problem.

3. *To Express What We Want*

Measuring performance would do more than just let us know which sector is better and promote cross-fertilization by facilitating a competitive neutrality regime. On an even higher level, it would encourage governments to better conceptualize what makes for a good prison—an exercise that’s long overdue.¹⁹⁹

Jon Vagg, for instance, argues that, in the U.K., private prisons “were a key factor in persuading the administration that standards were necessary, if only for the purpose of monitoring contractual compliance.”²⁰⁰ And that example isn’t just a fluke. Prisons have been operating for centuries,²⁰¹ and yet it was the experience of privatization that spurred the development of performance measures, as private-prison critics made arguments that privatization harmed quality and private-prison advocates made arguments to the contrary.²⁰² Now that performance measures exist, one can use them to evaluate both the private and the public sectors, to the benefit of both.

C. *For Performance-Based Contracting*

With performance measures, we can go further than just knowing how good public and private prisons are, implementing competitive neutrality, and formulating the proper goals of the prison system—important as all that is. We can also incorporate the performance measures into contracts and make compensation contingent on performance, finally giving prison providers strong incentives to deliver high quality.

1. *Limited Current Efforts*

Performance-based compensation is being implemented in the United States to a very limited extent. As noted above,²⁰³ 5% of the contract price at

¹⁹⁹ See DICKER, *supra* note 123, at 6 (“[P]ayment-by-outcome . . . compels commissioners to state explicitly the goals of policy.”).

²⁰⁰ VAGG, *supra* note 80, at 307.

²⁰¹ See, e.g., G. GELTNER, *THE MEDIEVAL PRISON: A SOCIAL HISTORY* (2008); RALPH B. PUGH, *IMPRISONMENT IN MEDIEVAL ENGLAND* (1968); Edward M. Peters, *Prison Before the Prison: The Ancient and Medieval Worlds*, in *THE OXFORD HISTORY OF THE PRISON: THE PRACTICE OF PUNISHMENT IN WESTERN SOCIETY* 3 (Norval Morris & David J. Rothman eds., 1995).

²⁰² See GAES ET AL., *supra* note 7, at xi, 153, 180; HARDING, *supra* note 72, at 22; cf. NISKANEN, *supra* note 185, at 217 (“[T]he case for the private supply of some public services is . . . to provide a yardstick to evaluate the performance of budget-maximizing monopoly bureaus.”).

²⁰³ See *supra* text accompanying note 172.

the Bureau of Prisons' Taft facility was performance-based. Taft was a demonstration project, which should give one a sense of how new this enterprise is.²⁰⁴

The U.K. is now on the forefront of performance-based compensation, which it calls “[p]ayment-by-outcome” or “payment-by-results.”²⁰⁵ The idea was floated in a 2008 Conservative Party Green Paper²⁰⁶ and, once the Conservative Party came into power, it was developed in a 2010 Green Paper from the Ministry of Justice.²⁰⁷ Payment-by-results is being introduced in three prisons: two private prisons, Peterborough²⁰⁸ and Doncaster,²⁰⁹ and a public prison, Leeds,²¹⁰ though the plan is to extend the model to all prisons by 2015.²¹¹ The measure is the twelve-month reconviction rate,²¹² compared to a matched comparison group. At Peterborough, performance-based “[p]ayments start when the reconviction rate of the intervention group is 7.5% less than that of the matched comparison group, with increasing returns up to a maximum rate of 13%.”²¹³ “The Peterborough pilot is the first in the world where private investors have assumed financial risk for reducing re-offending.”²¹⁴ In addition to having access to a range of prison programs to prevent recidivism, offenders at Doncaster are assigned case managers to support them during their sentence and after release, offering advice and help on employment, housing, and benefits issues.²¹⁵ (Earlier experience with payment-by-results was “primarily

²⁰⁴ Also, in Kansas, Senate Bill 14 rewards community corrections agencies for reductions in recidivism beyond a set target. See S.B. 14, 2007 Leg., Reg. Sess. (Kan. 2007); CONSERVATIVE PARTY, POLICY GREEN PAPER NO. 4, PRISONS WITH A PURPOSE: OUR SENTENCING AND REHABILITATION REVOLUTION TO BREAK THE CYCLE OF CRIME 74 (2008); Cullen et al., *supra* note 13, at 90 (listing Washington’s performance-based evaluation of treatment programs; Arizona’s Senate Bill 1476, which provides for performance-based compensation of probation departments; and other programs in California, Colorado, Illinois, and South Carolina).

²⁰⁵ DICKER, *supra* note 123, at 6.

²⁰⁶ See CONSERVATIVE PARTY, *supra* note 204, at 49, 72–75.

²⁰⁷ See MINISTRY OF JUSTICE, BREAKING THE CYCLE: EFFECTIVE PUNISHMENT, REHABILITATION AND SENTENCING OF OFFENDERS, 2010, Cm. 7972, at 38–39 (U.K.).

²⁰⁸ See DICKER, *supra* note 123, at 13.

²⁰⁹ See Wesley Johnson, *Payment-by-Results Project Bid to Cut Reoffending*, INDEPENDENT (Oct. 11, 2011), <http://www.independent.co.uk/news/uk/crime/paymentbyresults-project-bid-to-cut-reoffending-2368793.html>; John Biggin, *Innovative Rehabilitation—Payment by Results at Doncaster Prison*, GOV.UK (Oct. 13, 2011), <https://www.gov.uk/government/news/innovative-rehabilitation-payment-by-results-at-doncaster-prison>.

²¹⁰ Joe Inwood, *State-Run Leeds Prison to Be Paid on Results*, BBC NEWS (Oct. 27, 2011), <http://www.bbc.co.uk/news/uk-england-leeds-15479570>. Leeds Prison is also called Armley. *Id.*

²¹¹ *Id.*

²¹² See DICKER, *supra* note 123, at 13, 30 n.29.

²¹³ *Id.* at 13. At Doncaster, payments start when the reduction is 5%. Biggin, *supra* note 209.

²¹⁴ DICKER, *supra* note 123, at 13.

²¹⁵ There’s also a twenty-four-hour help line. Johnson, *supra* note 209; Biggin, *supra* note 209.

limited to the welfare to work market[,] where success [was] varied and limited.”)²¹⁶

A parallel program focused on finding jobs for offenders, called Job Deal, compensates providers based on employment rates.²¹⁷ Compensation is 70% fixed and 30% conditional—a third of the conditional payment is for an output measure, “successfully enrolling offenders” in the program; another third is for “a combination of outputs and processes” such as “helping clients open bank accounts”; and another third is “for achieving ‘hard outcomes.’”²¹⁸ Note, though, that even these “hard outcomes” are softer than they might seem, because they include finding a job but also include “enrolling in further learning.”²¹⁹ Some additional payment-by-results programs have also been proposed by the government or by the Social Market Foundation, focusing either on reoffending rates or on other outcomes or outputs like “drug use cessation or employment.”²²⁰

2. *The Range of Possible Contracts*

a. *General Considerations*

These examples suggest how performance-based contracts could be structured. The contract could provide that the contract price is not just the usual flat per-diem per prisoner,²²¹ but an incentive payment that—as a simple example—could vary (positively) with how many inmates find jobs or (negatively) with how many inmates are rearrested within two years.²²²

²¹⁶ CHRIS NICHOLSON, REHABILITATION WORKS: ENSURING PAYMENT BY RESULTS CUTS REOFFENDING 5 (2011); see also *id.* at 21–24 (discussing the experience with payment-by-results in the welfare to work context, characterizing the “Pathways to Work” program as unsuccessful and the “Employment Zones” program as reasonably successful).

²¹⁷ DICKER, *supra* note 123, at 13.

²¹⁸ *Id.* at 14.

²¹⁹ *Id.*

²²⁰ *Id.*

²²¹ Dolovich, *supra* note 5, at 474; see also Tennessee CCA 2007 contract, *supra* note 67, § C.3 (laying out schedule of per diems).

²²² See Kenneth L. Avio, *On Private Prisons: An Economic Analysis of the Model Contract and Model Statute for Private Incarceration*, 17 NEW ENG. J. ON CRIM. & CIV. CONFINEMENT 265, 294–95 (1991); Daniel L. Low, *Nonprofit Private Prisons: The Next Generation of Prison Management*, 29 NEW ENG. J. ON CRIM. & CIV. CONFINEMENT 1, 46 (2003); Kyle, *supra* note 162, at 2111–12; Gaes, *supra* note 36, at 23 (citing GAES ET AL., *supra* note 7).

Outcome measurements may not always be available for all dimensions of quality, so some measurement of inputs may continue to be necessary.²²³ But as far as possible, the ideal should be to make compensation contingent not on inputs like guard training, or even on outputs like the number of GEDs granted or the number of rehabilitative programs offered or ACA accreditation,²²⁴ but primarily on actual *outcomes* like the extent of unconstitutional conditions or how well prisoners are actually rehabilitated or how many prisoners get jobs.²²⁵

The amount of the bonus can be a flat fee, or it could be more complicated—in the case of recidivism bonuses, the bonus could be inmate-specific, depending on “the probability and social cost of recidivism for each inmate”—or it could even be determined by competitive bidding.²²⁶ It’s often charged that private prisons have little incentive to invest in rehabilitation,²²⁷ and in fact have an incentive to try to increase recidivism, so that they can get (at least some of) the same inmates back later; if this is so, the bonuses should be at least high enough to counteract this incentive so rehabilitating inmates is affirmatively attractive to prison firms.²²⁸

Though I focus here on *monetary* rewards and penalties, there are other possibilities. High performance could, instead of increasing a firm’s compensation in the individual contract, merely confer a reputational benefit,

²²³ Durham suggests that “process-oriented monitoring methods” continue to be used: “[A] system of frequent accounting of staffing levels can detect shortfalls in staffing that may lead to a diminution in service provision. . . . If the change in staffing levels is detected relatively quickly, efforts can be made to either restore institutional staff to initial levels or to alter the evaluation design.” Durham, *supra* note 20, at 66; *see also* DICKER, *supra* note 123, at 16 (suggesting intermediate outcomes such as drug misuse, stability of relationships, or becoming debt-free); Sidney A. Shapiro & Rena Steinzor, *Capture, Accountability, and Regulatory Metrics*, 86 TEX. L. REV. 1741, 1775, 1779 (2008); *cf.* U.S. GEN. ACCOUNTING OFFICE, GAO/T-GGD-97-151, PERFORMANCE-BASED ORGANIZATIONS: LESSONS FROM THE BRITISH NEXT STEPS INITIATIVE 7 (1997) (discussing why measuring inputs may be necessary in the context of British Next Steps agencies); Shapiro & Steinzor, *supra*, at 1779 (discussing why measuring inputs may be necessary in the context of the EPA and GPRA).

²²⁴ *See* MCDONALD ET AL., *supra* note 22, at 49 (“Correctional administrators . . . reported that 57 of the contracts in force at the end of 1997 required that facilities achieve ACA accreditation within a specified time.”).

²²⁵ *See* Kyle, *supra* note 162, at 2112–13.

²²⁶ Low, *supra* note 222, at 46; *see also infra* Part III.C.3.

²²⁷ The same charge can also be leveled against the public sector, where incentives generally aren’t strong. Currently, private prisons do invest in rehabilitative inputs as required by their contracts—something that isn’t always required in the public sector.

²²⁸ *See* Avio, *supra* note 168, at 150; Pozen, *supra* note 79, at 283–84; James Theodore Gentry, Note, *The Panopticon Revisited: The Problem of Monitoring Private Prisons*, 96 YALE L.J. 353, 362–63 (1986).

increasing its probability of winning future bids.²²⁹ One could give out certificates²³⁰ or “even simply publiciz[e] league tables of recidivism performance.”²³¹ Or one could reward good performers by giving them more flexibility in future contracts.²³²

b. Rewards or Penalties

Going back to monetary incentives, one can choose between penalties for bad performance and rewards for good performance²³³—or one could have both—though the difference needn’t be that important.

Consider a “rewards” contract that offers a \$1 per diem reward for each unit of quality on a hypothetical 0-to-10 scale, so the potential reward is \$0 to \$10. Suppose Acme Corrections Corp. expects to achieve a quality level of 5 at a total cost of \$35 per diem.²³⁴ Then it would be willing to submit a bid of \$30 or above for the project; it would just cover its costs with the \$30 payment plus the \$5 reward. (Recall that prison bids are bids on how much money the contractor will get from the government; a \$30 per diem winning bid means that the contractor will be paid \$30 per inmate-day.) Suppose bidding is competitive, other firms have similar technology, and Acme is the most efficient firm; then Acme wins the auction with its \$30 bid.²³⁵ (A less efficient firm, say one that would require \$36 per diem to achieve quality level 5, wouldn’t bid below \$31, so Acme, as a more efficient firm, would be automatically rewarded up front for its higher quality by having a better chance of winning the auction.²³⁶ The bids don’t tell us the true social cost, the true cost to the government, or the true quality—that requires waiting for the actual realized level of quality, which determines the level of the reward—but they do signal which firm is (or believes that it is) more efficient.²³⁷)

²²⁹ See CONSERVATIVE PARTY, *supra* note 204, at 73–74 (describing Avon Park Youth Academy in Florida as “a prison rewarded by results,” even though its only reward was having its contract renewed, “a decision clearly influenced” by its lower recidivism results); DICKER, *supra* note 123, at 25.

²³⁰ Burt S. Barnow, *The Effects of Performance Standards on State and Local Programs*, in *EVALUATING WELFARE AND TRAINING PROGRAMS* 277, 286 (Charles F. Manski & Irwin Garfinkel eds., 1992).

²³¹ Pozen, *supra* note 79, at 283; *accord* Cullen et al., *supra* note 13, at 86.

²³² See Barnow, *supra* note 230, at 286.

²³³ See Thomas, *supra* note 138, at 108–09.

²³⁴ This is taking into account the incentive effects of the \$1-per-unit reward. Perhaps earlier, with fixed-price contracts, Acme only achieved, say, a quality level of 3 at a total cost of \$32.

²³⁵ I discuss auction-theoretic considerations like the winner’s curse at *infra* text accompanying note 256.

²³⁶ See Gentry, *supra* note 228, at 363.

²³⁷ See also *infra* text accompanying notes 256–57.

Now consider an alternative “penalties” contract that offers a \$1 penalty for each unit of quality below 10 (i.e., 7 units of quality lead to a \$3 penalty). This contract has equivalent incentive effects to the previous one: a provider will invest in a unit of quality as long as its cost of doing so is under \$1.²³⁸ Therefore, these incentives, as before, make Acme expect to achieve the same quality level of 5, which we have seen carries a total cost of \$35 per diem. Now Acme is willing to submit a bid of \$40 or above for the project; it would just cover its cost with the \$40 payment minus the \$5 penalty. Again, with the competitive bidding assumptions listed above, Acme wins the auction with its \$40 bid.

So even though the contracts look different, they have essentially identical incentives, and any superficial differences between them are, roughly speaking, ironed out in the bidding process. The provider’s degree of risk aversion doesn’t change the result. The government can offer contracts with penalties, but then it will pay more to the winning bidder; or it can offer contracts with rewards, and the winning bidder will be satisfied with less. (One difference might be in the timing of the payments: if the base price is paid up front while rewards or penalties are processed some time later, the first contract is somewhat less valuable than the second because its payments are more delayed.)²³⁹

c. Controlling for Baselines

In the same way, it probably doesn’t make a huge difference whether the compensation takes into account the baseline level of quality.

Controlling for baselines is a huge issue in the literature on performance measures.²⁴⁰ For instance, an early paper on performance measures, by Gloria

²³⁸ Here, I’m abstracting away from behavioral factors that might make rewards more attractive than punishments. See Cass R. Sunstein, *Introduction* to BEHAVIORAL LAW AND ECONOMICS 1, 5 (Cass R. Sunstein ed., 2000); Christine Jolls et al., *A Behavioral Approach to Law and Economics*, in BEHAVIORAL LAW AND ECONOMICS, *supra* at 30–31; see also Cullen et al., *supra* note 13, at 85 (“[U]nless a manager is truly oppositional and incompetent, we would not favor the use of negative sanctions—sticks—to coerce compliance with efforts to reduce recidivism. In the long run, such meanness would risk creating collective defiance and a failed reform.”).

²³⁹ See *infra* text accompanying note 392.

²⁴⁰ See, e.g., DICKER, *supra* note 123, at 20 & fig.2 (discussing use of “performance of control groups” or a whole range of control methods); GAES ET AL., *supra* note 7, at 159 (citing Carolyn J. Heinrich, *Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness*, 62 PUB. ADMIN. REV. 712 (2002) (questioning, as characterized by Gaes, “whether outcome measures in the absence of a control or comparison group can provide meaningful information” in the context

Grizzle and coauthors, discussed methodological issues regarding what makes for a good performance measure.²⁴¹ A large part of the discussion focused on doing the proper econometric modeling to figure out the causal factors behind a performance measure.²⁴² Figuring out these causal factors is important for at least two reasons (beyond merely understanding the process). One is to have a sense of what input or output measures to use if the outcome measures aren't available in a given case.²⁴³ Another is to be able to properly assign credit, so providers who get a bad (or good) population of inmates aren't blamed (or praised) for bad (or good) results.²⁴⁴

Similarly, Gerald Gaes and his coauthors argue that “social scientists should push ultimate outcomes as far as they can be pushed,”²⁴⁵ but that, in light of the other factors that affect recidivism, “[i]t is also desirable to have more direct measures of intermediate changes to human behavior that precede desistance, and that may be influenced by criminal justice interventions.”²⁴⁶ They don't directly list desirable performance measures—they give an example of performance measures for the specific element of “Prison Security Performance,”²⁴⁷ though they stress that one should do a similar exercise for other elements of prison performance such as health care.²⁴⁸ The main characteristic of their approach is its emphasis on adequately modeling prison performance in terms of individual-level and institutional-level independent variables so that one can properly attribute credit where credit is due, avoid blaming prisons for factors beyond their control like the characteristics of the inmates, and figure out what inputs are actually important in producing prison performance.²⁴⁹ For instance, for health care, rather than measure (or in addition to measuring) the prevalence of a disease in the prison, which indicates the potential for transmission, it would be useful to use the number of

of the Job Training Partnership Act)); Barnow, *supra* note 230, at 281 (“[P]erformance management systems [could] measure outcomes relative to [a] standard. . . . [that is] set to take into account what would have occurred in the absence of the program . . .”); Kyle, *supra* note 162, at 2112 (controlling for “age, prior criminal history, and sex”); *id.* at 2113 & n.136 (controlling for crime rates).

²⁴¹ GLORIA A. GRIZZLE ET AL., BASIC ISSUES IN CORRECTIONS PERFORMANCE 4 (1982).

²⁴² *See id.* at 91.

²⁴³ *See infra* Part II.D.

²⁴⁴ *See* GRIZZLE ET AL., *supra* note 241, at 91.

²⁴⁵ GAES ET AL., *supra* note 7, at 7.

²⁴⁶ *Id.*

²⁴⁷ *Id.* at 142 tbl.10.1.

²⁴⁸ *See id.* at 141.

²⁴⁹ *See id.* at 144 (discussing differences with Logan model); *see also id.* at 4 (suggesting “develop[ing] an expected rate of crime for a community or an expected rate of misconduct for a prison based on characteristics of the people and inmates”).

cases in the incoming population as a baseline, and measure the number of new cases.²⁵⁰

Is all this necessary? Let's do our numerical example again: Consider the rewards contract discussed above, with a \$1 per diem reward for every unit of quality on a 0-to-10 scale;²⁵¹ the winning bidder, who expected to deliver quality level 5 at a cost of \$35, would have won the contract with a bid of \$30. Now consider a rewards contract that controls for the baseline level of quality; suppose the expected level of quality for this prison is 4, so a quality level of 5 would yield a reward of \$1.

The only effect of the quality adjustment is to reduce reward payments by \$4. A bidder who was willing to bid \$30 on the unadjusted contract would be willing to bid \$34 on the adjusted contract, to take into account the \$4 reduction in the expected reward. Either way, the payoff is the same to the contractor—and the price is the same to the government. The government saves \$4 on reward payments but pays it all out again in the base contract price that emerges from the auction. Jeremy Bentham argued against controlling for baselines two centuries ago:

I would make [the contractor] pay so much for every one that died, without troubling myself whether any care of his could have kept the man alive. To be sure he would make me pay for this in the contract; but as I should receive it from him afterwards, what it cost me in the long run would be no great matter. . . .

. . . [Under this system,] you need not doubt of his fondness of these his adopted children; of whom whosoever may chance while under his wing to depart this vale of tears, will be sure to leave one sincere mourner at least²⁵²

To be sure, the bidder has to have a way to figure out that the expected level of quality is 4. This requires two things. First, the bidder should have a belief about the proper model to predict the baseline quality level; different bidders can have competing beliefs about reality that lead them to different predictions. Second, it needs to have enough information about the population of inmates to plug into its model. Where either of these is absent, the contractor won't know how much to bid—this might lead to excessive

²⁵⁰ See *id.* at 38.

²⁵¹ See *supra* text accompanying notes 234–35.

²⁵² Gentry, *supra* note 228, at 362 n.52 (alterations in original) (quoting 1 JEREMY BENTHAM, PANOPTICON 71–73 (Dublin 1791)).

payments from the taxpayer's point of view or insufficient payments from the contractor's point of view—but the incentive effects will remain the same.

So while adjusting for the baseline is relevant for various reasons—it allows one to more accurately assign praise or blame, rank different facilities,²⁵³ and so on—it doesn't seem absolutely necessary for a compensation scheme to provide the proper incentives for improvement.

Moreover, risk aversion makes a difference here,²⁵⁴ but not in the way one would expect. Controlling for baselines might even *increase* risk, depending on the uncertainty in the calculation of the baseline.²⁵⁵

If the contractor gets too little, there is the concern that it might not be able to fund the project and might go bankrupt within the contractual term. But this is the same concern that happens with all bidding. Whether or not we adjust the payment for the baseline, the winning bid under a low-bid system will be subject to the “winner's curse.”²⁵⁶ As a simple example, consider many firms with identical technology. They each have slightly different models for predicting how profitable a prison will be, and firms with higher predictions will submit lower bids. At most one of these models is correct; everyone else's model is incorrect to some degree. The lowest bid will thus come from the bidder who makes the most wildly incorrect overestimate of his profits. Sophisticated bidders adjust their bids to take the winner's curse into account, but the winning bidder might either be unsophisticated or end up not having adjusted his bid enough. So the threat of contractors who go bankrupt—or of

²⁵³ See GAES ET AL., *supra* note 7, at 144 (discussing concern with rank-ordering institutions).

²⁵⁴ Recall that it didn't in the reasoning establishing the equivalence of reward and penalty contracts. See *supra* Part II.C.2.b.

²⁵⁵ Without controlling for baselines, the winning contractor gets a contract price of P and a performance-based reward R , bears costs of C , and his payoff is $P + R - C$; the variance of the payoff is $var(R) + var(C)$ if R and C are independent. Now let's control for baselines; for simplicity, assume this just involves subtracting an adjustment A from the reward, where A is determined by the expected baseline level of performance. The contract price becomes P' , and the contractor's new payoff is $P' + R - A - C$. If A has no randomness—everyone knows the government's formula and everyone knows the underlying data that the government is plugging into the formula—then $var(A) = 0$ and the variance of the new payoff is the same $var(R) + var(C)$. But if the data or the formula is somewhat uncertain, $var(A)$ is positive, so the variance of the new payoff is $var(R) + var(A) + var(C)$ if R , A , and C are independent, which is greater.

This doesn't necessarily have to happen. Suppose, for instance, that R , A , and C aren't independent, but instead there's some negative covariance among R , A , and C . Then the randomness of A might cancel out some of the randomness of R and C , and the adjustment can indeed reduce risk. The point in the text, though, is that this needn't be the case, and the adjustment, though often defended as a risk-reducing move for contractors, could end up doing the opposite.

²⁵⁶ See, e.g., PATRICK BOLTON & MATHIAS DEWATRIPONT, *CONTRACT THEORY* 283–85 (2005).

contractors who bid low and then try and hold the government up for more money²⁵⁷—is real. But, again, this happens regardless of whether we adjust for baselines. The solution is instead to require performance bonds, to rely on a track record of past performance (and restrict complete newcomers to small projects until they've proven themselves), or otherwise to try to weed out financially unsophisticated or untrustworthy parties.

d. Discrete vs. Continuous Measures

Note that, in the preceding example, the contract price varied continuously with the level of quality.²⁵⁸ Another possibility would have been to use a binary compensation scheme, where the reward or penalty is contingent on whether one reaches a particular target. This could look like “Get a fixed reward only if you achieve less than 50% recidivism.”²⁵⁹

These binary schemes, while easier to implement, are problematic in several ways. Providers who *don't* expect to be able to reach anywhere near the target have little incentive to try to achieve anything at all.²⁶⁰ Providers who *do* expect to be able to reach the target quite comfortably have little incentive to try to achieve anything additional.²⁶¹ Providers who may or may not be able to reach the target are subjected to more risk than they would bear under a continuous scheme.²⁶² Perhaps a large corporation might act somewhat

²⁵⁷ See Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. REV. 543, 574 (2000); Robert W. Poole, Jr., *Privatization*, in CONCISE ENCYCLOPEDIA OF ECONOMICS (2008), available at <http://www.econlib.org/library/Enc/Privatization.html>; Mary Sigler, *Private Prisons, Public Functions, and the Meaning of Punishment*, 38 FLA. ST. U. L. REV. 149, 155 (2010). See generally OLIVER HART, FIRMS, CONTRACTS, AND FINANCIAL STRUCTURE (1995) (discussing opportunistic behavior in contract relationships).

²⁵⁸ Well, the example as worded involved discrete jumps, but one can easily imagine the prorated version. The “continuous” scheme is also called a “distance travelled” scheme. DICKER, *supra* note 123, at 16 (internal quotation marks omitted); see *infra* text accompanying notes 422–25, 436.

²⁵⁹ See HARDING, *supra* note 72, at 68 (“*x* per cent of participants [in a remedial literacy class] reaching attainment level *y* in *z* months.”).

²⁶⁰ See DICKER, *supra* note 123, at 19 (explaining that a continuous measure “may incentivise providers to engage with high-risk offenders who are unlikely to achieve absolute desistance”); HARDING, *supra* note 72, at 68.

²⁶¹ On the other hand, incentives are very large for those who could be just under the cutoff but could also reach the cutoff; but even then, unless the cutoff is a magical point, it's probably more socially optimal to provide continuous incentives.

²⁶² Kyle also notes the following advantage of a sliding scale: it “would reduce the likelihood that private companies would receive an undeserved windfall—the farther in standard deviations from the mean the private prison is, the more likely a causal relationship that should be rewarded exists.” Kyle, *supra* note 162, at 2112. More accurately, this depends on the likely effect of rehabilitative measures versus the likely magnitude of unobserved factors: it could be that a truly exceptional performance in fact reflects an unusually (and unobservedly) good or rehabilitable crop of inmates.

risk-neutrally, so risk won't matter; but smaller firms or nonprofits may refrain from bidding, or may require more money to take the project, or may be reluctant to try high-expected-value but risky strategies.²⁶³

(Of course, one could also imagine intermediate reward schemes: for example, the reward could be almost flat for any level of recidivism above 50% and increase rapidly at or below 50%, for instance, "Get a reward of \$0.01 for every percentage-point reduction of recidivism below 100% and down to 50%, and then a reward of \$1.00 for every percentage-point reduction beyond 50%."²⁶⁴ British performance contracts, where payments don't start until the decrease in recidivism is 5% or 7.5%, and where payments are capped once the decrease is high enough, fit this mold.²⁶⁵ At this point I won't do anything more than signal the existence of such contracts, though the optimal slope of the compensation scheme is something I'll return to below when I discuss risk allocation.²⁶⁶)

The same is true of penalties that may occur during the contractual term. Governments can terminate their contracts²⁶⁷—this is a form of binary scheme—though this is a rare remedy that tends to be reserved for the most extreme abuses.²⁶⁸ Providing for graduated financial penalties for abuses of different severity is probably a better solution than merely providing for contract rescission, because draconian penalties are less likely to be used. Not that termination isn't appropriate in extreme cases—governments should always retain the ability to take over a prison if a contract is terminated.²⁶⁹ The need to retain a credible threat of termination is one reason to prefer that

²⁶³ See *infra* Part III.B.2. Some also mention the possibility that the public could see the continuous measure as being "too lenient." See, e.g., DICKER, *supra* note 123, at 20.

²⁶⁴ See DICKER, *supra* note 123, at 24 ("[C]reate a minimum threshold of achievement that providers must attain before payments commence."); *id.* at 25 (discussing a "target accelerator," where increases are rewarded at an increasing rate).

²⁶⁵ See *supra* note 213 and accompanying text.

²⁶⁶ See *infra* Part III.B.2.

²⁶⁷ See Tennessee CCA 2007 contract, *supra* note 67, § D.3 ("The State may terminate this Contract without cause for any reason."); *id.* § D.4 ("If the Contractor fails to properly perform its obligations under this Contract in a timely or proper manner, or if the Contractor violates any terms of this Contract, the State shall have the right to immediately terminate the Contract and withhold payments in excess of fair compensation for completed services.").

²⁶⁸ See Dolovich, *supra* note 5, at 495–500; *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1883–84.

²⁶⁹ See *supra* text accompanying note 181.

governments, not prison firms, own the prisons, since government ownership of the physical facility reduces termination costs.²⁷⁰

3. *The Feasibility of Merit Pay in the Public Sector*

Note, also, that while I've been primarily concentrating on incentives for private firms, there's no inherent reason why performance-based compensation can't also be considered for public prison wardens²⁷¹—consider the example of Leeds noted above²⁷²—especially if we simultaneously pursue competitive neutrality.²⁷³ As John Donahue says, “the fundamental distinction is between competitive output-based relationships and noncompetitive input-based relationships rather than between profit-seekers and civil servants per se.”²⁷⁴ Proposals to reward public servants for high performance aren't rare,²⁷⁵ and merit-based compensation in the public sector has increased in recent years,²⁷⁶ but it's still hard to find in corrections.²⁷⁷

Researchers differ on how feasible merit pay is in the public sector;²⁷⁸ I won't resolve the argument here, except to note that the Government Performance and Results Act of 1993 has a procedure by which agencies can make “proposals to waive administrative procedural requirements and controls, including specification of personnel staffing levels, limitations on compensation or remuneration, and prohibitions or restrictions on funding transfers . . . in return for specific individual or organization accountability to

²⁷⁰ See Levinson, *supra* note 24, at 90 (discussing the “possibility of [a] contractor’s bankruptcy which would require rapid, costly interim arrangements”).

²⁷¹ See Cullen et al., *supra* note 13, at 84; Rick Hills, *Merit Pay for Prison Wardens?*, PRAWFSBLAWG (Mar. 3, 2008), <http://prawfsblawg.blogspot.com/2008/03/tying-the-salar.html>.

²⁷² See *supra* text accompanying notes 210.

²⁷³ See *supra* text accompanying notes 183–97.

²⁷⁴ JOHN D. DONAHUE, *THE PRIVATIZATION DECISION: PUBLIC ENDS, PRIVATE MEANS* 82 (1989) (italics omitted).

²⁷⁵ See NISKANEN, *supra* note 185, at 201–09; Barnow, *supra* note 230, at 307–08; Lynn, *supra* note 165, at 11; cf. David N. Figlio & Lawrence W. Kenny, *Individual Teacher Incentives and Student Performance*, 91 J. PUB. ECON. 901, 903 (2007) (examining effects of teacher merit pay).

²⁷⁶ See Jon D. Michaels, *Privatization’s Progeny*, 101 GEO. L.J. 1023, 1048 & nn.124–25, 1049 (2013).

²⁷⁷ See Thomas, *supra* note 138, at 109.

²⁷⁸ Compare Harding, *supra* note 145, at 304 (“The financial incentive should drive performance in a way that is impossible in the state-funded public sector.”), and McDonald & Patten, *supra* note 155, at xxvii (“When structuring contracts, [governments] also have opportunities to create incentives and mechanisms for accountability that are more difficult to implement in existing public organizations.”), with GAES ET AL., *supra* note 7, at 151 (“There is certainly no reason why public administrators cannot award bonuses to the best performing public prison managers and their employees, while also demoting, firing, or transferring the managers who are substandard.”), and *id.* at 180 (“Contrary to the point of view of some scholars, we do not see how a contract offers an advantage over public provision.”).

achieve a performance goal.”²⁷⁹ Any such proposal, according to the statute, must “describe the anticipated effects on performance resulting from greater managerial or organizational flexibility, discretion, and authority, and . . . quantify the expected improvements in performance resulting from any waiver,”²⁸⁰ “precisely express the monetary change in compensation or remuneration amounts, such as bonuses or awards, that shall result from meeting, exceeding, or failing to meet performance goals,”²⁸¹ and be “endorsed by the agency that established the requirement.”²⁸² Just reading the statutory language—and this is a statute that purports to *encourage* flexibility—doesn’t exactly give one confidence that public-sector flexibility is easy to come by, at least in the federal system.

At the very least, though, to the extent performance-based compensation is a good idea in the private sector, it may well also be a good idea in the public sector.²⁸³ How feasible that is is a question of the relevant state or federal law.

D. What Measures to Choose

The earlier discussion of how to define recidivism²⁸⁴ 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.²⁸⁵

²⁷⁹ Government Performance and Results Act of 1993 § 5(a), 31 U.S.C. § 9703(a) (2006).

²⁸⁰ *Id.* § 9703(b).

²⁸¹ *Id.* § 9703(c).

²⁸² *Id.* § 9703(d).

²⁸³ Some of the disadvantages of performance-based compensation may apply with different force in the public than in the private sector. For instance, the concern that market incentives will discourage public-interested people from entering the industry, *see infra* Part III.B.1, seems to not apply at all to private providers, who are presumably already profit motivated.

²⁸⁴ *See supra* text accompanying notes 120–26.

²⁸⁵ *See infra* Part III.C.2. This section only covers what measures should rationally be chosen, not the real-world possibilities for manipulation in the choice of goals. That sort of strategic behavior is covered *infra* Part III.C.1.

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.”²⁸⁶ “‘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.”²⁸⁷ Similarly, “[r]eforming criminals” is nonoperational, while “[d]oubling the rate of inmate participation in prison industry programs” is operational.²⁸⁸ 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.²⁸⁹ 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.²⁹⁰ 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.²⁹¹

A useful way to explore this question is to examine some existing prison performance measures. Perhaps one of the oldest formal approaches²⁹² to measuring prison performance is the Correctional Institutions Environment Scale²⁹³ developed by Rudolph Moos in the late 1960s²⁹⁴ and often used in the

²⁸⁶ John J. DiIulio, Jr., *Measuring Performance When There Is No Bottom Line*, in PERFORMANCE MEASURES FOR THE CRIMINAL JUSTICE SYSTEM 142, 144 (1993).

²⁸⁷ *Id.*

²⁸⁸ *Id.* (internal quotation marks omitted).

²⁸⁹ See, e.g., Cullen et al., *supra* note 13, at 87 (“[T]he objectives should be stated in a concrete, unambiguous way: ‘The XYZ prison will reduce recidivism of released high-risk offenders so that no more than 20% are arrested within 1 year.’”).

²⁹⁰ See *supra* text accompanying notes 258–63.

²⁹¹ Of course, one should also set the weights to be put on the various measures in the index. See *infra* Part III.A; cf. GRIZZLE ET AL., *supra* note 241, at 80; Barnow, *supra* note 230, at 284 (“Even if the program has a single objective, it may be advantageous to use several measures as proxies if an ideal measure cannot be developed.”). Realistically, the number of measures shouldn’t be too large, lest it overwhelm decision-makers’ cognitive capacities. One can’t think about all things simultaneously.

²⁹² A survey article in 1975 reviewed 231 studies of particular performance measures, but at that time, in the authors’ opinions, there had apparently never been any comprehensive approach. (Presumably the Moos approach, if it was considered, was thought to be insufficiently comprehensive or not performance oriented.) The American Correctional Association had published comprehensive standards in the late 1970s, but they were primarily process oriented. See GRIZZLE ET AL., *supra* note 241, at 4 (citing DOUGLAS LIPTON ET AL., THE EFFECTIVENESS OF CORRECTIONAL TREATMENT: A SURVEY OF TREATMENT EVALUATION STUDIES (1975); AM. CORR. ASS’N, MANUAL OF STANDARDS FOR ADULT CORRECTIONAL INSTITUTIONS (1977)).

²⁹³ Michael Montgomery, *Performance Measures and Private Prisons*, in 3 PRISON PRIVATIZATION, *supra* note 142, at 187, 193.

1970s.²⁹⁵ The Moos scale contains several subscales: “Involvement,” “Support,” “Expressiveness,” “Autonomy,” “Practical Orientation,” “Personal Problem Orientation,” “Order and Organization,” “Clarity,” and “Staff Control.”²⁹⁶ 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.²⁹⁷ 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,²⁹⁸ and even to what extent they described a real phenomenon.²⁹⁹ 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.”³⁰⁰ Ultimately, the scale was “determined not to possess acceptable validity.”³⁰¹

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.³⁰² Only the fourth category is explicitly labeled “Intermediate Products,”³⁰³ but some of the other measures are also outputs, not outcomes—

²⁹⁴ Kevin N. Wright & James Boudouris, *An Assessment of the Moos Correctional Institutions Environment Scale*, 19 J. RES. IN CRIME & DELINQ. 255, 255 (1982).

²⁹⁵ See *id.* (citing sources using the Moos scale in the 1970s).

²⁹⁶ *Id.* at 257 (quoting RUDOLF H. MOOS, *EVALUATING CORRECTIONAL AND COMMUNITY SETTINGS* 41 (1975)).

²⁹⁷ *Id.*

²⁹⁸ *Id.* at 256; Elaine Selo, Book Review, 4 J. CRIM. JUST. 348, 349 (1976) (reviewing MOOS, *supra* note 296).

²⁹⁹ Wright & Boudouris, *supra* note 294, at 258.

³⁰⁰ *Id.* at 274.

³⁰¹ Montgomery, *supra* note 293, at 193.

³⁰² MARTHA R. BURT, *URBAN INST., MEASURING PRISON RESULTS: WAYS TO MONITOR AND EVALUATE CORRECTIONS PERFORMANCE*, at ii (1980).

³⁰³ *Id.* at 97–105.

see, for instance, the use of hospitalizations and sick days in the measures of inmate health.³⁰⁴

The mixing of output and outcome measures is fairly typical; John DiIulio criticizes the BOP's Key Indicators/Strategic Support System³⁰⁵ for also "indiscriminate[ly] mixing . . . process [i.e., input or output] and performance [i.e., outcome] measures."³⁰⁶ 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.)"³⁰⁷; 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)."³⁰⁸

The American Correctional Association's performance-based standards for correctional health care³⁰⁹ raise the same issue. Some of these are true outcomes, like "the rate of positive tuberculin skin tests"³¹⁰ or the suicide rate,³¹¹ though others are process measures or expected practices, like whether an offender "is informed about access to health systems and the grievance procedure."³¹² The Prison Social Climate Survey, which is based on inmate and staff surveys, likewise mixes outcomes (such as crowding³¹³ or safety³¹⁴)

³⁰⁴ *Id.* at 72.

³⁰⁵ See generally WILLIAM G. SAYLOR, FED. BUREAU OF PRISONS, DEVELOPING A STRATEGIC SUPPORT SYSTEM: MONITORING THE BUREAU'S PERFORMANCE VIA TRENDS IN KEY INDICATORS (1988).

³⁰⁶ DiIulio, *supra* note 286, at 150–52.

³⁰⁷ John J. DiIulio, Jr., *Recovering the Public Management Variable: Lessons from Schools, Prisons, and Armies*, 49 PUB. ADM. REV. 127, 129 (1989) (citing JOHN J. DI IULIO, JR., GOVERNING PRISONS: A COMPARATIVE STUDY OF CORRECTIONAL MANAGEMENT (1987)).

³⁰⁸ MTC INST., MEASURING SUCCESS: IMPROVING THE EFFECTIVENESS OF CORRECTIONAL FACILITIES 5 (2006).

³⁰⁹ AM. CORR. ASS'N, PERFORMANCE-BASED STANDARDS FOR CORRECTIONAL HEALTH CARE IN ADULT CORRECTIONAL INSTITUTIONS (2002). These standards are discussed in GAES ET AL., *supra* note 7, at 37–38.

³¹⁰ GAES ET AL., *supra* note 7, at 37.

³¹¹ *Id.* at 38.

³¹² *Id.* at 37.

³¹³ Michael W. Ross et al., *Measurement of Prison Social Climate: A Comparison of an Inmate Measure in England and the USA*, 10 PUNISHMENT & SOC'Y 447, 460–61 tbl.3 (2008).

with outputs (such as whether the prison is a pleasant place to work for staff³¹⁵).

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.³¹⁶ 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')."³¹⁷

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.³¹⁸ The "safety" category contains safety of inmates, safety of staff, dangerousness of inmates, safety of environment, and (again) staffing adequacy.³¹⁹

³¹⁴ *Id.* at 463, 466–67 tbl.5.

³¹⁵ See WILLIAM G. SAYLOR ET AL., FED. BUREAU OF PRISONS, PRISON SOCIAL CLIMATE SURVEY: RELIABILITY AND VALIDITY ANALYSES OF THE WORK ENVIRONMENT CONSTRUCTS 3–8 (1996); see also *supra* text accompanying note 87.

³¹⁶ Charles H. Logan, *Criminal Justice Performance Measures for Prisons*, in PERFORMANCE MEASURES FOR THE CRIMINAL JUSTICE SYSTEM, *supra* note 286, at 19; see GAES ET AL., *supra* note 7, at xi (calling Logan's approach "one serious attempt to develop a coherent theoretical and empirical approach to prison performance measurement"); *id.* at 5–8 (discussing Logan's model). Joan Petersilia has also developed performance measures for community corrections. See Joan Petersilia, *Measuring the Performance of Community Corrections*, in PERFORMANCE MEASURES FOR THE CRIMINAL JUSTICE SYSTEM, *supra* note 286, at 60, 77–78. But many of these are input measures ("Number and type of supervision contacts"), output measures ("Number of hours/days performed community service"), or outcome measures that can be easily gamed ("Number of arrests and technical violation[s] during supervision"). *Id.* at 77–78.

³¹⁷ Logan, *supra* note 316, at 27–32.

³¹⁸ *Id.* at 34.

³¹⁹ *Id.*

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.³²⁰ In all—over all eight dimensions—there are a few hundred measures.³²¹ Logan used this index to evaluate three women’s prisons in New Mexico and West Virginia.³²²

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”;³²³ what we measure “ought to be achievable and measurable mostly within the prison itself.”³²⁴ Second, including rehabilitation endorses the rehabilitative model of criminal punishment, and Logan makes it clear that his model is retributive, not rehabilitative.³²⁵ 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.”³²⁶

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.”³²⁷ (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.)³²⁸

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

³²⁰ *Id.* at 42–43.

³²¹ *See id.* at 42–57.

³²² *See* LOGAN, *supra* note 61, at 7–11, 13, 17; Logan, *supra* note 61, at 577–78, 583 fig.1.

³²³ Logan, *supra* note 316, at 24.

³²⁴ *Id.*

³²⁵ *Id.* at 19, 21, 24.

³²⁶ *Id.* at 25.

³²⁷ *Id.* at 26.

³²⁸ *Id.* at 29 n.7.

one can commit oneself to a particular form of performance measures.³²⁹ “[W]ithout declared goals, we cannot hold a jurisdiction accountable, and performance measurement is meaningless.”³³⁰

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.”³³¹ 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;³³² 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).”³³³ 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,³³⁴ 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.

³²⁹ John DiIulio thus seems incorrect when he states that Logan’s work “dispels the worry that any such measurement scheme is bound to be based exclusively on one or another moral or ideological view of the ‘ends of criminal justice’” and that his measures “encompass and satisfy every major school of thought about ‘what prisons are for.’” DiIulio, *supra* note 286, at 152.

³³⁰ GAES ET AL., *supra* note 7, at xii.

³³¹ DiIulio, *supra* note 307, at 129, 131 (citing JOHN E. CHUBB & TERRY M. MOE, *POLITICS, MARKETS, AND AMERICA’S SCHOOLS* (1990)); see also John E. Chubb, *Why the Current Wave of School Reform Will Fail*, PUB. INTEREST, Winter 1988, at 28.

³³² See GAES ET AL., *supra* note 7, at 10–16 tbl.1.1; see also *supra* text accompanying note 291.

³³³ John J. DiIulio, Jr., *Rethinking the Criminal Justice System: Toward a New Paradigm*, in PERFORMANCE MEASURES FOR THE CRIMINAL JUSTICE SYSTEM, *supra* note 286, at 1, 6 (italics omitted).

³³⁴ See *supra* text accompanying note 17.

Thus, some of Logan's dimensions, like "activity," which I'm inclined to call an output measure,³³⁵ might be an outcome measure given Logan's normative perspective. The same goes for variables like prison employees' job satisfaction³³⁶ (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 concentrating³³⁷ (which—unlike, say, overcrowding or physical safety³³⁸—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.³³⁹

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.³⁴⁰ But if we're going to engage in quality measurement, we might as well enforce quality directly by getting the rewards or penalties "right";³⁴¹ let the private firms worry about their own costs.³⁴²

³³⁵ See DiIulio, *supra* note 286, at 152 (distinguishing between certain "process measures" and certain "performance measures" within Logan's "security" dimension); see also Gaes, *supra* note 36, at 23 ("[J]urisdictions that buy prison services are most concerned about internal performance measures such as order, health, case management, program services, and safety.").

³³⁶ See *supra* text accompanying note 315.

³³⁷ Ross et al., *supra* note 313, at 464 tbl.4.

³³⁸ See *supra* text accompanying notes 313–14.

³³⁹ See *infra* Part III.C.2.

³⁴⁰ Cf. W. KIP VISCUSI ET AL., *ECONOMICS OF REGULATION AND ANTITRUST* 430–36 (4th ed. 2005) (discussing the theory of traditional rate-of-return regulation, primarily in the context of electric utilities).

³⁴¹ See *infra* Part III.A.

³⁴² Cf. Shapiro & Steinzor, *supra* note 223, at 1767 (questioning whether reducing regulatory cost to the private sector should be a GPRA performance measure for the FDA).

III. CONCERNS AND CRITIQUES

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 costs³⁴³—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.³⁴⁴

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, 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.

A. *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.³⁴⁵ But one can support performance measures without endorsing efficiency in any way—in fact, as a better way of achieving particular moral goals.

³⁴³ See Durham, *supra* note 20, at 66; *see also id.* at 67 (“At none of the sites we examined were attempts made by government to evaluate rehabilitative success.”) (quoting JUDITH HACKETT ET AL., ISSUES IN CONTRACTING FOR THE PRIVATE OPERATION OF PRISONS AND JAILS 48 (1987)).

³⁴⁴ Cf. Low, *supra* note 222, at 64; Alexander Volokh, *Prison Vouchers*, 160 U. PA. L. REV. 779, 834 (2012). One might also measure a random sample of inmates, *see* Low, *supra* note 222, at 46 n.298, though this might exacerbate risk issues. *See infra* Part III.B.2.

³⁴⁵ Sharon Dolovich critiques “comparative efficiency” analysis and stresses moral considerations, *see, e.g.*, Dolovich, *supra* note 3; Dolovich, *supra* note 5, though to my knowledge she hasn't opined on performance measures.

I myself have been critical of a focus on efficiency in the context of regulatory cost–benefit analysis,³⁴⁶ 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.³⁴⁷ 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.³⁴⁸ 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.³⁴⁹ 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.³⁵⁰ The political process is also likely to manipulate the numbers.³⁵¹ Moreover, concerns that are hard to quantify can be systematically slighted.³⁵²

In short, “[w]hile cost–benefit analysis may look like rationality, perhaps it’s merely rationalism.”³⁵³ And these are just the problems for people who *accept* the utilitarian basis of cost–benefit analysis. The problems for those

³⁴⁶ See Alexander Volokh, *Rationality or Rationalism? The Positive and Normative Flaws of Cost–Benefit Analysis*, 48 HOUS. L. REV. 79 (2011).

³⁴⁷ See *id.* at 82–83.

³⁴⁸ See *id.* at 84.

³⁴⁹ See *id.* at 85–86.

³⁵⁰ See *id.* at 86–88.

³⁵¹ See, e.g., Frank Ackerman & Lisa Heinzerling, *Pricing the Priceless: Cost–Benefit Analysis of Environmental Protection*, 150 U. PA. L. REV. 1553, 1580 (2002) (explaining that regulated industry has an incentive to overstate costs).

³⁵² See *id.* at 1579–80. This gives rise to potentially serious strategic behavior, which I address in *infra* Part III.C.2.

³⁵³ Volokh, *supra* note 346, at 88.

who reject utilitarianism as a moral philosophy are even greater.³⁵⁴ 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,³⁵⁵ or (equivalently) setting the separate rewards or penalties for every component of performance.³⁵⁶ Getting the prices “right,” in an efficiency sense, requires knowing the social value of the different components of performance;³⁵⁷ 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,³⁵⁸ measuring performance is an essential part of that debate, even though the choice of outcomes to measure is a value-laden enterprise;³⁵⁹ 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.³⁶⁰ Rather than

³⁵⁴ See *id.* at 88–91.

³⁵⁵ See *supra* note 291 and accompanying text.

³⁵⁶ These two approaches are identical. Let x_i be the i th component of performance and p_i be the reward for that component. Then the total performance-based component of compensation is $\sum p_i x_i$. Let P be the sum of the prices ($P = \sum p_i$). Then the performance-based component of compensation can be expressed as $P \sum (p_i/P) x_i = P \sum w_i x_i$, where $w_i = p_i/P$ is the weight placed on the i th component of performance and P is the price attached to the overall performance index $\sum w_i x_i$.

³⁵⁷ Not that the price necessarily has to be equal to the social value—paying the price requires incurring the deadweight losses involved in raising tax money, and making incentives so high-powered might make the contract too risky. See *infra* Part III.B.2 for a discussion of optimal risk allocation. But at least the optimal prices (or at least the *relative* optimal prices of the different components of performance), from an efficiency perspective, will probably bear some relation to social value.

³⁵⁸ See generally Volokh, *supra* note 6.

³⁵⁹ See *supra* text accompanying notes 329–32.

³⁶⁰ Cf. Christopher C. DeMuth & Douglas H. Ginsburg, *Rationalism in Regulation*, 108 MICH. L. REV. 877, 885 (2010) (reviewing RICHARD L. REVEZ & MICHAEL A. LIVERMORE, *RETAKING RATIONALITY: HOW COST–BENEFIT ANALYSIS CAN BETTER PROTECT THE ENVIRONMENT AND OUR HEALTH* (2008)) (stating that Richard Revez and Michael Livermore “regard regulatory cost-benefit analysis as a device for social

trying to achieve incentives that are correct in some abstract sense,³⁶¹ 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.³⁶² 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."³⁶³ Is adding this element of imperfect, numbers-based accountability better than not? The remaining sections in this Part address this question.

B. 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."³⁶⁴

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."³⁶⁵

engineering Our view of cost-benefit analysis is much more modest [W]e think that many important political questions . . . cannot be effectively decided by cost-benefit analysis.").

³⁶¹ See DiIulio, *supra* note 286, at 146.

³⁶² See Barnow, *supra* note 230, at 279.

³⁶³ *Id.* at 307; see also M. Todd Henderson & Frederick Tung, *Paying Bank Examiners for Performance*, REGULATION, Spring 2012, at 32, 36 ("We . . . make no attempt to offer firm prescriptions for the optimal ratio [between debt and equity]. The mix should induce regulators to care about bank profits but not at the expense of risk shifting to creditors.").

³⁶⁴ M. Todd Henderson & Frederick Tung, *Pay for Regulator Performance*, 85 S. CAL. L. REV. 1003, 1056-57 (2012).

³⁶⁵ *Id.* at 1057.

Henderson and Tung conclude, citing the crowding out literature,³⁶⁶ that this is possible, though not necessary: “public spiritedness and financial reward [might not be] mutually exclusive, at least up to a point.”³⁶⁷ Moreover, changing the mix of individuals “could be a good,” given the failures of the current crop of people.³⁶⁸

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,³⁶⁹ 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;³⁷⁰ moreover, corrections professionals move between the public and private sectors and presumably take their professionalism with them. Finally, as I discuss further below,³⁷¹ 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,³⁷² the concern that performance measures be based on factors that the relevant actor can actually control. Such concerns crop up frequently;³⁷³ 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

³⁶⁶ *Id.* at 1057 n.182 (citing Ernst Fehr & Armin Falk, *Psychological Foundations of Incentives*, 46 EUR. ECON. REV. 687, 688 (2002); Uri Gneezy & Aldo Rustichini, *A Fine Is a Price*, 29 J. LEGAL STUD. 1, 14 (2000)).

³⁶⁷ *Id.* at 1057.

³⁶⁸ *Id.*

³⁶⁹ *But see supra* Part II.C.3 (discussing possibilities for merit pay for public prison wardens).

³⁷⁰ *See* Volokh, *supra* note 6, at 178–85.

³⁷¹ *See infra* Part III.B.2.

³⁷² *See supra* text accompanying notes 316–28.

³⁷³ *See, e.g.,* DICKER, *supra* note 123, at 17; GRIZZLE ET AL., *supra* note 241, at 48–49; Petersilia, *supra* note 316, at 66.

measurable “can only partially, if at all, be affected by police behavior.”³⁷⁴ When he does favor a “micro-level measure of success” of whether the neighborhood is becoming safer and more orderly,³⁷⁵ 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.”³⁷⁶ The concern in the literature over controlling for baselines is similarly motivated.³⁷⁷

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 extent³⁷⁸) lower recidivism *is* a measure of the success of prisons.³⁷⁹ 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.³⁸⁰ 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

³⁷⁴ James Q. Wilson, *The Problem of Defining Agency Success*, in PERFORMANCE MEASURES FOR THE CRIMINAL JUSTICE SYSTEM, *supra* note 286, at 156, 159; *see also* DiIulio, *supra* note 333, at 1–2, 13.

³⁷⁵ Wilson, *supra* note 374, at 160–62.

³⁷⁶ *Id.* at 161.

³⁷⁷ *See supra* text accompanying notes 240–50.

³⁷⁸ *See* Camp et al., *supra* note 99; M. Keith Chen & Jesse M. Shapiro, *Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-Based Approach*, 9 AM. L. & ECON. REV. 1, 17–21 (2007); DiIulio, *supra* note 333, at 2; Francesco Drago et al., *Prison Conditions and Recidivism*, 13 AM. L. & ECON. REV. 103, 120–25 (2011); Daniel S. Nagin et al., *Imprisonment and Reoffending*, in 38 CRIME AND JUSTICE: A REVIEW OF RESEARCH 115, 115 (Michael Tonry ed., 2009); Rafael Di Tella & Ernesto Schargrodsky, *Criminal Recidivism After Prison and Electronic Monitoring* 28 (Nat’l Bureau of Econ. Research, Working Paper No. 15602, 2009), available at <http://www.nber.org/papers/w15602.pdf>; *see also* GAES ET AL., *supra* note 7, at 124 (citing ROBERT J. SAMPSON & JOHN H. LAUB, *CRIME IN THE MAKING: PATHWAYS AND TURNING POINTS THROUGH LIFE* (1993); Shawn D. Bushway et al., *An Empirical Framework for Studying Desistance as a Process*, 39 CRIMINOLOGY 491 (2001); Jeffrey Grogger, *The Effect of Arrest on the Employment and Earnings of Young Men*, 110 Q.J. ECON. 51 (1995); Jeffrey R. Kling, *The Effect of Prison Sentence Length on the Subsequent Employment and Earnings of Criminal Defendants* (Woodrow Wilson Sch., Discussion Paper No. 208, 1999)); *id.* at 129 (citing Alex R. Piquero et al., *Assessing the Impact of Exposure Time and Incapacitation on Longitudinal Trajectories of Criminal Offending*, 16 J. ADOLESCENT RES. 54 (2001)); *id.* at 136 (citing Gerald G. Gaes & Newton Kendig, *The Skill Sets and Health Care Needs of Released Offenders*, in PAPERS PREPARED FOR THE “FROM PRISON TO HOME” CONFERENCE 93 (2002), available at http://www.urban.org/UploadedPDF/410629_ReleasedOffenders.pdf).

³⁷⁹ DiIulio, *supra* note 333, at 5 (“[C]rime rates and recidivism rates are indeed important[, though not the only,] measures of the system’s performance, which ought to be continually used and refined.”).

³⁸⁰ *See* Barnow, *supra* note 230, at 281 (explaining that these are “gross outcome measures . . . in the sense that they do not necessarily reflect gains from the program”).

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.³⁸¹

As to the first concern, recall the earlier discussion about whether to control for baselines.³⁸² 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.³⁸³

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,³⁸⁴ 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.³⁸⁵ 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);³⁸⁶ thus, "\$1 for each quality unit" is less risky than "\$5 but only if you get five quality units."

³⁸¹ See HARDING, *supra* note 72, at 68 ("[T]he human variables are too volatile for any contractor to be expected to stand or fall by outputs alone . . ."); Lynn, *supra* note 165, at 12; Kyle, *supra* note 162, at 2112.

³⁸² See *supra* Part II.C.2.c.

³⁸³ See *supra* text accompanying note 255.

³⁸⁴ See *supra* text accompanying notes 234–38, 251.

³⁸⁵ See *supra* text accompanying note 340.

³⁸⁶ See *supra* text accompanying notes 258–70.

Do we care? Perhaps large corporations like CCA or The GEO Group, which are publicly traded³⁸⁷ and diversified across many contracts,³⁸⁸ can handle the risk; and they cover three-quarters of the industry.³⁸⁹ Smaller, privately held companies like MTC³⁹⁰ may be more sensitive to risk. Various potential entrants, especially nonprofits,³⁹¹ 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.)³⁹² This has potential implications for the competitiveness of the industry,³⁹³ possibilities for innovation,³⁹⁴ and the political influence that drives changes in criminal law.³⁹⁵

³⁸⁷ See *Who We Are*, CCA, <http://www.cca.com/who-we-are> (last visited Dec. 12, 2013) (noting that CCA joined the NYSE in 1994); *Historic Milestones*, GEO GROUP, <http://www.geogroup.com/history> (last visited Dec. 12, 2013) (noting that GEO joined the NYSE in 1996).

³⁸⁸ See *Who We Are*, *supra* note 387 (“CCA houses nearly 80,000 inmates in more than 60 facilities . . . CCA currently partners with all three federal corrections agencies . . . , many states[,] and local municipalities.”); *Who We Are*, GEO GROUP, http://www.geogroup.com/about_us (last visited Dec. 12, 2013) (“GEO’s operations include the management and/or ownership of 96 correctional, detention and residential treatment facilities encompassing approximately 73,000 beds.”).

³⁸⁹ See Volokh, *supra* note 9, at 1237 & n.182 (relying on data from 1999).

³⁹⁰ See *id.* (noting a 5%–8% share for MTC in 1999); *Overview*, MGMT. & TRAINING CORP., <http://www.mtctrains.com/about-mtc/overview> (last visited Dec. 12, 2013) (“Management & Training Corporation (MTC) is a privately-held company . . .”).

³⁹¹ For discussions of the possibility of nonprofit prisons, see Low, *supra* note 222, at 4, and Richard Moran, Op-Ed., *A Third Option: Nonprofit Prisons*, N.Y. TIMES, Aug. 23, 1997, § 1, at 23. Compare with discussions of the advantages of nonprofit schools. See Byron W. Brown, *Why Governments Run Schools*, 11 ECON. EDUC. REV. 287, 293–96 (1992); John Morley, Note, *For-Profit and Nonprofit Charter Schools: An Agency Costs Approach*, 115 YALE L.J. 1782, 1795–1810 (2006); *cf. also* Education: *Raising the Bar*, ECONOMIST, June 15–21, 2013, at 30 (discussing risk issues for schools and teachers resulting from educational accountability schemes).

³⁹² NICHOLSON, *supra* note 216, at 6 (“The working capital requirements of a [payment-by-results] system will cause problems for Small and Medium Sized Enterprises and the third sector [i.e., nonprofits] in bidding for contracts.”).

³⁹³ DICKER, *supra* note 123, at 24 (explaining that high incentives, through high risk, will “reduce the diversity of the market” by making it less attractive for nonprofits or small companies).

³⁹⁴ *Id.* at 23. On the relationship between market concentration and innovation, see Richard Gilbert, *Looking for Mr. Schumpeter: Where Are We in the Competition–Innovation Debate*, in 6 INNOVATION POLICY AND THE ECONOMY 159 (Adam B. Jaffe et al. eds., 2006), for an argument that the relationship is inconclusive.

³⁹⁵ See Volokh, *supra* note 9, at 1213–14 (arguing that the degree of concentration of the industry can affect the political influence the industry exerts); see also Volokh, *Privatization, Free Riding*, *supra* note 142, at 64; Volokh, *The Effect of Privatization*, *supra* note 142, at 10–11.

But the contract doesn't have to be especially high-stakes.³⁹⁶ 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.³⁹⁷ Flat-fee contracts are relatively low risk³⁹⁸ but also low incentive. Some moderate level of risk transfer will optimally balance incentives with risk.³⁹⁹ Thus, the incentive-based portion of the contract is only 10% of the contract price in U.K.'s Doncaster prison,⁴⁰⁰ and was only 5% in the Federal Bureau of Prisons' Taft demonstration project.⁴⁰¹ 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."⁴⁰²

For the cash-flow issue noted above,⁴⁰³ 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."⁴⁰⁴

b. Financing Nonprofits: Social Impact Bonds

The need to encourage the nonprofit sector calls for innovative funding mechanisms. Nonprofit prisons have been suggested⁴⁰⁵ though never implemented.⁴⁰⁶ But in light of the widespread concern that private prison firms will cut quality to save money,⁴⁰⁷ 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

³⁹⁶ DICKER, *supra* note 123, at 6.

³⁹⁷ *See supra* note 357.

³⁹⁸ Though not zero-risk: recall that the least risky contracts are cost-plus. *See supra* text accompanying note 385.

³⁹⁹ *See* DICKER, *supra* note 123, at 23–24; NICHOLSON, *supra* note 216, at 6–7; *see also* BOLTON & DEWATRIPONT, *supra* note 256, at 13 ("[W]hen both employer and employee are risk averse, they will optimally share business risk").

⁴⁰⁰ *See* Johnson, *supra* note 209.

⁴⁰¹ *See supra* text accompanying note 172.

⁴⁰² *See supra* text accompanying note 218.

⁴⁰³ *See supra* text accompanying note 392.

⁴⁰⁴ DICKER, *supra* note 123, at 24.

⁴⁰⁵ *See sources cited supra* note 391.

⁴⁰⁶ *See* Low, *supra* note 222, at 4–5 (suggesting creation of nonprofit prisons on "an experimental basis").

⁴⁰⁷ *See, e.g.,* Dolovich, *supra* note 5, at 474–80.

be a valuable signal of quality when quality itself is nonverifiable.⁴⁰⁸ (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.)⁴⁰⁹ Moreover, altruistic entrepreneurs will tend to be attracted to the nonprofit form.⁴¹⁰

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.⁴¹¹ 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.⁴¹²

Social impact bonds have been proposed as a funding mechanism for nonprofits.⁴¹³ Rather than contracting directly with a provider, the government contracts with a middleman. This middleman, a "social impact bond-issuing organization,"⁴¹⁴ has two functions. First, it hires the staff to provide the service. Second, it sells bonds to investors, particularly philanthropic ones;⁴¹⁵ 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.⁴¹⁶ 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,⁴¹⁷ and payments are capped when the difference reaches 13%.⁴¹⁸ The provider's

⁴⁰⁸ Edward L. Glaeser & Andrei Shleifer, *Not-for-Profit Entrepreneurs*, 81 J. PUB. ECON. 99, 102 (2001).

⁴⁰⁹ See *infra* Part III.C.2.

⁴¹⁰ Glaeser & Shleifer, *supra* note 408, at 102.

⁴¹¹ See Timothy Besley & Maitreesh Ghatak, *Government Versus Private Ownership of Public Goods*, 116 Q.J. ECON. 1343, 1347 (2001).

⁴¹² NICHOLSON, *supra* note 216, at 6–7.

⁴¹³ JEFFREY B. LIEBMAN, SOCIAL IMPACT BONDS: A PROMISING NEW FINANCING MODEL TO ACCELERATE SOCIAL INNOVATION AND IMPROVE GOVERNMENT PERFORMANCE 2 (2011); see also Cullen et al., *supra* note 13, at 88–89; Michaels, *supra* note 276, at 1052–58; Shifali Baliga, Note, *Shaping the Success of Social Impact Bonds in the United States: Lessons Learned from the Privatization of U.S. Prisons*, 63 DUKE L.J. 437 (2013).

⁴¹⁴ LIEBMAN, *supra* note 413, at 2 (internal quotation marks omitted).

⁴¹⁵ Though social impact bonds in the United States have been funded by non-philanthropic types such as Goldman Sachs. See *Social Impact Bonds: Being Good Pays*, ECONOMIST, Aug. 18–24, 2012, at 28.

⁴¹⁶ LIEBMAN, *supra* note 413, at 2.

⁴¹⁷ *Id.*

⁴¹⁸ See *supra* note 213 and accompanying text.

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.⁴¹⁹ 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.⁴²⁰

C. *Undesirable Strategic Behavior*

Perhaps the biggest disadvantage of using performance-based compensation is the strategic behavior it may spawn. This strategic behavior may come in several flavors. First, there is the possibility of manipulating the performance goals themselves. Second, effort may be distorted away from some dimensions and toward others. Third, effort may be distorted away from some groups of inmates and toward others. And fourth, performance measures may simply be falsified.

1. *Manipulating the Goals*

The Government Performance and Results Act of 1993⁴²¹ is one example of a recent effort to inject performance measures into government agencies that hasn't lived up to the hopes of its supporters.

One of the problems was that setting the performance goals was left to the agencies that were to be evaluated. Agencies "tr[ie]d to protect themselves by devising euphemistic performance goals in order to ensure that they [could] 'pass' their own grading criteria."⁴²² The Patent and Trademark Office, faced with rising backlogs, set itself progressively longer targets of "average total pendency" from year to year, rising from 27.7 months in fiscal year 2003 to

⁴¹⁹ See *infra* text accompanying note 459.

⁴²⁰ See NICHOLSON, *supra* note 216, at 16, 18; *Big Hurdles to Be Overcome if Social Impact Bonds to Move from Margins of Public Services, Says Think Tank*, SOC. MARKET FOUND. (July 31, 2013), <http://www.smf.co.uk/media/news/big-hurdles-be-overcome-if-social-impact-bonds-move-margins-public-services-says-think-tank/>; Tom Clougherty, *Pioneering Social Impact Bonds in the United Kingdom*, REASON FOUND. (Aug. 13, 2013), <http://reason.org/news/show/pioneering-social-impact-bonds>.

⁴²¹ Pub. L. No. 103-62, 107 Stat. 285 (codified as amended in scattered sections of 5 and 31 U.S.C.); see *supra* note 165.

⁴²² Shapiro & Steinzor, *supra* note 223, at 1744; see also *id.* at 1760 ("[A]gencies compelled to function in an antiregulatory, even hostile, political atmosphere are predictably reluctant to tell the truth to power. Instead, their goal has become convincing congressional and White House overseers that they are performing well despite budgets that are inadequate for effective implementation of their missions.").

29.8 months in 2004, 31.0 months in 2005, and 31.3 months in 2006.⁴²³ John DiIulio had warned of a similar danger: “that measurement-driven government workers will, so to speak, ‘set up the target in order to facilitate shooting.’”⁴²⁴ The similar problem was observed in the U.K., where “Next Steps agencies,” a type of performance-based organization, set their own targets, which often reflected merely an incremental improvement rather than an assessment of what was possible.⁴²⁵

Why would agencies set goals in such unambitious ways? Perhaps because agencies feared being punished for bad performance with budget cuts.⁴²⁶ Various politicians have indeed suggested that agencies’ funding be tied to their performance results,⁴²⁷ and agencies’ performance results have indeed been relevant to the administration’s budget proposals,⁴²⁸ so this fear may have been reasonable—though it’s also possible that performance scores have merely given political cover for cuts to programs that the administration wanted to defund for other reasons.⁴²⁹ On the other hand, the link between funding and performance results isn’t *that* tight,⁴³⁰ so agencies’ concern to look good may also have been a matter of good public relations.

⁴²³ Schoen, *supra* note 165, at 480.

⁴²⁴ DiIulio, *supra* note 286, at 154.

⁴²⁵ U.S. GEN. ACCOUNTING OFFICE, *supra* note 223, at 7.

⁴²⁶ Shapiro & Steinzor, *supra* note 223, at 1744.

⁴²⁷ See Schoen, *supra* note 165, at 464 (citing *The Results Act: Are We Getting Results?: Hearings Before the H. Comm. on Gov’t Reform & Oversight*, 105th Cong. 20 (1997) (statement of Rep. Richard K. Arney, H. Majority Leader)); *id.* at 465 (citing *Seven Years of GPRA: Has the Results Act Provided Results?: Hearing Before the Subcomm. on Gov’t Mgmt., Info., & Tech. of the H. Comm. on Gov’t Reform*, 106th Cong. 20 (2000) (statement of Rep. Pete Sessions, Chairman, H. Results Caucus)); *id.* at 466–67 (citing OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, THE PRESIDENT’S MANAGEMENT AGENDA (2002), available at <http://www.whitehouse.gov/sites/default/files/omb/budget/fy2002/mgmt.pdf>).

⁴²⁸ See Eileen Norcross & Joseph Adamson, An Analysis of the Office of Management and Budget’s Program Assessment Rating Tool (PART) for Fiscal Year 2008, at 25 (unpublished working paper), http://mercatus.org/sites/default/files/publication/20070725_Analysis_of_PART_for_FY_2008.pdf; Eileen Norcross & Kyle McKenzie, An Analysis of the Office of Management and Budget’s Program Assessment Rating Tool (PART) for Fiscal Year 2007, at 22 (May 2006) (unpublished working paper), http://mercatus.org/sites/default/files/publication/PDF_An_Analysis_of_the_Office_of_Management_of_Budgets_Program_Assessment_Rating_Tool_for_FY_2007.pdf.

⁴²⁹ See John B. Gilmour & David E. Lewis, *Does Performance Budgeting Work? An Examination of the Office of Management & Budget’s PART Scores*, 66 PUB. ADMIN. REV. 742, 751 (2006); Norcross & Adamson, *supra* note 428, at 29–30.

⁴³⁰ See, e.g., Teresa Currstine, *Reforming the U.S. Department of Transportation: Challenges and Opportunities of the Government Performance and Results Act for Federal-State Relations*, PUBLIUS, Winter 2002, at 25, 42; Jerry Ellig, *Has GPRA Increased the Availability and Use of Performance Information?* 5 (George Mason Univ. Mercatus Ctr., Working Paper No. 09-03, 2009).

The problem here is that agencies were allowed to think up their own performance goals; that they weren't required to meet those goals (and indeed, that often the performance information simply wasn't used in decisionmaking⁴³¹); and that the goals were binary rather than continuous outcome measures,⁴³² for example, that the EPA “will ‘achieve and maintain at least 95 percent of the maximum score on readiness evaluation criteria in each region’”⁴³³ or “‘complete an additional 975 Superfund-lead hazardous substance removal actions.’”⁴³⁴

These problems have easy fixes, though perhaps they weren't so easy in the context of the GPRA, where the problem was primarily giving performance incentives to public agencies. Prison contracts—or merit pay systems for public prison wardens⁴³⁵—should be set by the Department of Corrections or the relevant contracting authority; goals shouldn't be set by those who we want to comply with them. No one should be “required” to meet any performance standard, but compensation should be tied to these measures; providers' self-interest should take care of the rest. And adopting continuous outcome measures, rather than binary goals, reduces the ability to choose easy goals: one can game “achieve $x\%$ recidivism” by setting an appropriately high level of x , but it's harder to game the general effort of reducing recidivism where additional reductions are met with additional rewards.⁴³⁶

⁴³¹ See Schoen, *supra* note 165, at 466 (citing *10 Years of GPRA—Results, Demonstrated: Hearings Before the Subcomm. on Gov't Efficiency & Fin. Mgmt. of the H. Comm. on Gov't Reform*, 108th Cong. 4 (2004) (statement of Rep. Edolphus Towns, Member, Subcomm. on Gov't Efficiency & Fin. Mgmt. of the H. Comm. on Gov't Reform)); Ellig, *supra* note 430, at 1 (citing Jerry Ellig & Jerry Brito, *Toward a More Perfect Union: Regulatory Analysis and Performance Management*, 8 FLA. ST. U. BUS. REV. 1 (2009)); *id.* at 2 (citing U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-1026T, GOVERNMENT PERFORMANCE: LESSONS LEARNED FOR THE NEXT ADMINISTRATION ON USING PERFORMANCE INFORMATION TO IMPROVE RESULTS (2008)).

⁴³² See *supra* Part I.I.C.2.d.

⁴³³ Shapiro & Steinzor, *supra* note 223, at 1764 (quoting EPA, 2006–2011 EPA STRATEGIC PLAN: CHARTING OUR COURSE 67 (2006), available at <http://nepis.epa.gov/Adobe/PDF/P10011PK.PDF>).

⁴³⁴ *Id.* at 1765 (quoting EPA, *supra* note 433, at 67); see also *id.* at 1773 (“[A]ttain water quality standards for all pollutants and impairments in more than 2,250 water bodies [R]emove at least 5,600 . . . specific causes of water body impairment [I]mprove water quality conditions in 250 . . . impaired watersheds nationwide” (altered capitalization and third and fifth omissions in original) (quoting EPA, *supra* note 433, at 43)).

⁴³⁵ See *supra* Part I.I.C.3.

⁴³⁶ See also Barnow, *supra* note 230, at 287 (discussing “whether the size of the award should vary with the extent to which standards are exceeded”); *id.* at 291–92 (“The national standards are set, based on experience in prior years, so that approximately 75 percent of the nation's [providers] will exceed the standards”).

2. *Distortion Across Dimensions of Performance*

Everyone agrees that, in most areas, performance has multiple dimensions.⁴³⁷ Each dimension, in a performance-based contract, will have its price,⁴³⁸ and the relative prices of different dimensions will determine how the contractor will allocate his effort among them.⁴³⁹

So far, so good, as long as the set of performance measures is complete. But what if some dimensions of performance are unmeasurable?⁴⁴⁰ Just as cost-benefit analysis is accused of slighting the soft factors,⁴⁴¹ so might performance *measures* be biased in favor of the *measurable*. The result is that the contractor's work effort will be biased in the direction of increasing the measurable dimensions of performance.⁴⁴²

Consider a hypothetical example involving education. Suppose there are two measures of educational quality: "hard" (e.g., knowledge of facts) and "soft" (e.g., citizenship, critical thinking, socialization). Without hard accountability, it might be hard to give teachers serious incentives, so they will slack in their overall work effort, but divide their time between hard and soft education in a balanced way. With hard accountability, teachers can get much higher-powered incentives, but these incentives will tend to be skewed toward the hard measures of education. Thus, the teachers will provide more overall work effort, but their time will be skewed toward hard education.⁴⁴³

How serious is this problem? It depends how important it is to have a balance between hard and soft factors, how hard the soft factors really are to measure, and how harmful the status quo of low work effort is.⁴⁴⁴ It also

⁴³⁷ See *supra* text accompanying notes 291, 332–33.

⁴³⁸ See *supra* text accompanying note 356.

⁴³⁹ See Bengt Holmstrom & Paul Milgrom, *Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design*, 7 J.L. ECON. & ORG. (SPECIAL ISSUE) 24, 25 (1991) ("In general, when there are multiple tasks, incentive pay serves not only to allocate risks and to motivate hard work, it also serves to direct the allocation of the agents' attention among their various duties.").

⁴⁴⁰ See *supra* text accompanying note 352 (noting retributivism as a possible unmeasurable dimension).

⁴⁴¹ See *supra* text accompanying note 352.

⁴⁴² See GRIZZLE ET AL., *supra* note 241, at 50–51.

⁴⁴³ See Holmstrom & Milgrom, *supra* note 439, at 25 ("It would be better, . . . critics argue, to pay a fixed wage without any incentive scheme than to base teachers' compensation only on the limited dimensions of student achievement that can be effectively measured." (italics omitted)); see also Peter Smith, *On the Unintended Consequences of Publishing Performance Data in the Public Sector*, 18 INT'L J. PUB. ADMIN. 277, 284 (1995) (discussing "tunnel vision"); *Education: Raising the Bar*, *supra* note 391.

⁴⁴⁴ See Holmstrom & Milgrom, *supra* note 439, at 26 ("[T]he desirability of providing incentives for any one activity decreases with the difficulty of measuring performance in any other activities that make competing demands on the agent's time and attention.").

depends on whether the one type of education makes the other type easier or harder for the teacher; an excessively high-powered accountability system focusing, say, on standardized test scores could easily promote a “teaching to the test” strategy that can be antithetical to critical thinking (at the very least by taking up class time that could be otherwise used),⁴⁴⁵ this isn’t necessarily so, but it may be likely.⁴⁴⁶ Providing high-powered but skewed accountability may be beneficial in severely dysfunctional school systems where neither hard nor soft factors are taught well, but it may be harmful in better school systems.

Analogously, in the prison context, one can imagine two dimensions of quality: humane in-prison conditions and low recidivism after prison. Suppose one of these is harder to measure than the other. In-prison conditions could be harder to measure if effective monitoring is difficult;⁴⁴⁷ or perhaps recidivism is harder to measure if there aren’t good databases of offenders, especially if released inmates often commit their crimes in other states. Whichever one turns out to be less measurable, we can expect effort to be skewed toward the more measurable one.

Would it make a difference if prison policies were skewed toward humane conditions or toward reducing recidivism? If the two go together—if humane conditions are, on balance, effective at reducing recidivism⁴⁴⁸—then the inability to monitor both dimensions can be harmless. On the other hand, if bad prison conditions, on balance, reduce recidivism through a general deterrent effect,⁴⁴⁹ a focus on recidivism could lead to bad prison conditions—in which case there’s no guarantee that high-powered accountability would improve overall quality in the absence of effective in-prison monitoring. Since the

⁴⁴⁵ This assumes that test scores really are a true outcome measure, even if a partial one. Perhaps this is too charitable, though: it may be better to characterize test scores as proxy measures for a type of intelligence, and “teaching to the test” as a form of manipulation, as described below. See *infra* text accompanying note 457.

⁴⁴⁶ See Holmstrom & Milgrom, *supra* note 439, at 25; *id.* at 32–33 (explaining that the desirability of incentives for measurable tasks depends on whether measurable and unmeasurable tasks are complements or substitutes in an agent’s cost function).

⁴⁴⁷ See *infra* Part III.C.4.

⁴⁴⁸ See sources cited *supra* note 378.

⁴⁴⁹ See, e.g., Kelly Bedard & Eric Helland, *The Location of Women’s Prisons and the Deterrence Effect of “Harder” Time*, 24 INT’L REV. L. & ECON. 147, 159–61 (2004); Lawrence Katz et al., *Prison Conditions, Capital Punishment, and Deterrence*, 5 AM. L. & ECON. REV. 318, 331 (2003); Volokh, *supra* note 344, at 843–45. But see TOM R. TYLER, WHY PEOPLE OBEY THE LAW 64 (1990) (“The most important normative influence on compliance with the law is the person’s assessment that following the law accords with his or her sense of right and wrong; a second factor is the person’s feeling of obligation to obey the law and allegiance to legal authorities.”); Paul H. Robinson & John M. Darley, *The Role of Deterrence in the Formulation of Criminal Law Rules: At Its Worst When Doing Its Best*, 91 GEO. L.J. 949, 953–56 (2003).

precise determinants of recidivism aren't well understood, this shows the importance of properly monitoring whatever is considered desirable in the prison.⁴⁵⁰

In the extreme case, where some tasks remain completely unmeasurable and shirking on that task is highly detrimental to overall quality, we should junk the idea of high-powered incentives: the traditional input-and-output approach may then be optimal.⁴⁵¹

If an unmeasurable outcome is represented in the accountability scheme by some inputs or outputs as proxies, the possibilities for undesirable strategic behavior multiply. The previous examples involved *ignoring* the unmeasurable elements and maximizing the measurable component of performance, rather than maximizing overall performance. Replacing unmeasurable elements with proxies within the provider's direct control leads to pursuing the proxies for their own sake—which one can uncharitably call “manipulating” the proxy measures.

For example, consider recidivism rates, which I've been treating throughout as a true outcome measure. In reality, no one knows *true* recidivism rates; we don't know that a released inmate has committed a crime unless we catch him (and, depending on the recidivism measure we're using, unless we convict him or reincarcerate him).⁴⁵² So in reality, rather than using the unmeasurable dimension of recidivism, we're using the measurable proxy of, say, rearrest rates. If the relationship between rearrest rates and true recidivism is stable, using this proxy can be harmless;⁴⁵³ but more important still is that the contractor not be able to manipulate the rates in ways that don't correspond to true social improvements.

Thus, if in-prison misconduct is penalized, corrections officers will use their discretion very differently when deciding whether to write up an

⁴⁵⁰ See *infra* Part III.C.4; see also Cullen et al., *supra* note 13, at 84 (“Of course, some correctional managers might attempt to develop a painful prison in hopes of scaring offenders straight. We are confident that these efforts will fail and place managers at a disadvantage.” (citation omitted)).

⁴⁵¹ See Holmstrom & Milgrom, *supra* note 439, at 27 (“[I]ncentives for a task can be provided in two ways: either the task itself can be rewarded or the marginal opportunity cost for the task can be lowered by removing or reducing the incentives on competing tasks. Constraints are substitutes for performance incentives and are extensively used when it is hard to assess the performance of the agent.”).

⁴⁵² See *supra* text accompanying note 120.

⁴⁵³ Of course, the relationship between rearrest rates and true recidivism can change—for instance, enforcement agencies might, over time, reallocate resources from one type of crime to another. This raises the question of whether to control for baseline rates. See *supra* Part II.C.2.c.

offense.⁴⁵⁴ If urinalysis tests based on suspicion are rewarded, we can magically expect more inmates to seem suspicious. Perhaps the output (drug tests based on suspicion) seems to have a straightforward correlation with the outcome (inmate drug use, if one chooses to consider that an outcome⁴⁵⁵); but make it a subject of compensation, and you can't rely on that correlation anymore. Administrators will start pursuing the output for its own sake. (Random drug tests unrelated to suspicion remove that gaming problem, even if they are more expensive for the same level of deterrence.)

Similarly, in the context of community corrections, Joan Petersilia criticizes the use of recidivism rates as an outcome measure: if the number of arrests increases, is that bad because more people are committing offenses? Or is it good because probation officers are better at detecting technical violations and sending released offenders back to prison?⁴⁵⁶ If we decided that increased arrest rates were bad and attached penalties to that variable, we might find arrest rates plummeting, but merely because probation officers stopped supervising their charges very closely.

Recidivism may thus be a bad measure for the accountability of probation officers. But it can be a good measure for the accountability of prisons, provided that prisons leave supervision and rearrest to entirely separate actors. This is a reason to insist on the separation of prisons and probation officers, not granting contracts to criminal justice providers that are too integrated, and more generally preventing prisons from giving any incentives at all, even subtle ones, to probation officers.⁴⁵⁷ Similarly, the results of drug testing can be an acceptable measure, but random testing is better than testing based on suspicion. In-prison misconduct can be an acceptable measure, but it should be the type of serious misconduct that's least likely to be overlooked or characterized as something else.

We might even have to guard against other kinds of gaming: if prisons can affect where prisoners are released, for instance by partnering with post-release job placement programs that have good contacts in particular areas, they can try to have prisoners released in areas where policing is weaker. For understandable political economy reasons, a state Department of Corrections

⁴⁵⁴ See GAES ET AL., *supra* note 7, at 51. Currently, there is, to the contrary, some incentive for private firms to exaggerate infractions so as to prevent early releases. See Dolovich, *supra* note 5, at 518–23.

⁴⁵⁵ I prefer to think of drug use as neutral in itself, though one can want to control inmate drug use instrumentally for the sake of outcomes like violence or rehabilitation.

⁴⁵⁶ See Petersilia, *supra* note 316, at 66–67; see also GAES ET AL., *supra* note 7, at 23–24; *supra* note 76.

⁴⁵⁷ See Smith, *supra* note 443, at 286, 290–92 (discussing “suboptimization” and “measure fixation”).

might choose to ignore the welfare of people in other states and tie compensation only to an in-state measure of recidivism; then, the prison does better by finding out-of-state jobs for its inmates. A prison might also try to prevent recidivism by “paying offenders to desist,” but this might be controversial.⁴⁵⁸

Of course, even if we only use performance measures to reward providers, providers will inevitably have to translate these incentives into specific input- or output-based incentives to reward their own staff, at least in part—there are limits to the possibilities of stock options.⁴⁵⁹ And such incentives can sometimes backfire for the same reasons that input-based incentives can backfire at the prison level. At one CCA prison in Tennessee, the employee compensation policy discouraged “use-of-force incidents.”⁴⁶⁰ In general, this can be positive, but sometimes not: for nine straight months, CCA personnel stopped removing mentally ill inmate Frank Horton from his cell for showers, exercise, and mental health evaluations, because any attempt to do so would have been considered a “use of force” and could have affected their bonuses or pay raises.⁴⁶¹ Presumably, though, a provider motivated by good performance measures will have better incentives and better ability to monitor its own staff than the government has to monitor the provider.

3. *Distortion Across Types of Inmates*

One common complaint about high-powered outcome-based incentives is that they’ll lead to two related phenomena: “creaming”—only taking the easiest inmates—and “parking”—not providing services to the most difficult inmates.⁴⁶² There’s an easy way to prevent providers from taking the easiest inmates: insist that providers take all comers,⁴⁶³ limit opportunities for

⁴⁵⁸ DICKER, *supra* note 123, at 19.

⁴⁵⁹ On the use of stock options in private prisons, see Volokh, *supra* note 6, at 174.

⁴⁶⁰ *Braswell v. Corr. Corp. of Am.*, 419 F. App’x 662, 628 (6th Cir. 2011).

⁴⁶¹ *Id.* at 623–24.

⁴⁶² See DICKER, *supra* note 123, at 23; see also RICHARD A. MCGOWAN, *PRIVATIZE THIS?: ASSESSING THE OPPORTUNITIES AND COSTS OF PRIVATIZATION* 166 (2011); Barnow, *supra* note 230, at 287, 297–98, 305–06; Pozen, *supra* note 79, at 283; Kyle, *supra* note 162, at 2112; Inwood, *supra* note 210. For a recent example of parking in a non-prison context, see Mary Shinn et al., *Despite Backlogs, VA Disability Claims Processors Get Bonuses*, WASH. POST, Aug. 25, 2013, http://articles.washingtonpost.com/2013-08-25/world/41446536_1_claims-processors-new-claims-backlog.

⁴⁶³ See Gilroy, *supra* note 163 (“So literally, you have the private vendor take over the exact same population, and then use the same metrics you use to assess the public facilities.”); cf. Volokh, *supra* note 449, at 806–07 (arguing that requiring prisoners to take all comers makes sense to prevent prisons from systematically rejecting certain inmates).

providers to transfer inmates they don't like out of the prison, and have assigning agencies not discriminate either in favor of or against particular providers in assignment.⁴⁶⁴ And the bias toward treating easier inmates can be alleviated by mandating particular services for everyone. There remains, though, the concern that providers will be, for instance, more enthusiastic about providing rehabilitative services to those that can more likely benefit from them.

There are two lines of response to this concern. Clearly, paying the same rate, regardless of how hard the offender is to serve, will lead to parking; one can therefore provide payments that are inmate specific, where a harder-to-serve inmate's desistance from crime is rewarded more generously than an easier-to-serve inmate's.⁴⁶⁵ These payments can be based on the observable characteristics of the inmate; some characteristics might be illegal to consider while others can be better observed by the provider than by the government, so there will inevitably be some degree of mismatch.⁴⁶⁶ But a system of nonuniform rewards can generally alleviate parking.

The second line of response would question whether parking is even bad. Suppose some inmates are hard to rehabilitate, so prisons—in the presence of uniform rewards—will tend to spend less time trying to rehabilitate them. Is this bad? Some nonuniformity of rewards will be inevitable—presumably a murder by a released inmate will be penalized more heavily than a minor crime. But suppose there's a group of inmates whose recidivism is equally harmful. Wouldn't it be socially beneficial for the provider to concentrate its resources on the ones whose crimes can be prevented most cheaply, so that more inmates can be treated at the same cost?⁴⁶⁷ At least, so an efficiency framework might counsel. If one subscribes to a certain form of equity where everyone should have some amount of (even ineffective) rehabilitation, one might want to fall back on the solution I mentioned above: offering higher

⁴⁶⁴ See *supra* text accompanying notes 135–36.

⁴⁶⁵ See DICKER, *supra* note 123, at 24; NICHOLSON, *supra* note 216, at 6–7; David Boyle, *The Perils of Obsessive Measurement*, RSA (Nov. 1, 2010), <http://comment.rsablogs.org.uk/2010/11/01/perils-obsessive-measurement/>.

⁴⁶⁶ Cf. Volokh, *supra* note 344, at 806–07 (discussing inmate characteristics that institutions are able to consider).

⁴⁶⁷ One could argue—though I'm not doing so here—that even outright creaming, where prisons only accept the easier-to-treat inmates, might be beneficial, since it might be worthwhile to separate the “better” criminals from the harmful influence of the “worse” ones. See, e.g., Volokh, *supra* note 449, at 837–38.

payments for the harder-to-treat inmates⁴⁶⁸ or, if that can't be done reliably, mandating some amount of inputs or outputs.

4. *Falsifying Performance Measures*

Finally, when high-stakes compensation depends on numbers, there's an obvious incentive to falsify the numbers themselves.⁴⁶⁹ Reports of school cheating scandals are commonplace.⁴⁷⁰ Similarly, in the prison context, private providers plausibly prefer to underreport incidents, at least if they wouldn't inevitably become known.⁴⁷¹ Failure to report is grounds for contract termination, which can cut in the other direction, but contract termination is a strong remedy that's rarely used.⁴⁷² Public prisons, on the other hand, might have an incentive to overreport to get more funds but they also might have an incentive to underreport to make themselves look better compared to private prisons.⁴⁷³ Misconduct data are thus somewhat unreliable, especially if one wants to use them to compare different prisons.

Whichever way the incentives cut, the fact that compensation will inevitably be to some extent based on variables reported by the provider means that it's important to seriously invest in monitoring. Currently, monitoring practices vary quite a lot, "from minimal attention from a centrally located contract administrator to a combination of a contract administrator and one or more on-site monitors."⁴⁷⁴ The monitors themselves may have responsibility for more than one facility, which puts them on site at any particular prison once a quarter, once a week, or daily.⁴⁷⁵ Instead, contracts should provide for a full-time, on-site monitor⁴⁷⁶ with "unlimited access to the correctional facilities

⁴⁶⁸ DICKER, *supra* note 123, at 25.

⁴⁶⁹ See Smith, *supra* note 443, at 292; Boyle, *supra* note 465.

⁴⁷⁰ See, e.g., Emily Richmond, *Did High-Stakes Testing Cause the Atlanta Schools Teaching Scandal?*, ATLANTIC (Apr. 3, 2013, 11:50 AM), <http://www.theatlantic.com/national/archive/2013/04/did-high-stakes-testing-cause-the-atlanta-schools-cheating-scandal/274619/>.

⁴⁷¹ See JOEL DYER, THE PERPETUAL PRISONER MACHINE: HOW AMERICA PROFITS FROM CRIME 211, 221 (2000); Gaes et al., *supra* note 60, at 18; Low, *supra* note 222, at 39 (citing JOHN L. CLARK ET AL., REPORT TO THE ATTORNEY GENERAL: INSPECTION AND REVIEW OF THE NORTHEAST OHIO CORRECTIONAL CENTER ch. VII.B.2 (1998) (reporting that CCA's legal counsel warned administrators not to write reports about incidents because of potential legal liability); *id.* chs. VIII, XI; HARDING, *supra* note 145, at 323–24); *Developments in the Law—The Law of Prisons*, *supra* note 9, at 1884.

⁴⁷² See *supra* text accompanying notes 267–68.

⁴⁷³ See Gaes et al., *supra* note 60, at 18.

⁴⁷⁴ MCDONALD ET AL., *supra* note 22, at 50.

⁴⁷⁵ See *id.* at 50, 51 tbl.4.1.

⁴⁷⁶ See Thomas, *supra* note 138, at 109.

and assigned correctional units,”⁴⁷⁷ who isn’t the provider’s employee (even if the contract might mandate that the provider pay his salary as part of the deal).⁴⁷⁸ When prisoners are sent out of state, monitoring is more likely to be “on paper” rather than “in person”—which is one reason to keep one’s prisoners in state.

Because the capture of monitors is an enduring concern,⁴⁷⁹ other forms of monitoring are possible: a public-interest group could be given inspection rights,⁴⁸⁰ the surrounding community might be designated as a third-party beneficiary,⁴⁸¹ or the constitutional tort regime for prisons could be strengthened (rather than weakened, which is the current trend).⁴⁸²

A strong disclosure regime is also probably a good idea.⁴⁸³

One way of guaranteeing disclosure is to subject private prisons under contract with the federal government to the Freedom of Information Act,⁴⁸⁴ perhaps along the lines of the often-proposed Private Prison Information Act. Private prison firms themselves aren’t “agencies” for the purposes of FOIA,⁴⁸⁵ and the Bureau of Prisons isn’t covered if it hasn’t “created and retained” or doesn’t actually possess the documents.⁴⁸⁶ Even after these hurdles, much qualifying information, like contracts or incident reports, would be exempt under Exemption 4, which protects “trade secrets and commercial or financial

⁴⁷⁷ S.B. 2038, 2012 Leg., Reg. Sess., sec. 1, § 944.7115(8)(d) (Fla. 2012).

⁴⁷⁸ See *id.*; see also Nicole B. Cásarez, *Furthering the Accountability Principle in Privatized Federal Corrections: The Need for Access to Private Prison Records*, 28 U. MICH. J.L. REFORM 249, 293 (1995) (citing Ira P. Robbins, *The Legal Dimensions of Private Incarceration*, 38 AM. U. L. REV. 531, 752 (1989)) (explaining that Robbins’s Model Contract “calls for an employee of the contracting agency to have access to prison facilities and all records kept by the contractor at all times”); Low, *supra* note 222, at 39 (citing CLARK ET AL., *supra* note 471, ch. XI); Gilroy, *supra* note 163 (discussing the full-time monitor at each private prison in Ohio plus surprise inspections by the Correctional Institution Inspection Committee).

⁴⁷⁹ Cásarez, *supra* note 478, at 295; Dolovich, *supra* note 5, at 490, 493–95.

⁴⁸⁰ See Low, *supra* note 222, at 38.

⁴⁸¹ See Jody Freeman, *Extending Public Law Norms Through Privatization*, 116 HARV. L. REV. 1285, 1317 (2003).

⁴⁸² See generally Volokh, *supra* note 149.

⁴⁸³ See Cásarez, *supra* note 478, at 293–94 (noting that the American Correctional Association requires that certain records be maintained “for facility accreditation and the contracting agency”).

⁴⁸⁴ 5 U.S.C. § 552 (2012).

⁴⁸⁵ See *Forsham v. Harris*, 445 U.S. 169, 180 (1980) (stating that whether a private firm is subject to FOIA depends on whether it’s subject to extensive, day-to-day government control); see also Cásarez, *supra* note 478, at 268–79.

⁴⁸⁶ See *Kissinger v. Reporters Comm. for Freedom of the Press*, 445 U.S. 136, 152 (1980) (explaining that FOIA requires an agency to disclose only documents it has “created and retained”); see also Cásarez, *supra* note 478, at 279–84.

information . . . [that is] privileged or confidential.”⁴⁸⁷ Exemption 4 could be applied either if “disclosure could impair the reliability of data,”⁴⁸⁸ or if “disclosure would cause substantial competitive injury to the provider.”⁴⁸⁹ The competitive injury justification could be fairly broad—knowing the terms of a contract, for instance, can reveal the terms of the winning proposal to the winning firm’s competitors.⁴⁹⁰ Indeed, FOIA has been criticized as “a lawful tool of industrial espionage.”⁴⁹¹ On the other hand, says Cásarez, FOIA provides for the disclosure of “reasonably segregable portion[s]” of documents,⁴⁹² which “should include monitoring and reporting requirements.”⁴⁹³ Logan counsels against “saddl[ing] private prison operators with expensive monitoring requirements ‘far beyond those that exist for government prisons,’”⁴⁹⁴ but FOIA applicability would cut in the direction of establishing parity.

Similar legislative fixes are possible in the states: for instance, in Florida and Georgia, open records acts “already apply to private organizations that act on behalf of state agencies.”⁴⁹⁵ All of this (as well as any relevant public-law value) could also be imposed on private contractors by contract; Jody Freeman calls this process “publicization.”⁴⁹⁶

Another possibility is to ensure access to the prison by the public and the press.⁴⁹⁷ Bentham, who had smart things to say about the bidding process two centuries ago,⁴⁹⁸ also argued for “essentially unrestricted public access”⁴⁹⁹ to (private) facilities. His prison design

enables the whole establishment to be inspected almost at a view, it would be my study to render it a spectacle, as persons of all classes

⁴⁸⁷ 5 U.S.C. § 552(b)(4); Cásarez, *supra* note 478, at 284–91.

⁴⁸⁸ Cásarez, *supra* note 478, at 287 (citing *Critical Mass Energy Project v. NRC*, 975 F.2d 871, 878 (D.C. Cir. 1992)).

⁴⁸⁹ *Id.*

⁴⁹⁰ *See id.* at 289; *see also supra* text accompanying note 36.

⁴⁹¹ Cásarez, *supra* note 478, at 292 (quoting Stephen S. Madsen, Note, *Protecting Confidential Business Information from Federal Agency Disclosure After Chrysler Corp. v. Brown*, 80 COLUM. L. REV. 109, 113 (1980)) (internal quotation marks omitted).

⁴⁹² 5 U.S.C. § 552(b).

⁴⁹³ Cásarez, *supra* note 478, at 289.

⁴⁹⁴ *Id.* at 260 (quoting CHARLES H. LOGAN, PRIVATE PRISONS: CONS AND PROS 147 (1990)).

⁴⁹⁵ *Id.* at 296 (citing as examples FLA. STAT. ANN. § 119.011(2) (West 1985); GA. CODE ANN. § 50-18-70(a) (Michie 1994)).

⁴⁹⁶ Freeman, *supra* note 481, at 1285. The term is pronounced [püb’lī-kī-zā’shən]. *Id.* at n.1.

⁴⁹⁷ Cásarez, *supra* note 478, at 299 (citing Robbins, *supra* note 478, at 752–53).

⁴⁹⁸ *See supra* text accompanying note 252.

⁴⁹⁹ Durham, *supra* note 20, at 69.

would, in the way of amusement, be curious to partake of: and that not only on Sundays at the time of Divine service, but on ordinary days at meal times or times of work: providing therefore a system of inspection, universal, free, and gratuitous, the most effectual and permanent securities against abuse.⁵⁰⁰

I don't want to endorse watching prisoners as a source of amusement (and public access raises serious security and access-to-contraband issues), but the idea of at least some public access does seem to have some advantages in terms of accountability.

CONCLUSION

The failure of the comparative effectiveness studies, therefore, is completely understandable. Aside from the methodological problems, it's quite plausible that the results of prison privatization have been inconclusive because the changes in prison management that would lead to better performance are often neither permitted nor rewarded.

Using performance measures would change this by helping us do valid comparative studies, enabling the fair public-private competitions that are a hallmark of competitive neutrality, and pushing policymakers to clearly formulate what we want out of prisons. Using performance measures directly to drive compensation has the potential to radically alter prison outcomes by rewarding good performance and penalizing bad performance; this definitely has applicability for private prisons but could possibly be used for public prison wardens as well.

The critiques are serious, but I don't believe they undermine the experiment too seriously.

The information necessary to calculate the True Social Values in an efficiency framework may never be available, but we can approach the exercise with an air of humility, seeking only to improve incentives at the margins, not to achieve optimal social engineering.

The use of market incentives probably won't alter the public-interestedness of those who work at private prison firms, but it might alter the mix of people who choose to work in the public sector; on the other hand, combined with social impact bonds, performance-based compensation can also spur the

⁵⁰⁰ *Id.* (quoting JEREMY BENTHAM, A BENTHAM READER 200 (Mary Peter Mack ed., 1969)).

growth of nonprofit providers. Because small firms and nonprofits are particularly sensitive to risk, the incentives should only be moderately high-powered, to trade off incentives and risk tolerance.

Performance-based compensation will give rise to certain possibly undesirable strategic behavior. If providers can set their own goals, they'll be inclined to set them in ways that are easy to meet; this is why providers shouldn't set the goals at all, and in any event compensation should be based on the level of a continuous variable, not a binary goal. If some dimensions of quality are hard to measure, performance-based compensation will bias providers' effort toward the more measurable aspects of performance; this means that some reliance on inputs and outputs will still be necessary, having due regard for the need to avoid choosing measures that can be easily and undesirably manipulated by providers. Compensation schemes might lead providers to concentrate on treating certain inmates and neglect others; even if this is bad (which isn't clear), the problem can be alleviated by inmate-specific rewards. Finally, the levels of the measures themselves can be falsified, which points to the need for serious investments in monitoring and robust disclosure regimes.

These concerns are real, but the lesson to take from them is that more experimentation is required to see how much of a real-world effect they have and to what degree they really vitiate the promise of performance incentives. The status quo, where the level of experimentation is close to zero, is unlikely to be optimal.