Cream Skimming, Parking, and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts

Pierre Koning
Carolyn J. Heinrich

Abstract

As performance-based contracting in social welfare services continues to expand, concerns about potential unintended effects are also growing. We analyze the incentive effects of high-powered, performance-based contracts and their implications for program outcomes using panel data on Dutch cohorts of unemployed and disabled workers that were assigned to private social welfare providers in 2002 to 2005. We employ a difference-in-differences design that takes advantage of the fact that contracts gradually moved from partial performance-contingent pay to full (100 percent) performance-contingent contracting schemes. We develop explicit measures of selection into the programs and find evidence of cream skimming and other gaming activities on the part of providers, but little impact of these activities on program outcomes. Moving to a system with contract payments fully contingent on performance appears to increase job placements, but not job duration, for more readily employable workers. © 2013 by the Association for Public Policy Analysis and Management.

INTRODUCTION

Performance-based pay and incentives are more prevalent in the private than public sector, as are studies of their use and effectiveness (see Chiappori & Salanié, 2003; Prendergast, 1999, for surveys of the literature). Compared to private sector production technologies, public sector work more frequently involves complex, nonmanual work; multiple principals and group dynamics; political and environmental influences and interdependencies; and nonstandardized outputs that make precise measurement of performance challenging and costly (Blank, 2000; Brown & Potoski, 2003; Dixit, 2002; Heinrich & Marschke, 2010). The additional noise and special challenges in measuring public sector performance outcomes likely account at least in part for the predominance of low-powered incentive structures, which imply relatively weak incentives or consequences (monetary or nonmonetary) for achieving (or failing to achieve) target levels of performance (Burgess & Ratto, 2003; Heinrich, 2007).

In this context, our research presents a unique opportunity to investigate a special case in which the delivery of publicly funded social services moved to a system with very high-powered incentives, specifically, with payments to contractors fully (100 percent) contingent on their performance. We evaluate the effects of these incentives on client outcomes (job placements), as well as the unintended effects of

Journal of Policy Analysis and Management, Vol. 00, No. 0, 1–23 (2013)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.21695

fully performance-contingent contracting on the delivery of welfare-to-work (WTW) services in the Netherlands. In effect, this public sector setting presents a rare case where both the financial and service provision risks were wholly transferred to the (private) social welfare providers.

While performance-based pay has a long history, its use in public sector contracts for delivery of social welfare services by private providers is a more recent and increasingly common phenomenon. Some of the first experiences with performance-based incentive schemes in the public sector were in U.S. employment and training programs under the Job Training Partnership Act (JTPA) of 1982 (Barnow, 2000; Heckman, Heinrich, & Smith, 2002). More recently, governments in Australia and the Netherlands switched to fully privatized systems with substantial performance incentives (Bruttel, 2005; Finn, 2008; Struyven & Steurs, 2005). Contracting out public social welfare services to private providers is less widespread in the United Kingdom and Germany than in the United States—where it has been estimated that up to 80 percent of human services funding is contracted out (Martin, 2005)—but it is gradually becoming more important in these countries as well (Bernard & Wolff, 2008; Tergeist & Grubb, 2006; Winterhager, Heinze, & Spermann, 2006).

The expanded use of performance-based contracts in social welfare services is intensifying concerns about their potential unintended effects. Although there is some evidence that these contracts increase measured performance, they are often accompanied by various distortionary effects (Courty & Marschke, 2004; Courty, Kim, & Marschke, 2011; McBeath & Meezan, 2010). In some settings, this has led to a retraction of performance-based contract incentives, with governments reverting to fixed-payment schemes for private providers or even contracting back in (Hefetz & Warner, 2004). Other examples include the reappraisal of in-house provision of WTW services to Social Assistance recipients in the Netherlands and changes in the contract design and management of the Wisconsin Works program (Heinrich & Choi, 2007; Koning, 2012). These reversals and revisions undoubtedly reflect in part the challenges of getting incentives right and managing performance-based contracting arrangements. Still, there is little empirical evidence on the overall effectiveness of performance-based contracting schemes in social welfare programs that factors in both their intended and unintended effects and assesses their relative importance for program outcomes.

We aim to help fill this gap in the literature and break new ground by analyzing the implications of high-powered, performance-based contracts with Dutch WTW providers for unemployed and disabled worker outcomes, as well as their unintended effects. In making this contribution, we first observe the numbers of workers that were assigned to WTW providers over the period 2002 to 2005, including those that did not actually start programs, which allows us to develop explicit measures of preprogram selection. We test whether the gradual move to higher-powered (fully performance-contingent) contract incentives—where payments were made only for clients placed in jobs—changed the shares of clients that did not start programs. In doing so, we distinguish between clients that did not show up for the program (i.e., client-induced selection) and clients returned to the social benefit administration by providers (i.e., provider-induced selection). We then empirically assess the impact of fully performance-contingent contracting on workers' short- and long-term job placement rates, accounting for selection in the allocation of workers across contracts using difference-in-differences models.

This paper is one of the first to explicitly and empirically address a phenomenon that we characterize as the "parking" of hard-to-serve clients. In our analysis, parking is defined as not providing services to hard-to-place clients who have been assigned to WTW programs. In other words, parking might be described as a form of "cream-skimming" during the program, in which providers try to keep costs down by doing little to serve those with the poorest anticipated outcomes, while instead focusing resources on more able clients with better employment prospects. In the absence of detailed program administrative records, parking behavior, as well as spending per client, may be difficult to discern. If program staff direct the bulk of their resources to clients with better preprogram job prospects, we expect parking behavior to accelerate the rate of job placements among the more employable. This would thereby reduce their observed average job search duration, while simultaneously widening the gap in job placement rates between clients with better and worse job prospects. Accordingly, we develop a test statistic for parking that compares observed average job search durations of job finders in partial vs. full performance-contingent contracts, with the expectation that higher-powered incentives are more likely to encourage parking behavior.

Our analysis shows that the preprogram selection and parking activities were minimal among groups of workers where the risk of failing to place clients in jobs was lower (i.e., spread over a broader client base). Conversely, for smaller contracted worker groups with greater risks of nonpayment due to performance, the evidence points to more selection and parking under fully performance-contingent contracts. Although job placement rates for some workers increase under these contracts, job duration was unaffected. In addition, job placement rates for those with poorer employment prospects did not increase under these high-powered incentive contracts.

The remainder of the paper proceeds as follows. We first briefly review the literature on cream skimming, parking, and other gaming activities, taking special interest in the role of institutions in determining or incentivizing these phenomena. Next we describe the data and methods we use, while explaining the institutional context in which they originate. We then present the empirical analysis and findings and conclude in the final section.

LITERATURE REVIEW

Research to date suggests that the nature and incidence of cream skimming, parking, and other gaming activities are strongly influenced by institutional settings. As Heinrich and Marschke (2010) explain, incentives should be used (and should be more powerful) in organizations where individuals are able to respond to them. In some public sector settings, workers are highly constrained by procedural rules and regulations that allow little discretion for manipulating or improving program processes or by environmental conditions that interfere with outcomes (e.g., a high unemployment rate or an economically disadvantaged population). Thus, imposing performance-based contractual provisions subjects them to greater risk of lower compensation. In other organizational environments, program implementers may have more leeway to try out innovative ways of increasing program value and may therefore be more responsive to performance incentives in intended ways. In the typical, relatively constrained social service setting, however, unintended practices such as cream skimming (i.e., the selection of easier-to-place clients by providers) may increase if payments or rewards are performance based, as workers with poorer preprogram prospects of success will raise the risk of no (or lower) payments (Heckman, Heinrich, & Smith, 2002). Of course, this response is conditional on the extent to which providers are able to select clients, and to select them in ways that influence performance outcomes.

In U.S. employment and training programs, for example, program participation is voluntary, with decisions on the part of both potential participants and gatekeepers (program administrators) determining access to services. Heckman and

Smith (2004) analyzed multiple stages of the JTPA selection process and concluded that applicants' progression through these stages was less influenced by cream skimming on the part of program workers than by factors beyond the control of service providers (Heckman & Smith, 2004). Moreover, there is no strong evidence in the literature that social service providers routinely exploit the discretion they have to cream skim, and in some contexts, the incentives to do so may be mitigated by formal adjustments to performance standards that reduce the risks of serving specific hard-to-serve groups (Barnow & Heinrich, 2010; Courty, Marschke, & Kim, 2011; Heinrich, 1999). Courty, Marschke, and Kim (2011) show empirically how formal performance standard adjustments compel providers to factor in not only how client characteristics affect performance outcomes, but also how they influence performance standards.

In other settings, admission into programs may be compulsory, allowing little room for selection by providers. Examples include programs in Australia, the United Kingdom, and the Netherlands where participation in WTW programs has been compulsory for unemployed workers (Finn, 2008). In these programs, all assigned clients do not necessarily enroll, as some prefer to risk being sanctioned or may find jobs prior to the program on their own. In Australia and the United Kingdom, the assignment of workers to WTW providers is fully compulsory, implying that providers have no discretion to return workers to the social benefit administration (i.e., to decline to serve them). In contrast, providers in the Netherlands are allowed to return clients assigned to them, but with the caveat that returning too many workers may decrease their chances of being contracted to provide services in the future (Koning, 2008).

While cream skimming is probably less important and prevalent than suspected, particularly in programs with little leeway for selection by providers, concerns persist about the possibility of other gaming activities that occur in the implementation of programs, such as limiting access to services or parking activities (Bruttel, 2005; Finn, 2008). One view is that the parking of assigned clients after the start of a program—which results in a bare minimum of services for harder-to-serve clients—may represent a substitute for cream skimming, particularly in contexts where preprogram selection is effectively prohibited or restricted. WTW providers may still spend time in assessing the preprogram job prospects of workers, but use this information instead to determine how they will allocate "parking spaces." The experience of some providers, clients, and caseworkers, as well as anecdotal evidence, suggests an association between incentive-based contracts and parking activities (Finn, 2008).

This type of parking behavior has also been observed in U.S. public education systems that assess performance based on student achievement levels (test scores) relative to proficiency standards. In these systems, educators have been shown to exert or reallocate more effort to teaching "bubble kids," or those on the cusp of reaching the performance standard (i.e., whose individual performance is more likely to influence attainment of the standard). For example, Figlio and Rouse (2006) found that Florida schools that faced pressure to improve the fraction of students scoring above minimum levels on comprehensive assessment tests focused their attention on students in the lower portion of the achievement distribution and on the subjects tested and grade levels included in the exams, at the expense of higher achieving students and subjects not included in the test. And in their study of public school accountability systems in Chicago, Neal and Schanzenbach (2010) similarly concluded that a focus on proficiency standards left out students performing both far below and far above the standard. Still, empirical evidence on the importance of parking (in various forms) is limited, most likely due to difficulties in observing provider (or teacher or other service worker) efforts and resource allocations at the client level.

Among other empirical analyses of gaming of measured outcomes in public services provision were those conducted by Courty and Marschke (1997, 2004), who analyzed how JTPA providers manipulated the timing of client exits from employment and training programs to maximize their rewards. Courty and Marschke showed that providers were more likely to postpone the exit of poorly performing clients in relatively unsuccessful program years and to increase the number of exits of these clients in good years when they were confident of achieving the minimum performance standards required to secure an award. In effect, these types of gaming activities increase measured performance, but lower the efficiency of the system by diverting provider efforts to these gaming activities.

The incidence and types of these gaming activities appear to be strongly influenced by institutional settings. In the JTPA program, a combination of discretion over the timing of reporting on program outcomes and the use of performance thresholds triggered providers to game the system in these ways (Heckman et al., 2011). With linear contracts like those used in the Dutch WTW contracting system, gaming along these lines is less likely to benefit providers, particularly if client records are linked directly to the administration of benefits in a way that leaves no discretion in the timing of reporting successful program exits. Providers may still use their discretion, however, in determining the end date of the program, where program extensions may result in additional windfall gains of job placements. In general, the implications of high-powered incentive contracts, considering both their intended and unintended effects on social welfare, remain unclear.[1]

DATA AND INSTITUTIONAL CONTEXT

In the Netherlands (the setting of this study), the Disability Insurance (DI) and Unemployment Insurance (UI) programs are mandatory for workers seeking benefits. Both DI and UI are implemented by the social benefit administration. Similar to the New Deal in the United Kingdom, providers are expected to offer unemployed and disabled clients mediation, job training, or subsidized employment (WTW services) within 12 months after the start of their benefits.[2]In the period under investigation (2002 to 2005), the Dutch social benefit administration contracted out the delivery of these WTW services only to private job training service providers.[3]Later (in 2007), the privatization of social welfare services was reversed to some extent when the social benefit administration resumed delivery of social welfare services for clients with relatively good job prospects, leaving those with poorer job prospects to be contracted to private providers.

This paper uses registered data from the Dutch social benefit administration on procured WTW programs, including details of the contracts with private providers. Each year, the Dutch social benefit administration sorted workers using two classification levels in order to assign them to private providers in groups and un-

[1] One exception appears is the study by Burgess et al. (2004), who investigated the effectiveness of performance-related pay for public employment offices.

[2] Starting in 2005, the Netherlands relaxed the New Deal approach, reducing the number of WTW clients in 2005.

[3] In 2006, the delivery of social welfare services by the social benefit administration was replaced by a dual system, where clients could either opt for individual vouchers to choose (preferred) providers or were assigned to groups that were allocated to providers, as in the system that was compulsory until 2005. There also was a major DI reform of benefit conditions, where the DI scheme was split into a more generous scheme for fully and permanently disabled workers and a separate scheme for the partially and temporarily disabled workers.

Table 1. Gross worker types in study sample: Distribution among worker program types, worker groups, clients, and contract types.

Gross worker type	Worker program types	Worker groups	Total clients assigned	Fraction fully perf. contingent
Disabled immigrants, bad job prospects	9	1,157	10,583	0.000 (.)
Disabled, good job prospects	8	954	40,737	0.243 (0.439)
Disabled: mental impairments	6	805	31,517	0.000 (.)
Disabled: mild mental impairments	9	1,097	30,709	0.000 (.)
Disabled: work related impairments	5	49	722	0.000 (.)
Disabled: kidney patients	4	35	510	0.000 (.)
Disabled: visual and hearing impairments	4	152	1,988	0.000 (.)
Disabled: young workers, no work history	14	757	7,325	0.023 (0.151)
Unemployed immigrants	2	190	6,616	0.000 (.)
Unemployed, good job prospects	9	826	51,177	0.629 (0.483)
Unemployed, older than 50 years	9	537	20,358	0.000 (.)
Unemployed, bad job prospects	9	705	18,409	0.000 (.)
Unemployed: highly educated	2	52	962	1.000 (.)
Workers in graphical industry	5	49	722	0.000 (.)
Workers eligible for subsidized employment	1	89	767	0.000 (.)
Returned clients ("second chance" programs)	5	427	7,649	0.097 (0.296)
Trajectories aiming at self-employment	2	92	1,081	0.000 (.)
Other	7	586	12,720	0.508 (0.500)

der specific contract conditions. The explicit intent was to allow providers to specialize in particular types of services or groups of clients, while at the same time decreasing opportunities to cream skim (or cherry-pick the best individual clients).

The first level of classification is a broad characterization of workers into 18 different categories or "gross worker types." Some examples of these specific client groupings in the UI and DI programs include immigrants; those with good or bad job prospects (as defined through profiling by social benefit administration caseworkers); disabled workers with special impairments; and older workers or young disabled workers without any work history (see Table 1). We use this classification level, for which the categories do not change over the study period, to construct control dummies in our empirical models.

The second level of classification constitutes a more detailed description of the contract types and activities, which we call "worker program types." The designation of worker program types (106 in total) has varied over the years, depending on specific program needs at a given time. Examples include programs aimed at guiding particular groups of workers into self-employment, job training, job search, or job mediation. These worker program types are, in effect, combinations of gross worker types and specific program activities of gross worker types, defined distinctly for each year. Moreover, contract conditions specifying partial or 100 percent performance-contingent payments also vary at this (worker program type) level. It is important to emphasize that these contract conditions did *not* result from negotiations between the social benefit administration and providers; rather, they were set prior to the bidding process, with an increasing share of 100 percent performance-contingent contracts over time.

We also distinguish the actual group of workers that was assigned to a particular provider in a given region and at a specific time point. We call this a "worker group." The worker group is the primary unit of analysis in our study. There is no variation within worker groups in contract conditions.

Table 1 shows additional information on the structure, size, and characterization of clients in gross worker types, worker program types, and worker groups. Table 2 presents descriptive statistics of the characteristics of worker groups that were contracted out to WTW providers in 2002 to 2005, stratified by calendar years and contract payment schemes. The first four lines of Table 2 provide an overview of the structure of the data (the classification of the worker groups), whereas lines 5 to 16 show sample statistics of the worker groups.

As seen in Table 2, there is a steady decline in the number of WTW clients in the period under investigation. Specifically, the numbers of newly assigned program participants dropped from about 105,000 clients in 2002 to approximately 67,000 in 2003 and from about 48,000 in 2004 to 23,000 in 2005. Two possible explanations for this pattern have been identified: (1) business cycle effects and (2) a relaxation of the New Deal strategy from 2005 onward, where the social benefit administration no longer aimed to fully treat clients within 12 months after the start of their UI or DI spell (Finn, 2008; Groot et al., 2006). The new procurement system in 2002 also applied to clients who entered in prior years (when the New Deal strategy had not yet been implemented), suggesting a backlog of clients may have been waiting for services in 2002.

Provider- and Client-Induced Selection

During the contracting process, individual clients within a particular worker group were not known yet by the WTW provider. Providers were only informed of worker characteristics at the gross worker type level and of the expected group size and contract conditions. Thus, providers did not know at the time of the contract award which clients would be assigned to them; this depended on the regional supply of relevant workers at a specific point in time.

After contracts were awarded, providers had to contact the clients assigned to them and make a reintegration (service) plan with the clients. This reintegration plan included a list of proposed activities to get the client back to work, as well as the rights and duties of the provider and the client, and had to be approved by the social benefit administration. In practice, reintegration plans were not formulated for all individual clients and were not approved in all cases. Thus, we distinguish

Table 2. Sample statistics of contracted worker groups by year and contract type: Observations and population averages (standard deviations in parentheses).

	Years				Contract types		
	2002	2003	2004	2005	Partial perf. contingent	Fully perf. contingent	Total
1. Number of gross worker types	16	13	10	11	17	6	18
2. Number of worker program types	52	20	17	17	93	13	106
3. Number of worker groups	3,733	1,292	1,690	1,795	7,441	1,069	8,510
4. Number of assigned clients	105,623	66,861	47,910	23,436	193,361	50,469	243,830
Contract and client characteristics							
5. Participants per gross worker group	28.3	51.8	28.4	13.1	26.0	47.2	28.7
	(42.3)	(62.8)	(32.7)	(19.3)	(37.2)	(66.0)	(42.5)
6. Participants per net worker group	24.2	43.0	23.7	10.9	22.1	38.5	24.1
	(36.0)	(51.5)	(27.5)	(15.9)	(31.7)	(53.2)	(35.5)
7. Fully performance-contingent contract	0.000	0.351	0.296	0.546			0.126
	(0.017)	(0.477)	(0.457)	(0.498)			(0.331)
8. Disability insurance (DI)	0.610	0.413	0.545	0.370	0.592	0.246	0.520
	(0.487)	(0.492)	(0.498)	(0.483)	(0.491)	(0.431)	(0.500)
9. Average payment per client (euros)	–	3,690	3,601	3,272	3,953	2,898	3,579
		(1,167)	(784)	(1,293)	(1,015)	(917)	(1,104)
Outcomes							
10. Client-induced selection	0.118	0.147	0.150	0.156	0.128	0.165	0.136
	(0.073)	(0.075)	(0.107)	(0.135)	(0.089)	(0.088)	(0.090)
11. Provider-induced selection	0.028	0.022	0.014	0.013	0.023	0.020	0.022
	(0.034)	(0.029)	(0.024)	(0.035)	(0.032)	(0.030)	(0.031)
12. Duration of program participation, job finders	272.8	291.8	276.3	207.7	288.5	215.2	273.0
	(89.4)	(78.9)	(87.6)	(97.1)	(90.1)	(59.0)	(89.7)
13. Duration of program participation, nonfinders	467.5	564.8	526.2	418.0	514.3	451.7	501.3
	(91.0)	(102.0)	(109.0)	(131.6)	(111.5)	(105.2)	(113.1)
14. Job placement (total)	0.327	0.334	0.368	0.351	0.316	0.429	0.339
	(0.121)	(0.124)	(0.135)	(0.194)	(0.129)	(0.116)	(0.134)
15. Job placement: contract of six to 12 months	0.257	0.257	0.263	0.257	0.240	0.329	0.258
	(0.107)	(0.106)	(0.117)	(0.162)	(0.112)	(0.100)	(0.115)
16. Job placement: contract > 12 months	0.070	0.077	0.105	0.094	0.076	0.100	0.081
	(0.059)	(0.047)	(0.073)	(0.085)	(0.065)	(0.054)	(0.063)

gross numbers of clients that were assigned to a provider from *net* numbers of clients for which reintegration plans were submitted and approved by the social benefit administration. The resulting difference between gross and net participation, shown in Table 2, therefore reflects both provider- and client-induced selection. As to the former, providers were allowed to send back some clients if they viewed them as unsuitable for the program. We suggest that this form of selection is most likely to be associated with cream skimming. Over the years, the population-weighted fraction of provider-induced selection ranged from 1 to 3 percent of the gross worker types. Although there was no formal limit on provider-induced selection, providers were aware that high rates of this type of selection would diminish their prospects of future contracts.

We define *client-induced* selection as the balance of assigned clients who did not participate in programs, but were also not returned by the service provider. This group likely consisted of some clients that already had found a job by the time of program start or for whom reintegration plans were not approved by the social benefit administration. In both cases, clients were no longer assigned to their respective job training service provider, and thus, did not affect providers' future prospects for contracts. Client-induced selection increased from 11 percent in 2002 to 16 percent in 2005. In addition, client-induced selection was about four percentage points higher for fully performance-contingent contracts than partial performance-contingent contracts, suggesting that clients assigned to the former type of contract may have been more likely to find jobs in the time between assignment and program start.

Partial and Full Performance-Contingent Contracts

In 2002, the new conservative government of the Netherlands announced its plans to move to a fully performance-contingent payment system for social welfare organizations. As Table 2 (line 7) shows, the share of fully performance-contingent contracts for worker groups gradually increased from close to zero in 2002 to more than half of the contracts in 2005 (weighted by the number of program participants), as the social benefit administration gradually implemented these plans. However, these high-powered contracts were never effectuated for all gross workers types. In particular, Table 1 shows that these contracts were confined primarily to gross worker types with better preprogram prospects. Typically, about 50 percent of the *partial* performance-contingent contracts were paid as a fixed amount, with the other 50 percent paid at placement.

Table 3 presents the estimation results of a simple probit model that predicts the use of high-powered contract incentives in the years 2003 to 2005. In this model, we specify 2002 as a baseline year (where fully performance-contingent contracts were nonexistent) and include average values of the job placement rates and risk proxies by gross worker type for 2002. In light of the 2002 change in government that prompted the social benefit administration's new contracting practices, we argue that these baseline observations are exogenous with respect to later program outcomes (e.g., job placement rates). We also control for the providers' market share of worker group observations, the (expected) reward per client, and an indicator for worker groups with DI recipients only.

The estimation results shown in Table 3 are consistent with a standard risk-incentives framework (Burgess & Ratto, 2003). First, we find the probability of fully performance-contingent contracts in 2003 to 2005 increases with the 2002 average job placement rate by gross worker type. In particular, a 10 percentage point increase in the baseline job placement rate increases the probability of a fully performance-

Table 3. Probit estimation results for fully performance-contingent worker groups, 2003 to 2005 (marginal effects; standard errors in parentheses).

	Coefficient	Standard error
Average job placement rate (at gross worker type level in 2002)	1.372*	(0.659)
Risk proxy (or standard deviation of the 2002 average job placement rate)	−0.119*	(0.083)
Provider market share	−0.095	(0.111)
(Expected) Reward per client	−0.090*	(0.039)
Disability insurance (dummy)	0.026	(0.060)
Year = 2004	−0.054	(0.071)
Year = 2005	0.066	(0.100)
Number of observations	3,959	
Pseudo-R-squared	0.484	
Log-likelihood	−1,010.2	

Note: We also included region dummies in our model, but do not report the corresponding estimation coefficients, which were all statistically insignificant.

*Statistical significance at $\alpha = 5$ percent or less.

contingent payment scheme by about 14 percentage points. In a framework with risk-adverse providers (i.e., with convex utility curves) and in the absence of other adjustments to performance benchmarks for client mix, this result suggests that risks are more likely to be transferred to providers in the form of performance-contingent pay when job placement probabilities for worker groups are higher; that is, a lower risk premium is required for worker groups with higher probabilities of success. Or alternatively, WTW providers may spend less effort in helping clients if there is no risk premium to compensate them for clients' poorer job prospects. In general, institutional knowledge suggests that higher payments (i.e., premiums) offered per placement for a given worker group signals the difficulties associated with placing these groups of workers into jobs and the risks for job placement rate outcomes associated with serving them.

We also included the standard deviation of the 2002 average job placement rate (of the observed worker group per client) as a proxy for risk involved with the contracted group. Consistent with economic intuition, Table 3 shows that this variable is negatively associated with the probability of high-powered contracts. Risk premiums per client are expected to be lower with larger worker groups, implying a lower variance of average placement rates. It is also possible, however, that providers with larger market shares were more likely to bid for and be awarded these groups of workers, given that there was no opportunity to negotiate on risk premiums (with exogenous contract conditions). We did not find a relationship between the provider market share and the likelihood of fully performance-contingent contracting schemes. Furthermore, the probit model estimates do not indicate that the concentration of (large) WTW providers for these contract types increased over time.

An alternative test for concentration effects of providers would involve a direct comparison between measures of market concentration of partial vs. full performance-contingent contract types. Therefore, we calculated Herfindahl-Hirschman Indices (HHI), using the market shares of all providers per worker program type j and per year t. The resulting (average) HHI scores are fairly constant over the years, ranging between 0.25 and 0.30. We also do not find a statistically significant difference between the HHI averages of worker program types in partial vs. full performance-contingent contracts, suggesting that large-scale providers were not expanding their market shares in worker groups served under high-powered incentive contracts.

Parking, Gaming, and Job Placement

Over the time period of this study, we observe a gradual increase in job placement rates of new worker groups that join WTW programs. This holds for both job placements with employer-employee contracts of six to 12 months and contracts for more than 12 months. Worker job placement rates were measured at the end of the program for all (net) assigned clients. Programs typically lasted one or two years; for clients that did not succeed in finding a job, outcomes were measured at the end of program participation. As job placements were typically realized in the first month of program spells, any variation in job placement rates due to variation in the length of program participation is likely small. We control for gross worker type (with dummies) to account for variation in program participation length by gross worker type.

In our analyses, we explore whether shorter job search durations of clients in fully performance-contingent contracts correspond to the presence of parking activities in these programs, that is, whether providers concentrate on shortening job search durations among clients with more favorable job prospects, while letting hard-to-place clients languish in the programs. Figure 1 (A and B) depicts the distributions of (average) job search durations for successful and unsuccessful worker groups by contract type, respectively. These distributions suggest that the average job search durations of both job finders and nonjob finders were shorter (with less dispersion) for the high-powered contracts. More specifically, it appears that most WTW programs for unsuccessful participants end after two years for worker groups under fully performance-contingent contracts, whereas they continue for as long as three years for worker groups under partial performance-contingent contracts. A more formal analysis is required, however, to identify the role of contract incentives in these patterns of effects, as job search durations are undoubtedly driven by participant characteristics as well.

METHODS AND MODEL SPECIFICATION

We now turn to our empirical analysis of how the shift to 100 percent performance-contingent contracting has influenced the responses (intended and unintended) of providers in implementing these programs and overall program effectiveness. The higher-powered incentive contracts were intended to increase client job placement rates. However, we also expect preprogram cream skimming in the form of provider-induced selection—and to a lesser extent client-induced selection—to increase under fully performance-contingent contracts, along with parking behavior (measured using information on the average successful job search duration per worker group in our sample).

It is essential to adequately control for unobserved heterogeneity and selection bias in our empirical analysis of performance-based contracting (Chiappori & Salanié, 2003). For example, the allocation of contract (worker program) types may be driven by expected job prospects of the gross worker types. We therefore use a difference-in-differences design to control for worker characteristics affecting both the payment scheme and program outcomes, including gross worker type dummies to control for preprogram differences in worker characteristics. In the

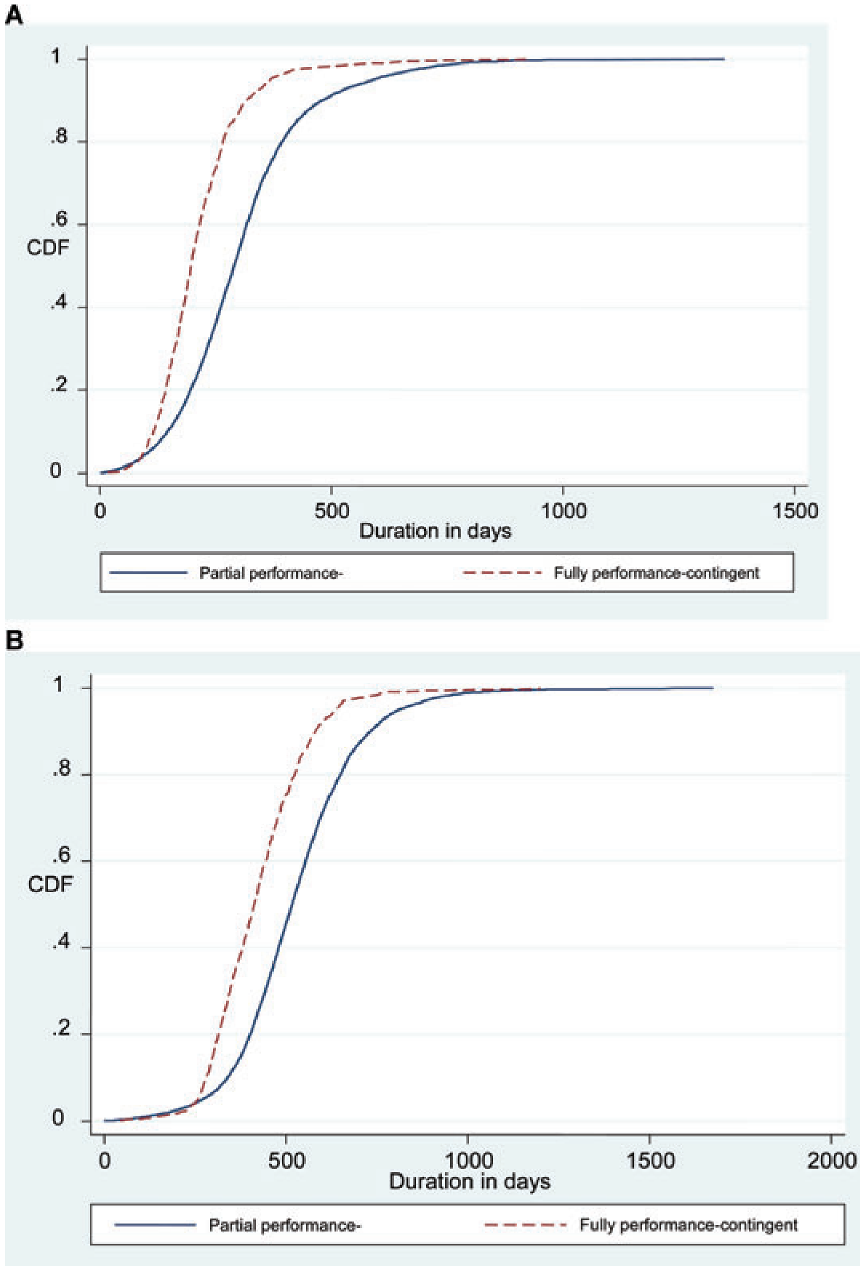

Figure 1. (A) Cumulative Density Function Distributions of Average Successful Durations (in Days): Worker Groups Under Partial vs. Fully Performance-Contingent Contracts (2002 to 2005). **(B)** Cumulative Density Function of Average Unsuccessful Durations (in Days): Worker Groups Under Partial vs. Fully Performance-Contingent Contracts (2002 to 2005).

time period under consideration, the workers in these classification groups remained more or less similar, even as some gross worker groups gradually moved to fully performance-contingent contracts. In our baseline model, we start by assuming that the treatment (fully performance-contingent) and comparison (partial performance-contingent) groups are affected equally by calendar time effects. Next, we relax the equal-trends assumption to test the robustness of our results.

We formally model three outcomes—provider-induced selection (S^P), client-induced selection (S^C), and job placement rates (JPR)—using similar (linear) specifications:

$$S^P_{ijk,t} = a^P 100\%PC_{ij,t} + \boldsymbol{X}_{ijk,t}\beta^P + \gamma^P t + \sum\nolimits_r \delta^P I(r = i) + \varepsilon^P_{ijk,t} \tag{1}$$

$$S^C_{ijk,t} = a^C 100\%PC_{ij,t} + \boldsymbol{X}_{ijk,t}\beta^C + \gamma^C t + \sum\nolimits_r \delta^C I(r = i) + \varepsilon^C_{ijk,t} \tag{2}$$

$$\begin{aligned} JPR^C_{ijk,t} = {} & a^{JPR} 100\%PC_{ij,t} + \boldsymbol{X}_{ijk,t}\beta^{JPR} + \gamma^{JPR} t + \eta^P S^p_{ijk,t} + \eta^C S^C_{ijk,t} \\ & + \sum\nolimits_r \delta^{JPR} I(r = i) + \varepsilon^{JPRi}_{jk,t}, \end{aligned} \tag{3}$$

with worker groups indexed by k ($k = 1 \ldots K$), worker program types indexed as j ($j = 1 \ldots J$), and gross worker types indexed as i ($i = 1 \ldots I$). The variable t indicates the year in our sample ($t = 1 \ldots T$), and I is an indicator for the event that is expressed in parentheses. The variables α^P, α^C, and α^{JPR} describe the effects of high-powered contracts on provider-induced selection, client-induced selection, and job placement rates, respectively. β^P, β^C, and β^{JPR} are vectors denoting the effects of the matrix $\boldsymbol{X}$ on provider- and client-induced selection and the job placement rates, respectively. The matrix $\boldsymbol{X}$ includes a dummy for the scheme (DI or UI), the (expected) reward per client, and the log of the worker group size and regional dummies. γ^P, γ^C, and γ^{JPR} describe the effect of (linear) time trends, whereas δ^P, δ^C, and δ^{JPR} are dummy vectors accounting for preprogram differences in gross worker types. In equation 3, we also add provider- and client-induced selection as controls (${S^p}_{ijk,t}$ and ${S^C}_{ijk,t}$) to take into account possible selection effects that occur prior to the start of the trajectories. Finally, ε^P, ε^C, and ε^{JPR} are i.i.d. residual terms with mean zero and variance ${\sigma_P}^2$ and ${\sigma_C}^2$.

We estimate equations (1) to (3) with standard GLS techniques, adjusting for clustering in worker program types j. The size of the worker groups varies substantially over time, affecting the variance of (pre-) program outcomes as well. Individual clients have a higher probability of being part of large groups than smaller ones, so equal weights for all observations may yield inefficient coefficient estimates. We therefore estimate the equations with worker group size as relative (analytic) weights (see Koning, 2008).

Our key assumptions in the identification of effects of fully performance-contingent contracts are that (i) the composition of worker groups is well captured by the (constant) gross worker group types that are used, and (ii) worker groups in the treatment and comparison groups are equally affected by time trends. Conditional on these two assumptions, we can identify the effect of the payment schemes in the difference-in-differences estimation.

We test for the sensitivity of our results to these assumptions by allowing the time trends to differ between worker groups belonging to classes of gross worker groups that (at some point in time) have fully performance-contingent contracts, and those belonging to 10 classes of gross worker groups that remained under partial performance contingent contracts. This estimation might be characterized as *triple-differencing*; that is, we relax the common trends assumption and control for preprogram differences between worker types, as well as differences in time trends, and between those under full vs. partial performance-contingent contracts. One caveat is that the effects of high-powered contracts are identified from a smaller subsample of treated gross worker groups (than the more restrictive double-difference model).

Finally, we also estimate a set of models that includes year-specific fixed effects, rather than linear time trends, to allow more flexibility in the time path that outcomes are expected to follow under partial performance-contingent contracts. We report these results as an additional specification test that allows a flexible baseline.

EMPIRICAL ANALYSIS FINDINGS

Preprogram Selection and Job Placement Rate Effects

Table 4 shows the estimation results of our benchmark model for client-induced selection, provider-induced selection, and job placement rates. This table also shows the estimation results that allow for different time trends or triple-differenced estimates of the effects of fully performance-contingent contracting, as well as the flexible baseline (year-specific fixed-effects) model. Next, Table 5 presents the model results that distinguish the effects of fully performance-contingent contracting by worker group size (smaller or larger than 50 clients) and UI vs. DI clients. This table shows the decomposition of job placements into short-term contracts (six to 12 months) and job contracts of more than 12 months.

Focusing first on the two types of selection effects (provider-induced and client-induced), the coefficient estimates of fully performance-contingent contracts have the expected (positive) signs, but are only weakly significant for client-induced selection (see Table 4). As for provider-induced selection, the magnitude and range of estimates are small. In particular, the coefficient estimate of 0.572 of a percentage point (of worker groups) translates into provider-induced selection of about 0.12 clients on average (given 24 clients in an average worker group). This suggests that the potential for (any) cream skimming from this source is likely limited. We also find a strong downward time trend for provider-induced selection, suggesting that providers became more aware of the consequences of provider-induced selection over the years and may have attempted to reduce it. Moreover, this trend is considerably stronger for gross worker types under high-powered contracts.

As for client-induced selection, our coefficient estimates of the effect of fully performance-contingent contracts for double and triple differenced models are 2.331 and 3.016 percentage points, respectively (implying a 12.56 percent average increase in client-induced selection). These effects are more substantial than those of provider-induced selection, as the (average) share of client-induced selection is much larger than that of provider-induced selection (see lines 10 and 11 of Table 2). It is well possible that client-induced selection was used in more subtle ways to cream skim among assigned clients. There are well-documented means by which programs (by design) and program administrators can discourage participation in public programs (that are subsequently recorded as individual decisions not to participate) (Orbach, 2006).

Table 5 presents the estimates of fully performance-contingent contracts on provider- and client-induced selection for subsamples with different worker group sizes (more or less than 50 clients) and UI or DI beneficiaries. These estimates suggest that for both selection measures, the effects of high-powered contracts have

Table 4. GLS estimation results of provider- and client-induced selection and job placement rates (effects in percentage points; standard errors in parentheses).

	Provider-induced selection			Client-induced selection			Job placement rate		
	Diff.-in-diff.	Triple-diff.	Flexible baseline	Diff.-in-diff.	Triple-diff.	Flexible baseline	Diff.-in-diff.	Triple-diff.	Flexible baseline
Fully perform.-contingent contract	0.572	0.066	0.716**	2.331**	3.016*	2.130	2.130*	2.052*	2.081**
	(0.390)	(0.479)	(0.380)	(1.378)	(1.405)	(1.456)	(0.976)	(0.895)	(1.109)
Disability Insurance	0.831*	0.530*	0.857*	5.147*	5.556*	5.201*	−11.45*	−11.50*	−11.98*
	(0.222)	(0.207)	(0.243)	(1.118)	(1.245)	(1.151)	(2.093)	(2.212)	(2.191)
(Expected) Reward per client	−0.067	−0.062	−0.046	−0.302	−0.308	−0.300	0.579	0.579	0.391
	(0.077)	(0.079)	(0.078)	(0.294)	(0.287)	(0.334)	(0.420)	(0.422)	(0.461)
Log gross worker group size	0.106	0.128	0.133	0.294	0.263	0.259	0.093	0.097	0.070
	(0.089)	(0.083)	(0.090)	(0.287)	(0.289)	(0.315)	(0.315)	(0.332)	(0.326)
Time trend	−0.754*	−0.424*		0.632**	0.185		2.761*	2.811*	
	(0.086)	(0.160)		(0.378)	(0.812)		(0.463)	(0.528)	
Time trend treatment group		−0.483*			0.656			−0.075	
		(0.193)			(0.995)			(0.911)	
Year = 2003			−0.972*			0.965			2.667*
			(0.225)			(0.784)			(0.899)
Year = 2004			−1.642*			1.224			6.855*
			(0.226)			(1.068)			(1.035)
Year = 2005			−2.214*			2.206			6.141*
			(0.271)			(1.539)			(1.608)
Provider-induced selection							−10.07*	−10.11*	−9.919*
							(4.557)	(4.528)	(4.676)
Client-induced selection							−2.576	−2.571	−2.450
Constant	3.836*	4.389*	2.707*	13.581*	12.831*	13.912*	12.197*	12.283*	19.731*
	(0.926)	(1.060)	(0.962)	(3.931)	(4.156)	(5.078)	(5.671)	(5.750)	(6.384)
Number of obs.	8,508	8,508	8,508	8,508	8,508	8,508	8,302	8,302	8,302
F-test gross worker type	19.23	8.22	7.61	14.50	13.11	15.17	113.18	77.85	101.97
dummies; $F(18,104)$	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.00$
R-squared	0.106	0.108	0.107	0.151	0.151	0.151	0.341	0.341	0.343

Note: The regressions also included six region dummies as controls in all specifications.

*Statistical significance at $\alpha = 5$ percent or less

**Statistical significance at 10 percent.

Table 5. GLS estimation results of fully performance-contingent contracting on provider- and client-induced selection and job placement rates by worker group size and UI and DI worker groups (diff.-in-diff. estimates in percentage points; standard errors in parentheses).

		Worker group size		Benefit scheme	
	Full sample	< 50 clients	UI	DI	
Provider-induced selection	0.572	1.176*	0.419	−0.079	1.420*
	(0.391)	(0.363)	(0.402)	(0.176)	(0.184)
Client-induced selection	2.233*	5.057*	1.640	−0.100	5.498*
	(1.378)	(2.419)	(1.161)	(1.233)	(1.946)
Job placement rates	2.130*	−0.119	2.678*	2.825*	1.200
	(0.980)	(1.669)	(0.920)	(1.084)	(1.700)
Job contracts six to 12 months	2.766*	1.518	3.070*	2.338*	3.336*
	(0.850)	(1.089)	(0.942)	(0.929)	(0.805)
Job contracts > 12 months	−0.636	−1.637	−0.391	0.487	−2.135**
	(0.727)	(1.091)	(0.617)	(0.389)	(1.249)

Note: We include the same predictor variables as in the model specification presented in Table 4.

*Statistical significance at $\alpha = 5$ percent or less

**Statistical significance at 10 percent.

been strongest in the (smaller) DI groups. It may well be that the risks associated with having disabled clients with poor job prospects in smaller groups were exacerbated by the high-powered incentives. And as these risks were not (fully) compensated with higher risk premiums (or adjustments in benchmarks for client mix), WTW providers were apparently more likely to cream skim. Training and mediation costs were substantial for this group, whereas the risk of not receiving any (flexible) payments (due to failure to place) was substantial as well.

Overall, these findings on preprogram selection suggest that the effects of high-powered contract incentives on provider behavior were probably limited to worker groups for which the risks of poor outcomes were greatest and most evident. This is consistent with the fact that high fractions of clients sent back could damage the reputation and future contracting prospects of providers. Providers had little time for more in-depth assessments of clients' a priori chances of finding work in the few weeks between assignment of worker groups and program start.

Ultimately, however, the public may be more interested in knowing the implications of high-powered incentive contracts for program outcomes (or client success). Focusing on the results for client job placement rates in Tables 4 and 5, we find that fully performance-contingent contracting increased job placement rates, with the effect concentrated on short-term placements (with contracts from six to 12 months) and for larger groups. In particular, we find the overall effect of high-powered contracts is to raise the job placement rate by 2.13 percentage points (with little difference when we account for time trends with triple differencing or allow for a flexible baseline). The positive effect of fully performance-contingent contracting appears to be confined to short-term employment contracts only, suggesting that providers with high-powered incentives focused primarily on short-term employment contracts for these clients. The effect is also larger (and only statistically significant) for contracts serving worker groups of 50 or greater clients. For the DI worker groups, it also appears that job placement rates were significantly lower for those in long-term job contracts, suggesting the possibility that WTW providers may have parked some of these clients with lower prospects for workforce success.

Table 4 also shows that preprogram provider-induced selection decreased job placement rates. In the context of cream skimming, we expected this form of selection to increase the job placement rates of the remaining worker groups. We therefore performed two additional robustness checks. First, a possible explanation for the negative coefficient is that provider-induced selection is higher for low-quality WTW providers with poorly managed programs. We thus reestimated equation 3 with provider dummies added. The resulting coefficient for provider-induced selection subsequently turned positive, increasing to 2.218 (4.438), whereas the estimated impact of fully performance-contingent contracts remained almost unchanged at 2.831 (1.400). For the second robustness check, we reestimated equation 3 without preprogram provider- and client-induced selection as controls to test for possible endogeneity problems that might bias our estimated effects of fully performance-contingent contracts on job placement rates. Using this specification, the coefficient also remained unchanged, likewise confirming that high provider-induced selection in these worker groups is associated with low-quality providers.

Parking Effects

Thus far our empirical findings suggest that fully performance-contingent contracts increase preprogram selection for smaller worker groups with poorer job prospects (disabled clients). We speculate that they may also potentially increase parking by WTW providers during the programs as an alternative means of selection. If true, this might also account for our finding that high-powered incentive contracts appear to reduce job placement rates for smaller (DI) worker groups in job contracts greater than 12 months.

We expect parking to be more likely if the value added of serving hard-to-place clients is low or when the program costs of serving them are high. Parking behavior typically cannot be observed directly, and spending per client is also difficult to discern. We therefore use the data at hand to derive an indirect test for parking by focusing on its potential effect on job search durations of clients. More specifically, we characterize parking as providers concentrating their effort on clients with better job prospects, which in turn would cause the employment prospects for this group to improve, while lowering them for hard-to-place clients who receive less attention. Stated differently, the job search durations of easier-to-place clients would become shorter, while those of hard-to-place clients would be unchanged or possibly even lengthen.[4]

We implement a formal test for parking that compares the average job search duration for clients in full vs. partial performance-contingent contracts who have found a job. Conditioning on job placement rates, we expect job duration to be shorter for worker groups served under fully performance-contingent contracts. In the Appendix, we present the technical details of this test.[5]Specifically, as the average job search duration of job finders is a function of the job placement rate (which also may differ between worker groups under full vs. partial performance-contingent contracts), comparing the average durations of completed spells may give

[4] Our implicit assumption is that cream skimming operates through a priori chances of finding employment, rather than on program value-added. In the literature, this is referred to as the "common-effect" model, in which the value-added of programs is equal for all clients (Heckman et al., 1997). Heckman et al. justify this assumption by the fact that the variance in value-added is small compared to the variance in a priori job finding probabilities.

[5] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://ww3.interscience.wiley.com/cgi-bin/jhome/34787.

Table 6. GLS estimation results of (log) program duration (effects in percentage points; standard errors in parentheses).

	Diff.-in-diff.		Triple-diff.		Flexible baseline	
Fully performance-contingent contract	−3.400	(5.906)	−7.977	(7.736)	−0.932	(5.886)
Disability insurance (dummy)	8.573*	(3.504)	5.441	(4.078)	9.495*	(3.664)
(Expected) Reward per client	0.051	(1.053)	0.114	(0.534)	−0.131	(0.540)
Log gross worker group size	1.400**	(0.770)	1.674*	(0.776)	2.276*	(0.947)
Time trend	−2.674*	(1.472)	0.325	(2.821)		
Time trend for treatment group			−4.410	(3.456)		
Year = 2003					−6.903**	(3.905)
Year = 2004					−6.689*	(3.313)
Year = 2005					−8.141	(5.012)
F-test gross worker type dummies; *F*(18,104)	$F = 9.02$ $P = 0.000$		$F = 7.49$ $P = 0.000$		$F = 4.61$ $P = 0.000$	
Number of observations	5,965		5,965		5,965	
R-squared	0.119		0.120		0.120	

Note: The regressions also included six region dummies as controls in all specifications.

*Statistical significance at $\alpha = 5$ percent or less

**Statistical significance at 10 percent.

biased results. Using our model specified in equation (3), however, we can calculate this bias. We demonstrate this by first estimating the impact of fully performance-contingent contracting on completed program durations in a *naive* model, and then correcting the estimates for any biases.

We use a difference-in-differences (diff.-in-diff.) specification for the log average value of program spells that lead to work, as we did for the other outcome measures:

$$ln(DUR)_{ijk,t} = a^{DUR}100\%PC_{ij,t} + \boldsymbol{X}_{ijk,t}\beta^{DUR} + \gamma^{DUR}t + \sum_i \delta^{DUR}I(r = i) + \varepsilon^{DUR}_{ijk,t}, \tag{4}$$

with *DUR* as the average program duration of workers in worker groups that have found a job at the end of the program. Further, the notation of indices, the right-hand side variables, coefficient values, and estimation strategy is similar to that of equations (1) to (3). We thus assume ε^{DUR} to be i.i.d. with $(0, \sigma^2{}_{DUR})$.

Table 6 shows the estimation results and test statistics for program duration differentials between worker groups under full vs. partial performance-contingent contracts. In addition, Table 7 contains coefficient estimates of (log) program duration differentials for the samples of UI and DI workers and small and large worker groups, together with estimates of bias due to differences in job placement rates. As expected, we find that worker groups under high-powered contracts have shorter program duration (than those under partial performance-contingent contracts), but the effects are small and insignificant in most cases. Also, the effect estimates have the same order of magnitude as the bias estimates that result from differences in the job placement rates of worker groups under these two types of contracts. Thus, we cannot reject the null hypothesis of no parking. We also see in Table 6, as expected,

Table 7. GLS estimation results of program duration by worker group size and UI and DI worker groups (diff.-in-diff. effect in percentage points; standard errors in parentheses).

	Effect of fully Performance-contingent contracts on log program duration		Bias estimate	
	Coefficient estimate	Standard error	Coefficient estimate	Standard error*
Full sample	−3.398	(5.906)	−4.791*	(2.235)
< 50 clients	−8.699	(6.932)	−0.247	(3.738)
≥ 50 clients	−7.959	(7.761)	−6.375*	(2.196)
UI groups	2.667	(2.662)	−5.543*	(2.135)
DI groups	−19.11*	(3.053)	−2.936	(4.165)

Standard errors of the bias estimates are obtained by the Delta method.

*Statistical significance at $\alpha = 5$ percent or less.

that program durations are overall longer for the (harder-to-place) DI workers, although the coefficient is only statistically significant in the difference-in-differences model.

When we examine these relationships separately for UI and DI worker groups, we see that the estimated effect of a fully performance-contingent contract on DI program duration is statistically significant, suggesting that (log) program duration is 19 percentage points lower for DI worker groups under high-powered contracts (and cannot be fully explained away by differences in job placement rates). We suggest a plausible interpretation of these results is that providers were more likely to park the most difficult to serve in the DI groups, resulting in shorter program durations for those who were placed in jobs. Thus, similar to preprogram selection, it appears that parking is more prevalent among worker groups where program costs per client are greater. Parking may be one way to lower the (downward) financial risks in some contracts, particularly those for DI groups.

CONCLUSION

We have explored both the intended and unintended effects of performance-based contracting (with high-powered incentives) in social welfare services, using unique data on groups of workers that were procured by the Dutch social benefit administration—some of whom were under fully performance-contingent contracts. The results of our analysis are generally consistent with expectations of the standard theoretical framework of risk and incentives, where incentives are most likely to work when the (downward) financial risks of performance-contingent payments for providers are not too large to induce gaming (selection and parking) behavior (Burgess & Ratto, 2003; Skedinger & Widerstedt, 2007). It is possible that the Dutch benefit administration designed the contracts with this knowledge in hand, as they used high-powered incentive contracts primarily for worker groups in which clients had higher preprogram prospects for finding work and job placement risks per client were low. While our analysis shows that the overall (additional) preprogram selection and parking activities associated with the fully performance-contingent contracts were probably not extensive, they were clearly not inconsequential for some groups.

For the UI groups, we find that the fully performance-contingent contracts contributed to about a three percentage point increase in job placements, with no evidence of preprogram selection and parking. Although this effect is smaller than those reported in previous studies—with estimates ranging from five to 10 percentage points for these worker groups (Finn, 2008)—the high-powered contracts seem to have improved overall outcomes.

Alternatively, fully performance-contingent contracting was not effective in raising job placement rates in DI worker groups, with evidence pointing to parking—that is, the concentration of provider efforts on easier-to-place clients—and client-induced selection. Both forms of selection by providers are seemingly driven by risk and cost considerations, which are particularly relevant for harder-to-place DI clients. Although it is difficult to assess the exact magnitude of these effects, parking could well have harmed both the efficiency of the program—in terms of job placement rates—and equity in access to services.

The other concern that follows from our analysis is related to the duration of jobs that were facilitated through contracts with high-powered performance incentives. Additional job placements were limited primarily to shorter-term employment arrangements of no more than 12 months. With the data at hand, we cannot infer whether this was to the detriment of the efficiency of the WTW programs, particularly for UI clients. The program might be viewed as highly successful if short-term jobs were a stepping stone to longer-term employment, and if shorter program lengths decreased the risk of lock-in effects for clients. The literature on the role of temporary employment as a stepping stone to better (longer lasting) jobs reports mixed findings on this question (see Heinrich, Mueser, & Troske, 2009), and other evidence seems to suggest that short- and long-term effects of programs are not strongly correlated (Heckman, Heinrich, & Smith, 2002).

In general, the findings of this study suggest that in contexts where the risks to service providers of failing to meet contract expectations are greater, due to factors such as client characteristics that portend barriers to successful outcomes, high-powered, performance-based contracts are more likely to induce unintended effects such as parking. Alternatively, there may be some small gains to contractual arrangements that put a greater portion of provider compensation at risk when the uncontrollable risks to provider contract outcomes are lower, although these benefits of performance-based contracting appear to be limited to short-term outcomes. We do not find that higher-powered performance-based contracts increase job durations, suggesting that they may be less effective in increasing long-run impacts of these types of public programs in their current form. Designing performance-based contracts that simultaneously motivate improved short-term and long-run outcomes remains a challenge for policymakers as well as academics (Heckman et al., 2011).

Interestingly, we see the policy decisions made in the Netherlands after 2007, which reverted contracts for clients with relatively good job prospects to in-house service delivery, as potentially at odds with our study findings. Whereas high-powered, performance-based contracts for these groups might marginally improve outcomes, those for the more difficult to serve were more likely to generate unintended negative effects. We suggest that a more appropriate policy change might have been to reduce risk for providers (moderating the level of performance contingency in contracts or adjusting performance expectations for client mix) or to revert to in-house service delivery for the higher risk groups to reduce cream skimming and parking.

That said, the social welfare implications of parking will ultimately rest on public values that are embodied in a program mission or goals, such as preferences for ensuring equity in access to publicly funded services vs. for increasing efficiency (i.e., maximizing the returns to taxpayer investments). For clients in the UI program, job placement outcomes (in short-term contracts) improved under high-powered incentive contracts, suggesting that the use of fully performance-contingent contracts could be welfare enhancing. Moreover, if the clients who are parked in programs are least likely to benefit from services, then these practices could likewise be welfare improving from a societal perspective that emphasizes efficient use of resources. The differences in administrative costs associated with using full vs. partial performance-contingent contracts appear to be negligible. Still, if societal preferences place a high value on equality in access to program resources and services (as in public education systems), or if harder-to-serve clients might be more likely to realize benefits from WTW services in the long run, then high-powered incentive contracts (and the parking they induce) could be welfare reducing. The limited empirical evidence we generate sheds some light on these questions, but knowledge of the relative weights attached to these (or other) public values are also necessary to guide policy decisionmaking.

PIERRE KONING is Professor of Labor Market and Social Security, Department of Economics, VU University Amsterdam, De Boelelaan 1105, 1081-HV Amsterdam, The Netherlands.

CAROLYN J. HEINRICH is Sid Richardson Professor of Public Affairs and affiliated Professor of Economics, The University of Texas at Austin, P.O. Box Y, Austin, TX 78713-8925.

REFERENCES

Barnow, B. A. (2000). Exploring the relationship between performance management and program impact: A case study of the JTPA. Journal of Policy Analysis and Management, 19, 118–141.

Barnow, B. A., & Heinrich, C. J. (2010). One standard fits all? The pros and cons of performance standard adjustments. Public Administration Review, 70, 60–71.

Bernard, S., & Wolff, J. (2008). Contracting out placement services in Germany. Is assignment to private providers effective for needy job-seekers? IAB Discussion Paper 5–2008. Nuremberg, Germany: IAB.

Blank, R. M. (2000). When can public policy makers rely on private markets? The effective provision of social services. Economic Journal, 110, C34–C49.

Brown, T. L., & Potoski, M. (2003). Managing contract performance: A transaction costs approach. Journal of Policy Analysis and Management, 22, 275–297.

Bruttel, O. (2005). Contracting-out and governance mechanisms in the public employment service. Wissenschaftszentrum Berlin für Sozialforschung (WZB), Discussion paper 2005–109. Berlin, Germany: WZB.

Burgess, S., & Ratto M. (2003). The role of incentives in the public sector: Issues and evidence. Oxford Review of Economic Policy, 19, 285–300.

Burgess, S., Propper, C., Ratto, M., & Tominey, E. (2004). Incentives in the public sector: Evidence from a government agency. CMPO Working Paper Series No. 04–103. Bristol, UK: CMPO.

Chiappori, P., & Salanié, B. (2003). Testing contract theory: A survey of some recent work. In M. Dewatripont, L. Hansen, & P. Turnovsky (Eds.), Advances in economics and econometrics—Theory and applications. Eight World Congress of the Econometric Society, Econometric Society monographs (pp. 115–149). Cambridge, UK: Cambridge University Press.

Courty, P., & Marschke, G. (1997). Measuring government performance: Lessons from a federal bureaucracy. American Economic Review Papers and Proceedings, 87, 383–388.

Courty, P., & Marschke, G. (2004). An empirical investigation of gaming responses to performance incentives. Journal of Labor Economics, 22, 23–56.

Courty, P., Kim, D., & Marschke, G. (2011). Curbing cream-skimming: Evidence on enrollment incentives. Labor Economics, 18, 643–655.

Dixit, A. (2002). Incentives and organisations in the public sector: An interpretative review. Journal of Human Resources, 37, 696–727.

Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? Journal of Public Economics, 90, 239–255.

Finn, D. (2008). The British "welfare market." Lessons from contracting out welfare to work programmes in Australia and the Netherlands. Portsmouth, UK: University of Portsmouth. Retrieved from http://www.jrf.org.uk/sites/files/jrf/2306-welfare-unemployment-services.pdf.

Groot, I., Hollanders, D., Hop, J., & Onderstal, S. (2006). Werkt de re-integratiemarkt? SEO rapport 946. Amsterdam, The Netherlands: SEO.

Heckman, J., & Smith, J. (2004). The determinants of participation in a social program: Evidence from a prototypical job training program. Journal of Labor Economics, 22, 243–298.

Heckman, J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. Review of Economic Studies, 64, 487–535.

Heckman, J., Heinrich, C. J., & Smith, J. (2002). The performance of performance standards. Journal of Human Resources, 38, 778–811.

Heckman, J. J., Heinrich, C. J., Courty, P., Marschke, G., & Smith, J. (2011). The performance of performance standards. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Hefetz, A., & Warner, M. (2004). Privatization and its reverse: Explaining the dynamics of the government contracting process. Journal of Public Administration Research and Theory, 14, 171–190.

Heinrich, C. J. (1999). Do government bureaucrats make effective use of performance management information? Journal of Public Administration Research and Theory, 9, 363–393.

Heinrich, C. J. (2007). False or fitting recognition? The use of high performance bonuses in motivating organizational achievements. Journal of Policy Analysis and Management, 26, 281–304.

Heinrich, C. J., & Choi, Y. (2007). Performance-based contracting in social welfare programs. The American Review of Public Administration, 37, 409–435.

Heinrich, C. J., & Marschke, G. M. (2010). Incentives and their dynamics in public sector performance management systems. Journal of Policy Analysis and Management, 29, 183–208.

Heinrich, C. J., Mueser, P. R., & Troske, K. R. (2009). The role of temporary help employment in low-wage worker advancement. In D. Autor (Ed.), Labor market intermediation (pp. 399–436). Chicago, IL: The University of Chicago Press.

Koning, P. (2008). Not-for-profit provision of job training and mediation services: An empirical analysis using contract data. De Economist, 156, 221–239.

Koning, P. (2012). Contracting welfare-to-work services: Use and usefulness. Economics Letters, 114, 349–352.

Martin, L. L. (2005). Performance-based contracting for human services: Does it work? Administration in Social Work, 29, 63–77.

McBeath, B., & Meezan, W. (2010). Governance in motion: Service provision and child welfare outcomes in a performance-based, managed care contracting environment. Journal of Public Administration Research and Theory, 20, i101–i123.

Neal, D., & Whitmore Schanzenbach, D. (2010). Left behind by design: Proficiency counts and test-based accountability. Review of Economics and Statistics, 92, 263–283.

Orbach, B. Y. (2006). Unwelcome benefits: Why welfare beneficiaries reject government aid. Minneapolis, MN: University of Minnesota Law School.

Prendergast, C. (1999). The provision of incentives in firms. Journal of Economic Literature, 37, 7–63.

Skedinger, P., & Widerstedt, P. (2007). Cream skimming in employment programs for the disabled? Evidence from Sweden. International Journal of Manpower, 28, 694–714.

Struyven, L., & Streurs G. (2005). Design and redesign of a quasi-market for the reintegration of jobseekers: Empirical evidence from Australia and the Netherlands. Journal of European Social Policy, 15, 211–229.

Tergeist, P., & Grubb, D. (2006). Activation strategies and the performance of employment services in Germany, the Netherlands and the United Kingdom. OECD Social Employment and Migration Working Papers 42. Paris, France: OECD Publishing.

Winterhager, H., Heinze, A., & Spermann, A. (2006). Deregulating job placement in Europe: A microeconometric evaluation of an innovative voucher scheme in Germany. Labour Economics, 13, 505–517.

APPENDIX: TESTING PARKING BEHAVIOR

Although parking behavior typically cannot be observed directly, the data at hand allow us to derive an indirect test for parking by focusing on its effect on job search durations of clients. The idea is that providers concentrate their effort on clients with better job prospects, causing the prospects for this group to improve, whereas for the hard-to-place, they remain constant or even decrease. Stated differently, parking shortens the job search durations of easy-to-place clients, while keeping constant or even lengthening those of hard-to-place clients. Thus, if one was comparing two providers with equal job placement rates, we would expect the one with more parking to place clients into jobs more quickly (i.e., they have shorter durations in the program). Eventually, the other provider will catch up (in terms of its placement rate), as it more gradually finds jobs for the hard-to-place clients.

Thus, an obvious test statistic for parking would entail a comparison of the average job search duration for clients in worker groups under partial vs. full performance-contingent contracts that have found a job. In this parking test, we expect job search durations to be shorter for worker groups under fully performance-contingent contracts. To formalize and develop such a test, we first introduce some notation:

- τ = program length, which is the same for all worker groups.
- t = time, with $0 \leq t \leq \tau$.
- $P(t)$ is the job placement rate of a worker group at time t. We only observe P at time $t = \tau$.
- $\theta(t)$ is the job finding hazard in a worker group at time t. Thus, $P(t) = 1 - \exp[-\int_0^t \theta(s)ds]$.
- $E(t \mid t \leq \tau)$ is the expected job search duration of clients that have found a job before τ.

We start by writing down the definition of the expected truncated job search duration:

$$E(t|t = \tau) = \frac{\int_0^\tau [1 - P(t)]dt - (1 - P(\tau))\tau}{P(\tau)}. \tag{A.1}$$

As the right-hand side of equation (A.1) is a function of $P(\tau)$, a simple comparison between the truncated average durations of worker groups under partial vs. full performance-contingent contracts may give biased results. In particular, if one of the worker groups has a higher (average) placement rate, this would result in a lower value of the truncated average duration. To test for parking, we thus have to control for any differences that are due to differences in $P(\tau)$. We do so by performing a first-order expansion log value of E $(t|t \leq \tau)$ with respect to $P(\tau)$:

$$\frac{\partial \ln E(t|t{=}\tau)}{\partial P(\tau)} = \frac{\tau}{E(t|t \leq \tau)}\left(1+\frac{\partial \gamma}{\partial P(\tau)}\right) - \frac{1}{P(\tau)}. \tag{A.2}$$

with

$$\gamma = \frac{\int_{0^\tau} [1 - P(t)]dt}{\tau}. \tag{A.3}$$

In our sample, we observe the variable values of τ, $E(t|t \leq \tau)$, and $P(\tau)$, but not γ. Still, the derivative of γ with respect to $P(\tau)$ can be obtained by first rearranging equation (A.1):

$$\gamma = \frac{P(\tau)\mathrm{E}(t|t=\tau) + (1 - P(\tau))\tau}{\tau} \tag{A.4}$$

and then estimating the observed values of γ as a polynomial function of $P(\tau)$. In particular, we can use three polynomials with the data at hand. From the resulting coefficient estimates, we can calculate the derivative value of γ with respect to the observed values of $P(\tau)$. This completes the information we need to calculate the left-hand side of equation (A.2).

Finally, to calculate the bias that is due to difference in job placement rates by contract type, we multiply the derivative in equation (A.2) with the estimated difference, α^{JPR}, that is obtained from equation (3):

$$\frac{\partial \ln \mathrm{E}(t|t{=}\tau)}{\partial P(\tau)} \cdot a^{\mathrm{JPR}}. \tag{A.5}$$

In sum, our estimation strategy to detect parking effects entails the comparison of worker groups under partial vs. full performance-contingent contracts, while correcting for any biases due to differences in job placement rates. To obtain the bias estimates, we first need to calculate the values of γ, then estimate γ as a polynomial function of $P(\tau)$, and finally calculate equation (A.5).