# Cream-skimming, Parking and other Intended and Unintended Effects of Performance-Based Contracting in Social Welfare Services

Pierre Koning

CPB NETHERLANDS BUREAU OF ECONOMIC POLICY ANALYSIS
P.O. BOX 80510
2508 GM, THE HAGUE
THE NETHERLANDS

Carolyn J. Heinrich

UNIVERSITY OF WISCONSIN-MADISON
LA FOLLETTE SCHOOL OF PUBLIC AFFAIRS
1225 OBSERVATORY DRIVE
MADISON, WI 53706

March 2010

We thank the Dutch Social Benefit Administration (UWV), in particular Peter Rijnsburger and Jaap Spitz, for gaining access the contracted cohort data that were used. Wim Zwinkels and Margaretha Buurman are gratefully acknowledged for gaining access to more detailed information on the contracts that were procured by UWV. Alp Esener provided us with excellent research assistance in the earlier stages of this research project. Finally, participants at the La Follette School of Public Affairs Seminar series and the CEPR workshop on "The Role of Incentives, Information and the Private Sector in the Delivery of Public Services" in Brussels on 10-11 December 2009 are thanked for useful comments and suggestions.

**Abstract**

In a growing number of countries, the delivery of social welfare services is contracted out to private providers, and increasingly, using performance-based contracts. Critics of performance-based incentive contracts stress their potential unintended effects, including cream-skimming and other gaming activities intended to raise *measured* performance outcomes. We analyze the incentive effects of performance-based contracts, as well as their impacts on provider job placement rates, using unique data on Dutch cohorts of unemployed and disabled workers that were assigned to private social welfare providers in 2002-2005. We take advantage of variation in contract design over this period, where procured contracts gradually moved from partial performance-contingent pay to contracts with 100%-performance contingent reward schemes, and analyze the impact of these changes using panel data that allow us to control for cohort types and to develop explicit measures of selection into the programs. We find evidence of cream-skimming and other gaming activities on the part of providers but little impact of these activities on job placement rates. Overall, moving to a system with contract payments fully contingent on performance appears to increase job placements for more readily employable workers, although it does not affect the duration of their jobs.

*Keywords:* social welfare, performance contracting

*JEL-codes:* I38, H11, H53.

## 1. INTRODUCTION

The majority of empirical studies on the use and effectiveness of performance incentives address performance-related pay in private firms, reflecting their greater prevalence in the private sector relative to the public sector (see Prendergast (1999) and Chiappori and Salanié (2003) for survey studies). Compared to private sector production technologies, public sector work is more likely to involve complex, non-manual work, multiple principals and group dynamics, political and environmental influences and interdependencies, and non-standardized outputs that make precise measurement of performance challenging and costly (Blank 2000; Dixit 2002; Heinrich and Marschke 2010). In light of the larger role of random components in performance outcome measurement in the public sector, incentive structures are more likely to be low-powered (Burgess and Ratto 2003; Heinrich 2007). For this reason, our research is relatively unique in investigating a special case in which the delivery of publicly-funded social services moved to a system with very high-powered incentives, that is, with payments to contractors 100% contingent on their performance. We are particularly interested in evaluating not only the effects of these incentives on client outcomes (job placements), but also the unintended effects of 100%-contingent performance-based contracting ('No Cure No Pay') on the delivery of social welfare services. In other words, this public sector setting for the study in the Netherlands provides us with a unique case where both the financial incentives and the risks of social welfare provision were fully transferred to the social welfare providers.

Although 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 phenomenon, and one that is becoming increasingly common. Some of the first experiences with performance-based incentive schemes in the public sector date from U.S. employment and training programs, beginning with the Job Training Partnership Act (JTPA) of 1982 and continuing with the Workforce Investment Act (WIA), which replaced JTPA in 2000 (Barnow, 2000; Heckman et al. 2002). More recently, governments in Australia and the Netherlands switched to fully privatized systems with substantial performance incentives (Bruttel 2004; Struyven and Steurs 2005; Finn 2008). Contracting out public social welfare services to private providers is less widespread in the United Kingdom and Germany but is gradually becoming more important in these countries as well (Bernard and Wolff 2008; Tergeist and Grubb 2006; Winterhager et al. 2006).

As the literature on contracting social welfare services has grown, concerns about the unintended effects of performance-based contracting have risen as well. There is some consensus that these contracts have increased *measured* performance outcomes, but they are frequently accompanied by various distortionary effects, such as cream-skimming and other gaming activities. In some settings, this has led to a retraction in the use of performance-based contract incentives, with some governments reverting to fixed-payment schemes for private providers or even contracting back in (Hefetz and Warner 2004). Examples here include the reappraisal of in-house provision of welfare-to-work services to Social Assistance recipients in the Netherlands and changes in the contract design and management of the Wisconsin Works program (Koning 2009; Heinrich and Choi 2007). Clearly, these reappraisals and retractions 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. For example, providers that cream-skim workers with high a priori job placement rates may well be present, but if these groups of workers are also more likely to benefit from programs (Heckman et al. 2002), the efficiency costs of cream-skimming may be low.

We aim to help fill this gap in the literature by undertaking an integrative analysis of the distortive effects of performance-based contracting with private social welfare service providers, as well as their impact on the overall effectiveness of performance schemes. For this purpose, we use unique data on Dutch cohorts of unemployed and disabled workers that were assigned to private welfare-to-work (WTW) providers in performance-based contracts in 2002-2005. Over this period, procured contracts gradually moved from No Cure Less Pay (NCLP) to No Cure No Pay (NCNP) reward schemes, the latter of which constituted a fully (100%) performance-contingent contract, with payments made only for clients placed in jobs. We analyze the impact of this change on job placement rates as an overall measure of program effectiveness and investigate the importance of possible distortive effects of the incentives, using data from multiple contract periods with controls for cohort (worker) types. To account for selection in the process of allocating cohorts of workers and NCNP contracts, we estimate two-stage treatment selection models and fixed-effect models that control for cohort characteristics affecting both the payment scheme and program outcomes.

This study makes several important contributions to the literature. First, we observe the numbers of workers that were assigned to WTW providers but did actually not start programs, allowing us to develop explicit measures of pre-program selection that may constitute a form of cream-skimming. We test whether the additional incentives associated with NCNP (fully performance-contingent) contracting have changed the shares of clients that do not start programs, distinguishing between clients that do not show up at the program (i.e. client-induced selection) and clients that are being returned to the social benefit administration by providers (i.e.

provider-induced selection). We also explore (empirically) whether client- and/or provider-induced selection is related to the job placement rates of contractors.

Second, to the best of our knowledge, this paper is one of the first to explicitly address in an empirical analysis the issue of the "parking" of hard-to-place WTW clients. Within the context of our analysis, parking is defined as not serving hard-to-place clients who have been assigned to WTW programs. In other words, parking might be characterized as a form of cream-skimming *during* the program, in which providers try to maximize placement rates and keep costs down by focusing resources on the most able clients while doing little to serve those with the poorest job prospects. As this activity increases the job placement rates of clients with better a priori job prospects, we expect this to reduce the observed average job search duration of the subsample of job finders, while also widening the gap in job placement probabilities between clients with better and worse job prospects. We therefore develop a test statistic for parking that exploits observed average job search durations of job finders in NCLP and NCNP cohorts.

Finally, our analysis addresses the importance of other gaming activities that increase the emphasis on short-term job placements for performance outcomes. We investigate whether the move to NCNP contracting affects the duration of job placements among job finders, where jobs lasting for at least one year are considered to be more durable. In particular, we analyze the effectiveness of NCNP incentives in raising total job placement rates, with the expectation that they will be more successful in increasing short-term job placements, or even lead to substitution of short-term for long-term placements.

The results of our analysis are consistent with theoretical predictions that incentives are most likely to work when the financial risks of performance-contingent payments for providers are not too large to induce selection and parking behavior. Our analysis also shows that the overall pre-program selection and parking activities were probably minimal. NCNP contracting appeared to increase job placements for some cohorts, but it did not affect job duration and was not effective in raising job placement rates for those in disability insurance cohorts. Taking a broader perspective, our results are consistent with the classic economic trade-off between risk and incentives. That is, for larger cohorts where the risk of failing to place clients in jobs is spread over a broader client base (i.e., the per-client risk is smaller), we find less of an effect of NCNP incentives on selection and parking behavior. Conversely, for smaller contracted cohorts with greater risks of non-payment due to performance, the evidence points to more selection and parking.

This paper proceeds as follows. Section 2 briefly reviews the literature on cream-skimming, parking and other gaming activities, with special interest in the role of institutions in determining or incentivizing these phenomena. Section 3 describes the cohort and contract data and methods we use, while explaining the institutional context in which they originate. Section 4 presents the empirical analysis and findings, and section 5 concludes.

## 2. 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, however, 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), and 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, or the selection of easier-to-place clients by providers, are expected to increase if payments or rewards are performance-based, as workers with bad a priori job prospects will increase the risk of no (or lower) payments (Heckman et al. 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. In their study of cream-skimming in JTPA programs, Heckman and Smith (2004) analyzed multiple stages of the JTPA selection process, including eligibility, awareness, application, acceptance and formal enrollment. They concluded that applicants' progression through these stages toward enrollment was less influenced by cream-skimming on the part of program workers than by factors beyond the control of service providers (Heckman and 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-place" groups (Heinrich et al. 1996; Barnow and Heinrich, forthcoming; Courty et al. 2009). Courty et al. (2009) 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. Thus, incentives to cream-skim on observed characteristics that are accounted for in the adjustments are likely to be reduced, although cream-skimming may still occur on unobservable characteristics within the hard-to-place groups.

In other settings, admission into programs may be compulsory, allowing little room for selection by providers. Examples of such programs include those in Australia, the UK and the Netherlands, where participation in welfare-to-work (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 UK, the assignment of worker cohorts is fully compulsory for WTW providers as well, 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 2007).

The picture that emerges from these studies is that cream-skimming is probably less important and prevalent than suspected or stressed by policymakers, particularly in programs where selection occurs at earlier stages of awareness and application (with little leeway for selection by providers). Still, concerns have arisen about the possibility of other gaming activities that occur in the implementation of programs, such as parking in WTW-programs (Bruttel 2004; Finn 2008). For example, one view is that the parking of assigned clients (after the start of a program)—which typically results in a bare minimum of services for harder-to-place clients—may represent a substitute for cream-skimming, particularly in contexts where pre-program selection is effectively prohibited or restricted. Moreover, WTW providers can spend time in assessing the a priori job prospects of workers and use this information to determine how they will allocate "parking spaces", rather than using it for selecting clients prior to the program. The experiences of some providers, clients and caseworkers, as well as anecdotal evidence, suggests an association between incentive-based contracts and parking activities. Still, empirical evidence on the importance of parking is almost nonexistent, likely due to difficulties in observing provider efforts and budget allocations.

Among the early empirical analyses of gaming of measured program outcomes in social 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. In more recent work, Courty and Marschke (2008) present a broader framework for detecting increases in distortions in performance measures associated with contract incentives, showing that the alignment between a performance measure and the true goal of the organization may decrease after a performance measure is introduced, as bureaucrats respond by exploring all strategies for raising it, not just those that also increase program value.

Heckman et al. (2002) also show that short-term outcomes of programs may bear very little relationship to long-term outcomes. Thus, performance standards relying on readily available data on outcomes, such as job placement rates and wages at placement, may not necessarily advance the longer-term goals of public programs (such as employment retention or longer-run earnings gains). In many of these cases, the incompleteness of performance measures is inherent in social services provision, where longer-term outcomes are more difficult to measure, rather than an artifact of institutional settings (see also Dias and Maynard-Moody 2006).

Similar to cream-skimming, it seems clear that the incidence and type of gaming activities are strongly determined by institutional settings. In the JTPA program, where the vast majority of studies have been undertaken to date, 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. With linear contracts, however, like the NCLP and NCNP systems, 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. Still, even with linear incentives and without discretion in reporting successful programs, providers may use their discretion in determining the end date of the program, that is, the reporting of "unsuccessful programs". In particular, extending programs may result in additional windfall gains of job placements on the one hand (a gaming behavior), while creating additional program costs on the other. The impact of optimizing the duration of program lengths (for unsuccessful clients) on job placement rates is not clear. And taking an even broader perspective, it remains unclear what the overall impact of the unintended effects of performance-based contracting in social welfare services may be relative to its intended effects.[1]

[1] Thus far, the only notable exception appears to be the study by Burgess et al. (2004), who investigated the effectiveness of performance-related pay for public employment offices. They find the effectiveness of incentive effects to be confined to small offices, but do not address the issues of parking and gaming.

## 3. DATA AND INSTITUTIONAL CONTEXT

In the Netherlands (the setting of this study), the Disability Insurance (DI) and Unemployment Insurance (UI) programs are mandatory for all workers. 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 partially disabled clients mediation, job training or subsidized employment within twelve months after the start of their benefit spell.[2] In the period under investigation (2002-2005), the Dutch social benefit administration[3] contracted out the delivery of these WTW-services to private job training service providers only. As of 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 and were free to choose (preferred) providers or choose to be assigned to cohorts—like in the system that was compulsory until 2005. In 2007 the privatization of social welfare services was reversed to some extent—that is, the social benefit administration resumed the delivery of social welfare services for their clients with relatively good job prospects. Thus, both of these changes to the system caused the number of procured cohorts (and subsequently the number of clients) with NCNP contracts to decrease substantially from 2006.

**< TABLES 1 AND 2 HERE >**

Table 1 provides an overview of the characteristics of cohorts that were contracted out to WTW-providers in 2002-2005 and were registered by the social benefit administration, stratified with respect to calendar years and contract payment schemes (NCNP and NCLP). Table 2

[2] As of 2005, however, the New Deal approach has been relaxed in the Netherlands. This is mirrored by a decrease in the number of WTW-clients in 2005.

presents information on the structure, size and characterization of the difference cohort types and cohorts. As is apparent in this table, each year the Dutch social benefit administration defined cohort types that were assigned to private providers in separate groups and with separate contract conditions, so as to allow for specialized services and decrease opportunities to cream-skim within cohorts.

In the data, cohorts can be specified to three digit levels. The first digit level indicates a broad characterization of cohort types that holds for all years in our data and includes 18 categories. These "gross" cohort types are observed in almost all years of our sample (see the second column of Table 2). Gross cohort types are characterized by specific client groups in the UI and DI scheme, but some smaller groups are defined by other strata. These client groups include immigrants, those with good or bad job prospects (after being profiled by social benefit administration caseworkers), disabled workers with special impairments, and older workers or young disabled workers without any work history.

The second digit level entails a more detailed description of the contract types and activities that have been contracted out in subsequent years. In particular, the designation of program types for the gross cohort types has varied over time, depending on the specific needs pertinent at that time. Examples include programs aimed at guiding workers into self-employment, job training, job hunting and job mediation. The resulting cohort types that started a program were defined distinctly for each year. Moreover, contract conditions—particularly those regarding the payment schemes, NCLP and NCNP—also varied according to these contract types. Thus, it is important to stress that contract conditions did not result from negotiations between the social benefit administration and providers. Rather, these were set

[3] In Dutch: Uivoeringsinstituut Werknemersverzekeringen (UWV).

prior to the bidding process, with an increasing share of NCNP contract types over time.[4] In total, there were 106 cohort types that were contracted out in 2002-2005, with most the observations in 2002—the year in which the most clients were contracted out as well. As WTW-providers complained about the administrative burden that came with the bidding process involving so many cohort types, the number of cohort and contract types was reduced substantially in 2003.

Finally, the third digit level indicates the actual cohort that was assigned to a particular provider, in a specific region and at a specific time point.[5] In effect, the contract conditions for these cohorts were defined by contract types, with within-type-variation in the actual cohort outcomes potentially stemming from cohort size, region and the provider. As Table 2 shows, many cohorts (particularly the larger ones) consisted of (UI) workers with better job prospects. In addition, we observe a substantial increase in the average cohort size in 2003 (from 28 to 52)—resulting from the aforementioned reduction in cohort types. In 2004, however, the average cohort size returned to its 2002 level. We also see a general decline in the number of WTW-clients in the period under investigation. In particular, the numbers of newly assigned program participants dropped from about 105,000 clients in 2002 to about 67,000 in 2003, and from about 48,000 in 2004 to 23,000 in 2005. We suggest two possible explanations for this pattern: business cycle effects and a relaxation of the New Deal strategy from 2005 onward, where the social benefit administration no longer aimed to fully treat clients within twelve months after the start of their UI or DI spell (Groot et al. 2006; Finn 2008). It also should be noted that the new procurement system in 2002 also applied to cohorts that entered in earlier

[4] Koning (2007) discusses the bidding process for cohorts in more detail.
[5] Koning (2007) studies the performance of for-profit and non-profit WTW-providers, taking cohorts per provider per region as a relevant unit of investigation and aggregating cohorts that started at different dates within a year.

years, when the New Deal strategy had not taken off yet, implying a backlog of clients may have been waiting for services in 2002.

***NCLP and NCNP contracts***

During the contracting process, individual clients within a particular cohort were not known yet by the WTW provider. Providers were only informed about the average characteristics of the cohort types, the expected cohort size, and contract conditions. As Table 1 shows, the share of NCNP contracting for contract types gradually increased, from close to zero in 2002 to more than half of the contracts in 2005 (weighted by the gross cohort size). Table 2 shows that NCNP contracts were confined to four gross cohort types, mostly consisting of (UI) workers with better job prospects. This largely explains why the average (expected) payment of NCNP contracts was lower than NCLP contracts in the period—we return to this issue in the next section. On average, about 50 percent of the NCLP contracts were paid as a fixed amount, with the other 50 percent paid at placement.

***Provider and client-induced selection***

In awarding contracts to particular providers, client addresses were transferred to the provider, who in turn had to contact the clients and make a reintegration plan together with them. 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. In order for the program to get started, the reintegration plan had to be approved by the social benefit administration. In practice, reintegration plans were not formulated for all individual clients and/or were not approved in all cases. Thus, we distinguish "gross" cohorts that were assigned to a provider from "net" cohorts 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 1, 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% of the gross cohorts. Although there was no formal limit on the percentage of 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 the programs but were not returned by the service provider either. This group most likely consisted of clients that already had found a job by the time the trajectory started or for whom the 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% in 2002 to 16% in 2005. In addition, it was about 4 percentage points higher for NCNP contracts than NCLP contracts, suggesting that clients assigned to NCNP contracts may have been more likely to find jobs in the time span between assignment and the actual start of the program.

***Parking, gaming and job placement***

Over the years, we have observed a gradual increase in the job placement rates of new cohorts that join the WTW programs. This holds for both job placements with employer-employee contracts of 6 to 12 months and contracts for more than 12 months. The definitive placement rates per cohort were measured at the end of the program for all (net) assigned clients. Programs typically lasted one to two years; for clients that did not succeed in finding a job, outcomes were measured at the end of program participation. Table 1 also shows that the share of NCNP contracts—with higher placement rates—has increased over time. We find more marked differences between payment forms and the average duration of both "successful" job finders and "unsuccessful" clients (who haven't found employment within the time span of the program).

We also explore whether the shorter job search durations of clients among NCNP contracts corresponds to the presence of parking activities in these programs—that is, whether providers concentrate on shortening job search durations among clients with high a priori job chances, while letting hard-to-place clients languish in the programs. Heretofore, figures 1 and 2 depict the distributions of (average) job search durations for successful and unsuccessful NCLP and NCNP cohorts, respectively. These figures suggest that the average job search durations of both job finders and non-job finders were shorter and less dispersed for the NCNP contracts. More specifically, it appears that most WTW programs for unsuccessful participants end after two years for NCNP cohorts, whereas for NCLP cohorts, they may participate for as long as three years. Still, a more formal analysis is needed to distinguish the role of payment schemes in these patterns of effects, as job search durations are undoubtedly driven by participant characteristics as well.

**< FIGURES 1 AND 2 HERE >**

# 4. EMPIRICAL ANALYSIS

***General research approach***

Above we defined and described different stages in the contracting and provision of WTW services in the Netherlands. In this section, we analyze how the shift to NCNP contracting has influenced the responses (intended and unintended) of providers in implementing these programs, and how this in turn has affected overall program effectiveness.

We first analyze the decision process of the social benefit administration in choosing between NCLP and NCNP payment schemes. Second, we address pre-program selection responses—both provider and client-induced—to the new performance-based incentives. Supposing that pre-program cream-skimming would increase under NCNP incentives, we expect provider-induced selection—and to a lesser extent client-induced selection—to increase. Third, we study the effect of NCNP contracting on parking behavior for clients that actually start their programs and develop a test that uses information on the average successful job search duration per cohort in our sample. Finally, we estimate the (overall) effect of payment schemes on performance outcomes—both for shorter employer-employee contracts and longer ones—focusing special attention on the role of pre-program selection. In particular, if NCNP contracting has increased such selection, this may have affected the placement rates in an indirect way as well.

As Chiappori and Salanié (2003) point out in their survey study, a unifying theme in empirical studies on the effects of contracting is the necessity to control adequately for unobserved heterogeneity and selection bias. Clearly, this holds for each stage in the estimation procedure. For example, the allocation of contract types may be also driven by the (a priori) expected job prospects of the cohort types. We therefore use two research strategies to control for possible selection effects. The first strategy takes explicit account of the allocation process of NCNP contracts, by estimating two-stage treatment selection models. In this approach, we use data from 2002 (a baseline year) as exclusion restrictions, assuming, for instance, that placement risks of contracts affect the choice of payment schemes but not job placement outcomes themselves in a direct way. Thus, we do not rely only on the distributional (normality) assumption of the selection model itself, but rather on exclusion restrictions to estimate the effects of performance-based incentives. The second research approach entails the use of gross cohort type fixed effects (FE) to control for cohort characteristics affecting both the payment scheme and program outcomes. Using (net) cohort types as fixed effects here would yield more precise controls, but then the effects of payment schemes—which vary between contract types only—would no longer be identified.

In both of these research strategies, one should be aware that the size of observed cohorts in our data varies substantially over time, affecting the variance of (pre-) program outcomes as well. Individual clients have a higher probability of being part of large cohorts than smaller ones, so equal weights for all observations may yield inefficient coefficient estimates. We therefore estimate model specifications with cohort size as relative (analytic) weights as well (see also Koning 2007). Within the context of (conventional) fixed effects (FE) estimation, however, rescaling is cumbersome, since observed weights are not constant within cohort contracts. We therefore follow Mundlak (1978), who proposes a specification for fixed effects with pooled data—that is, we re-specify the cohort-specific effect as a function of averages per gross cohort types. Using this approach on pooled data yields estimates that are equivalent to those with FE estimation, at least for those variables with variation within gross cohort types. In other words, in contrast to the conventional fixed effects specification, the pooled model allows for the use of cohort size as relative ('analytic') weights.

***Assignment of NCNP contracts and provider responses***

The use of performance-based pay is typically modeled as a classic trade-off between risk and incentives, a perspective that may also well apply to the choice of WTW contracts and the associated risk stemming from the uncertainty of cohort job placement rates (Burgess and Ratto 2003). Adopting a standard concave utility function for WTW providers, where providers are more risk averse than the social benefit administration, it can easily be shown that the impact of this risk will be higher for lower (expected) job placement rates. As risk is transferred to the provider in the form of performance-contingent pay, a higher risk premium will be required for cohorts with lower job placement probabilities, and the use of NCNP contracts will be less attractive. We also expect risk premiums per client to be lower if cohorts are larger, implying a lower variance of average placement rates. Analogously, it could be argued that providers with high market shares will require lower risk premiums. Within the context of the Dutch WTW-system, however, it should be noted that payment schemes were fixed in advance, so there was no negotiating on risk premiums. Thus, if large providers are overrepresented in NCNP cohorts, this could be due to the fact that these providers were more likely to bid for and/or be awarded these cohorts, with contract conditions being exogenous.

In testing these hypotheses, we first estimate a probit model that explains the use of the NCNP payment scheme (*NCNP*) among cohorts for 2003 to 2005, as a function of cohort and (assigned) provider characteristics. In particular, we specify the use of *NCNP* as:

$$(1) \qquad \Pr\,(\,NCNP_{ijk,\,t} = 1\,) = \Phi\,(\,\boldsymbol{X}_{i.,\,t=2002}\,b_{NCNP} + \boldsymbol{Y}_{ij,\,t}\,d_{NCNP}\,)$$

with $i$ as an index for the gross cohort type ( $i = 1,..,I$ ) , $j$ as an index for the cohort type ( $j = 1,..,J$ ), $k$ indicating the actual cohort ( $k = 1,..K$ ) and $t$ denoting years ( $t = 2002,..,2005$). $\Phi$ is the cumulative standard normal distribution function. Further, $\boldsymbol{X}$ represents average values of the job placement rates and average risk proxies per gross cohort type in 2002, which we are using as a baseline year (that is, where NCNP contracts were still almost absent and the social benefit administration decided on the assignment of NCNP contracts for the next year). We argue that these baseline observations are exogenous with respect to later program outcomes, such as the job placement rates. Alternatively, $\boldsymbol{Y}$ includes explanatory variables that vary over time and per cohort type $j$. These are the market share of the assigned provider of the actual cohort observation, the (expected) reward per client, and an indicator variable for cohorts with disability insurance recipients only. Finally, it should be noted that payment conditions only vary between (106) contract types. Thus, standard errors for the estimates have to be adjusted for contract-type clustering effects.

**< TABLE 3 HERE >**

Generally, almost all of the coefficient estimates in Table 3 corroborate the standard risk-incentives framework. First, we find the probability of NCNP contracts in 2003-2005 to increase with respect to the average job placement rate per gross cohort type in 2002. In particular, a 10 percentage point increase in the job placement rate increases the probability of a NCNP payment scheme about 14 percentage points. The proxy for risk involved with the contracted cohort, that is, the standard deviation of the average job placement rate of the observed cohort per client (in

2002), is negatively associated with the probability of NCNP payment schemes. This is again in line with standard risk premium arguments. Interestingly, higher payments offered (per placement) for a given cohort is negatively related to the likelihood of a NCNP payment scheme, likely reflecting that higher payments per placement signal the difficulties of placing these cohorts into jobs and the risk for placement rate outcomes associated with serving them. In contrast, we do not find a relationship between the provider market share and the likelihood of NCNP schemes. This result is not surprising, in light of our earlier point that payment schemes were determined ex ante by cohort characteristics and not by the providers assigned to these cohorts. It may still be that selection of larger providers into NCNP contracts has occurred in the bidding process, after the release of contract type descriptions and corresponding payment schemes. That said, our results do not suggest that selection in the bidding process was important; that is, the probit model estimates do not indicate that the concentration of (large) WTW providers for these contract types has increased over time.

An alternative test for concentration effects of providers would involve a direct comparison between measures of market concentration of NCLP vis-à-vis NCNP contract types. Therefore, we calculated Herfindahl-Hirschman Indices (HHI), using the market shares of all providers per cohort type $j$ and per year $t$.[6] Over the years, the resulting (average) HHI scores are fairly constant, ranging between 0.25 and 0.30. We also do not find a significant difference between the HHI averages of NCLP and NCNP cohort types. Again, this suggests that large-scale providers did not expand their market shares in NCNP cohorts, crowding out smaller ones. Finally, we also compared the persistency of market shares of providers in gross cohort types that switched from NCLP to NCNP contracting in 2003-2005 to those that continued as NCLP contracts. We found 57.6 percent of the providers in the first category to be awarded with contracts in 2003-2005 for the same gross cohort types, and 50 percent in the second category to be awarded with contracts for the same gross cohort types. This suggests that persistency rates were higher in the NCNP gross cohort types.

***Pre-program selection effects***

As discussed above, we measure two types of selection effects with our data: provider-induced selection and client-induced selection. In addition to direct selection by sending clients back, providers may have also used more subtle ways to selectively treat clients, such as exerting less effort in working with certain clients or even by discouraging them to fill in reintegration plans and/or start the program. As WTW providers were not formally held responsible for such indirect client-induced selection, this would not necessarily diminish their future prospects of being assigned new cohorts. Providers may therefore have tried to influence client-induced selection as well. We model both of these potential effects, using the following specifications:

$$(2) \qquad S^{P}_{ijk,\,t} \;=\; a^{P}\, NCNP_{\,ij,t} \;+\; \mathbf{Z}_{ijk,\,t}\, b^{P} \;+\; \varepsilon^{P}_{\,ijk,\,t}$$

and

$$(3) \qquad S^{C}_{ijk,\,t} \;=\; a^{C}\, NCNP_{\,ij,t} \;+\; \mathbf{Z}_{ijk,\,t}\, b^{C} \;+\; \varepsilon^{C}_{\,ijk,\,t}$$

where $S^P$ and $S^C$ denote the observed provider and client-induced selection shares of the assigned clients to a provider. Both these selection measures are observed per cohort $k$ belonging to

[6] We also calculated HHI values per gross cohort type. Obviously, as gross cohort types are more heterogeneous than cohort types, this resulted in lower HHI-values, but again there were no significant differences between NCLP

cohort type $j$ and gross cohort type $i$. The variables $a^P$ and $a^C$ describe the effect of *NCNP* contracting on provider and client-induced selection, respectively. Again, it should be noted here that the payment scheme dummy *NCNP* varies with (gross) cohort type only. $b^P$ and $b^C$ are vectors denoting the effects of the matrix **Z** on provider- and client-induced selection, respectively. Note that the matrix **Z** includes variables that all vary over $i$, $j$, $k$ and $t$. Finally, $\varepsilon^P$ and $\varepsilon^C$ are i.i.d. residual terms with mean zero and variance $\sigma_P^{\,2}$ and $\sigma_C^{\,2}$.

Equations (2) and (3) can be estimated with standard GLS techniques, using cohort size as analytic weights. In addition, we have indicated earlier that we essentially follow two estimation strategies to control for selection effects. The first strategy is to estimate treatment selection models for NCNP, with equation (1) in the first stage and $S^P$ and $S^C$ (equations (2) and (3)) in the second stage. Thus, equations (2) and (3) are extended with coefficient estimates of the Mill's ratios that are obtained from equation (1). As equation (1) includes the exogenous baseline parameters for 2002 and is estimated for observations from 2003 to 2005, the time span for observations that are used to estimate equations (2) and (3) is limited to this period as well.

The second estimation strategy is a fixed effects estimation of equations (2) and (3), with gross cohort types as the relevant units. Thus, our basic assumption in this model is that $\varepsilon^P$ and $\varepsilon^C$ consist of a time invariant, fixed gross cohort effect that may be correlated with *NCNP* and a residual term that is assumed to be uncorrelated.

**< TABLE 4 HERE >**

Table 4 shows the estimation results for equations (2) and (3), using GLS and the treatment selection model and fixed effects approaches. The coefficient estimates of *NCNP* have

and NCNP gross cohort types.

the expected (positive) signs in all cases, although only about half are statistically significant at $\alpha<0.05$. As for provider-induced selection, both the magnitude and the range of estimates are small (0.6 to 0.8 percentage -point), suggesting the potential for (any) cream skimming from this source is robust but also likely limited. As the (average) share of client-induced selection is much more substantial than that of provider-induced selection, the room for any differences between NCLP and NCNP cohorts seems higher as well. Still, the fixed effect estimates suggest that WTW-providers do not seem to affect these outcomes to a large extent. The results also show that provider-induced selection is higher in DI cohorts, which is probably due to the higher degree of specialization of services that is needed for these cohorts. Finally, we only find the year dummy coefficients to be significant for provider-induced selection, with a downward trend over time. It may be that providers became more aware of the consequences of high provider-induced selection rates over the years, therefore trying to keep these numbers low.

**< TABLE 5 HERE >**

Table 5 presents the (pooled) fixed effect estimates of NCNP schemes on provider and client-induced selection for subsamples with different cohort sizes (more or less than 50 clients) and benefit schemes (UI or DI).[7] These estimates suggest that for both selection measures, the effect of NCNP schemes has been strongest in the (smaller) DI cohorts. It may well be that the risks associated with having DI clients with poor job prospects in smaller cohorts were exacerbated by the NCNP scheme. Training and mediation costs are substantial for this group, whereas the risk of not receiving any (flexible) payments (due to failure to place) was substantial, too. Still, one should bear in mind that the estimated effect of about 5 percentage-points was relevant for only 8 percent of the DI cohorts, so the overall effect was limited.

Together, these findings on pre-program selection suggest that the effects of NCNP contracting (compared to NCLP) on the behavior of providers was probably limited. We offer several potential explanations for this result. First, high fractions of clients being sent back could damage the reputation and future contracting prospects of providers. Second, providers have little leeway to exploit any informational advantages over the social benefit administration. In particular, the time span between the assignment of cohorts and the actual start of the program was only a few weeks. It is unlikely that this period was long enough for providers to make an accurate assessment of the a priori chances of their clients finding work. Furthermore, the contracted cohorts were intended to consist of fairly homogenous groups of workers, which likewise limited the room for selection behavior as well.

***Parking effects***

We argued earlier that parking might be an alternative device for pre-program selection, with the key difference that parking occurs after clients are assigned to providers. In the case of strict performance-based contracting arrangements (like NCNP), parking is unlikely to raise (average) job placement rates but may be beneficial 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 may also be difficult to discern. Still, the data at hand allow us to derive a test on parking indirectly, essentially by focusing on its potential effect on job search durations of clients. That is, parking is characterized as providers concentrating their effort on clients with better job prospects, causing the prospects for this group to improve, while

[7] We also estimated model variants for sub samples, using the two stage selection approach. This led to similar, but less efficient coefficient estimates, as cohort characteristics as size and benefit scheme largely explain the

hard-to-place clients receive less attention (doing little to affect or even decrease their job prospects). Stated differently, the job search durations of easy-to-place clients will become shorter, while those of hard-to-place clients will be unchanged or may even lengthen. In effect, our implicit assumption is that cream skimming operates through the a priori chances of finding employment, rather than on the value-added of programs. 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). As Heckman et al. (1997) argue, this assumption can be justified by the fact that the variance in value-added is small compared to the variance in a priori job finding probabilities.

The test for parking that we implement entails a comparison of the average job search duration for clients in NCLP and NCNP cohorts that have found a job. Under the parking hypothesis, we expect job duration to be shorter for NCNP cohorts. In the appendix to this paper, we present the technical details of this test. Specifically, as the average job search duration of job finders is a function of the job placement rate (which may differ between NCLP and NCNP cohorts), a simple comparison between the truncated average durations—controlling for gross cohort type fixed effects—may give biased results. We therefore control for job placement rate differentials by employing (and controlling for) a first order Taylor series expansion, so as to distinguish the variation in truncated average job finding durations that can be associated with parking only.

**< TABLE 6 HERE >**

Table 6 shows the estimated test statistics for parking differentials between NCLP and NCNP cohorts, with and without controlling for differentials in job placement rates (for

---

probability of NCNP payment schemes as well.

subsamples of our data). The table also contains coefficient estimates of (log) program length differentials between NCLP and NCNP cohorts, (again) after controlling for gross cohort type fixed effects and year dummies. Clearly, the observed program lengths are shorter for NCNP cohorts, with estimates ranging from 8 to 21 percent. The coefficient estimates obtained for program lengths are comparable in magnitude to those for differentials in the job search durations of job finders (that is, our test statistic for parking that does not correct for job placement rates). As the windfall gains of additional job placements within the program for NCNP contracts were larger, one may expect program lengths to be longer for NCNP contracts. Instead, it appears that NCNP contracting encouraged providers to concentrate on short term services for their clients, being less costly and requiring shorter program lengths. In UI cohorts, the opportunities for providers to respond accordingly were probably greater, which is reflected in the larger effect of NCNP contracting we find for this subsample. We return to this issue below when discussing the estimation results for job placement rates.

For the full sample, we find that we cannot reject the null hypothesis of "no parking." Although the average job search duration of job finders is significantly shorter in NCNP cohorts (about 12%, or 35 days), controlling for job placement rates yields a coefficient estimate that is insignificant. Focusing on the subsamples in our data, however, we find evidence for parking in smaller cohorts (less than 50 clients) and weak evidence for DI cohorts, with estimated effects of 25 percent and 22 percent, respectively. Thus, similar to pre-program selection, it appears that parking is confined to cohorts where placement risks and program costs per client were greater. Parking may be one way to lower the financial risks in some contracts—particularly those for smaller DI cohorts—casting doubts on the likelihood of strong unintended effects of incentives induced by NCNP contracting.

***Job placement rate effects***

In estimating the impact of NCNP payment schemes on provider job placement rates, we essentially follow the same two estimation strategies as for the selection models in equations (3) and (4). An important difference now is that pre-program selection itself may have affected the composition of clients in the *net* cohorts—at least to the extent that this in turn has affected the job placement rates, as compared to the *gross* cohorts. We thus add both provider and client-induced selection as additional controls, and specify the job placement rate (*JPR*) per cohort *k* at time *t*, for cohort type *j* and gross cohort type *i* as:

$$(4) \qquad JPR_{ijk,\,t} = a^{JPR}\, NCNP_{\,ij,t} \; + \; \mathbf{Z}_{ijk,\,t}\, b^{JPR} \; + d^{P}\, S^{p}_{\,ijk,\,t} \; + \; d^{C}\, S^{C}_{\,ijk,\,t} \; + \; \varepsilon^{JPR}_{\,ijk,\,t}$$

In equation (4) $a^{JPR}$ and $b^{JPR}$ describe the effect of NCNP contract types and other cohort characteristics **Z** on JPR, respectively. Moreover $d^P$ and $d^C$ denote the effect of provider and client-induced selection on the job placement rates of the (remaining) clients in the cohort. $\varepsilon^{JPR}$ is an i.i.d. residual with mean zero and variance $\sigma_{JPR}^{2}$. Similar as the selection models, we estimate treatment selection models for NCNP in the first stage, and equation (4) in the second stage, adding Mill's ratios that are obtained from equation (1). The second research strategy is the gross cohort type fixed effects estimation of the model.

**< TABLES 7 AND 8 HERE >**

Table 7 shows the estimation results of equation (5) from GLS and the two estimation strategies that take into account potential endogeneity effects. Next, Table 8 highlights the effects on (two) types of job placement rates—into short term and more durable jobs—and distinguishes the effects of NCNP with respect to cohort size (smaller or larger than 50 clients) and type of scheme (UI versus DI). It should be noted that all coefficient estimates in Table 8 derive from the (pooled) fixed effects estimation results.

The comparison of GLS and the other estimates in Table 8 makes apparent that the NCNP contracts were mostly used for cohorts of clients with better job prospects; without controlling for these selection effects, the NCNP effect would be overestimated. The overall picture that emerges is that the NCNP contracting has increased job placement rates, but with the effect being concentrated in short term placements (with contracts from 6 to 12 months) and for (larger) UI cohorts. In particular, we find the overall effect of NCNP contracts on the job placement rates is equal to 7.5 percentage points in the treatment selection model and 2.5 percentage points in the fixed effects model. One may argue that parking decreased the effectiveness of the NCNP incentives in the smaller DI cohorts. We further find the positive effect of NCNP on the UI cohorts to be confined to short-term employment contracts only, suggesting that providers with NCNP incentives focused on short-term employment contracts for their clients. This confirms our earlier findings that indicated that NCNP cohorts had shorter program durations. It is probable that the opportunities for finding temporary work and cutting costs were greater in cohorts with clients with better job prospects. At the same time, however, we also find evidence that providers took the opportunity to increase the number of short-term employment contracts in DI cohorts, at the cost of fewer longer-term contracts.

Table 7 shows that most (other) coefficient estimates of the controls in our sample are not statistically significant. In addition, we unexpectedly find in one model that pre-program provider-induced selection decreased the job placement rates of WTW providers. In case of cream-skimming, we would have expected this form of selection to increase job placement rates of the remaining cohort instead. A possible explanation for this result would be that provider-induced selection is high for low-quality WTW providers employing low-quality workers and/or with poorly managed programs. We therefore re-estimated equation (4) with (pooled) gross cohort type fixed effects, extending the model with the addition of provider dummies as well. The resulting coefficient for provider-induced selection subsequently turns positive, increasing to 2.218 (4.438), whereas the estimated impact of NCNP remains almost unchanged (with an estimate of 2.831 (1.400)). Thus, it seems that high provider-induced selection in these cohorts is indeed associated with low-quality providers.

## 5. CONCLUSIONS

In this paper we have explored both the intended and unintended effects of performance-based contracting in social welfare services, using rather unique data on cohorts that were procured by the Dutch social benefit administration. The results of our analysis generally fit expectations of the standard theoretical framework of risk and incentives, with incentives most likely to work when the financial risks of performance-contingent payments for providers are not too large to induce selection and parking behavior (Burgess and Ratto 2003). It is possible that the Dutch benefit administration was aware of this, as they used NCNP payment schemes primarily for cohort types where clients had higher a priori chances of finding work and job placement risks per client were low. At the same time, our analysis shows that the overall (additional) pre-program selection and parking activities were probably low. For the UI cohorts, we find that the NCNP payment scheme contributed to about a 3 percentage point increase in job placements, whereas there is no evidence for pre-program selection and parking. Although this effect is smaller than those obtained in previous research—with estimates ranging from 5 to 10 percentage points (Finn 2008)—for these cohorts, the NCNP schemes seem to have been generally successful.

The downside of the story, however, is that NCNP contracting was not effective in raising job placement rates in DI cohorts, with evidence pointing to additional parking and provider- and client-induced selection here. Both forms of selection by providers are seemingly driven by risk and cost considerations, which are particularly relevant for hard-to-place DI recipients. Although it is difficult to assess the exact magnitude of these effects, parking may well have harmed both the efficiency of the program—in terms of job placement rates—and its equity in access to services.

The other concern that follows from our analysis is related to the durability of jobs that were found in NCNP programs. Additional job placements were almost fully confined to temporary employment arrangements of twelve months at maximum. 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. So far, however, other evidence seems to suggest that short- and long-term effects of programs are not strongly correlated (Heckman et al. 2002), casting doubt on the long-term effectiveness of the NCNP payment scheme.

In general, the findings of this study suggest that in contexts where the risks to social welfare service providers of failing to meet contract expectations are greater, due to factors such as client characteristics that portend barriers to successful outcomes, performance-based contracts are less effective and are more likely to induce unintended effects such as parking that result in inequity in access to services. 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. That said, these benefits of performance-based contracting appear to be limited to short-term outcomes, and thus, their effectiveness in increasing the long-run impact of public programs such as welfare-to-work services remains questionable.

## REFERENCES


Baker, G. P., 1992. Incentive Contracts and Performance Measurement, Journal of Political Economy 100(3), 598-614.

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(1), 118–141.

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.

Blank, R.M., 2000. When can public policy makers rely on private markets? The effective provision of social services, Economic Journal 110(462), C34-49.

Brown, T. L., Potoski, M., 2003. Managing Contract Performance: A Transaction Costs Approach, Journal of Policy Analysis and Management 22(2), 275-297.

Bruttel, O., 2005. Contracting-out and Governance Mechanisms in the Public Employment Service, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Discussion paper 2005-109.

Burgess, S., Ratto M., 2003. The Role of Incentives in the Public Sector: Issues and Evidence, Oxford Review of Economic Policy 19(2), 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.

Chiappori, P., Salanié B., 2003. Testing Contract Theory: A Survey of Some Recent Work, in: Dewatripont, M., Hansen, L., Turnovsky, P. (Eds), Advances in Economics and Econometrics—Theory and Applications, Eight World Congress of the Econometric

Society, Econometric Society Monographs, Cambridge, Cambridge University Press, 115-49.

Courty, P., Marschke, G., 2004. An Empirical Investigation of Gaming Responses to Performance Incentives, Journal of Labor Economics 22(1), 23-56.

Courty, P., Kim, D., Marshke, G., 2009. Curbing Cream-Skimming: Evidence on Enrollment Incentives, Journal of Labor Economics, forthcoming.

Dixit, A., 2002. Incentives and Organisations in the Public Sector: An Interpretative Review, Journal of Human Resources 37(4), 696-727.

Finn, D., 2008. The British ‘welfare market’. Lessons from contracting out welfare to work programmes in Australia and the Netherlands, University of Portsmouth, 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?, Amsterdam, SEO rapport 946.

Hart, O., Schleifer A., Vishny, R.W., 1997. The Proper Scope of Government: Theory and an Application to Prisons, The Quarterly Journal of Economics 112(4), 1127-1161.

Heckman, J., Smith, J., Taber, C., 1996. What Do Bureaucrats Do? Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program, In: Advances in the Study of Entrepreneurship, Innovation and Economic Growth Volume 7, Reinventing Government and the Problem of Bureaucracy, Gary Libecap, ed., Greenwich, CT: JAI Press, 191-218.

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(4), 487-535.

Heckman, J., Lalonde, R., Smith, J., 1999. The Economics and Econometrics of Active Labor Market Programs, in: Ashenfelter, O., Card, D (Eds.*), Handbook of Labor Economics, Volume 3A*, Amsterdam: North Holland, 1865-2097.

Heckman, J., Heinrich, C., Smith, J., 2002. The Performance of Performance Standards, Journal of Human Resources 38(4), 778-811.

Heckman, J. and J. Smith, 2004, The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program, *Journal of Labor Economics* 22(2), 243-298.

Heinrich, C. J., 1999. Do Government Bureaucrats Make Effective Use of Performance Management Information?, Journal of Public Administration Research and Theory 9(3), 363-393.

Heinrich, C. J., Marschke, G. M., 2010. Incentives and their Dynamics in Public Sector Performance Management Systems, Journal of Policy Analysis and Management 29(1), 183-208.

Heinrich, C., Choi Y., 2007. Performance-Based Contracting in Social Welfare Programs, The American Review of Public Administration 37(4), 409-435.

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(2), 171-190.

Johnson Dias, J., Maynard-Moody, S., 2006. For-Profit Welfare: Contracts, Conflicts, and the Performance Paradox, Journal of Public Administration Research and Theory 17, 189-211.

Koning, P., 2008. Not-for-profit Provision of Job Training and Mediation Services: An Empirical Analysis Using Contract Data, De Economist 156(3), 221-239.

Koning, P., 2009, Contracting Welfare-to-work services: Use and Usefulness, CPB Discussion paper 135.

Mundlak, Y., 1978. On the Pooling of Time Series and Cross Section Data, Econometrica 46(1), 69-85.

Ladd, H., 1999. The Dallas School Accountability and Incentive Plan: An Investigation of its Impact on Student Outcomes, Economics of Education Review 18(1), 1-16.

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(8), 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(3), 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, OECD Publishing.

Van Slyke, D., 2003. The Mythology of Privatization in Contracting for Social Services, Public Administration Review 63(3): 296-315.

Weber, A., 2008. Individual Incentives in Program Participation: Splitting Up the Process in Assignment and Enrolment, IZA Discussion paper 3404.

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

Parking behavior typically cannot be observed directly. Still, the data at hand allow us to derive an indirect test for parking, essentially 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 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 NCLP and NCNP cohorts that have found a job. Under this parking hypothesis, we expect job search durations to be smaller for NCNP cohorts. To formalize and develop such a test, we first introduce some notation:

- $\tau$ = program length, which we assume is the same for all cohorts.
- $t$ = time, with $0 \leq t \leq \tau$ .
- $P(t)$ indicates the job placement rate of a cohort at time t. We only observe $P$ at time $t = \tau$.
- $S(t) = 1 - P(t)$, which is the fraction of clients that have not found a job at time $t$ (or: 'the survival function').
- $\theta(t)$ is the job finding hazard in a cohort at time $t$. Thus, $S(t) = \exp[ - \int_0^t \theta(s)ds ]$
- $E(t|t \leq \tau)$ is the expected job search duration of clients that have found a job before $\tau$ .

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

$$(\text{A.1}) \quad \text{E}\,(t|t \leq \tau) \quad = \quad \frac{\int_0^{\tau} S(t)\, dt - S(\tau)\,\tau}{1 - S(\tau)}$$

As the RHS of equation (A.1) is a function of $P(\tau)$, a simple comparison between the truncated average durations of NCLP and NCNP cohorts may give biased results. In particular, if one of the cohort types has a higher placement rate, this would result in a lower value of the survival rate $S(\tau)$, regardless of whether this difference is associated with parking behavior. *We therefore define "no parking" as a situation where all hazard rates $\theta(t)$ during the program are affected proportionately by the higher placement rate, perhaps with a (small) percentage increase in $\alpha$.* It can easily be shown that the new survival function would then be $S(t)^{1-\alpha}$ for all $t$. Next, we perform a first order Taylor expansion on the log value of E $(t|t \leq \tau)$ , with $\alpha = 0$ as a point of reference. The first order derivative with respect to $\alpha$ for $\alpha = 0$ is:

$$(\text{A.2}) \quad \frac{\partial \ln \text{E}(t|t \leq \tau)}{\partial \alpha} \quad = \quad -\alpha - \frac{\alpha\, S(\tau)}{1\text{-}S(\tau)} \quad = \quad - \frac{\alpha}{1\text{-}S(\tau)}$$

and the derivative of (1) with respect to $P(\tau)$ for $\alpha = 0$ is:

$$(\text{A.3}) \quad \frac{\partial P(\tau)}{\partial \alpha} = -\frac{\partial S(\tau)}{\partial \alpha} = \alpha S(\tau)$$

Dividing (A.2) and (A.3) yields

$$(\text{A.4}) \quad \frac{\partial \ln \mathrm{E}(t|t \le \tau)}{\partial P(\tau)} = -\frac{1}{S(t)\,(1-S(\tau)\,)}$$

Equation (A.4) shows that the (log) truncated expected value of job search durations decreases with respect to the job placement rate at time $\tau$, assuming constant concavity of the survival function (i.e., without parking). Thus, when comparing the (log values of) the truncated durations $\mathrm{E}(t|t \le \tau)$ for cohorts with NCLP and NCNP, we control for differences in job placement rates as well. In doing this, we first calculate the average derivative value of equation (4) for each gross cohort type *i* in our sample, measured for the "clean" baseline year 2002 when NCNP was still absent.[8] Similarly, we derive the average job placement rate per gross cohort type *i* , also for 2002 only. Next, the "no-parking" effect of job placement rate differentials is estimated as the product of the average derivative value of equation (A.4) and the difference between the job placement rate of cohort *ijt* and the average job placement rate per gross cohort type in 2002. We subtract this value from $\ln \mathrm{E}(t|\, t \le \tau)$ and then estimate the "clean" impact of

[8] If we would perform a Taylor approximation within cohort types instead, we would not have the within variation that is needed to estimate the impact of NCNP contracting.

NCNP on this value for 2003-2005, while controlling for gross cohort type fixed effects and year dummies.

## TABLES AND FIGURES

**Table 1: Sample statistics of contracted cohort data: observations and population averages (Standard deviations in parentheses)**

| | Years | | | | Contract types | |
|---|---|---|---|---|---|---|
| | **2002** | **2003** | **2004** | **2005** | **NCLP** | **NCNP** |
| # Gross cohort types | 16 | 13 | 10 | 11 | 17 | 7 |
| # Cohort types | 52 | 20 | 17 | 17 | 93 | 13 |
| # Cohorts | 3,733 | 1,292 | 1,690 | 1,795 | 7,441 | 1,069 |
| # Assigned clients | 105,623 | 66,861 | 47,910 | 23,436 | 193,361 | 50,469 |
| Participants per cohort (gross) | 28.3 (42.3) | 51.8 (62.8) | 28.4 (32.7) | 13.1 (19.3) | 26.0 (37.2) | 47.2 (66.0) |
| Participants per cohort (net) | 24.2 (36.0) | 43.0 (51.5) | 23.7 (27.5) | 10.9 (15.9) | 22.1 (31.7) | 38.5 (53.2) |
| NCNP contracting | 0.000 (0.017) | 0.351 (0.477) | 0.296 (0.457) | 0.546 (0.498) | | |
| Disability Insurance | 0.610 (0.487) | 0.413 (0.492) | 0.545 (0.498) | 0.370 (0.483) | 0.592 (0.491) | 0.246 (0.431) |
| Payment per client (euros) | – | 3,690 (1,167) | 3,601 (784) | 3,272 (1,293) | 3,953 (1,015) | 2,898 (917) |
| Client-induced selection | 0.118 (0.073) | 0.147 (0.075) | 0.150 (0.107) | 0.156 (0.135) | 0.128 (0.089) | 0.165 (0.088) |
| Provider-induced selection | 0.028 (0.034) | 0.022 (0.029) | 0.014 (0.024) | 0.013 (0.035) | 0.023 (0.032) | 0.020 (0.030) |
| Trajectory duration, job finders | 272.8 (89.4) | 291.8 (78.9) | 276.3 (87.6) | 207.7 (97.1) | 288.5 (90.1) | 215.2 (59.0) |
| Trajectory duration, non finders | 467.5 (91.0) | 564.8 (102.) | 526.2 (109.0) | 418.0 (131.6) | 514.3 (111.5) | 451.7 (105.2) |
| Job placement (total) | 0.327 (0.121) | 0.334 (0.124) | 0.368 (0.135) | 0,351 (0.194) | 0.316 (0.129) | 0.429 (0.116) |
| Job placement: contract of 6-12 months | 0.257 (0.107) | 0.257 (0.106) | 0.263 (0.117) | 0.257 (0.162) | 0.240 (0.112) | 0.329 (0.100) |
| Job placement: contract > 12 months | 0.070 (0.059) | 0.077 (0.047) | 0.105 (0.073) | 0.094 (0.085) | 0.076 (0.065) | 0.100 (0.054) |

## Table 2: Gross cohort types

| Gross cohort type | # Years | Cohort types | Cohorts | Clients (gross) | Fraction NCNP |
|---|---|---|---|---|---|
| Disabled immigrants, bad job prospects | 3 | 9 | 1,157 | 10,583 | 0.000<br>(.) |
| Disabled, good job prospects | 3 | 8 | 954 | 40,737 | 0.243<br>(0.439) |
| Disabled: mental impairments | 2 | 6 | 805 | 31,517 | 0.000<br>(.) |
| Disabled: mild mental impairments | 4 | 9 | 1,097 | 30,709 | 0.000<br>(.) |
| Disabled: work related impairments | 1 | 5 | 49 | 722 | 0.000<br>(.) |
| Disabled: kidney patients | 4 | 4 | 35 | 510 | 0.000<br>(.) |
| Disabled: visual and hearing impairments | 3 | 4 | 152 | 1,988 | 0.000<br>(.) |
| Disabled: young workers, no work history | 3 | 14 | 757 | 7,325 | 0.023<br>(0.151) |
| Unemployed immigrants | 2 | 2 | 190 | 6,616 | 0.000<br>(.) |
| Unemployed, good job prospects | 4 | 9 | 826 | 51,177 | 0.629<br>(0.483) |
| Unemployed, older than 50 years | 3 | 9 | 537 | 20,358 | 0.000<br>(.) |
| Unemployed, bad job prospects | 3 | 9 | 705 | 18,409 | 0.000<br>(.) |
| Unemployed: highly educated | 2 | 2 | 52 | 962 | 1.000<br>(.) |
| Workers in graphical industry | 1 | 5 | 49 | 722 | 0.000<br>(.) |
| Workers eligible for subsidized employment | 1 | 1 | 89 | 767 | 0.000<br>(.) |
| Returned clients ("second chance" programs) | 2 | 5 | 427 | 7,649 | 0.097<br>(0.296) |
| Trajectories aiming at self-employment | 2 | 2 | 92 | 1,081 | 0.000<br>(.) |
| Other | 3 | 7 | 586 | 12,720 | 0.508<br>(0.500) |

**Table 3: Probit estimation results for NCNP cohorts (2003-2005)**
**(Marginal effects; standard errors in parentheses; *, ** and ***
indicate significance at 10%, 5% and 1%, respectively)**

| | **Coefficient estimate** | **Standard deviation** |
|---|---|---|
| Average job placement rate (per gross cohort type in 2002) | 1.372*** | (0.659) |
| Risk proxy (per gross cohort type in 2002) | −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 insignificant.

**Table 4: Estimation results of provider and client-induced model (Effects in %-points; standard errors in parentheses; *,** and *** indicate significance at 10%, 5% and 1%)**

| | Provider-induced selection | | | Client-induced selection | | |
|---|---|---|---|---|---|---|
| | GLS | Two stage selection model | Fixed effects; Mundlak | GLS | Two stage selection model | Fixed effects; Mundlak |
| NCNP payment scheme | 0.572** | 0.847*** | 0.702* | 2.417* | 5.908** | 2.209 |
| | (0.240) | (0.191) | (0.398) | (1.362) | (2.760) | (1.530) |
| Disability Insurance (dummy) | 0.748*** | 1.231*** | 1.123*** | 0.761 | 4.102*** | 2.818*** |
| | (0.204) | (0.168) | (0.191) | (1.264) | (1.525) | (0.809) |
| (Expected) reward per client | −0.040 | −0.109 | −0.055 | −0.376 | −0.101 | −0.282 |
| | (0.067) | (0.086) | (0.076) | (0.341) | (0.753) | (0.338) |
| Log gross cohort size | −0.069 | −0.012 | 0.118 | 0.234 | −0.199 | 0.144 |
| | (0.089) | (0.085) | (0.089) | (0.442) | (0.312) | (0.317) |
| Year = 2003 | −0.745*** | | −0.958*** | 2.268* | | 0.988 |
| | (0.225) | | (0.263) | (1.269) | | (0.830) |
| Year = 2004 | −1.560*** | −0.848*** | −1.641*** | 2.684 | 0.134 | 1.292 |
| | (0.184) | (0.152) | (0.239) | (1.859) | (1.429) | (1.116) |
| Year = 2005 | −1.720*** | −1.287*** | −2.237*** | 2.144 | −1.029 | 1.598 |
| | (0.257) | (0.196) | (0.284) | (1.524) | (1.068) | (1.388) |
| Constant | 2.126*** | 1.658 | 11.705*** | 15.626*** | 7.589 | 34.824 |
| | (0.745) | (1.143) | (4.477) | (4.420) | (10.761) | (23.214) |
| Mill's lambda | | −0.207** | | | −2.122 | |
| | | (0.094) | | | (1.712) | |
| Rho | | −0.077** | | | −0.229 | |
| | | (0.036) | | | (0.183) | |
| R-squared | 0.091 | | 0.103 | 0.083 | | 0.139 |
| Log likelihood | | 7,617.3 | | | 2,750.1 | |

Note that the regressions also included (6) region dummies as controls in all specifications.

**Table 5: Fixed effect estimation results of NCNP on provider and client-induced selection (Effects in %-points; standard errors in parentheses; *, ** and *** indicate significance at 10%, 5% and 1%)**

| | Provider-induced selection | | Client-induced selection | |
|---|---|---|---|---|
| | coefficient estimate | standard error | coefficient estimate | standard error |
| < 50 clients | 1.177*** | (0.413) | 4.588* | (2.445) |
| ≥ 50 clients | 0.519 | (0.417) | 1.115 | (1.151) |
| DI cohorts | 1.515*** | (0.313) | 5.285*** | (1.981) |
| UI cohorts | 0.047 | (0.266) | −0.353 | (1.121) |

**Table 6: Program length differentials and test statistics for NCLP-NCNP parking difference, using job search durations of job finders (Standard errors in parentheses; *, ** and *** indicate significance at 10%, 5% and 1%, respectively)**

| | Effect of NCNP on Log Program Length | | Test statistic, corrected | |
|---|---|---|---|---|
| | coefficient estimate | Standard error | coefficient estimate | Standard error |
| Full sample | –0.154*** | (0.049) | –0.113 | (0.088) |
| < 50 clients | –0.158*** | (0.045) | –0.249** | (0.113) |
| ≥ 50 clients | –0.163*** | (0.048) | –0.067 | (0.081) |
| DI cohorts | –0.086*** | (0.038) | –0.221* | (0.122) |
| UI cohorts | –0.209*** | (0.051) | –0.031 | (0.081) |

Coefficient estimates follow from (pooled) FE estimation with gross cohort types as fixed effects and controlling for year effects.

**Table 7: Estimation results for job placement rate model (Effects in %-points; standard errors in parentheses; *, ** and *** indicate significance at 10%, 5% and 1%, respectively)**

| | GLS | | Treatment selection model | | Fixed effects (Mundlak) | |
|---|---|---|---|---|---|---|
| | coefficient estimate | standard error | coefficient estimate | standard error | coefficient estimate | standard error |
| NCNP payment scheme | 11.115*** | (1.999) | 7.489* | (3.947) | 2.555* | (1.436) |
| Provider-induced selection | −13.331 | (8.835) | −10.995 | (8.222) | −10.867* | (4.768) |
| Client-induced selection | 2.331 | (4.333) | −0.006 | (5.310) | −2.623 | (2.986) |
| Disability Insurance (dummy) | −4.578** | (1.795) | −4.949*** | (1.769) | −8.005*** | (2.761) |
| (Expected) reward per client | −0.839 | (0.582) | 0.298 | (0.764) | 0.399 | (0.461) |
| Log gross cohort size | 0.258 | (0.532) | 0.388 | (0.769) | 0.194 | (0.348) |
| Year = 2003 | −3.955 | (2.402) | | | 2.039* | (1.163) |
| Year = 2004 | 0.762 | (2.104) | 4.371*** | (1.600) | 6.041*** | (1.103) |
| Year = 2005 | −5.754* | (2.990) | 2.117 | (2.848) | 5.292*** | (2.019) |
| Expected job placement rate in 2002 per gross cohort. | | | 80.205*** | (20.01) | | |
| Constant | 47.333*** | (6.681) | 8.671 | (13.847) | 17.621 | (35.602) |
| Mill's lambda | | | −3.542 | (2.206) | | |
| Rho | | | −0.315 | (0.188) | | |
| N | 8,302 | | 3,819 | | 8,302 | |
| R-squared | 0.177 | | | | 0.318 | |
| Log likelihood | | | 4,390.4 | | | |

Note that the regressions also included (6) region dummies as controls in all specifications.

**Table 8: Estimation results of effects of NCNP on different job placement rate types and for sub samples (Effects in %-point; standard errors in parentheses; *, ** and *** indicate significance at 10%, 5% and 1%)**

| | Job contracts: 6-12 months | | Job contracts: more than 12 months | | Total effect on job placement | |
|---|---|---|---|---|---|---|
| | coefficient estimate | standard error | coefficient estimate | standard error | coefficient estimate | standard error |
| Cohorts with < 50 clients | 1.723 | (1.154) | −1.552 | (1.113) | 0.171 | (1.594) |
| Cohorts with ≥ 50 clients | 3.614*** | (1.135) | −0.336 | (0.664) | 3.277** | (1.269) |
| DI cohorts | 3.387*** | (1.079) | −2.218* | (1.332) | 1.117 | (2.044) |
| UI cohorts | 2.707*** | (0.917) | 0.544 | (0.391) | 3.250*** | (1.012) |
| Total group | 3.208** | (1.306) | −0.653 | (0.753) | 2.556* | (1.436) |

**Figure 1: Distribution of average successful durations: NCLP and NCNP cohorts (2002-2005)**

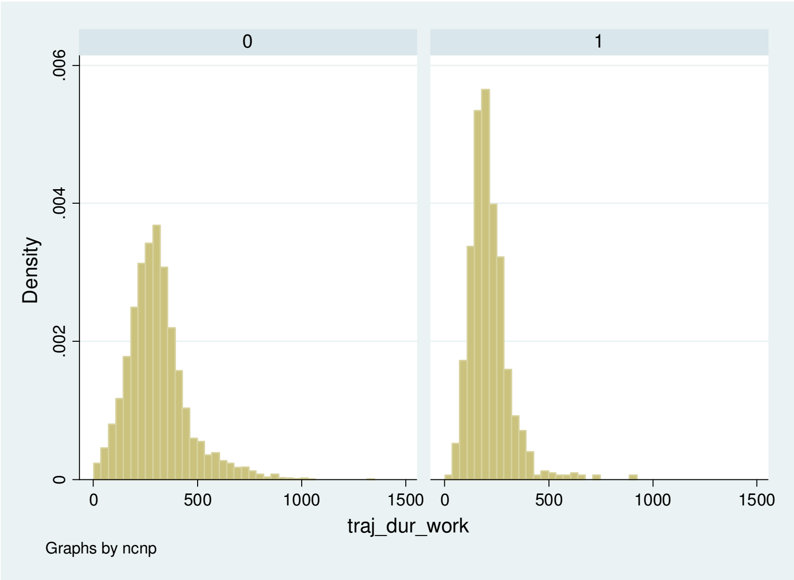

**Figure 2: Distribution of average unsuccessful durations: NCLP and NCNP cohorts (2002-2005)**