Creaming and Dumping: Who on Whom

by
Karolina Socha-Dietrich\textsuperscript{1} and Peter Zweifel\textsuperscript{2}

\textsuperscript{1}University of Southern Denmark
\textsuperscript{2}University of Zurich
Creaming and Dumping: Who on Whom?

Karolina Socha-Dietrich

COHERE, Department of Business and Economics, University of Southern Denmark

Peter Zweifel

Professor Emeritus, Department of Economics, University of Zurich;

COHERE, Department of Business and Economics, University of Southern Denmark
Abstract

In several countries, public healthcare providers purchase services from private providers to shorten waiting times. Some private providers in turn combine public service with practice in their own facilities. According to the existing literature, they are viewed as cream-skimming profitable (low-severity) public patients to the benefit of private practice, causing cost of treatment in the public sector to increase. This is particularly problematic when public provider payment is prospective. However, two facts seem to be neglected. First, cream skimming involves effort and thus does not occur in all circumstances. Second, public providers might have an incentive to select patients too, resulting in dumping of the least profitable (high-severity) patients on the private sector. This paper derives the conditions under which both creaming and dumping are predicted to occur.

**Keywords** Creaming; Dumping; Waiting lists; Public Service; Dual Practice

**JEL Classification** I11 I18
I. Introduction

Patient selection by healthcare providers occurs in two ways: providers may prefer patients with expected payment greater than expected cost of treatment (cream skimming), and they may get rid of patients with expected payment less than expected cost (dumping) (Ma, 1994; Newhouse, 1996; Ellis, 1998). The incentive to discriminate against high-severity patients arises under prospective payment, which makes providers bear the risk of excessive cost, e.g. Diagnosis Related Group (DRG) payment (Ma, 1994; Newhouse, 1996). This incentive cannot be fully neutralized by risk adjustment because the physician diagnosing a patient is likely to have more information about the patient’s future cost than the information contained in the risk-adjustment formula (Newhouse, 1989; Chalkley and Malcomson, 2000). Patient selection both in the guise of cream skimming and dumping is more than a theoretical possibility but has been found in empirical research (Newhouse and Byrne, 1988; Newhouse, 1989; Ellis and McGuire, 1996).

Patient selection is usually viewed as a problem characterizing market-oriented rather than public healthcare systems (Le Grand, 1991; Ellis, 1998). Even when discrimination against patients is illegal in a market-oriented system, providers may effectively reject patients by claiming that they lack the facilities necessary to treat severe cases (Ma, 1994) or by advising medical staff to convince unprofitable patients to seek care elsewhere (Newhouse, 1996). Providers might also cream skim by concentrating on relatively profitable DRGs (Ma, 1994) or by avoiding the addi-
tional effort associated with the treatment of more severe patients (Newhouse, 1996; Barros and Olivella, 2005). As a consequence, providers are predicted to differ substantially with regard to the share of expensive patients treated, while severely ill patients face problems of access to health care (Ellis, 1998). Admittedly, extensive regulation in the public system is designed to inhibit exploitation of unpriced risk heterogeneity. In addition, providers are tied up with a patient population in a given geographical area, while regulators decide on range of facilities and services provided.

Yet, risk selection might still occur through the purchasing of services from the private sector by public providers. These arrangements have been introduced with the aim of shortening waiting lists in the public sector (González, 2005). Patients to be transferred to private facilities have to satisfy certain eligibility criteria, usually length of their waiting time (Vrangbæk et al., 2007; González, 2005; Wiley, 2005). In turn, some providers combine public service with private practice. These dual-job practitioners (DPs) are suspected to manipulate patient transfers in a way that they receive in private practice only the least severe and thus most profitable of the public patients (González, 2005). DPs are said to have the incentive because they receive a fixed salary in the public sector, which serves to insulate them from its financial performance. At the same time, they benefit from a low cost of treatment in their private practice given prospective payment. This causes public institutions to face an increased risk of deficit as a result of an increase in the average severity and
costliness of their patients, whereas private institutions appropriate the surplus from treating the low-severity public patients. This reallocation of surplus is argued to be inefficient since it undermines the ability of public providers to perform special functions in the public interest, such as education and research (González, 2005). Thus, authorities should reconsider purchasing treatment from private providers or disallow the dual job holding.

The present paper re-examines these claims. It proposes a model that takes into account five facts which seem to have been neglected in the existing literature. First, patient selection generally requires effort and time that alternatively could be used for treating patients or for leisure activities. Second, a decrease in the average severity of patients transferred to private facilities by DPs necessarily leads to an increase in the average severity of patients retained in the public sector, causing non-contractible work effort for single-job practitioners (SPs) in public facilities to rise. Therefore, SPs who work exclusively in public hospitals may have an interest in dumping high-severity patients on private providers. Third, SPs can be viewed as residual claimants of the surplus generated by public facilities. While this surplus does not accrue to them as a profit, it does provide resources for research, new equipment, and additional facilities (Rickman and McGuire, 1999). Therefore, SPs have a vested interest in keeping the average cost of treatment low in the presence of prospective payment. This interest is shared by hospital managers, who may instruct SPs to dump high-severity patients on private providers as a way to lower
average cost of treatment. Fourth, the inclination of DPs to cream skim is predicted to be weak if their involvement in the private sector is limited or if their total income is high. Fifth, the effectiveness of patient selection must exceed a threshold value for selection to be lucrative.

In sum, the actions of the two groups of providers (DPs and SPs) have opposite effects on the severity of patient cases treated and hence average cost of treatment, resulting in their (partial) cancellation. Hence, transfers of public patients to private providers, even in combination with allowing dual job holding, does not necessarily drive up average case severity in the public sector and might in fact decrease it.

The reminder of this paper is structured as follows. The theoretical model is presented in Section II, where subsection B derives the optimality conditions for cream skimming and dumping. Finally, a conclusion is provided in the least section.

II. Two models of provider behavior

This section contains two models. One depicts the behavior of providers who work exclusively in a public hospital (single-job practitioners, SPs). The other model applies to providers who combine public service with private practice (dual-job practitioners, DPs). Practitioners working exclusively in the private sector are disregarded because they cannot transfer patients between the two settings. Both types aim at maximizing their utility, which is increasing in leisure and income. The
decision variable is selection effort. For the SP, one has

\[ U_p = U_p(I_p, L_p), \quad \text{with} \quad \frac{\partial U_p}{\partial I_p} > 0, \quad \frac{\partial U_p}{\partial L_p} > 0, \quad \frac{\partial^2 U_p}{\partial I_p^2} < 0, \quad \frac{\partial^2 U_p}{\partial L_p^2} < 0, \quad (1) \]

where \( I_p \) is income and \( L_p \), leisure. Income \( I_p \) in the public sector is assumed to be fixed, e.g. in the guise of a fixed monthly salary. In particular, it does not vary with treatment effort exerted. Yet, this effort does vary as a function of length of treatment and intensity (stressfulness, respectively) of the work, which in turn depends on the severity of patients treated (Barros and Olivella, 2005). Accordingly, leisure is defined as total time \( T \) minus treatment effort \( S_p \) (proportional to average severity of cases) minus other types of effort \( E_p \). In particular, \( E_p \) can be seen as effort exerted on patient selection,

\[ L_p = T - E_p - S_p, \quad E_p > 0. \quad (2) \]

Turning to the dual practitioner, the utility function of the DP is fully analogous to the one of the SP,

\[ U_d = U_d(I_d, L_d), \quad \text{with} \quad \frac{\partial U_d}{\partial I_d} > 0, \quad \frac{\partial U_d}{\partial L_d} > 0, \quad \frac{\partial^2 U_d}{\partial I_d^2} < 0, \quad \frac{\partial^2 U_d}{\partial L_d^2} < 0, \quad (3) \]

where \( I_d \) and \( L_d \) are the DP’s income and leisure, respectively. However, his or her income \( I_d \) has two components, one arising from involvement in the public sector and the other, from private practice. The first component is given by \( k_p I_p \), with \( I_p \) denoting maximum attainable income in the public sector and \( k_p \), the degree of the DP’s involvement. Together with income from private practice, \( I_d \) is given by

\[ I_d = k_p I_p + k_d (r - S_d). \quad (4) \]
Here, $k_d$ symbolizes involvement in the private sector. It is worth noticing that $k_p + k_d \neq 1$. For example, $k_p = 1$ and $k_d > 0$ characterize a DP, who while working full time in the public sector, spends additional time in private practice. The parameter $r$ is total payments for treatment of patients transferred from the public sector. They are reduced by $S_d$, the patients’ mean severity, on the assumption that cost is proportional to severity. Accordingly, DPs’ leisure is given by

$$L_d = T - E_d - (k_p S_p + k_d S_d) > 0,$$

(5)

with $E_d$ denoting other effort (again in particular on patient selection) and $(k_p S_p + k_d S_d)$, total treatment effort.\(^1\)

A. Cream skimming by DPs versus dumping by SPs

It might be worthwhile to recall a few basic facts about the policy of purchasing treatment for public patients from the private sector. Public patients referred to a hospital contact initially their designated public provider, where upon diagnosing a decision is taken whether a patient should join a waiting list. To be transferred to a private provider, patients need to satisfy certain eligibility criteria such as the expected length of their waiting time. Private facilities contracting with the public payer cannot refuse to treat any of the transferred patients who are part of the agreed

\(^1\) For the sake of simplicity, the relative importance of treatment effort in and of income from the private sector are both reflected by the parameter $k_d$. A more general formulation does not change results.
contingent. Usually, choice between the public and the private facilities is left to eligible patients. Alternatively, patients may leave the choice to the public provider in charge, who books treatment according to the available operational capacity and the prioritization rule.

The decision variable of the SP is selection effort \( E_p \geq 0 \). For simplicity, following Ma (1994), SPs are assumed to be able to predict exactly the amount of resources required by a patient. In this situation, they have an incentive to dump severe cases on private facilities for two reasons. First, as long as effort \( E_p \) is smaller than the reduction in treatment effort \( S_p \) achieved, they benefit from an increase of leisure [see eq. (2)]. Second, SPs can be seen as residual claimants to surplus generated by the public provider which can be used for research, new equipment, or facilities (Rickman and McGuire, 1999). While the existing literature on patient selection does not describe how selection is performed in practice, SPs can control the composition of patients they remain in charge of by manipulating transfers to private facilities. One way is to break rules governing the booking of patients for treatment, in particular medical prioritization. Another way is to use their informational advantage to influence patients in their choice between the public and private setting. However, all of this requires effort, denoted by \( E_p \).

Let \( e_p > 0 \) be the marginal effectiveness of selection effort \( E_p \), decreasing in effort \( E_p \). Therefore, mean severity of patients remaining in the public sector decreases by \( e_p E_p \). However, DPs try to reduce the mean severity of patients trans-
ferred to their private practice by exerting effort $E_d > 0$, with marginal effectiveness $e_d$, resulting in a reduction by $e_d E_d$. Therefore, one has

$$\frac{\partial(e_p E_p)}{\partial E_p} > 0, \quad \frac{\partial(e_d E_d)}{\partial E_d} > 0, \quad \frac{\partial e_p}{\partial E_p} \cdot \frac{\partial e_d}{\partial E_d} < 0 \quad \text{and} \quad \frac{\partial^2(e_p E_p)}{\partial E_p^2}, \frac{\partial^2(e_d E_d)}{\partial E_d^2} < 0. \quad (6)$$

Moreover, for both the SP and the DP, cream skimming and dumping by the other necessarily leads to an increase in the severity of patients treated. Hence,

$$S_p = S - e_p E_p + e_d E_d, \quad (7)$$

$$S_d = S + e_p E_p - e_d E_d, \quad (8)$$

where $S$ is mean severity in the population.

B. Optimization by the two types of practitioner

According to the theory of labor economics, dual job holding is a response to income constraints such as limited demand for labor and/or standardized work contracts. If there were no constraints in any of the two jobs, individuals would focus exclusively on the preferred one (Perlman, 1966; Shisko and Rostker, 1976). The healthcare sector abounds with regulations regarding hours of work, which may explain the prevalence of DPs who combine work in public and private practice.

In terms of the two models, these regulations justify treating the two involvement parameters $k_p$ and $k_d$ as exogenous, leaving $E_p$ and $E_d$ as the decision variables of the SP and DP, respectively.
From eq. (2), for the SP in the public sector the first-order condition for an interior solution reads,

\[
\frac{dU_p}{dE_p} = \frac{\partial U_p}{\partial L_p} \frac{\partial L_p}{\partial E_p} = 0,
\]

(9)
since income \(I_p\) does not depend on effort. In view of eqs. (2) and (7), eq. (9) can be rewritten to become

\[
\frac{\partial U_p}{\partial L_p} \left( \frac{\partial (e_p E_p)}{\partial E_p} - 1 \right) = 0.
\]

(10)

Since \(\partial U_p/\partial L_p > 0\), this boils down to

\[
\frac{\partial (e_p E_p)}{\partial E_p} = \frac{\partial e_p}{\partial E_p} E_p + e_p = (\eta_p + 1)e_p = 1,
\]

(11)

where \(\eta_p = \frac{\partial e_p}{\partial E_p} e_p < 0\) is the elasticity of selection effectiveness w.r.t. selection effort.\(^2\) Therefore, \(e_p > 1\) at an interior optimum, while \(e_p \leq 1\) everywhere leads to a boundary optimum \(E_p = 0\). As a consequence, the SP has an incentive to select patients only if the efficiency of dumping exceeds the critical value \(\hat{e}_p = 1\). Otherwise, the SP makes no effort to dump high-severity patients; only for \(e_p > 1\) does the gain from selection outweigh the effort involved.

In full analogy to eq. (9), the first-order condition w.r.t. effort for a DP reads,

\[
\frac{dU_d}{dE_d} = \frac{\partial U_d}{\partial I_d} \frac{\partial I_d}{\partial E_d} + \frac{\partial U_d}{\partial L_d} \frac{\partial L_d}{\partial E_d} = 0.
\]

(12)

The distinguishing feature is that for the DP, the mean severity of patients does not only affect leisure, but also income from private practice. In view of eqs. (3) to

\(^2\) Also, \(\eta_p > -1\), as according to eq. (6), \(0 > \partial e_p/\partial E_p = e_p(1 + \eta_p)\) and \(e_p > 0\).
(5), and (8), the first-order condition reads,

\[
\frac{\partial U_d}{\partial I_d} (-k_d \frac{\partial S_d}{\partial E_d}) + \frac{\partial U_d}{\partial L_d} \left[ -1 - k_p \frac{\partial (e_d E_d)}{\partial E_d} - k_d \left( - \frac{\partial (e_d E_d)}{\partial E_d} \right) \right] = \frac{\partial U_d}{\partial I_d} (-k_d) \left( - \frac{\partial (e_d E_d)}{\partial E_d} \right) + \frac{\partial U_d}{\partial L_d} \left[ (k_d - k_p) \frac{\partial (e_d E_d)}{\partial E_d} \right] = 0.
\]

(13)

Using the elasticity \( \eta_d = \frac{\partial e_d}{\partial E_d} e_d \leq 0 \) but \( > -1 \), due to \( \frac{\partial (e_d E_d)}{\partial E_d} > 0 \) (‘more effort leads to more effect’) from eq. (6) and the marginal rate of substitution between income and leisure \( \mu_d = \frac{\partial U_d}{\partial I_d} / \frac{\partial U_d}{\partial L_d} > 0 \), this can be rewritten as

\[
\mu_d k_d (\eta_d + 1) e_d + (k_d - k_p) (\eta_d + 1) e_d - 1 = 0
\]

or

\[
(\eta_d + 1) e_d \left[ (1 + \mu_d) \rho - 1 \right] - \frac{1}{k_p} = 0,
\]

(14)

where \( \rho = k_d / k_p \). It is of particular interest to know when there is positive selection effort by the DP, i.e., when eq. (14) has a solution. Dividing through by \( k_p \) permits us to discuss the DP’s behaviour based on the relative involvement in the two jobs. Indeed, \( \rho = 0 \) (no involvement in private practice) implies \( E_d^* = 0 \) (an optimum on the boundary), as there is no solution to the first-order condition in view of \( \eta_d + 1 > 0 \). This is consistent with the fact that a DP without involvement in the private practice is actually an SP. As long as the optimum stays on the boundary, \( e_d \) and \( \eta_d \) remain unchanged while \( \rho \) is being increased. For sufficiently large \( \rho \) the term in square brackets becomes positive, and the first term can eventually compensate the
remaining negative addend. Consequently, there also exists a \( \rho \) above which the DP will exert selection effort.

The question now arises as to how long an increase in private practice involvement causes the DP to step up selection effort. To simplify notation, consider an exogenous change \( dr > 0 \). From the first-order condition one obtains

\[
\frac{\partial}{\partial E_d} \frac{\partial U_d}{\partial E_d} dE^*_d + \frac{\partial}{\partial \rho} \frac{\partial U_d}{\partial E_d} dr = 0.
\] (15)

If the solution of the first-order condition is stable, there \( \frac{\partial^2 U_d}{\partial E_d^2} < 0 \) and consequently, the sign of \( \frac{\partial E^*_d}{\partial \rho} \) is given by the sign of \( \frac{\partial}{\partial \rho} \frac{\partial U_d}{\partial E_d} \) at the optimum. From eq. (13) at fixed \( k_p \) one gets

\[
\frac{\partial}{\partial \rho} k_p \frac{\partial U_d}{\partial E_d} = \frac{\partial U_d}{\partial I_d} \left( \frac{\rho}{\mu_d} \frac{\partial \mu_d}{\partial \rho} + 1 + \mu_d \right) (\eta_d + 1) e_d,
\] (16)

with all factors positive apart from the expression in big brackets with indefinite sign. Its first addend is negative under the reasonable assumption that the DP does not lose money by working in the private sector\(^3\), \( r - S_d > 0 \). The remaining addends are positive by definition. As by assumption we are in a parameter range, where eq. (14) has a solution, \( \frac{\partial E^*_d}{\partial \rho} \geq 0 \) and generally, \( \frac{\partial E^*_d}{\partial \rho} > 0 \) at least on a finite interval. Hence, if there is a solution to the first-order condition, when the DP gets more involved in private practice, he or she is predicted to increase selection effort at least for this range of \( \rho \). Whether there will cease to be solutions of eq. (14) for some larger values of \( \rho \), i.e., whether the DP will stop feeling inclined to exert

\(^3\) Growing \( \rho \) at fixed \( k_p \) means less leisure and more income, leading to a decrease in \( \mu_d \).
selection effort depends on the details of the utility and efficiency functions. It appears reasonable to assume that \( \mu_d \) approaches zero as \( \rho \) grows large, as this corresponds to more work and more income, i.e., to a lower valuation of income relative to leisure. Then, if eq. (14) remains solvable for large \( \rho \), \((\eta_d + 1)e_d = \partial(e_d E_d)/\partial E_d \) must decrease (eventually to zero) to compensate for the growing \( \rho \).

To the contrary, if the first-order condition is only solvable up to a certain maximal value of \( \rho \) there must be two solutions (they might coincide in a borderline case) for eq. (14) with \((\eta_d + 1)e_d\) evaluated at \(E_d = 0\),

\[
\rho_{\min}(1 + \mu_d)_{\rho = \rho_{\min}} = \frac{1}{k_p[(\eta_d + 1)e_d]_{E_d = 0}} = \rho_{\max}(1 + \mu_d)_{\rho = \rho_{\max}},
\]  

implying

\[
\frac{(1 + \mu_d)_{\rho = \rho_{\min}}}{(1 + \mu_d)_{\rho = \rho_{\max}}} = \frac{\rho_{\max}}{\rho_{\min}}.
\]

Additionally there must occur a sign change of eq. (16),

\[
\frac{\rho}{\mu_d} \frac{\partial \mu_d}{\partial \rho} < -(1 + \mu_d)
\]

for \( \rho > \rho_0 \), necessitating a sufficiently fast decay of \( \mu_d \) with growing \( \rho \). In any case, it must be pointed out that \( \rho \) cannot become arbitrarily large. \( L_d \) must stay positive implying\(^4\)

\[
(T - E_d)/k_p - (S_p + \rho S_d) > 0.
\]

\(^4\) As \( \mu_d \to 0 \) for growing \( \rho \) any additional activity is very costly in terms of effort for the DP and therefore will not take place.
Already in the absence of a solution of the first-order condition, i.e., while $E_d^*$ is still zero, the work performed in the two jobs, $S_p + \rho S_d$, can violate this condition. In that case there would be no selection effort for any $\rho$.

In sum, the DP is predicted to exert positive selection effort:

- only above a minimum value of involvement in the private practice - see discussion of eq. (17);

- only if the positivity requirement on leisure is not violated, which can make such an effort entirely unattractive - see eqs. (5) and (20).

- that increases with increasing involvement in the private practice - eq. (16) and

  - keeps growing if the rate of substitution between income and leisure decreases rather slowly with increasing involvement,

  - or starts decreasing and cease altogether at some maximum involvement.

![Graph](image)

**FIG. 1:** $E_d$ as a function of $\rho$ if eq. (14) has solutions for $\rho > \rho_{\text{min}}$. 
In fig. 1, the solid line depicts the case (a), where the solutions stop existing for \( \rho > \rho_{\text{max}} \), which happens if \( \mu_d \) decreases sufficiently fast with growing \( \rho \). The dashed line represents the case (b), where there is no change of sign in \( \partial E^*_d / \partial \rho \) and where there are always solutions to the first order condition. The dotted lines show three possible positions where \( L_d > 0 \) is violated. The rightmost dotted line, does not affect case (a), but in (b) the solution of eq. (14) cannot be maintained to the right of the dotted line. The middle dotted line also affects (a) by cutting it off. The leftmost dotted line is a case with no solution that does not violate \( L_d > 0 \).

Additionally, the first-order condition can also be interpreted in a different way:

For \( \left[ \frac{\partial U_d}{\partial I_d} k_d + \frac{\partial U_d}{\partial E_d}(k_d - k_p) \right]_{E^*_p, E^*_d} > 0 \) there can exist a nontrivial solution

\[
0 < \left\{ \frac{\partial U_d}{\partial I_d} k_d + \frac{\partial U_d}{\partial E_d}(k_d - k_p) \right\}_{E^*_p, E^*_d} = \left( \frac{\partial (e_d E_d)}{\partial E_d} \right)_{E^*_p, E^*_d} < e_d \big|_{E_d=0}
\]

if \( e_d \big|_{E_d=0} \) exceeds a finite threshold for cream skimming. The exact functional form of this threshold, however, depends on \( e_d \) and its derivative at \( E^*_d \). Due to the positivity requirement - eq.(20) - the value of \( \rho \) needed for the DP to become interested in cream skimming may, in fact, become so large that it is above the point, where the DP is willing to exert any kind of additional effort. The result is that the DP will not cream skim for any value of \( \rho \).

C. Conclusions

The present paper focuses on a public healthcare system where public patients are transferred from public to private providers in the aim of shortening waiting lists.
in the public sector. The existing literature claims that under such a policy, dual-job practitioners (DPs), who combine public service with private practice, manipulate patient transfers in a way that only the least severe and thus most profitable of the public patients are treated in private practice. This is said to result in an increase in average severity and costliness of patients retained in the public sector (González, 2005). The analysis presented in this paper shows this not necessarily to be the case. Patient selection costs time and effort that alternatively could be used for treating patients or for leisure activities. Specifically, there exists a finite threshold for the DP’s involvement in private practice below which patient selection is not lucrative. Beyond that threshold, selection effort is predicted to increase with increasing involvement in private practice under realistic conditions - but only to some degree. A critical condition is a sufficiently high effectiveness of DPs’ cream-skimming effort. Moreover, physicians working exclusively in the public sector (SPs) in turn have an incentive to dump high-severity patients on private sector. This is because high-severity patients entail disutility for them due to additional work and an erosion of margins under DRG-based prospective payment.

The analysis suggests that the effects of cream-skimming efforts by the DPs and the dumping efforts by the SPs on average severity of patients remaining in the public sector can cancel each other at least to some extend. While it would go beyond the present paper to quantitatively model the net effect of the two types of effort, a preliminary assessment seems possible. In healthcare systems where
hospital care is provided predominantly in the public sector, there are many more SPs than DPs. This has a twofold effect for the mean severity and hence treatment cost of patients treated in the public sector. First, DPs have comparatively limited leverage in influencing average severity and cost in the public sector. Second, this means that their cream-skimming effort has low effectiveness, making it less likely that DPs ever engage in patient selection, as shown in this paper. This analysis is subject to several limitations. First, neither the DP nor the SP are viewed as having a special professional ethics, which might be intrinsic or induced be reputation effects. Second, the two types of providers do not react to each others actions, although they usually can identify each other, calling for a game-theoretic approach. Still, the present contribution calls attention to the fact that cream skimming and dumping are not a one-way street, as is often surmised in the existing literature.


Ma C-tA, McGuire TG. Optimal health insurance and provider payment. American


