Workplace drug abuse policy

By: Thomas G. McGuire and Christopher J. Ruhm

"Workplace Drug Abuse Policy" (with Thomas McGuire), *Journal of Health Economics*, Vol. 12, No. 1, April 1993, 19-38.

Made available courtesy of Elsevier: http://www.elsevier.com/

***** Note: Figures may be missing from this format of the document**

Abstract:
An estimated 70 percent of illicit drug users are in the workforce. This paper studies workplace policies relating to drug abuse treatment and testing in a labor market with asymmetric information about worker proclivities to abuse drugs and to incur costs of workplace accidents. Drug abuse has a moral hazard component related to worker choice of treatment or other deterrent activities, and a selection component related to drug testing. We characterize the type and frequency of workers treated and tested in labor market equilibrium. Labor market incentives will generally lead to too little treatment and too much testing.

Article:

1. Introduction

Although rates of drug abuse have fallen recently in the U.S., the percentage of the workforce who are drug abusers is actually increasing, as older "drug-naive" cohorts are replaced by younger "drug-exposed" cohorts. Household surveys indicate that 70 percent of users of illicit drugs are in the workforce. According to the Institute of Medicine (1990), the annual productivity loss to U.S. employers from drug abuse exceeds $33 billion. Four-fifths of respondents in a recent national survey of CEOs, human resource executives, governors, and mayors, stated that alcohol and drug abuse constitute "significant" or "very significant" problems at their work-place (Hansen, 1988).

Employer policy towards drug abuse includes several major components. First, private health insurance coverage for drug abuse treatment has become a common feature of group health insurance plans. By 1988, 90 percent of full-time workers had health insurance and 74 percent had plans which included some type of drug abuse coverage (U.S. Bureau of Labor Statistics, 1989). Second, many employers offer in-house counseling services. In a 1989 survey of 2,000 employers, A. Foster Higgins and Co (1990) found that 48% percent of employers had an Employee Assistance Plan, and another 10 percent planned one for the next year. Finally, drug testing, either prior to or during employment, is on the rise. 60 percent of establishments with 5000 or more employees and 43 percent of those with 1000 or more workers had drug testing programs in 1988 (U.S. Bureau of Labor Statistics, 1989). According to a recent Gallup Organization (1988) survey, 14 percent of large companies without a drug testing program said they planned to implement one within the next year and 12 percent of the remainder planned to do so at some future date. "Protection of the company's safe work record and accident reduction" were cited as the main reasons for doing so.

Despite the increased importance of workplace drug policies, there has been little attention devoted to their market-wide impact and to the implications for public policy. Employers are willing to spend money to rehabilitate drug abusers through health care services and in-house pro-grams, and at the same time they are willing to spend money to avoid employing drug users. Are employees receiving enough/too much rehabilitation? Are employers doing enough/too much drug testing? The goal of the paper is to address these questions with a simple model of a competitive labor market featuring the moral hazard and adverse selection issues inherent in workplace policy. Our starting point is the assumption that the proclivity of a worker to abuse drugs and impose costs on the firm is an unobserved characteristic which differs across workers (the adverse selection element) and which may be partly under the worker's control (the moral hazard element). We will refer in the paper to "drug abuse," but we mean our analysis to apply to abuse of alcohol as well.

Throughout the paper we employ a spot labor market which is competitive, implying that risk-neutral workers will be paid their expected marginal products, given the information publicly available. This limitation on the form of labor contracts is discussed in section 2, while section 3 derives the expected marginal products for the model we construct.

We consider worker moral hazard in sections 4 and 5. Moral hazard has been frequently studied in the context of labor markets. The pioneering analysis of Becker and Stigler (1974) postulated that employment bonds might be used to prevent worker malfeasance.[1] More recently, a large literature on the principal-agent problem has investigated the design of contracts in situations where an input of the agent (worker) is unobserved (e.g. see Grossman and Hart, 1983; Shapiro and Stiglitz, 1984; or Ma, 1988). Our analysis focuses on workers' choice of engaging in treatment or deterrent activities which would reduce the cost of drug use.

The "disease" component of drug use limits the role for deterrence and implies the need to consider selection effects. Adverse selection has received considerable attention when evaluating the consequences for sorting behavior of alternative wage contracts. For instance, Weiss (1980) discusses how firms may pay relatively high wages as a way to recruit and retain high-ability workers. Our analysis assumes that firms are restricted in their ability to use wages as a mechanism for sorting across heterogeneous labor. Instead, we introduce adverse selection (in section 6) by permitting employers to use preemployment drug testing to distinguish between workers with a high and low likelihood of a drug-related work-place accident.

Within the confines of the model we analyze, we conclude that labor market incentives will generally lead to too little treatment and to too much testing. The criterion for making this judgment is the maximization of net social product. There can be too little treatment because workers are unable to internalize all the benefits of treatment if, because of asymmetric information, productivity gains must be shared with other workers. We show that this sharing takes place even when treatment/deterrence activities by workers can be observed by firms. There can be too much testing because testing is most attractive (least costly) to the workers who are at the lowest risk for drug abuse. Firms then have incentives to institute drug-testing programs in order to reduce the number of drug abusers they employ. This creates a negative externality for non-testing firms, because it reduces the quality of their applicant pool. As a result, the private value to workers of sending the signal that they are drug-free (by their willingness to work in settings with drug testing) generally exceeds the social value of the job sorting efficiencies that testing can bring.[2]

Many of the important social aspects of drug abuse are obviously not captured in our labor market model. We take a very simple view about what it means to be "addicted," and we ignore costs of drug abuse external to labor contracts. These and other omissions are summarized in a final section where we sketch out some directions for research on drug abuse and labor markets.

2. Worker liability and contingent contracts

One economic solution to the problem of drug abuse would be to make abusers liable for any work (or non-work) costs their actions impose. Penalties would take the form of fines paid to those suffering damages. Individuals would then determine their level and frequency of drug use in such a way that marginal social benefit and marginal costs are equalized.

There are several reasons why a system of strict liability will be unworkable. First, drug use will generally be unobservable, and so direct penalties for the associated costs cannot be imposed. An alternative is to base the penalties on a set of outcomes which are correlated with drug use. For instance, we focus on workplace accidents, which are assumed to be more common among drug users than non-users. In principle, it would be possible to levy substantial fines on persons involved in workplace accidents. If properly calibrated, the penalties could lead workers to equalize the expected marginal costs and marginal benefits of drug use, as in Becker (1968) and Stigler's (1970) analysis of the economics of crime. The resulting outcome may be extremely inequitable however, since persons having accidents, despite abstaining from drugs, will pay (possibly large) fines, while those drug users who are lucky enough to avoid accidents will escape any penalty.

Second, the required system of penalties may be illegal or impracticable. This will be of particular importance when accident probabilities are low but the costs of accidents are extremely large. For instance, the costs of accidents by airline pilots or nuclear power plant operators may be many times their total lifetime earnings. It may be impossible to impose a monetary fine which sufficiently deters these activities (since the worker would simply declare bankruptcy) and even criminal sanctions may be inadequate.

Third, the usual economic calculus of utility maximization may break down when considering addictive activities such as drug use. Even where economic models lead to predictions of "rational" drug use (e.g. Becker and Murphy, 1988), the individual is making a choice which is consistent with ex ante preferences. It is possible, however, that ex ante and ex post preferences differ, depending on whether or not, ex post, the worker becomes a drug user (Glazer and Weiss, 1991). Further as Akerlof (1991) has recently argued, drug addiction may be the end result of a sequence of "near-rational" decisions which individually cause only small losses, but which, when combined, have a large negative impact. Finally, it is not clear to what extent substance use is a choice variable and to what extent it should be considered a possibly time-varying endowment (like health) over which the individual has no control. If the use of drugs is similar to the consumption of other types of goods, individuals will respond rationally to rewards and penalties which are put into place. Conversely, if there is a substantial "disease" component, efforts to deter use or to pass the costs on to the individual abuser will be less successful. Attention may then shift to minimizing the costs of the behavior either through placing individuals in positions where the damages associated with drug use are minimized or by using external interventions (e.g. subsidized treatment) in an attempt to change behaviors.

The inability of firms to hold workers strictly liable for drug-related expenses implies that companies may attempt to use contractual arrangements whereby the actions of employees reveal their type. For example, the firms may write multiperiod contracts in which individuals post bonds (in the form of reduced wages) which are returned after evidence of satisfactory work performance. Drug users would refuse these contracts, since the risk of forfeiting the bond is substantial, while the expected compensation of nondrug users is maximized by accepting such a long-term employment arrangement. Alternatively, the firm could offer a variety of single-period contracts where pay is contingent on accident outcomes. Sorting would occur because drug-free individuals would be willing to accept larger wage reductions, in the event of an accident, than would drug users.

Firms are actively involved in costly drug-testing and frequently provide treatment benefits whose costs are paid for by all employees (e.g. in the form of higher insurance premiums). This provides evidence that employers are unable to write contracts which completely sort their workforces, and is consistent with the widely acknowledged limits to bonding as a solution to these types of problems (e.g. see Akerlof and Katz, 1989; and Dickens et al., 1989). Concerns over fairness may further limit the ability of firms to pay different wages to workers with identical observed characteristics, even if the individuals would be willing to precommit to different contingent pay contracts.[3] Finally, it may be administratively expensive to provide a sufficiently wide range of contingent contracts, especially when there are multiple worker (and possibly accident) types. In what follows, firms are assumed to be unable to require workers to post a "performance bond" in any form.

3. Paying expected marginal product

We assume there are two types of workers, α's and β's. The α workers do not abuse drugs, but nonetheless have the probability α of causing a work-related accident during a period for non-drug related reasons. The β's are drug abusers (in the absence of treatment or deterrent activities); their probability of causing a work-related accident is β, where $\beta > \alpha$.[4] The β workers make up a small share λ of the total workforce.

Aside from accidents, all workers produce a marginal product valued at m per period, which is assumed to be constant. When a work-related accident occurs, the firm incurs a cost of c. Expected productivities are thus $m-\alpha c$ for the α workers and $m-\beta c$ for the β workers.

A worker may be able to affect the probability of having a workplace accident caused by drug abuse. He or she can seek professional help for a drug abuse problem and more generally may be able to engage in preventive behavior which reduces the likelihood of drug abuse. For simplicity, we refer to any costly activity undertaken to reduce the probability of a drug-related accident as "treatment." In our model, the α's are already drug-free, and while we allow α's to choose treatment, doing so does not reduce their probability of an accident. We assume that if a β worker is treated, his or her accident probability falls to α. The assumption that treatment is completely effective is made for simplicity. We show below that our qualitative conclusions hold so long as treatment is effective to any degree. The shares of all α and β workers choosing treatment are defined to be γ_α and γ_β respectively.

To keep the analysis within a static framework, we assume events occur in the following sequence: (1) Firms announce a one-period wage contract which pays the worker his or her expected marginal product (EM P) contingent on information available to the firms at the start of the period. This includes information on accident histories and (possibly) on treatment undertaken by the worker and results of drug tests. (2) Workers choose whether to receive (and pay for) treatment and then generate a prior work history of one period.[5] (3) Firms pay workers in the period under study according to the prespecified wage contract.[6]

Consider now the wages firms will offer in this model. Firms are risk neutral. EMP is equal to productivity m minus the worker's expected probability of an accident times the cost of an accident c. Firms know work history from a prior period and they may know if the worker has taken treatment, depending on whether treatment is assumed to be observable or not observable to firms. In equilibrium, firms' conjectures about the shares of workers taking treatment, γ_α and γ_β, are self-fulfilling.

If treatment is unobserved by firms, a worker's EMP can depend only on whether there was an accident in the prior period. Let $W_{U,A}$ and $W_{U,N}$ be the wages (or EMPs) paid to workers when treatment is unobserved, conditional on an accident and no accident in the prior period respectively. Appendix B summarizes the definitions used in the model so far, and the calculations necessary for $W_{U,A}$ and $W_{U,N}$, based on formulae for conditional probabilities. Notice that the share of α workers getting treatment does not affect the portion of workers with a β risk of accident, λ_U, or the EMPs when treatment is unobserved (because α workers are already drug free). The share of β workers getting treatment does affect the EMP of the pool of workers and so does have an effect on wages.

When treatment is observed by the firm, the probability of an accident for any worker getting treatment is known to be α. Conditional on γ_α and γ_β, the portion of all untreated workers with a β risk of accident is λ_N, defined in Appendix B. When treatment can be observed, the EMP of a worker known not to have treatment will depend on whether or not there was an accident in the previous period, and be equal to $W_{N,A}$ if so and $W_{N,N}$ if not. For treated workers, accident history is irrelevant (since the worker is known to have an α risk), and EMP is simply $W_{T,*}$. With this wage structure, we can .now characterize the labor market equilibria in the case of treatment unobserved and treatment observed by firms.

4. Treatment unobserved by firms

The labor market is in equilibrium if workers obtaining treatment cannot increase expected wages (utility) if they did not receive treatment, while workers who are untreated cannot increase expected wages by obtaining treatment. The equilibrium levels of treatment depend on whether or not firms observe treatment. If treatment cannot be observed by firms, wage outcomes will only depend on accident histories. In this case, the incentive for β's to take treatment is the lower risk of a work-place accident. α's do not have this incentive, since their risk of accident is not lowered by treatment.

Letting treatment costs equal υ for both types of workers, then for any a worker choosing treatment, the following condition must hold:

$$\alpha W_{U,A} + (1\text{-}\alpha) W_{U,N} \text{-} \upsilon \geq \alpha W_{U,A} + (1\text{-}\alpha) W_{U,N}. \qquad (1)$$

This will obviously never be satisfied for $\upsilon > 0$, since α workers do not benefit from treatment which is unobserved by the firm. Therefore, in any equilibrium, no α's get treatment, and $\gamma_\alpha = 0$.

The condition for a β worker to choose treatment is:

$$\alpha W_{U,A} + (1\text{-}\alpha) W_{U,N} \text{-} \upsilon \geq \beta W_{U,A} + (1\text{-}\beta) W_{U,N}. \qquad (2)$$

The weak inequality in (2) is reversed for any β not taking treatment. Rewriting (2), for any β in treatment,

$$(\beta\text{-}\alpha)(W_{U,N}\text{-}W_{U,A}) \geq \upsilon,$$

$$(\beta\text{-}\alpha)[\phi_U\alpha + (1\text{-}\phi_U)\beta\text{-}\Phi_U\alpha\text{-}(1\text{-}\Phi_U)\beta]c \geq \upsilon,$$

$$(\Phi_U\text{-}\phi_U)c(\beta\text{-}\alpha)^2 \geq \upsilon, \qquad (3)$$

whereas for any untreated β,

$$(\Phi_U\text{-}\phi_U)c(\beta\text{-}\alpha)^2 \leq \upsilon.^{7} \qquad (4)$$

Since the left-hand side of conditions (3) and (4) varies with the share of β workers who decide to take treatment, γ_β serves a potentially equilibrating function. Referring to Appendix B, $\Phi_U - \phi_U$ represents the difference in the firm's estimate of the probability that an individual is an α, conditional on whether or not he or she had an accident in the prior period. This magnitude determines the penalty a worker bears from having an accident, and is affected by the share of β's taking treatment. When few β's take treatment $\Phi_U - \phi_U$ is large, and the worker bears a heavy penalty for an accident. This is why when costs of treatment υ are low enough, some β's will take treatment. When almost all β's have taken treatment, however, $\Phi_U - \phi_U$ is very small, and the worker bears a small penalty for an accident. This is why not all β's will take treatment. Intuitively, as the pool improves by β's taking treatment, the informational content of accident histories diminishes. (See Appendix A for proof.) In the extreme case, where γ_β approaches one, almost all workers with accidents will be unlucky α's, and the probability of an accident being caused by an untreated β is small.

With treatment unobserved by firms, there are two possible equilibria. If treatment costs are sufficiently high, no workers will be treated. Conversely, with lower treatment costs, some but not all β's receive treatment. With positive treatment costs, there will never be an equilibrium where all β workers seek treatment. α workers never find it worthwhile to get treatment when treatment is unobserved. These equilibria are described formally in Appendix A.

We can now consider whether when treatment is unobserved by firms, there is too little or too much treatment (or worker deterrent activity). Treatment is socially efficient if the gains, in terms of reduced accident expenses, are greater than or equal to treatment costs. Since there is no reduction in accident probabilities for the α's, treatment is never efficient for them. Treatment reduces accident probabilities for β workers by β-*α*. Treatment benefits exceed treatment costs, and treatment is efficient for β workers if:

$$(\beta\text{-}\alpha)c \geq \upsilon. \qquad (5)$$

If the inequality is reversed, treatment is not socially efficient.

"Overtreatment" occurs when any α individuals obtain treatment or if β workers do so when (5) does not hold. "Undertreatment" occurs when (5) holds and some β workers are untreated. Since both labor market equilibria imply $\gamma_\beta < 1$, there is undertreatment whenever (5) holds. The reason for undertreatment is easy to understand. β workers can free-ride on the less accident-prone α workers as long as firms imperfectly observe worker type. If any β's do take treatment, the private gains to other β's from treatment are reduced as the productivity of the average worker is raised.

Can there ever be overtreatment? α's never get treatment. Therefore treatment will only be excessive if some β's seek treatment when (5) does not hold. Rewriting (3), β workers choose treatment if:

$$[\Phi_U - \phi_U](\beta-\alpha)(\beta-\alpha)c \geq v, \tag{6}$$

Since $\Phi_U - \phi_U$ and β-α are both less than one, the left-hand-side of (5) is always greater than that of (6). This guarantees that β workers never seek treatment which is socially undesirable. Overtreatment will not occur in equilibrium. We can thus conclude that when treatment is unobserved by firms, there is likely to be too little treatment.[8]

5. Treatment observed by firms

When employers can observe whether a worker has been treated, treatment acts to signal worker type, in addition to reducing the accident probabilities of β workers. This raises the possibility that α as well as β workers will seek treatment. The equilibrium condition for any a worker taking treatment is:

$$W_{T,*} - v \geq \alpha W_{N,A} + (1-\alpha) W_{N,N}. \tag{7}$$

The inequality is reversed for any α not taking treatment.

Similarly, for any β in treatment,

$$W_{T,*} - v \geq \beta W_{N,A} + (1-\beta) W_{N,N}. \tag{8}$$

The inequality is reversed for any β declining treatment. Substituting in the expressions for the relevant EMPs from Appendix B, a β will take treatment if:

$$[1 + \beta(\Phi_N - \phi_N) - \Phi_N]c(\beta-\alpha) \geq v. \tag{9}$$

An α will do so if:

$$[1 + \alpha(\Phi_N - \phi_N) - \Phi_N]c(\beta-\alpha) \geq \mathrm{v}. \tag{10}$$

Since β > α, the left-hand side of (9) always strictly exceeds the left-hand side of (10), which implies that β's benefit more from treatment than α's. Intuitively, β's get the signal value and the productivity effect; α's only get the signal value.[9]

If treatment costs υ are sufficiently high, neither (9) nor (10) will be satisfied, and no worker will seek treatment. This is one possible equilibrium. A more interesting situation occurs when treatment costs are low enough that at $\gamma_\alpha = 0$ and $\gamma_\beta = 0$, (9) is satisfied, and a β finds it worthwhile to buy treatment. What are the possible equilibria in this case? Recall that the expected wage increase from treatment is always strictly larger for a β than an α, since the β benefits by having a lower accident probability as well as from the signal effect enjoyed by the α. In other words, if any α finds treatment worthwhile, all β's will also find it worthwhile. Thus, no equilibrium with some α's getting treatment is possible unless all the β's also obtain treatment. But if all the β's get treatment, there is no signalling value to treatment (since all the untreated workers are α's), and no α would get treatment. By this reasoning, an equilibrium cannot exist when just some of the α's get treatment.

One possible equilibrium with low treatment costs is for all workers to receive treatment, $\gamma_\alpha = 1$ and $\gamma_\beta = 1$. In this case, since there are no untreated workers, it is firms' beliefs about the productivity of the first untreated worker that determine the wage a worker would be paid without treatment and these beliefs could support an equilibrium with all workers treated. For example, $\gamma_\alpha = \gamma_\beta = 1$ would be an equilibrium if firms believe that the first untreated worker is a β. However, if a single α were not treated, and this were recognized by firms, the equilibrium would not be reattained. Further-more, there is no dynamic process consistent with our model that leads to this equilibrium. For the remainder of the paper, we will therefore focus on equilibria which are attainable through a dynamic adjustment process.

Another equilibrium exists with low treatment costs when some β's and no α's get treatment. In particular, define $\gamma_\beta{}^*$ as the value of γ_β which makes (9) an equality when $\gamma_\alpha = 0$. If (9) holds with equality, both the treated and the untreated β's are satisfied. Because the benefit of treatment is strictly less for the α's than for the β's, (10) must not hold, implying that the α's are happy not getting treatment. Appendix A proves $\gamma_\alpha = 0$, $\gamma_\beta = \gamma_\beta{}^*$ is an equilibrium when treatment costs are sufficiently low. The two attainable equilibria, with treatment observed, thus parallel the equilibria in the treatment unobserved case, despite the fact that treatment has a signalling value to α as well as β workers.

When treatment is observed, there is again likely to be too little treatment. At $\gamma_\beta{}^* < 1$, too few β*'s* will be treated when treatment is efficient. Except for the unattainable equilibria mentioned above, α's are never treated in equilibrium, and there will be no overtreatment of α's. Rewriting (9), a β worker obtains treatment when:

$$[1-(1-\beta)\Phi_N-\beta\phi_N](\beta-\alpha)c \geq v, \tag{11}$$

Since $1 - (1-\beta)\Phi_N - \beta\phi_N$ is less than one, the left-hand side of (5) is greater than that of (11), implying that β's will never be overtreated. Thus, undertreatment is again the general case.[10] As above, after a certain number of β's are treated, it becomes attractive for the remaining β's to free ride on the pool of untreated α's, rather than to obtain treatment.[11]

To summarize, if employers cannot observe underlying worker type, the market equilibria will generally be characterized by too little worker treatment or deterrent activities. This is true whether or not firms can observe worker treatment activities. We want to emphasize that the β's free-riding behavior constitutes an externality. No firm in our model will subsidize a worker's treatment to restore socially efficient incentives. Employers are observed subsidizing some of workers' treatment costs in the form of employer-provided health insurance, presumably because of the private benefits accruing to the firm from such a policy. Since the social benefits of treatment exceed the private benefits enjoyed by the worker and employer, there is the basis for a case for public subsidy of treatment costs or regulations requiring minimum standards for employer-based insurance for drug-abuse services.

We next consider the labor market when firms can take an action to identify worker type.

6. Drug testing

Testing for drug use among current employees and job applicants is an increasingly important and controversial policy adopted by firms to limit drug-related employment costs. Drug testing has been questioned on several grounds. It invades workers' privacy and imposes economic penalties on behavior only stochastically related to work performance. Furthermore, depending on the drug and the testing procedures, drug testing is not completely accurate or reliable.

We consider a simple model of drug testing in which firms can perfectly distinguish between α and β workers, at drug test cost per worker of *t*. If a firm tests, it pays workers their individual expected marginal product less *t*. If a firm does not test, it pays the EMP of the pool of workers not seeking employment from firms with testing policies.[12] To illustrate the potential inefficiency associated with testing, we begin with a brief consideration of the case in which all firms have the same cost of a workplace accident, *c*. Any testing which emerges in such a model must be socially inefficient since testing simply redistributes the β workers. Without a detailed consideration of equilibria, we show, however, that testing can take place in such a model.

The major purpose of this section is to consider the more interesting model where there is a second type of employer for whom drug use has no effect on the risk of a workplace accident. With two types of firms, testing can increase net social product by channeling the β workers to the "risk-free" jobs. We describe equilibria in this model and compare conditions for competitive equilibrium with conditions for efficiency.

To focus on the selection effect of testing, we assume that workers' type is fixed. We ignore treatment and assume testing has no deterrent effect on drug abuse. It is sufficient in this circumstance to consider a model in which the wage setting and drug testing choices are made simultaneously by workers and firms. Since the forms of equilibria with testing are similar to those with treatment, we present our reasoning about equilibria intuitively, and omit formal proofs.

Non-testing firms pay EMP which depends on the portion of nontested workers λ_{NT} who have a β risk of accident. λ_{NT} can be defined as:

$$\lambda_{\mathrm{NT}}=\frac{\lambda(1-\eta_\beta)}{\lambda(1-\eta_\beta)+(1-\lambda)(1-\eta_\alpha)},$$

with η_α and η_β the shares of α and β workers, respectively, being tested for drugs. The EMP per worker in a nontesting firm is:

$$W_{\mathrm{NT}}=m-[(1-\lambda_{\mathrm{NT}})\alpha+\lambda_{\mathrm{NT}}\beta]c=\mathrm{m}-\beta c+(\beta-\alpha)(1-\lambda_{\mathrm{NT}})c.$$

Drug users will never choose to work for a testing firm, since the wage they receive *(m-βc-t)* must be less than W_{NT}. An α worker, however, takes a job in a firm with a testing policy if:

$$m-\alpha c-t\geq W_{\mathrm{NT}}. \qquad (12)$$

Since no β workers choose to be tested *($\eta_\beta = 0$)*, λ_{NT} equals one *(λ)* when η_α equals one (zero), and $dW_{NT}/d\eta_\alpha <$ 0.

α workers can gain from testing by being certified as low-risk workers. The key feature of this model is that as more α's are tested, the value of testing to the remaining α's increases. This occurs because as α's are pulled out of the pool of untested workers (and certified as α's), the average productivity in the pool falls, and so the wage differential between untested workers and certified drug-free individuals increases.

If test costs are low enough *($t \leq \lambda(\beta-\alpha)c$)* so that it pays the first α to be tested, all α's will seek testing and the equilibrium is $\eta_\alpha = 1$. Testing is obviously socially inefficient since testing occurs at a cost, but leads to no increase in expected output.

Next consider the more interesting situation where there is an alternative form of employment to which the β workers are better suited. To do so assume there is another type of firm in which production takes place with no risk of accident — the marginal product of both α and β workers is constant at m' in this "riskless" production. Assume that β workers are efficiently employed in the riskless industry, but when pooled with the α workers, would be paid more in the risky industry, which implies that

$$m-\beta c<m'<m-[(1-\lambda)\alpha+\lambda\beta]c. \qquad (13)$$

Further assume that α workers are efficiently employed in the risky industry, even if they are required to pay for drug testing, which occurs if

$$m' < m\text{-}\alpha c\text{-}t. \tag{14}$$

(14) implies that α workers will never work in the risk-free industry and (13) indicates that β workers will never seek testing.

The fraction of β's in the pool of untested workers employed by the risky firm is:

$$\lambda_{\mathbf{NT}} = \frac{\lambda(1\text{-}\rho_\beta)}{\lambda(1\text{-}\rho_\beta)+(1\text{-}\lambda)(1\text{-}\eta_\alpha)},$$

where η_α is the share of α workers tested for drugs, and p_β is the share of β workers in the risk-free industry. (Recall that α workers never go to the risk-free industry, and β workers are never tested.) Wages paid to the pool of untested workers in the risky industry are:

$$W_{\mathrm{NT}} = m\text{-}[(1\text{-}\lambda_{\mathrm{NT}})\alpha + \lambda_{\mathrm{NT}}\beta]c.$$

Workers at the risk-free firms earn m', while drug-tested α's receive *m-αc-t.*

The equilibrium condition for β workers describes their choice to work for the risk-free or risky firms. A β will work for a risk-free firm if:

$$m' \geq W_{\mathrm{NT}}; \tag{15}$$

and choose employment in a risky firm if the inequality is reversed.

The equilibrium for α workers describes their choice of whether or not to be drug tested by risky firms that offer tests for drugs. *α* workers will be drug tested if:

$$m\text{-}\alpha c\text{-}t \geq W_{\mathrm{NT}}; \tag{16}$$

and will not be tested if the inequality is reversed.

There are two potential equilibrating forces in this market, the share of β workers going to the risk-free industry, p_β, and the share of α workers being tested for drugs, η_α. As p_β rises, β workers leave the pool of untested workers in the risky industry, the average productivity of the pool rises, and it becomes less attractive for other β's to leave. Conversely, as α's leave the pool by being tested, the average productivity of the pool falls, and it becomes more attractive for α's to seek testing.

When testing costs are high (specifically, when the pooled wage W_{NT} with no α's tested exceeds the wage an α worker can get by testing, *m-αc-t),* no α's will seek to be tested, and all β's will work in the risky sector (because of assumption (14) above). Thus, the high test cost equilibrium is $\eta_\alpha = 0$, $p_\beta = 0$. When testing costs are low (when the pooled wage W_{NT} with all α's tested is below the wage of a confirmed α, *(m-αc-t)),* there is an equilibrium with all α's tested and all β's working in the risk-free sector. With no workers in the pool of untested workers in the risky sector, it is employers' beliefs that set W_{NT}. Any belief that the productivity of an untested worker is less than *m-αc-t* will support the equilibrium. Thus, the low test cost equilibrium is $\eta_\alpha = 1$, $p_\beta = 1$.

An intermediate case occurs where testing costs are high enough so that if no α's are tested, none want to be tested, but at the same time low enough so that if all α's are tested, all α*'s* want to be tested (this occurs when $\lambda(\beta\text{-}\alpha)c \leq t \leq (\beta\text{-}\alpha)c$). If test costs are within this range, there is also a third equilibrium where just some of the α's are tested. This occurs when enough α's get tested that (16) holds with equality. β workers prefer employment in the risky industry whenever $\eta_\alpha < 1$, (since $m'<m\text{-}\alpha c\text{-}t=W_{NT}$); therefore, this third equilibrium is described by $\eta_\alpha < 1$, $p_\beta = 0$. Note that this final equilibrium is unattainable within our model. If one additional α worker were tested, the value of testing increases to the remaining α's and a new equilibrium would occur with all α's being tested. Similarly, if one fewer α obtained a drug test, the value of testing would decline for all α's and the equilibrium would occur with no drug testing.

To summarize, the attainable equilibria in this model are with all or none of the α workers getting tested, and all or none of the β workers switching into riskless employment. The efficiency of testing can be considered by comparing total products with and without testing, for the equilibrium in which testing occurs. Thus we compare the total product when all α's are tested and all β's work in the risk-free sector, to the equilibrium with no testing, i.e., $\eta_\alpha = p_\beta = 1$ *vs.* $\eta_\alpha = p_\beta = 0$.

The requirement for testing to increase social product is:

$$\lambda m' + (1-\lambda)(m-\alpha c-t) > m-[(1-\lambda)\alpha + \lambda\beta]c; \qquad (17)$$

the average product net of testing costs of workers sorted between the two sectors by testing must exceed the average product of all workers pooled in the risky sector. Rewriting (17), testing increases net social product if the costs of testing are less than the gain in productivity for the β's, weighted by the proportion of α and β workers in the economy:

$$t < \frac{\lambda}{1-\lambda}[m'-(m-\beta c)]. \qquad (18)$$

One implication of (18) is that if the share of β workers is very low (the "few bad apples" model), then λ will be small, and testing will tend to be socially inefficient.

While testing may increase social product in equilibrium, it may also reduce it. We have assumed that α's have a higher productivity in the risky industry, even after paying for testing $(m' < m-\alpha c-t)$. Noting that the right-hand side of (17) is W_{NT} when $\eta_\alpha = 0$ and $p_\beta = 0$, it is straightforward to show that the private net gain to an α from testing $[(m-\alpha c-t)-W_{NT}]$ always exceeds the social net gain. Intuitively, the α workers see the wage increase they receive, but this is not a social gain. The social gain is the β worker shifted to a more productive employment. In general, there cannot be undertesting (α's will always seek testing if it is efficient), but there can be overtesting. The α's may seek testing when the social condition for efficiency is not met.

7. Conclusions

This paper has considered the effects of workplace policies relating to drug abuse treatment and testing in a labor market with asymmetric information about worker proclivities to abuse drugs and incur costs of workplace accidents. Two important observations about these policies are highlighted. First, when worker type cannot be observed, treatment or deterrent activities undertaken by a worker, creates a *positive externality* for other members of the workforce by raising average productivity levels. Since these benefits are not fully received by the person paying for them, there will generally be undertreatment in the labor market equilibrium. Second, asymmetric information may create a *negative externality* from drug testing, as workers engage in costly testing to signal to firms that they abstain from drugs. To the extent these costly signals are acquired without any corresponding increase in total productivity, overtesting may result.

Our approach omits consideration of social costs and benefits outside the labor market, and our conclusions about too little treatment and too much testing must be regarded with this in mind. Presumably, consideration of the wider effects of drug abuse would reinforce the conclusion that there was too little treatment, so it seems quite safe to say that functioning of labor markets adds to the case for public subsidies for drug treatment facilities and for mandated coverage of drug treatment under private health insurance. It seems unlikely to us that any reasonable modification of the theoretical approach taken here will affect this conclusion materially. Our conclusion that there is excessive drug testing is undoubtedly more susceptible to revision once other factors are brought into the picture.

A number of potentially important theoretical extensions of our work deserve consideration. These include: (1) recognizing the potential deterrent effect of testing and selection effect of treatment; (2) considering labor market equilibria when treatment and testing take place simultaneously; (3) expanding the admissible form of wage contracts; (4) including more realistic descriptions of testing and treatment (e.g., testing with error, and treatment effects which vary with observed and unobserved worker characteristics); and, (5) considering multi-period models where workers' drug use is integrated into a utility framework. Potential empirical work corresponds to all of these topics.

Appendix A

This appendix first proves that the informational value of an accident $\Phi_U - \phi_U$ falls with an increase in the share of β's being treated, γ_β, when λ is small. Next it fully describes the equilibria discussed in sections 3 and 4 of the text.

The informational value of an accident

When the share of β's in the original pool, λ, is small, the information content of a worker accident is inversely related to the proportion of β's who are treated. Conversely, if λ is near 1, more β's being treated can increase the informational value of an accident. We focus on the case where drug abuse occurs among a few "bad apples" in a workforce, so λ will be regarded as small.

Since $d\lambda_U/d\lambda_\beta < 0$, $d[\Phi_U - \phi_U]d\gamma_\beta < 0$ if $d[\Phi_U - \phi_U]/d\lambda_U > 0$. For ease of notation, drop the U subscript, let $X = 1-\lambda$, and $Z = \alpha(1-\lambda) + \beta\lambda$.

We can now write:

$$\Phi - \phi = (1-\alpha)X/(1-Z) - \alpha X/Z$$

$$= X(Z-\alpha)/(1-Z)Z$$

Taking the derivative:

$$d(\Phi-\phi)/d\lambda = \frac{(1-Z)Z[X'(Z-\alpha)+XZ']-X(Z-\alpha)[Z'-2ZZ']}{(1-Z)^2Z^2}. \tag{A.1}$$

(A.1) is positive when the numerator is positive. Noting that $X' = -1$ and $Z' = \beta - \alpha$, this will be the case when:

$$(1-\lambda)(\beta-\alpha)[Z(1-Z)-(Z-\alpha)(1-2Z)] > (Z-\alpha)(1-Z)Z. \tag{A.2}$$

The left-hand side of (A.2) is always positive, since $Z > Z-\alpha$ and $1-Z > 1-2Z$. The right-hand side is also positive but approaches zero as λ becomes small (since Z approaches α as λ approaches 0). Thus, for sufficiently small λ, (A.2) must be satisfied and $d[\Phi_U - \phi_U]/d\gamma_\beta < 0$. As λ approaches 1, Z approaches β. *It* is straightforward to show that, for sufficiently large λ, (A.2) will not be satisfied and the derivative reverses sign.

Treatment unobserved

Case 1: When $v \geq c(\beta-\alpha)^2 Q$, for $Q = [(1-\lambda)(\beta-\alpha)\lambda]/[(1-\alpha)(1-\lambda)+(1-\beta)\lambda][\alpha(1-\lambda)+\beta\lambda]$, equilibrium is unique at $\gamma_\beta = 0$.

Proof: At $\alpha_\beta = 0$, $\Phi_U - \phi_U = Q$, which implies $(\Phi_U - \phi_U)c(\beta-\alpha)^2 \leq \upsilon$ at $\alpha_\beta = 0$. Therefore, the inequality (4) holds at $\gamma_\beta = 0$, in which case no β gains from obtaining treatment. Therefore, $\gamma_\beta = 0$ is an equilibrium. Since $d[\Phi_U - \phi_U]/d\lambda_\beta < 0$ (with relatively few "bad apples"), (4) must also hold for any $\gamma_\beta > 0$. Thus any $\gamma_\beta > 0$ cannot be an equilibrium since no β would seek treatment. The equilibrium at $\gamma_\beta = 0$ is therefore unique.

Case 2: When $v<c(\beta\text{-}\alpha)<Q$, equilibrium occurs at γ_β* where $(\Phi_U - \phi_U)c(\beta\text{-}\alpha)^2 = \upsilon$, $1 > \gamma_\beta^* > 0$.

Proof: γ_β* is an equilibrium because for any β in treatment, (3) holds with equality, while for untreated β, (4) also holds with equality. Since $d[\Phi_U - \phi_U]/d\gamma_\beta < 0$, for any $\gamma_\beta > \gamma_\beta^*$, (3) is violated, and so γ_β cannot be an equilibrium. Similarly, for any $\gamma_\beta < \gamma_\beta^*$, (4) is violated and γ_β is not an equilibrium. Thus, the equilibrium is unique.

Treatment observed

Case 1: When $v \geq c(\beta\text{-}\alpha)(X + \beta Q)$, where $X = (1\text{-}\beta)\lambda/[(1\text{-}\alpha)(1\text{-}\lambda)+(1\text{-}\beta)\lambda]$, and Q defined as above, equilibrium occurs with $\gamma_\alpha = \gamma_\beta = 0$.

Proof: At $\gamma_\alpha = \gamma_\beta = 0$, $[1+ \beta(\Phi_N - \phi_N)\text{-}(\Phi_N]c(\beta\text{-}\alpha)< \upsilon$, (9) therefore does not hold, so no β's will take treatment. Since the left-hand side (LHS) of (9) exceeds the LHS of (10), no α will seek treatment. The LHS of (9) and (10) fall with γ_β since $d\Phi_N/d\gamma_\beta > 0$. Thus, there is no value of for which any γ_β would seek treatment. Since the LHS of (9) exceeds the LHS of (10), no value of $\gamma_\alpha > 0$ can be an equilibrium, since if any α's seek treatment, all β's must also seek treatment.

Case 2: When $v < c(\beta\text{-}\alpha)(X + \beta Q)$ one labor market equilibrium occurs at $\gamma_\alpha = 0$ and $\gamma_\beta^* > 0$ such that (9) is fulfilled with equality. Depending on the structure of firm beliefs, this equilibrium need not be unique. A second equilibrium may occur at $\gamma_\alpha = \gamma_\beta = 1$.

Proof: $[1+ \beta(\Phi_N - \phi_N)\text{-}\Phi_N]c(\beta\text{-}\alpha) \leq \upsilon$ at $\gamma_\alpha = \gamma_\beta = 0$. Therefore some β's have incentives to obtain treatment. γ_β* exists because the LHS of (9) falls with $\gamma_\beta - \gamma_\beta^*$ is an equilibrium because for any β in treatment, (3) holds with equality, while for untreated β, (4) also holds with equality. Since the LHS of (9) exceeds the LHS of (10), no α will seek treatment.

Also note that at $\gamma_\alpha = \gamma_\beta = 1$, λ_N is undefined. Let λ* indicate the employer's beliefs about λ_N, when $\gamma_\alpha = \gamma_\beta = 1$. Firms can plausibly believe that λ* takes any value in the range $0 \leq \lambda^* \leq \lambda$. For any $\lambda^* > 0$, $\gamma_\alpha = \gamma_\beta = 1$ will be an equilibrium with sufficiently low treatment costs. For example, assume $\lambda^* = \lambda$. β workers then lose by moving from treatment to no treatment if $v < c(\beta\text{-}\alpha)(X + \beta Q)$. Similarly, α workers lose by the change if $v<c(\beta\text{-}\alpha)(X+\alpha Q)$. Thus for $v<c(\beta\text{-}\alpha)(X+\alpha Q)$ and $\gamma^* = \gamma$, $\gamma_\alpha = \gamma_\beta = 1$ is an equilibrium. The proof is analogous for any λ* which fulfills $0 < \lambda^* < \lambda$.

Appendix B

Definitions

$m =$ productivity of all workers if no accident
$\alpha =$ probability of an α worker having an accident
$\beta =$ probability of a β worker having an accident
$c =$ cost of an accident to the firm
$\lambda =$ initial share of β workers among all workers
$\gamma_\alpha =$ share of α workers getting treatment
$\gamma_\beta =$ share of β workers getting treatment
$\lambda_U =$ share of β-risk workers in the total workforce after γ_β share of β's get unobserved treatment
$\lambda_N =$ share of β-risk workers among those with no treatment after γ_α and γ_β share of α's and β's respectively get observed treatment
$\phi =$ probability a worker is an α given an accident occurred in the prior period
$\Phi =$ probability a worker is an α given no accident in the prior period

Derived variables

$$\lambda_U = \lambda(1-\gamma_\beta)$$

$$\lambda_N = \frac{\lambda(1-\gamma_\beta)}{(1-\gamma_\beta)+(1-\lambda)(1-\gamma_\alpha)}$$

$$\phi = p(\alpha/A) = \frac{\alpha(1-\lambda)}{\alpha(1-\lambda)+\beta\lambda}$$

$$\Phi = p(\alpha/N) = \frac{(1-\alpha)(1-\lambda)}{(1-\alpha)(1-\lambda)+(1-\beta)\lambda}$$

In the case of treatment unobserved we have ϕ_U and Φ_U, replacing λ by λ_U in the above conditional probability formulae. When treatment is observed, we have ϕ_N and Φ_N, replacing λ by λ_N.

Conditional expected marginal products

$W_{U,N} = m-[\Phi_U\alpha+(1-\Phi_U)\beta]c$ treatment unobserved, no accident
$W_{U,A} = m-[\phi_U\alpha+(1-\phi_U)\beta]c$ treatment unobserved, accident
$W_{N,N} = m-[\Phi_N\alpha+(1-\Phi_N)\beta]c$ treatment observed and declined, no accident

$W_{N,A} = m-[\phi_N\alpha+(1-\phi_N)\beta]c$ treatment observed and declined, accident
$W_{T,*} = m-\alpha c$ treatment observed and taken

Notes:

1 For extensions, see Lazear (1979, 1981).

2 An analogous example of a positive externality is Levine's (1991) investigation of the potential benefit of mandating just-cause employment security regulations. In his analysis, firms switching from at-will to just-cause policies will disproportionately attract persons who do not work hard but whose shirking is difficult to detect. Since the firms will not take into account the benefits to other employers (of an improved applicant

pool), the change may not be privately beneficial, even when there are net social gains obtained by their doing so. Social welfare may then be maximized if all firms are required to adopt just-cause policies.

3 See Akerlof and Yellen (1990) or Frank (1984) for discussions of wage equity and productive efficiency.
4 Drug-abusing employees are more likely to cause work accidents or lower productivity, to make claims for disability or health insurance for medical treatment not related to drugs, and to have higher rates of absenteeism. All of these costs are modeled here as a work-related accident imposing a cost on the firm. See McGuire, Ruhm and Shatkin (1991) for a review and discussion.

5 Workers pay for treatment, but the employer may pay for part of the cost of treatment through higher wages. In this way employers may subsidize treatment as they do with employer-based health insurance. Employer-based health insurance does not appear in our model because there is no tax subsidy. In our analysis we will emphasize that there are external benefits to treatment for drug abuse not enjoyed by either the firm or the worker. The firm has no incentive to consider these in its wage offer. Our conclusion that there is "too little" subsidy for drug treatment by employers could be offset by other factors causing workers and employers to settle on a fringe benefit package with "too much" health insurance. One potential cause is that employer-paid health insurance premiums are not taxable as income to the worker.

6 This sequence ignores the worker's payment in the prior period, but this is inconsequential for the model. If treatment is unobserved, all workers must be paid the same wage in the prior period. If treatment is observed, workers taking treatment are paid more in the prior period. We have solved our model in this case and the results are qualitatively identical to the "one-period" model discussed here.

7 Partially effective (rather than completely effective) treatment can be considered by assuming that the accident probabilities of drug users entering treatment programs falls to α with probability p and remains at β with probability $(1-p)$. The expected accident probability of a treated β worker is $\delta = p\alpha + (1-p)\beta$, with $\alpha \le \delta \le \beta$, and the reduction in accident probability is $(\delta-\alpha)$ rather than $(\beta-\alpha)$. Expressions for Φ_U and ϕ_U become more complicated but the difference $\Phi_U - \phi_U$ continues to be positive.

8 The conclusion holds with partially effective treatment. This can be seen by substituting $(\delta-\alpha)$ for $(\beta-\alpha)$ in equations (5) and (6). δ is defined in footnote 7 above.

9 β's also benefit more from partially effective treatment than do α's, so similar results are obtained in this case.

10 With partially effective treatment, $(\delta-\alpha)$ is substituted for $(\beta-\alpha)$ in equations (5) and (11). Expressions for Φ_N and ϕ_N change but undertreatment is again the result.

11 It is easily shown a β worker receives more rewards for treatment when the latter is observed than when it is not. Expected compensation is therefore equalized (with and without treatment) with a higher γ_β in the treatment observed than in the treatment unobserved cases. Thus, the extent of undertreatment is reduced. but not eliminated. when firms observe treatment.

12 Equivalently, the firm could offer voluntary drug tests and make wage offers which are dependent on test results and whether or not the individual chooses to take the test. In the equilibria of our model, drug users never work for firms which test. Our results are consistent with the practice of testing firms not to hire individuals testing positive.

References

A. Foster Higgins & Co., 1990, Mental health and substance abuse benefits survey, Research and Survey Services, Author.
Akerlof, George A., 1991, Procrastination and obedience, American Economic Review 81, No. 2, 1-19.

Akerlof, George A. and Lawrence F. Katz, 1989, Workers trust funds and the logic of wage profiles, Quarterly Journal of Economics 64, No. 3,525-536.
Akerlof, George A. and Janet Yellen, 1990, The fair-wage effort hypothesis and unemployment, Quarterly Journal of Economics 55, No. 2,255-283.
Becker, Gary S., 1968, Crime and punishment: An economic approach, Journal of Political Economy 76, March/Apr., 169-217.
Becker, Gary S. and George J. Stigler, 1974, Law enforcement, malfeasance, and the compensation of enforcers, Journal of Legal Studies 3, Jan., 1-18.
Becker, Gary S. and Kevin M. Murphy, 1988, A theory of rational addiction, Journal of Political Economy 96, No. 4,675-700.
Bureau of Labor Statistics, U.S. Department of Labor, 1989, Survey of employer anti-drug programs, Report 760 (U.S. Government Printing Office, Washington, DC) Jan.
Dickens, William T., Lawrence F. Katz, Kevin Lang and Lawrence H. Summers, 1989, Employee crime and the monitoring puzzle, Journal of Labor Economics 7, No. 3,331-347.
Frank, Robert F., 1984, Are workers paid their marginal products? American Economic Review 74, Sept., 549-571.
Gallup Organization, The, 1988, Drug testing at work: A survey of American corporations, Sponsored by Hoffman-LaRoche.
Glazer, Jacob and Andrew Weiss, 1991, Conflicting preferences and economic behavior, Unpublished, Department of Economics, Boston University.
Grossman, Sanford J. and Oliver D. Hart, 1983, An analysis of the principal-agent problem, Econometrica 51, No. 1,7-45.
Hansen, William M., Mercer Meidinger Inc., 1988, Substance abuse in the workplace, conducted for Marsh and McLennan Companies, Inc., New York, NY.
Institute of Medicine, 1990, Treating drug problems, Vol. I (National Academy Press, Washington, DC).
Lazear, Edward, 1981, Agency, earnings profiles, productivity, and hours restrictions, American Economic Review 71, Sept., 606-620.
Lazear, Edward, 1979, Why is there mandatory retirement? Journal of Political Economy 87, Dec., 1261-1284.
Levine, David I., 1991, Just-cause employment policies in the presence of worker adverse selection, Journal of Labor Economics 9, No. 3,294-305.
Ma, Ching-To, 1988, Unique implementation of incentive contracts with many agents, Review of Economic Studies 55,555-572.
McGuire, Thomas G., Christopher J. Ruhm and Barbara F. Shatkin, 1991, Defining the public interest in workplace drug abuse policy, National Institute on Drug Abuse, Services Research Monograph Series, No. 1, Washington, DC.
Shapiro, Carl and Joseph S. Stiglitz, 1984, Equilibrium unemployment as a worker discipline device, American Economic Review 74, June, 433-444.
Stigler, George J., 1970, Optimum enforcement of laws, Journal of Political Economy 78, March/ Apr., 526-536.
Stigler, George J. and Gary S. Becker, 1977, De gustibus non est disputandum, American Economic Review 67, No. 1,76-90.
Weiss, Andrew, 1990, Job queues and layoffs in labor markets with flexible wages, Journal of Political Economy 88, No. 3, 526-538.