HARVARD UNIVERSITY
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Faculty Research Working Papers Series

Health Benefits and Wages: Minimizing Total Compensation Cost

Nolan Miller

March 2005

RWP05-029

(Revision of RWP01-023)

Health Benefits and Wages: Minimizing Total Compensation Cost

Nolan H. Miller*

March 11, 2005

Abstract

This paper studies the role of health benefits in an employer's compensation strategy, given the overall goal of minimizing the total compensation expense (wages plus health-insurance cost) for a fixed number of workers. The employer's basic benefit package consists of a base wage and a moderate health plan. It may also offer the option of upgrading to a generous health plan for an additional surcharge. Optimally, the base wage is set in order to balance the total wage bill against the expected cost of health care. In setting the charge for generous coverage the employer acts as a monopolist who sells generous health plans to its employees. The cost-minimization approach is shown to be less vulnerable to adverse selection than other common approaches to the pricing of health benefits, but it may result in excluding some healthy workers from employment.

*John F. Kennedy School of Government, Harvard University, Cambridge, MA, 02138. I thank Chris Avery, Pedro Barros, Suzanne Cooper, David Dranove, Bill Encinosa, Karen Eggleston, Eduardo Loyo, Albert Ma, Tom McGuire, Joe Newhouse, Richard Zeckhauser and participants at the Boston University/Department of Veterans Affairs Fifth Biennial Conference on the Industrial Organization of Health Care for helpful comments. Financial support of the Department of Veterans Affairs is gratefully acknowledged.

1 Introduction

In the United States today, employer-provided health benefits comprise an important part of employees' total compensation. In 1999, approximately 71% of the non-elderly population of the United States received private health insurance, and for 90% of them this insurance was employment based.[1] The prevalence of employer-based health insurance in the U.S. is largely attributable to health insurance payments made through one's employer receiving beneficial tax treatment, and to employer-based group insurance being less expensive than individual (non-group) insurance, due to lower per capita administrative costs. Consequently, to the extent that workers value health plans, the employer's purchasing advantage can make providing these benefits an efficient means of compensation.

This paper analyzes how an employer should design and price its health benefits when it cannot observe workers' (privately known) health care needs, given the goal of minimizing the expected total cost of attracting a certain number of workers. In this paper, the employer offers two plans, a moderate plan, such as an HMO, and a more generous plan, such as a PPO or indemnity plan. The employer's compensation scheme therefore consists of a net wage paid to all employees, which we call the base wage, and an additional surcharge imposed on those who elect the more generous health plan.[2] Each aspect of this scheme presents important trade-offs that the employer must manage.

In setting the base wage, the employer must be mindful of the fact that workers choose to accept employment based on the total utility derived from their wages and health benefits combined. Thus, while lowering the base wage may benefit the employer by lowering its wage bill, doing so also tends to increase health care costs, because those employees who are still willing to accept employment at the lower wage are those who get a relatively high benefit from their health plans, i.e., they have high expected health care needs. The employer's cost-minimizing scheme must balance these two forces.

Setting the surcharge for generous coverage also presents several trade-offs. When an employee

[1] From the EBRI March 1999 Current Population Survey, reported in Gruber (2000).

[2] Use of the term surcharge is intended to emphasize that it is a charge in addition to the charge for the moderate plan. It is not intended to imply that this charge is unfair. In principle, the employer could also offer a no-insurance option. However, we argue below that whenever the employer is willing to offer a health plan to any of its employees, it will not benefit from simultaneously offering a no-insurance option.

elects generous coverage, he pays the surcharge to the employer, which effectively decreases that employee's wage. However, since the generous plan is more expensive than the moderate one, this wage savings is partly offset by the additional cost of the employee's health insurance. In addition, as the employer increases the surcharge, fewer workers choose the generous plan, and this effect must also factor into the employer's decision. This paper argues that in optimally balancing these effects, the cost-minimizing employer should act as a monopolist who sells "health plan upgrades" to its employees. The optimal program involves equating the appropriate concepts of marginal revenue and marginal cost.

Analyzing the behavior of a cost-minimizing employer yields insight into the theoretical basis for two real-world concerns about employer-based health insurance. One concern is that the employer will design its benefits package so as to discourage unhealthy workers from accepting employment, or more generally, to control the mix of employees it attracts.[3] While exclusion of unhealthy workers is the more commonly expressed concern, this paper finds find that a cost-minimizing employer may also benefit from excluding the healthiest workers, since doing so allows it to lower the wages it pays to all of its employees.

A second concern with employer-provided health benefits is that when the employer offers plans of differing intensity, it exposes itself to adverse selection, which may jeopardize the plans' viability. Given the choice of plans, those who expect to have the highest health care needs are drawn to the generous option. This increases the average cost associated with that plan, leading the insurer to demand a higher per-worker premium from the employer. If the employer passes this price increase on to its employees, further adverse selection will result. Those in the generous plan with the lowest expected health care needs will choose to switch to the moderate plan, again increasing the average cost of caring for those who remain in the generous plan. In extreme cases, this phenomenon results in the so-called "premium death spiral" with a high premium leading to adverse selection leading to an even higher premium, the process continuing until the generous plan is no longer sustainable.

Adverse selection and premium spirals have received a great deal of attention in the literature, most often in contexts in which employers follow simple, non-optimal pricing rules.[4] This paper's

[3] Exclusion of unhealthy workers when the perpetrator is a health care provider is studied in the literature on "dumping." See Ellis (1998).

[4] Cutler and Reber (1998) document a premium spiral that occurred at Harvard during the mid 1990s. Ellis and Aragao (2001) document a premium spiral that occurred at Boston University. See also Feldman and Dowd (1982,

analysis of cost-minimizing employers shows that these phenomena arise as consequences of the employer's naïve approach to the benefits-design problem, and that adopting the cost-minimizing approach will tend to decrease their severity. A sophisticated employer should recognize that self-selection will affect the cost of any wage-benefit scheme and incorporate these costs into its planning. Seen in this light, the employer is not a passive force that may fall victim to adverse selection. Rather, it should design its compensation scheme in order to control selection and minimize its overall compensation cost.

This paper shows that the generous plan is viable when the employer follows the cost-minimization rule (i.e., it is chosen by a positive mass of workers) as long as there are some consumers for whom the marginal benefit of the generous plan is greater than its marginal cost.[5] Thus, the cost-minimizing employer offers the generous plan whenever it is socially beneficial to do so. Further, the cost-minimizing compensation scheme is less vulnerable to adverse selection than the commonly employed equal lump-sum contribution rule (in which the employer pays a fixed dollar amount toward the employee's health plan regardless of which plan he chooses) in the sense that there are situations in which the generous plan will be viable under cost minimization approach but not under the equal lump-sum contribution rule. In such cases, offering the generous plan is both socially efficient and cost effective for the employer, yet employers following the equal lump-sum contribution rule fail to take advantage of this opportunity.

This paper adopts a "total compensation" view of health benefits according to which workers' employment decisions are driven by the total value of the compensation package offered to them, i.e., both wages and insurance contribute to inducing the worker to accept employment. Although we believe our implementation of the idea is novel, the view itself is not new. Pauly (1997) espouses this approach, as does the labor economics literature on compensating differentials, summarized in Rosen (1986), which holds that in a competitive market relative wages adjust so that a worker in a particular job is just compensated for any extraordinary benefits (or costs) he receives.

A fundamental difference between this paper and previous academic work on health benefits is in the explicit treatment of employers as acting strategically in order to minimize expected

1991) and Cutler and Zeckhauser (1998).

[5] Although the cost-minimizing scheme offers the generous plan to some workers, it need not offer it to all workers for whom the marginal benefit outweighs the marginal cost.

compensation cost. Two notable exceptions are Levy (1998) and Dranove, Spier, and Baker (2000). Levy (1998) examines the possibility that employers might use employee contributions for health plans as a means of separating workers who want health insurance from those who don't, allowing the employer to avoid buying health care coverage for those who do not value it. Dranove, Spier, and Baker (2000) argue that a possible explanation for why many employers require their employees to contribute toward the cost of their health insurance is that these charges are an attempt to induce married employees to acquire health insurance from their spouses' employers instead.

The paper continues with a description of the model in Section 2. Section 3 discusses cost-minimization with full information, and Section 4 derives and analyzes cost-minimizing compensation programs with private information. Section 5 discusses commonly used pricing rules, and Section 6 concludes. Some of the proofs are contained in the Appendix.

2 The Model

2.1 Basic Structure

Potential workers differ in expected health care costs but are otherwise identical.[6] Let $c \in [0, \gamma]$, $\gamma > 0$, denote a worker's type, where c is normalized to be the expected cost of health care for this worker if enrolled in the generous plan. The cost of care for a worker of type c if enrolled in the moderate plan is given by αc, where $\alpha \in (0, 1)$. The consensus in the literature is the α is somewhere between 0.8 and 0.9 in the case where the moderate plan is an HMO and the generous plan is a PPO.[7]

Denote the dollar-valued benefit derived by a worker of type c enrolled in the moderate plan as $m(c)$, where $m'(c) > 0$ for $c \in (0, \gamma)$. We assume that $m(c) > \alpha c$ for all c, i.e., for every type of worker the benefit from moderate coverage is greater than its cost. Let $g(c)$ denote the additional benefit derived by enrolling in the generous plan, where $g(c)$ is strictly increasing and

[6] The basic structure of the model is similar to Cutler and Reber (1998).

[7] The 0.9 figure is put forth in Cutler and Reber (1998). Compared to an HMO, the Congressional Research Service reports α is approximately 0.82 when the generous plan is a fee-for-service plan, 0.89 when the generous plan is a preferred provider organization, and 0.88 when the generous plan is a point-of-service plan (Congressional Research Service, 1997).

strictly convex: $g'(c) > 0$ and $g''(c) > 0$ for $c \in (0, \gamma)$.[8] In addition, assume $g(0) = 0$.

The employer is risk neutral, and it is only concerned with minimizing its expected compensation cost. While each worker knows his type, for the bulk of the analysis we assume that the employer is unable to observe workers' types.

In order to rule out degenerate solutions to the employer's problem in which only the least healthy (type-γ) workers are hired, we assume that the employer hires a workforce of mass s from a population of total mass 1, where $0 < s < 1$. Let $F(c)$ be the cumulative distribution function of types in the population, where $F(0) = 0$, $F(\gamma) = 1$. Distribution $F()$, assumed to be common knowledge, is differentiable with continuous density function $f(c) = F'(c)$. Frequently, it is useful to consider the density of workers' types conditional on c being greater than some number x, in which case we write $f(c|x)$, where $f(c|x) = \frac{f(c)}{1-F(x)}$ for $x \leq c \leq \gamma$. Similarly, let $F(c|x) = \int_x^c f(s|x)\, ds$.

2.2 The Insurance Market

The employer contracts with an insurer to provide health insurance for its employees. This paper makes two assumptions regarding insurance providers. First, we assume that, based on observable characteristics of workers such as age, sex, and current health status, the insurance company is able to generate an unbiased estimate of the cost of caring for any individual worker (or group of workers).[9] Because the insurer is risk neutral, it is effectively as if the insurer can observe each worker's expected cost of care, i.e., his type.[10]

Second, we assume that insurance companies operate in a perfectly competitive insurance industry. This assumption implies that, in equilibrium, the expected profit earned on any plan must be zero. Thus, for example, if the insurer provides the employer's generous plan, the total premium paid to the insurer in any equilibrium must equal the expected cost of caring for those enrolled in the generous plan. Similarly, if a worker purchases insurance as an individual (i.e., without going

[8] These assumptions imply the relevant "single-crossing property," ensuring that a type-c worker never prefers a less generous plan than a lower-type worker. Convexity simplifies workers' participation decisions but can be relaxed without affecting most of the qualitative results.

[9] This assumption is consistent with industry practice, where insurers price policies based on an assessment of health risk using demographic data, but could also involve physical examinations of the members of the group. See Pauly and Herring (1999) for a discussion of the underwriting of individual and group policies.

[10] In the equilibrium derived later, the insurer knows that the healthiest workers elect the generous plan and the sickest workers elect the moderate one. Thus, even without observing the particular workers enrolled in a plan, the insurer will be able to generate unbiased estimates of the cost of offering a particular plan based its knowledge of the population and the number of workers electing each plan.

through his employer), the premium charged to that individual must equal the expected cost of caring for him. Hence competition implies the individual (non-group) insurance market is perfectly experience rated.[11]

For the sake of simplicity, we ignore any administrative costs which insurers may incur. In reality, the per-capita administrative cost of coverage is higher for purchasers of individual coverage than it is for purchasers of group coverage (i.e., employers). This results in additional "loading factor" being added to the price of individual coverage, which can make it significantly more expensive to cover a worker through an individual policy rather than a group policy.[12] While important in practice, explicitly modeling the loading factor does not significantly impact the results of the paper.

While the zero-profit condition implied by the perfect-competition assumption appears strong, it is a necessary requirement for any equilibrium, given perfect competition in the insurance market. An alternative formulation of the insurance industry's breakeven constraint might require that the insurer break even overall but need not break even on each individual plan. That is, total payments to the insurer must equal the expected cost of care across those enrolled in both plans, but there may be cross-subsidization between the plans. However, such cross subsidization seems inconsistent with perfect competition. When the moderate plan subsidizes the generous plan, one might expect another insurer to offer the moderate plan at a lower cost.[13]

Before considering workers' employment decisions, we first characterize the socially optimal allocation of workers to plans and show that, in a perfectly experience-rated individual insurance market, workers make surplus-maximizing (i.e., socially optimal) decisions. Since a type-c worker earns surplus $m(c) - \alpha c$ from his health plan if he enrolls in the moderate plan or $m(c) + g(c) - c$

[11] While there is strong empirical support for the presence of experience rating in individual insurance markets, some authors have argued that these markets are less than perfectly experience rated. Pauly and Herring (1999), for example, argue that there may be more pooling (i.e., less experience rating) in the individual insurance market than previously thought. The analysis of situations in which there is less than perfect experience rating is similar, and many of the conclusions presented here remain qualitatively the same. A complete analysis of the employer's cost-minimization problem when there is no experience rating in the individual market is available from the author.

[12] Diamond (1992) cites a Congressional Budget Office estimate that, compared to large group coverage (provided by employers with 10,000 or more employees), individual insurance costs about 1.3 times as much. Similar figures are reported by Feldstein (1999, p. 189) and Phelps (1997, p. 345).

[13] Further, as will become clear, our cost-minimizing employer is concerned with its total compensation cost, not how this cost is divided among those receiving the moderate and generous plans. Any insurer who is willing o cross-subsidize its plans will do so only if the employer purchases both plans from it. However, in this case cross-subsidization does not affect the total cost of health insurance remains the same. Thus, as long as the insurer earns zero profit overall, whether or not it cross-subsidizes has no affect on the employer's optimal strategy.

if he enrolls in the generous plan, the surplus-maximizing allocation of workers to plans is for a type-c worker to elect generous coverage if and only if $g(c) \geq (1-\alpha)c$.

To focus in the interesting case where both plans are efficiently provided to some workers, we assume that

$$\begin{aligned} g'(0) &< 1-\alpha, \text{ and} \\ g(\gamma) &> (1-\alpha)\gamma. \end{aligned} \tag{1}$$

Together, these assumptions ensure that there exists a unique worker type $c_I \in (0,\gamma)$ who is indifferent between the two plans on the individual market:

$$g(c_I) = (1-\alpha)c_I. \tag{2}$$

Under the efficient allocation, workers of type $0 \leq c < c_I$ elect the moderate plan and workers of type $c_I \leq c \leq \gamma$ elect the generous plan, where without loss of generality we adopt the convention that workers indifferent between the two plans choose the generous one.

It is straightforward to show that in a perfectly experience-rated individual market workers' individual decisions are socially efficient. Since a perfectly experience-rated individual insurance market charges a type-c worker αc if he elects the moderate plan and c if he elects the generous plan, the worker's utility-maximization problem is the same as society's welfare-maximization problem.

2.3 Workers' Employment Decisions

Whether a particular worker accepts employment depends on the employer's offer and the worker's outside option. Workers who do not accept employment earn alternative wage w_0 (regardless of type) and may purchase insurance from the individual market. As discussed above, workers of type $0 \leq c < c_I$ elect moderate coverage on the individual market, while those of type $c_I \leq c \leq \gamma$ elect generous coverage. Thus, workers' reservation utility function, $u(c)$, is given by:

$$u(c) = \begin{cases} w_0 + m(c) - \alpha c & \text{if } 0 \leq c < c_I, \text{ and} \\ w_0 + m(c) + g(c) - c & \text{if } c_I \leq c \leq \gamma. \end{cases} \tag{3}$$

If the employer offers the moderate plan along with wage w, employment is acceptable to workers for whom $w+m(c) \geq u(c)$. Using (3), this implies that a type-c worker offered moderate coverage has reservation wage, $w_{res}(c,m)$, given by:

$$w_{res}(c,m) = \begin{cases} w_0 - \alpha c & \text{if } 0 \leq c < c_I, \text{ and} \\ w_0 + g(c) - c & \text{if } c_I \leq c \leq \gamma. \end{cases} \tag{4}$$

Similarly, if the employer offers the generous plan along with wage w, employment is acceptable to workers for whom $w + m(c) + g(c) \geq u(c)$, and the reservation wage for a type-c worker is given by:

$$w_{res}(c,g) = \begin{cases} w_0 - \alpha c - g(c) & \text{if } 0 \leq c < c_I, \text{ and} \\ w_0 - c & \text{if } c_I \leq c \leq \gamma. \end{cases} \tag{5}$$

It is straightforward to see that $w_{res}(c,g)$ is decreasing in c. While $w_{res}(c,m)$ is decreasing in c for $0 \leq c \leq c_I$, it is possible that, for large enough c, $w_{res}(c,m)$ increases with c. The reason is that workers of type $c > c_I$ are given moderate coverage by the employer but would have elected generous coverage on their own. Thus, in order to induce them to accept employment they must be compensated $g(c) - c$ in addition to w_0. As c increases, this additional compensation changes as $\frac{\partial w_{res}(c,m)}{\partial c} = g'(c) - 1$, which may be positive if $g'(c)$ is sufficiently large, i.e., if the value of generous coverage increases rapidly as workers become less healthy.

Since $g(c)$ is convex, $g'(c)$ is large for large c, and so $w_{res}(c,m)$ could eventually increase in c. In this case, it could be that some workers are willing to accept employment with moderate coverage but workers with higher expected health care needs are not, which complicates the analysis. However, since $w_{res}(c,m)$ can only increase when the benefit to generous coverage is large, these situations tend to be exactly when the employer can benefit from making generous coverage available, so it is unlikely that the employer would not want to offer the generous plan in this case. Consequently, we assume that $g'(c) < 1$ for all c,[14] which implies that $w_{res}(c,m)$ decreases in c.[15]

[14]Since the worker's total benefit from generous coverage is $m(c) + g(c)$, patient benefit increases with c at a rate of $m'(c) + g'(c)$. Thus assuming $g'(c) < 1$ does not rule out the possibility that patient benefit increases faster than its cost.

[15]While the possibility of increasing reservation wages complicates the analysis, the qualitative results presented in this paper continue to hold. The full analysis of the case where reservation wages increase in c is available from the author.

When $w_{res}(c, m)$ is everywhere decreasing in c, if any worker is willing to accept the employer's offer then all workers with higher expected health care needs are also willing to accept the employer's offer. The result is that in the model under consideration the employer is not able to design its compensation scheme so that healthy workers are willing to accept employment but unhealthy workers are not. The reason for this is that the employer is constrained to compensate either with wages (which appeal to all workers equally) or health insurance, which is more appealing to the unhealthy than the healthy. Thus, to the extent that exclusion of unhealthy workers is a real problem, employers must use some other instrument to bring it about. If, for example, the employer were able to compensate with a benefit that appealed more to the healthy than the unhealthy, such as subsidized membership at a health club, then this benefit, coupled with a wage just high enough to induce health workers to accept employment, would not be sufficient to induce unhealthy workers to accept employment.

3 Cost Minimization with Full Information

Before considering the employer's problem when workers' types are unobservable, we first briefly discuss the employer's problem when it can observe the health status of potential employees and can base its compensation package on this information.

In the U.S., employers who purchase health coverage for their employees do so with pre-tax dollars, while workers who purchase insurance from the individual market generally do so with after-tax dollars. Thus, a worker who receives his reservation wage w_0 and purchases the moderate health plan has final wealth $w_0 - \alpha c$, which costs the employer $\frac{w_0}{1-t}$ dollars, where t is the individual income tax rate (where $0 < t < 1$). If the employer purchases the health plan instead of the employee, it can lower the employee's after-tax wage to $w_0 - \alpha c$, leaving the employee as well off as before and lowering the employer's cost to $\frac{w_0 - \alpha c}{1-t} + \alpha c = \frac{w_0 - t\alpha c}{1-t} < \frac{w_0}{1-t}$. [16] This cost advantage provides one of the primary motives for employer-provided health insurance.

For the purposes of this analysis, we ignore the question of whether the employer requires workers electing the moderate plan to pay part of the cost of their health insurance. Employees care only about their net wage and the health plan they receive, and under Section 125 of the

[16] We ignore the corporate tax since both wages and health care expenditures are tax deductible for corporations.

Internal Revenue Code an employee's contribution for health insurance made through his employer can be made tax deductible (Gruber 2000). Thus, each dollar the employee must contribute for the moderate plan reduces the worker's net after-tax wage by $1-t$ dollars. In order to compensate the worker, the employer must increase his after-tax wage by $1-t$ dollars, which costs the employer 1 dollar. Thus, whether or not the employee contributes for the moderate plan does not affect the employer's overall cost.[17] To simplify the analysis, we assume that the employer takes advantage of Section 125. The results change only slightly if it does not.[18]

Workers' after-tax reservation-wage functions are given by (4) or (5), depending on whether the worker is given moderate or generous coverage. Consider a worker of type c, where $c < c_I$. Such a worker chooses moderate coverage from the non-employment insurance market. If the employer hires the worker and gives him generous coverage, the worker must earn wage $w_0 - g(c) - \alpha c$ after tax, which implies total employment cost of $\frac{w_0 - g(c) - \alpha c}{1-t} + c$, including the cost of health care. If the employer offers this employee moderate coverage, he must be paid $w_0 - \alpha c$, with total cost $\frac{w_0 - \alpha c}{1-t} + \alpha c$. Hence the employer prefers that the type-c worker receive generous coverage whenever:

$$g(c) \geq (1-t)(1-\alpha)c. \tag{6}$$

A similar calculation shows that the same condition determines whether a worker of type $c \geq c_I$ should be given generous coverage. Let c_F be the lowest-cost employee for which the employer prefers generous coverage to moderate, i.e., c_F satisfies (6) with equality. Comparing (6) with (2) shows that $c_F < c_I$. That is, with full information the employer gives generous coverage to some employees for whom the incremental benefit is less than its incremental cost. The reason for this is that, due to the tax advantage afforded employer-provided health benefits, the employer's cost of providing generous coverage is less than the true cost.

The previous argument establishes that with full information the employer should give each

[17] A number of reasons have been suggested for why an employer may want to charge employees for the moderate plan, including that doing so may help separate those who desire health insurance from those who do not (Levy 1998), that it might drive married workers to choose their spouse's insurance instead which reduces the employer's cost (Dranove, Spier and Baker 2000), and that workers are more likely to appreciate the value of their health plans if they must pay something for them (Richard Zeckhauser, personal communication).

[18] Despite the tax-subsidy, not all employers take advantage of Section 125. If the employer does not, then it should pay the entire cost of the moderate plan. The analysis of the optimal charge for generous coverage also changes slightly in this case, since the charge is now paid out of after-tax dollars.

worker it hires a wage/insurance bundle that makes him just indifferent between working and not. However, the employer must also decide which workers to hire. Simplifying the above expressions, the minimum cost of hiring a type-c worker is given by:

$$\begin{array}{lll} \frac{w_0-\alpha ct}{1-t} & \text{if} & c < c_F \\ \frac{w_0-g(c)+(1-\alpha-t)c}{1-t} & \text{if} & c_F \leq c < c_I \\ \frac{w_0-ct}{1-t} & \text{if} & c_I \leq c \leq \gamma. \end{array}$$

It is straightforward to show that the minimum cost of hiring an employee is decreasing in c.[19] This is because the employer can only take advantage of the tax subsidy for compensating with health benefits up to the cost of the worker's health care. Consequently, it has an incentive to hire only those employees with the highest health care needs.[20] If it could, it would minimize its cost by hiring only type-γ employees. This degenerate solution is clearly not sensible, and it is ruled out by the assumption that the employer must hire a workforce of mass $s > 0$. However, the degenerate solution is instructive in that it highlights the employer's incentives to compensate with subsidized health benefits rather than wages whenever possible, an incentive that remains even in the asymmetric information case.

Let c_S be such that $1 - F(c_S) = s$. With full information, the employer hires workers whose types lie between c_S and γ and compensates them as described above.

4 Cost Minimization with Private Information.

While the full information case provides a useful benchmark, in practice employers rarely base what they charge employees for health insurance on health status. Encinosa and Selden (2001) report that in 1993 only 0.6% of workers receiving employment-related single coverage faced contributions that depended directly on health status, and less than 5% faced contributions that depended on anything at all, including smoking status, gender, age, or income.

[19] The only difficult case is for $c_F \leq c < c_I$. However, note that $\frac{w_0-g(c)+(1-a-t)c}{1-t}$ is concave and bounded above by $\frac{w_0-\alpha ct}{1-t}$, and that these two quantities are equal at c_F. Thus the minimum cost of hiring a type-c worker decreases for $c_F \leq c < c_I$.

[20] If high health-cost workers are less productive than low health-cost workers, then there will be a countervailing incentive that pushes in the other direction. The employer may be willing to hire healthy workers if their increased productivity is worth the additional cost of hiring them.

There are a number of reasons why employers might choose not to vary their insurance charges with health status. For example, when employers do not vary their charges for insurance, low-cost workers subsidize high-cost workers, providing insurance against becoming a high-cost type that is lost when the employer bases its charges on health status (Cutler and Zeckhauser, 1998). Along these lines, there are strong equity-based arguments for why workers should not be charged more for insurance simply because they are sick, and employers may be loathe to violate these norms.[21] In addition, while insurance companies are very efficient at estimating the expected cost of caring for a particular individual, employers may be less able to do so, or else able to do so only by incurring significant costs, making basing wages on health status too costly to be practical. For these reasons, the remainder of the analysis is concerned with the case in which the employer does not base the compensation package it offers to a particular worker on that worker's health status.

If the employer either does not know workers' types or is unable to act upon this knowledge, then it can neither condition workers' wages upon their types nor assign them to health plans. Since in this case the only information upon which the employer can base its benefits package is the employee's choice of moderate or generous coverage, the employer's compensation plan will consist of a choice between moderate coverage and a higher wage or generous coverage and a lower wage. The employer's task is then to choose wages for workers electing each health plan in order to minimize the expected compensation cost of its workers, subject to the constraints that each worker voluntarily chooses employment and chooses the health plan that maximizes his net benefit from employment.

Equilibrium Requirements

In our model, three parties make decisions: the employer, the workers, and the insurers. Our chief interest is in understanding the influence of the employer's compensation scheme on the resulting equilibrium and its behavior. Consequently, we impose the following structure on the game. The employer moves first, choosing its compensation practice. Then, after observing the employer's wage offerings, workers and insurers simultaneously choose their strategies.

Since the workers and insurers observe the employer's action before choosing their own strate-

[21] Worker advocacy groups, such as labor unions, may also have an interest in preventing employers from discriminating among their members.

gies, the equilibrium concept we employ is subgame perfect Nash equilibrium. The equilibrium requirements (ER) are:

ER1: The employer chooses a menu of wage-health plan pairs in order to minimize its expected compensation cost (including both wage and health insurance costs) given that workers and insurers react optimally to this compensation scheme.

ER2a: Given the employer's compensation scheme, the workers choose whether to accept employment, and those who accept employment choose the insurance/wage pair that maximizes their utility.

ER2b: Given the employer's compensation scheme, the insurers compete among themselves to offer insurance plans to the employer and to individual (non-group) purchasers. All plans offered in equilibrium must earn zero profit.

Because insurers and workers move simultaneously, equilibrium requirements **ER2a** and **ER2b** are imposed at the same time and define the equilibrium of the subgame created by the employer's choice of compensation scheme. Condition **ER1** then requires that the employer choose the compensation plan that induces its most preferred subgame equilibrium.

This definition of equilibrium is identical to that adopted by Cutler and Reber (1998), except that the present paper requires voluntary participation, whereas their analysis did not. As in Cutler and Reber (1998), the employer is, effectively, a Stackelberg leader. For any compensation scheme the employer chooses there is a subsequent equilibrium in which workers optimize and insurers earn zero profit. The employer's task is therefore to choose the compensation policy that induces its most preferred equilibrium.

The Stackelberg structure of the game follows from the nature of the markets involved. Through its choice of compensation policy, the employer influences workers' insurance choices and insurers' behavior. However, due to competition in the insurance market, an individual insurer is not able to influence the employer's behavior. If a particular insurer were to attempt to do so by, for example, committing to certain prices in order to gain an advantage over the insurer, this would leave the door open to other insurers, who, by tailoring their offerings to the employer's desires, would render the committing firm's offer irrelevant. Similarly, individual workers are unable to influence the employer's behavior.

Worker Behavior

Let w_m denote the after-tax take-home wage paid to workers who accept employment and elect moderate coverage, net of any contribution the employee must pay for his health plan. Similarly, let w_g denote the after-tax take-home wage paid to workers who elect generous insurance, again net of any required contribution for the employee's health plan. It is often convenient to think of w_m as the base wage offered to all employees and $w_m - w_g$ as the surcharge to employees for generous coverage, measured in terms of the reduction in after-tax wages, and we will employ the terms "base wage" and "surcharge" throughout the paper to denote these quantities. Without loss of generality, assume $w_g \leq w_m$.

Since $w_{res}(c, m)$ is decreasing in c, the set of workers who find the base wage and moderate plan acceptable is an interval $[c_m, \gamma]$, where c_m satisfies:

$$w_m = w_{res}(c_m, m). \tag{7}$$

Due to the fact that the employer must hire mass s of workers, the employer faces the additional constraint that $1 - F(c_m) \geq s$. In the case in which $1 - F(c_m) > s$, the employer hires a random sample of mass s from the interval $[c_m, \gamma]$.

A worker who chooses the moderate plan earns utility $w_m + m(c)$, and a worker who chooses the generous plan earns utility $w_g + m(c) + g(c)$. Hence, a type-c worker prefers generous coverage if and only if:

$$w_m - w_g \leq g(c).$$

Since $g(0) = 0$ and $g(c)$ is increasing, define c_g to be the lowest-cost worker who elects generous coverage. That is,

$$w_m - w_g = g(c_g), \tag{8}$$

unless $w_m - w_g > g(c)$ for all c, in which case let $c_g = \gamma$, or $w_m - w_g < g(c)$ for all c, in which case $c_g = 0$. Thus, given wages w_m and w_g, the workers who accept employment self-select into two groups. Workers of type $c_m \leq c < c_g$ choose moderate coverage at wage w_m, and workers of type $c_g \leq c \leq \gamma$ choose to pay a surcharge and receive generous coverage, leaving them net wage w_g. Based on this, we can prove the following result, which is useful later on.

Proposition 1 *If w_m and w_g satisfy (7) and (8), then the surplus a worker earns from employment is non-negative and strictly increasing in c.*

Proof. *See the Appendix.* ■

The only participation constraint that binds belongs to type c_m, i.e., the healthiest worker-type that accepts employment. All other workers earn a positive surplus from employment, and that surplus increases in c. This is an example of the informational rent earned by less attractive agents in any pooling equilibrium. Unhealthy workers benefit from the presence of healthy workers because high wages must be offered to all in order to induce those who get little benefit from health care to accept employment.

When workers' types are private information, almost all workers are able to appropriate some of the surplus created by employment. Because of this, Proposition 1 also establishes that the employer's cost is strictly lower when the employer is able to observe employees' types than when this information is private, since in the full information case no employee earns a rent.

The Employer's Optimization Problem

We now turn to the employer's cost minimization problem. Given that the workers the employer actually hires are a random sample from among those who accept employment, the employer's objective is to minimize the expected average cost of its employees. For brevity, we will often refer to this as the employer's cost, dropping the adjectives "expected" and "average".

If workers of type $c_m \leq c < c_g$ choose moderate coverage and workers of type $c_g \leq c \leq \gamma$ choose generous coverage, perfect competition among insurers implies that, in equilibrium, the employer's cost of the moderate and generous plans are $\int_{c_m}^{c_g} \alpha c f(c|c_m)$ and $\int_{c_g}^{\gamma} c f(c|c_m)\, dc$, respectively.[22] Therefore, anticipating equilibrium behavior by the workers and insurers, the employer's

[22] Cutler and Reber (1998, p. 437) assume the same cost structure. This is also the relevant cost structure if the employer self-insures, as do many large employers. Marquis and Long (1999) analyze self-insurance in seven states and find that 56% of employers with more than 500 employees self-insured in 1997.

cost minimization problem is:

$$\min_{0\leq w_g\leq w_m} \int_{c_m}^{c_g} \left(\frac{1}{(1-t)}w_m + \alpha c\right) f\left(c|c_m\right) dc + \int_{c_g}^{\gamma} \left(\frac{1}{(1-t)}w_g + c\right) f\left(c|c_m\right) dc, \tag{9}$$

subject to (7) and (8), and $1 - F\left(c_m\right) \geq s$.

The employer's problem can be simplified by substituting in for the constraints. Since there is a one-to-one relationship between pairs (w_m, w_g) and (c_m, c_g) given by (7) and (8), we can treat the employer's problem as choosing cut-off values c_m and c_g, with $c_g \geq c_m$. Substituting from (7) and (8) into (9) and rearranging yields the following unconstrained version of the employer's problem:

$$\begin{aligned}\min_{0\leq c_m\leq c_S;c_m\leq c_g\leq\gamma} \phi\left(c_m, c_g\right) \quad = \quad & \alpha \int_{c_m}^{c_g} cf\left(c|c_m\right) dc + \int_{c_g}^{\gamma} cf\left(c|c_m\right) dc \\ & + \frac{1}{(1-t)} w_{res}\left(c_m, m\right) - \frac{1}{(1-t)} g\left(c_g\right) \int_{c_g}^{\gamma} f\left(c|c_m\right) dc.\end{aligned} \tag{10}$$

The first two terms of $\phi\left(c_m, c_g\right)$ comprise the expected cost of health care, the third term is the cost of the base wage paid to all workers, and the final term is the portion of wages recovered as surcharges for the generous health plan. Throughout the paper, we assume a unique solution to the employer's problem, which we denote $\left(c_m^*, c_g^*\right)$.

The solution to (10) defines the cost-minimizing scheme that bundles health insurance with employment. However, it is possible that the employer achieves lower expected cost by not offering its employees health insurance. If the employer chooses not to offer health coverage, it minimizes cost by offering wage w_0, which is the minimum wage acceptable to a worker who is not given health insurance. In this case, each worker costs the employer $\frac{w_0}{1-t}$ dollars, and so it is optimal for the employer to bundle health insurance and employment only if:

$$\phi\left(c_m^*, c_g^*\right) \leq \frac{w_0}{1-t}. \tag{11}$$

If $c_m \leq c_I$, expanding and simplifying (11) yields:[23]

$$-\alpha c_m^* - g\left(c_g^*\right)\left(1 - F\left(c_g^*|c_m^*\right)\right) + (1-t)\left(\alpha \int_{c_m^*}^{c_g^*} cf\left(c|c_m^*\right)dc + \int_{c_g^*}^{\gamma} cf\left(c|c_m^*\right)dc\right) \leq 0. \qquad (12)$$

The first two terms of (12) represent the wage savings due to offering health insurance. When the employer offers health coverage, the base wage for all employees is $w_m^* = w_0 - \alpha c_m^*$, instead of w_0, and wages for those who elect generous coverage are lowered by an additional $g\left(c_g^*\right)$. The remainder of the right-hand side of (12) is due to the employer having to purchase insurance for its employees, which is costly. However, since such purchases are tax advantaged, the impact of this cost is discounted by the tax rate.

The key question in deciding whether it is worthwhile to bundle health coverage with employment is whether the wage savings from doing so is greater than the cost of health coverage, net of the tax subsidy. Proposition 2 characterizes the relationship between the tax rate and whether the employer will choose to bundle health insurance with employment.

Proposition 2 *There exists $t^* \in (0,1)$ such that the employer prefers to bundle health insurance with employment when $t > t^*$, prefers not bundle health insurance with employment when $t < t^*$, and is indifferent between bundling and not when $t = t^*$.*

Proof. *See the Appendix.* ■

To get a sense of the intuition for Proposition 2, recall that when health insurance is not bundled with employment every worker costs the employer $\frac{w_0}{1-t}$. Under bundling, the healthiest worker who accepts employment receives exactly his reservation utility. Because of the tax subsidy, the employer's cost of hiring this employee using money and a health plan is less than its cost if it did so using money alone. Thus, relative to non-bundling, the employer saves money on healthy workers. At the same time, Proposition 1 establishes that higher-cost employees receive strictly more than their reservation utilities under bundling. This rent is an additional cost borne by the employer when it bundles health insurance with employment. When the tax rate is high, the employer's savings due to the tax subsidy outweighs the cost of high-cost employees' rents, and it

[23] If $c_m^* > c_P$, the term $-\alpha c_m^*$ is replaced by $g\left(c_m^*\right) - c_m^*$. Since $g\left(c_m^*\right) - c_m^* < 0$ in this case, the analysis when $c_m > c_P$ is similar to the analysis when $c_m^* \leq c_P$.

is worthwhile for the employer to bundle health insurance with employment. When the tax rate is low, the value of the tax subsidy is small and the rent-cost incurred when health insurance is bundled with employment dominates. In this case it is optimal for the employer not to bundle.[24]

Proposition 2 also establishes that it is never in the employer's interest to offer a no-insurance option in addition to the moderate and generous options, i.e., to offer all three plans at prices such that each plan is chosen by a positive mass of workers. Cost minimization implies that either the employer offers only the no-insurance option (i.e., it does not offer any health insurance to its employees), or else the employer offers only the moderate and generous options, at the prices defined by the solution to (10).

To see why simultaneously offering all three options cannot benefit the employer, consider with the optimal two-option plan defined by c_m^* and c_g^*. If the employer wishes to hire a worker but give him no health insurance, the worker must be paid w_0, which costs the employer $\frac{w_0}{1-t}$. In this case, every type of worker is indifferent between this offer and his outside option. However, under the cost-minimizing plan, almost all workers the employer hires strictly prefer the employer's offer to their outside opportunity. Therefore, no worker currently employed by the firm would prefer the no-insurance option. Workers of types $0 \leq c \leq c_m^*$ are indifferent between the no-insurance option and declining employment. However, even if these workers were to accept employment, hiring them could only decrease the employer's expected total compensation cost if $\frac{w_0}{1-t} < \phi\left(c_m^*, c_g^*\right)$, which is the same condition that defines when the employer prefers not to bundle health insurance with employment. Thus, when the tax rate is high enough that the employer is willing to offer health insurance to its employees, adding a no-insurance option to the employer's benefit package can only increase its cost. On the other hand, when the tax rate is low, the employer minimizes cost by offering only the no-insurance option.

4.1 Setting the Base Wage

We begin by considering how the employer should set its base wage. As suggested above, the critical trade-off here is that lowering w_m lowers the employer's wage bill, but worsens the average

[24] If the employer could fill its workforce by only hiring type-γ workers, then it would be in its interest to bundle for any positive tax rate. Since no worker earns a rent in this case, the savings due to the tax subsidy implies that bundling always saves money.

health status of its workers.

If the employer wishes to offer a base wage that, along with moderate coverage, is acceptable to the entire population, it should choose $w_m = w_{res}(0, m) = w_0$. However, if it is only interested in attracting some subset of the population, $[c_m, \gamma]$, where $c_m > 0$, it could reduce the base wage to $w_m = w_{res}(c_m, m)$. This lowers the employer's wage bill, but also worsens the employee mix, since workers are now drawn from the interval $[c_m, \gamma]$ instead of $[0, \gamma]$.

Let c_g^* be the optimal value of c_g, as determined by (16). Taking into account that c_m cannot be larger than c_S, the effective upper bound on c_m is $b = \min\left\{c_S, c_g^*\right\}$. The cost-minimizing value of c_m solves:

$$\begin{aligned}\min_{c_m \in [0,b]} \phi\left(c_m, c_g^*\right) &= \alpha \int_{c_m}^{c_g^*} cf\left(c|c_m\right) dc + \int_{c_g^*}^{\gamma} cf\left(c|c_m\right) dc \\ &\quad - \frac{g\left(c_g^*\right)}{(1-t)} \int_{c_g^*}^{\gamma} f\left(c|c_m\right) dc + \frac{w_{res}\left(c_m, m\right)}{(1-t)}.\end{aligned} \tag{13}$$

Assume (13) has a unique solution and denote it c_m^*. Differentiating $\phi\left(c_m, c_g^*\right)$ with respect to c_m yields optimality condition:

$$-\alpha c_m^* f\left(c_m^*|c_m^*\right) + f\left(c_m^*|c_m^*\right) K + \frac{1}{(1-t)} \frac{\partial w_{res}\left(c_m^*, m\right)}{\partial c} \begin{cases} \geq 0 & \text{if } c_m = 0 \\ = 0 & \text{if } 0 < c_m < b \\ \leq 0 & \text{if } c_m^* = b \end{cases}, \tag{14}$$

where

$$K = \alpha \int_{c_m}^{c_g^*} cf\left(c|c_m^*\right) dc + \int_{c_g^*}^{\gamma} cf\left(c|c_m^*\right) dc - \frac{1}{(1-t)} g\left(c_g^*\right) \frac{1 - F\left(c_g^*\right)}{1 - F\left(c_m^*\right)},$$

and (from (4)) $\frac{\partial w_{res}(c_m^*, m)}{\partial c} = -\alpha$ if $0 \leq c_m^* < c_I$, and $\frac{\partial w_{res}(c_m^*, m)}{\partial c} = -1$ otherwise.

Increasing c_m has three effects on the employer's total compensation cost. First, the employer avoids the health care cost of the workers who are no longer employed. This is captured in the term $-\alpha c_m f\left(c_m|c_m\right)$ in (14). Second, the employer must pay the employment cost for these workers' replacements, who are drawn randomly from the distribution $f\left(c|c_m\right)$. This effect is captured by $Kf\left(c_m|c_m\right)$, where K can be further decomposed into $\alpha \int_0^{c_g^*} cf\left(c|c_m^*\right) dc + \int_{c_g^*}^{\gamma} cf\left(c|c_m^*\right) dc$, the expected health care cost of the replacement workers, and $-\frac{1}{(1-t)} g\left(c_g^*\right) \frac{1-F\left(c_g^*\right)}{1-F\left(c_m^*\right)}$, the expected cost recovery due to the fact that a random replacement worker elects the generous plan with probability $\frac{1-F\left(c_g^*\right)}{1-F\left(c_m^*\right)}$, in which case his wage cost is reduced by $\frac{1}{(1-t)}g\left(c_g^*\right)$. The third term corresponds to the fact that increasing c_m allows the employer to lower the base wage to $w_m = w_{res}\left(c_m^*, m\right)$, which reduces wage expenditure at a rate of $\left|\frac{1}{(1-t)}\frac{\partial w_{res}(c_m^*,m)}{\partial c}\right|$ per unit mass of workers.

Of particular interest is the question of whether the employer will choose $c_m > 0$, excluding some low-cost workers and lowering the base wage to $w_m = w_{res}\left(c_m, m\right) < w_0$.

Proposition 3 *The cost-minimizing choice of c_m is strictly positive provided that i) the density of healthy workers is sufficiently small (i.e., $f\left(c\right)$ is small near $c = 0$), or ii) the tax rate is sufficiently high.*

Proof. *See the Appendix.* ■

Intuitively, the employer may want to exclude some healthy workers because doing so lowers the base wage that must be paid to all employees. However, at the same time, exclusion of healthy workers implies that the employer's workforce becomes, on average, less healthy. This leads to higher health care costs. If there are relatively few healthy workers (i.e, $f\left(c\right)$ is small near $c = 0$), then lowering the base wage excludes relatively few workers, and so replacing these workers with less healthy ones does not have a large impact on the average health of the employer's workforce. Consequently, in such cases the employer will find it worthwhile to lower its base wage. On the other hand, the chief cost to the employer of increasing the average health care needs of its employees is that it must pay more for their health insurance. However, the tax-deductibility of the employer's health insurance expenditures mitigates this effect: the higher the tax rate, the less the cost to the employer of its workforce's higher health care costs. Thus, higher tax rates also help to induce the employer to exclude some healthy workers.

Proposition 3 shows that the employer chooses $c_m^* > 0$ when the tax rate is sufficiently high, and, as established in Proposition 2, the employer also prefers to bundle health insurance with employment in this case. Thus, when the tax rate is high enough, the employer both excludes some workers and cannot benefit from offering them a no-insurance option. This establishes that exclusion of healthy workers is a real possibility. To the extent that excluding low-cost workers leads to lower base wages, (almost) all workers are harmed by this practice.

Throughout the paper, we have assumed that health status does not affect productivity. If healthy workers are more productive than unhealthy workers, then this will make excluding low-cost workers less attractive. The employer may be willing to increase its wage cost in order to attract healthy workers. However, to reap the productivity gain, the employer must pay higher wages to all of its employees. Depending on the magnitude of the productivity difference and the distribution of healthy and unhealthy workers, this may or may not be worthwhile.

4.2 Setting the Surcharge for Generous Coverage

Next, we consider the employer's optimal choice of the surcharge for generous insurance. Conveniently, the optimal level of c_g depends on c_m only inasmuch as the employer must choose $c_g \geq c_m$. To see why, fix c_m at c_m^*, divide $\phi(c_m^*, c_g)$ by $(1 - F(c_m^*))$, and rewrite (10) as:

$$\begin{aligned} \min_{0 \leq c_m^* \leq c_g \leq \gamma} \frac{\phi(c_m^*, c_g)}{1 - F(c_m^*)} \quad = \quad & \alpha \int_{c_m^*}^{c_g} cf(c)\, dc + \int_{c_g}^{\gamma} cf(c)\, dc \\ & + \frac{1}{(1-t)} w_{res}(c_m^*, m)(1 - F(c_m^*)) - \frac{1}{(1-t)} g(c_g) \int_{c_g}^{\gamma} f(c)\, dc. \end{aligned} \tag{15}$$

Having normalized by $1 - F(c_m^*)$, $\frac{\phi(c_m^*, c_g)}{1 - F(c_m^*)}$ depends on c_m^* only through the base wage paid to all employees, $w_{res}(c_m^*, m)(1 - F(c_m^*))$. However, since this term does not depend on c_g, it has no effect on the employer's optimal choice of c_g. Therefore, if there is an interior solution to the employer's problem, this solution does not depend on c_m^*.

Proposition 4 *The cost-minimizing choice of c_g is characterized by:*

$$-\frac{1}{(1-t)} g'\left(c_g^*\right)\left(1 - F\left(c_g^*\right)\right) + \left(\frac{1}{(1-t)} g\left(c_g^*\right) - (1-\alpha) c_g^*\right) f\left(c_g^*\right) \begin{cases} \geq 0 & if \quad c_g^* = c_m^* \\ = 0 & if \quad c_m^* < c_g^* < \gamma \\ \leq 0 & if \quad c_g^* = \gamma \end{cases} . \tag{16}$$

The surcharge for generous coverage is $w_m - w_g = g\left(c_g^*\right)$.

Proof. *Follows directly from the first-order condition for (15) subject to the constraint that $c_m^* \leq c_g^* \leq \gamma$.* ■

Consider an interior solution, i.e., one where (16) holds with equality. The first term on the left-hand side of (16) captures the fact that increasing c_g increases the surcharge that can be imposed on all workers who elect generous coverage, since the size of the surcharge is determined by the marginal benefit of generous coverage earned by the lowest-cost employee who elects it. Increasing c_g also causes some workers who used to elect generous coverage to elect moderate coverage instead. These workers must then be paid a higher wage but incur lower expected health costs. The difference in these two, the impact on total compensation of shifting the marginal worker to the moderate plan, is captured by the second term on the left-hand side of (16).

Under the cost-minimizing plan the surcharge for electing generous coverage is $g\left(c_g^*\right)$. One way to interpret this is that the employer is a monopolist who sells health care plans to its workers at price $g\left(c_g^*\right)$. To see why, fix w_m, and let $p = w_m - w_g$ be the price of generous insurance. Demand for generous health coverage is then given by $D(p) = 1 - F\left(g^{-1}(p)\right)$. The employer's insurance cost is:

$$C(p) = \alpha \int_{c_m}^{g^{-1}(p)} cf(c)\,dc + \int_{g^{-1}(p)}^{\gamma} cf(c)\,dc.$$

Hence for fixed w_m the employer's problem is equivalently written:

$$\max_{p \geq 0} \frac{1}{(1-t)} D(p)\,p - C(p)\,. \tag{17}$$

Differentiating with respect to p and setting the result equal to zero,

$$\frac{1}{(1-t)}\left(1 - F\left(g^{-1}(\tilde{p})\right) - f\left(g^{-1}(\tilde{p})\right)\frac{dg^{-1}(\tilde{p})}{dp}\tilde{p}\right) + (1-\alpha)\frac{dg^{-1}(\tilde{p})}{dp} g^{-1}(\tilde{p})\, f\left(g^{-1}(\tilde{p})\right) = 0,$$

at the optimal price, $\tilde{p}$. Let $\tilde{c}_g$ be the lowest-cost type who elects generous coverage at price $\tilde{p}$. Substituting in that $\frac{dg^{-1}(p)}{dp} = \frac{1}{g'(c_g)}$ and $g^{-1}(p) = c_g$, and rearranging,

$$-\frac{1}{(1-t)} g'(\tilde{c}_g)\left(1 - F(\tilde{c}_g)\right) + \left(\frac{1}{(1-t)} g(\tilde{c}_g) - (1-\alpha)\,\tilde{c}_g\right) f(\tilde{c}_g) = 0 \tag{18}$$

Note that (18) is identical to (16), and therefore $\tilde{c}_g = c_g^*$. Seen this way, $g\left(c_g^*\right)$ is the price paid by each worker who elects generous coverage, and (16) is the monopolist's optimality condition requiring that, at the optimum, the marginal revenue due to selling the generous plan to slightly sicker people,

$$-\frac{1}{(1-t)}g'\left(c_g^*\right)\left(1-F\left(c_g^*\right)\right)+\frac{1}{(1-t)}g\left(c_g^*\right)f\left(c_g^*\right),$$

is just equal to the marginal cost of providing them with more generous care, $(1-\alpha)\,c_g^* f\left(c_g^*\right)$.

To push the monopoly analogy a bit farther, just as a monopolist can increase its revenue by price discriminating, so could the employer reduce its total compensation cost if it charged high-cost employees more for their health plans than low-cost employees. However, as noted earlier, few employers attempt to charge unhealthy employees more for health insurance.

Corner solutions to (15) are also possible. The case where all employees receive generous coverage, i.e., $c_g^* = c_m^*$, is possible if the marginal benefit to "selling" generous coverage is always greater than its marginal cost. This will tend to be the case if the marginal benefit to generous coverage is large relative to its marginal cost for all c (i.e., $g(c)$ is large relative to $(1-\alpha)\,c$ and $g'(c)$ is small).

The other corner solution, where $c_g^* = \gamma$, is of particular importance because it represents the situation where the forces of adverse selection are so strong that the employer does not find it worthwhile to offer the generous plan. Evaluating (16) at $c_g^* = \gamma$, this situation is possible only if:

$$g(\gamma) \leq (1-t)(1-\alpha)\,\gamma. \tag{19}$$

That is, the employer will not offer generous coverage if the benefit derived from it it is small even for those who value it the most. However, condition (19) can only be satisfied if it would not be socially desirable to offer the generous plan to any worker, even under full information.

Proposition 5 *If $g(c) > (1-\alpha)\,c$ for some $c > 0$, then the employer's cost minimizing plan offers the generous plan to some workers, i.e., $c_g^* < \gamma$.*

Proof. *If $g(c) > (1-\alpha)\,c$ for some $c > 0$, then $g(\gamma) > (1-\alpha)\,\gamma > (1-t)(1-\alpha)\,\gamma$ by convexity of $g(\cdot)$.* ■

Proposition 5 says that if there is some worker for whom the incremental benefit of generous coverage is greater than its incremental cost then the cost-minimizing employer offers it. That is, if the generous plan is socially desirable, then it is individually desirable for the cost-minimizing employer. Since (1) assumes that the generous plan is worthwhile for some workers, the employer's optimal plan will always give generous coverage to some workers. However, Proposition 5 does not imply that the employer's plan will give generous coverage to all workers for whom it is worthwhile, i.e., that $c_g^* = c_I$. As noted above, the employer's incentive to act as a monopolist will give it a tendency to restrict the quantity of workers receiving generous coverage in order to drive up the price.

5 Other Rules and Adverse Selection Spirals

The cost-minimization approach focuses on setting the surcharge for the generous plan (the difference in the employee's out-of-pocket cost for the two plans) to induce the employees to sort themselves as the employer desires. The size of the surcharge determines how employees will sort themselves into plans and the equilibrium premiums the plans will charge. The employer must then subsidize each plan in order to bring about the desired difference in employee cost. Thus, the surcharge and employees' and insurers' equilibrium behavior determine the employer's subsidies.

Firms' actual approaches to pricing their health plans typically involve the opposite order. Employers choose how much to subsidize the plans, and employees' and insurers' equilibrium behavior determine the difference in the employees' cost, i.e., the surcharge. The two most widely used rules have the employer either paying an equal lump sum subsidy (ELS) for the employee's insurance, regardless of which plan is chosen, or the employer subsidizing an equal percentage of the premium (EPP) of whichever plan the employee chooses.[25] In either case, the worker bears the remainder of the premium cost for whichever plan he elects.

Since the premium competitive insurers charge employers for a particular plan is proportional to the average cost its enrollees, under either the ELS or EPP subsidy rule the cost to the worker of changing plans depends on enrollment choices of other workers. This dependence leaves the system vulnerable to adverse selection and, in extreme cases, the type of premium "death spiral" experienced by Harvard University in the mid 1990s, during which Harvard's most generous health plan became so expensive that Harvard stopped offering it.

[25] According to the Kaiser/HRET Employer Health Benefits Annual Survey (2002), 17% of workers had employers followed an ELS rule in 2002 (down from 28% in 1999) and 37% of workers had employers who followed an EPP rule in 2002 (up from 29% in 1999).

Cutler and Reber (1998) study the Harvard case. In the early 1990s, Harvard's health plan offerings included the generous Blue Cross/Blue Shield PPO as well as several less-generous HMO's. In 1995, the university changed its premium-contribution policy from a generous EPP-subsidy rule to an ELS-subsidy rule which imposed on workers a larger portion of the incremental cost of more generous care. As workers switched from the PPO to the HMOs, the difference in the average cost of the two plans increased. Under the ELS subsidy rule, this caused the out-of-pocket cost of the PPO to increase as well, which lead to further adverse selection. In just three years, declining enrollment forced the most generous plan, the Blue Cross/Blue Shield PPO, to be withdrawn because it could no longer be offered at a reasonable price.

Under an ELS subsidy rule such as the one Harvard used, the price paid by a worker who moves from the moderate plan to the generous plan is given by the difference in the average cost of those enrolled in each of the plans. This price, which we denote $p(c_g)$, is given by:[26]

$$p(c_g) = \int_{c_g}^{\gamma} c \frac{f(c)}{1 - F(c_g)} dc - \alpha \int_{c_m^*}^{c_g} c \frac{f(c)}{F(c_g) - F(c_m^*)} dc.$$

Since such expenditures are made with pre-tax dollars, the worker's after-tax cost of the generous plan is $(1-t)\,p(c_g)$. Increasing c_g increases the average cost of both the generous and moderate plans. Consequently, whether $p(c_g)$ increases or decreases generally depends on the distribution of health costs. Note, however, that the price of the generous plan under the ELS rule is always greater than the true incremental resource cost of the marginal employee: $p(c_g) > (1-\alpha)\,c_g$. This is the fundamental reason why the ELS rule promotes adverse selection.

The equilibrium under the ELS rule involves those for whom $g(c) \geq (1-t)\,p(c_g)$ electing generous coverage, while those for whom $g(c) < (1-t)\,p(c_g)$ elect moderate coverage. The generous plan is not viable under the ELS subsidy rule whenever $g(c_g) < (1-t)\,p(c_g)$ for all c_g. This is likely to be the case when workers' incremental value of the generous plan is small, the tax rate is low, or $p(c_g)$ is uniformly large, which is more likely to happen when the population is not highly concentrated on a single worker type (i.e., $f(c)$ is relatively large near both $c = 0$ and $c = \gamma$).[27]

[26] Cutler and Reber (1998, p. 437) denote this quantity the "premium-out-of-pocket," or "P_{oop}."

[27] If $f(c)$ is highly concentrated near a single point (e.g., $f(c)$ is normal with a low variance), then $p(c_g)$ will tend

As noted in (19) above, under the cost-minimizing scheme the generous plan remains viable (i.e., is offered and is accepted by a positive mass of employees) whenever $g(\gamma) > (1-t)(1-\alpha)\gamma$. Comparing the ELS and cost-minimizing rules leads directly to Proposition 6.

Proposition 6 *If the generous plan is viable under the ELS subsidy rule, then it is viable under cost minimization as well. However, the converse is not true. It may efficient to provide the generous plan to some workers, and that plan may be viable under the cost-minimization rule but not the ELS rule.*

Proof.

$$\begin{aligned}(1-t)\,p(c_g) &= (1-t)\left(\int_{c_g}^{\gamma} c\frac{f(c)}{1-F(c_g)}dc - \alpha\int_{c_m^*}^{c_g} c\frac{f(c)}{F(c_g)-F(c_m^*)}dc\right)\\ &> (1-t)\left(\gamma - \alpha\int_{c_m^*}^{c_g} c\frac{f(c)}{F(c_g)-F(c_m^*)}dc\right)\\ &> (1-t)(1-\alpha)\gamma.\end{aligned}$$

Hence if $(1-t)\,p(c_g) < g(c_g)$ *for some* c_g*, then* $(1-t)(1-\alpha)\gamma < g(\gamma)$*, but the reverse implication does not hold.* ■

Proposition 6 establishes that there are times when an employer subsidizing according to the ELS rule would "lose" its generous plan to a premium death spiral, whereas the generous plan would have continued to be offered had the employer instead followed the cost-minimization approach. This is most likely to occur when $p(c_g)$ is uniformly large, i.e., when there is substantial dispersion in the distribution of health types.[28] In such cases the employer's total compensation cost is lower when the generous plan is offered than when it is not. Thus, under the ELS-subsidy rule, there are cases in which the death of the generous plan is not only inefficient and unnecessary, but also unprofitable. The cost-minimizing plan would have sustained the generous plan and lowered total compensation cost.

Despite this finding, it is not necessarily the case that more employees receive generous coverage under the cost-minimization rule than under the ELS rule. Further, the cost-minimization approach

to be small. When the distribution does not have this feature (e.g., the distribution has "fat tails" or multiple points of high density that are far from each other), $p(c_g)$ is more likely to be uniformly large.

[28] Note that viability of the generous plan depends on the distribution of health types under the ELS rule but not under the cost-minimization rule.

is not necessarily superior from a welfare perspective. When a cost-minimizing employer chooses to offer the generous plan, it does so in order to lower its compensation cost, not for any efficiency reasons.

The ELS rule is just one rule employers use. Other rules are less susceptible to adverse selection spirals. For example, Cutler and Zeckhauser (1998) contrast Harvard's practices with those used by the Group Insurance Commission of Massachusetts (GIC), which designs the health insurance options for state employees' health benefits. By Massachusetts law, the state must use a fairly generous EPP rule. Since employees' out-of-pocket cost is only a small fraction of the difference in the cost of the two plans, this tends to dampen the effects of adverse selection. Consequently, it is more likely that the generous plan remains viable under such a rule. Indeed, Cutler and Zeckhauser explain that, because of its contribution rule and other cost-containment measures, the GIC has kept its most generous plan viable.

6 Discussion

This paper has developed a number of basic principles for how an employer should design its compensation scheme to minimize its cost. Two robust qualitative features emerge from the analysis. First, in setting the surcharge for generous coverage, the employer should act as a monopolist who sells health insurance upgrades to its employees. Second, the employer may benefit from lowering its base wage and excluding the healthiest workers from employment.

At the most general level, these principles advocate a change in approach to the benefits design problem: employers' primary focus should be on managing self-selection (both in terms of which workers choose employment and which employees choose generous coverage) in order to minimize total compensation cost. Such an approach can ameliorate common benefits-related problems employers face. For example, by explicitly considering the cost effects of differing selection, the employer controls it, reducing the likelihood and severity of adverse selection and premium spirals.

This paper has assumed that all workers are equally productive, that the size of the employer's workforce is exogenous, and that the employer has all of the bargaining power in the game with employees and insurers. While relaxing any of these assumptions would change the results somewhat, doing so would not affect the fundamental forces that drive the employer's decision.

As mentioned earlier, if healthy workers are more productive than unhealthy ones, then this factor will tend to counteract the employer's benefit from lowering its base wage. In this world, the employer's wage savings will have to be weighed against the fact that it worsens the mix of workers it attracts, both in terms of their expected health care needs and their expected productivity. The decrease in expected productivity is just another cost that must be weighed against the wage savings.

This analysis has taken the size of the employer's workforce as fixed. When the employer's labor demand is endogenous, then the number of employees it hires will be determined by setting the marginal benefit of another employee equal to its marginal cost. Since, by definition, the cost-minimizing employer minimizes the cost of hiring an additional employee, we would expect that if an employer were to adopt the cost-minimization approach instead of the ELS or EPP subsidy rules, the subsequent cost savings would lead to an increase in labor demand.

The analysis in this paper also relies on the assumption that the employer is a Stackelberg leader, i.e., that it chooses can choose policies that influence its employees' insurance decisions and, through them, the insurance-market equilibrium. This assumption is most natural when the employer is a medium or large employer and there is vigorous competition in the insurance market and labor markets. In cases where the insurance or labor sectors have more bargaining power or other employers compete for the same workers, we might expect slightly different conclusions. Nevertheless, to the extent that the employer is able to influence the terms of the negotiation, its goals and incentives should remain the same. The extent to which it is able to accomplish these goals will depend on the relative strength of its bargaining position. The precise role of the relative bargaining power of employers, workers, and insurers in determining the equilibrium of the employment-based health insurance market is the subject of future study.

Finally, this paper has been somewhat critical of the ELS approach. To be fair, in addition to the liabilities discussed here the ELS approach also has strong advantages. The principle underlying the ELS approach, which is an integral part of "Managed Competition" advocated by Enthoven (1988, 2003) and others, is that making employees bear the entire additional cost of a more expensive health plan will give them strong incentives to choose less expensive plans, and this increased price sensitivity will increase insurers' incentives to lower their prices.[29] Implicit in the theoretical argument for the ELS rule is the assumption that expensive plans are that way either because they are inefficient or the market is not fully competitive. Selection, i.e., the tendency of the plans to attract different mixes of health risks, does not play a major role in the discussion.

In this paper, we have argued that the employer should follow the cost-minimization rule, choosing the price for the generous plan that, taking selection into account, results in the lowest overall compensation cost. While we have not explicitly considered the beneficial supply-side effects of charging a high surcharge for the generous plan, the logic of the cost-minimization approach extends readily to this case. Supply-side incentives depend on employee's price sensitivity in selecting plans, which in turns depends on the price they are charged for the generous plan. Thus, when supply-side incentives are a concern, the cost-minimizing employer should choose the surcharge in order to minimize its overall compensation cost, taking into account the effects of the surcharge on both selection and supply-side incentives. While the ELS rule is an effective and easily implementable first cut at the problem of insufficient supply-side incentives, its neglect of selection may increase employers' total compensation cost and the likelihood of extreme adverse selection.

References

[1] Congressional Research Service (1997) "Managed Care: A Primer" CRS Report for Congress *97-913.* Available from the CRS web at http://www.senate.gov/~dpc/crs/reports/pdf/97-913.pdf.

[2] Cutler, D., and S. Reber (1998). "Paying for health insurance: the trade-off between competition and adverse selection," *Quarterly Journal of Economics* **113(2):** 443-466.

[3] Cutler, D., and R. Zeckhauser (1998). "Adverse selection in health insurance," in A. Garber, ed., *Frontiers in Health Policy Research*, Vol. 1, (MIT Press, Cambridge, MA) 1-31.

[4] Diamond, P. (1992) "Organizing the Health Insurance Market," *Econometrica* **60**(6): 1233-1254.

[29] Indeed, in their discussion of the Harvard case Cutler and Reber (1998) argue that increased supply-side competition resulting from adoption of the ELS rule reduced Harvard's premiums by 5 to 8 percent.

[5] Dranove, D., Spier, K., and L. Baker (2000) "'Competition' among employers offering health insurance," *Journal of Health Economics* 19: 121-140.

[6] Ellis, R. (1998) "Creaming, skimping and dumping: provider competition on the intensive and extensive margins," *Journal of Health Economics* 17: 537-555.

[7] Ellis, R., and F. Aragao (2001) "Death spirals, switching costs, and health premium payment systems," mimeo (Boston University).

[8] Encinosa, W. and T. Selden (2001) "Designing Employer Health Benefits for a Heterogeneous Workforce: Risk Adjustment and Its Alternatives," *Inquiry* 38: 270-279.

[9] Enthoven, A. (1988) *Theory and Practice of Managed Competition in Health Care Finance,* (North-Holland, Amsterdam).

[10] Enthoven, A. (2003) "Employment-Based Health Insurance is Failing: Now What?" *Health Affairs Web Exclusives,* W3-237, http://www.healthaffairs.org.

[11] Feldstein, P. (1999) *Health Care Economics* (Delmar Publishers, Albany).

[12] Feldman R. and B. Dowd (1991) "Must adverse selection cause premium spirals?" *Journal of Health Economics* 10: 349-357.

[13] Feldman R. and B. Dowd (1982) "Simulation of a health insurance market with adverse selection," *Operations Research* 30(6): 1027-1042.

[14] Gruber, J. (2000) "Health insurance and the labor market," in A. Culyer and J. Newhouse, eds., *Handbook of Health Economics*, Vol. 1a, (Elsevier, Amsterdam) 645-706.

[15] Kaiser Family Foundation and Health Research and Educational Trust (2002), "2002 Annual Employer Health Benefits Survey," http://www.kff.org/insurance/ehbs-archives.cfm.

[16] Levy, H (1998) "Who pays for health insurance? worker contributions to health insurance premiums," Industrial Relations Section Working Paper #398, Princeton University.

[17] Marquis, S. and S. Long (1999) "Recent trends in self-insured employer health plans," *Health Affairs* 18(3): 161-166.

[18] Pauly, M. (1997) *Health Benefits at Work,* (Univ. of Michigan Press, Ann Arbor).

[19] Pauly, M. and Herring, B. (1999) *Pooling Health Insurance Risks,* (AEI Press, Washington, D.C.).

[20] Phelps, C. (1997) *Health Economics,* (Addison-Wesley, Reading, Massachusetts).

[21] Rosen, S. (1986) "The theory of equalizing differences," in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, Vol. 1, (Elsevier, Amsterdam) 641-692.

A Proofs of the Propositions

Proof of Proposition 1: *Let c_m^* and c_g^* satisfy (7) and (8). If $0 < c_g^* < \gamma$ and $c_g^* \geq c_I$, the utility increment, $v(c)$, earned by a worker of type c as a result of employment is:*

$$v(c) = \begin{cases} 0 & \text{if } 0 \leq c \leq c_m^* \\ \alpha(c - c_m^*) & \text{if } c_m^* \leq c \leq c_I \\ (c - \alpha c_m^*) - g(c) & \text{if } c_I \leq c < c_g^* \\ (c - \alpha c_m^*) - g(c_g^*) & \text{if } c_g^* \leq c \leq \gamma \end{cases} . \tag{20}$$

If $0 < c_g^* < \gamma$ and $c_g^* < c_I$, function $v(c)$ is

$$v(c) = \begin{cases} 0 & \text{if } 0 \leq c \leq c_m^* \\ \alpha(c - c_m^*) & \text{if } c_m^* \leq c < c_g^* \\ \alpha(c - c_m^*) + g(c) - g(c_g^*) & \text{if } c_g^* \leq c < c_I \\ (c - \alpha c_m^*) - g(c_g^*) & \text{if } c_I \leq c \leq \gamma \end{cases} . \tag{21}$$

In either case, $v(c)$ is increasing in c.■

Proof of Proposition 2: *Let $c_m(t)$ and $c_g(t)$ denote the optimal plan when the tax rate is t. Steps 1 - 3 establish the existence of t^*.*

1. **For t sufficiently close to 1, bundling health insurance with employment is optimal.**

   Consider the feasible compensation plan that sets $c_m = c_g = c_S$. Evaluating (12) for this *plan shows that for t sufficiently close to 1, $\frac{w_0}{1-t} > \phi(c_S, c_S, t) \geq \phi(c_m^*, c_g^*, t)$.*

2. **For t sufficiently close to 0, the employer prefers not to bundle health insurance with employment.** *Consider the case where $c_m < c_I$. The proof for the case where $c_m \geq c_I$ is similar. Let $c_m(t)$ and $c_g(t)$ denote the optimal choice of c_m and c_g as functions of t. Note that $c_m(t) \leq c_S$ by assumption. First, consider those who receive moderate coverage. With no bundling, types $[c_m(t), c_g(t)]$ cost the employer:*

$$\int_{c_m}^{c_g} \frac{w_0}{1-t} f(c|c_m)\, dc = \frac{w_0}{1-t}\left(F(c_g) - F(c_m)\right).$$

With bundling, types $[c_m(t), c_g(t)]$ cost the employer:

$$\begin{aligned}
& \int_{c_m(t)}^{c_g(t)} \left(\frac{w_0 - \alpha c_m(t)}{1-t} + \alpha c\right) f(c|c_m(t))\, dc \\
= \quad & \frac{w_0 - \alpha c_m(t)}{1-t}\left(F(c_g(t)) - F(c_m(t))\right) + \int_{c_m(t)}^{c_g(t)} \alpha c f(c|c_m(t))\, dc.
\end{aligned}$$

The cost of these workers with no bundling is less than their cost with bundling whenever:

$$\begin{aligned}
\frac{w_0}{1-t}\left(F(c_g(t)) - F(c_m(t))\right) &\leq \frac{w_0 - \alpha c_m(t)}{1-t}\left(F(c_g(t)) - F(c_m(t))\right) + \int_{c_m}^{c_g} \alpha c f(c|c_m)\, dc \\
\frac{\alpha c_m(t)}{1-t}\left(F(c_g(t)) - F(c_m(t))\right) &\leq \alpha \int_{c_m(t)}^{c_g(t)} c f(c|c_m(t))\, dc.
\end{aligned}$$

Since the right-hand side is the average cost of those with types between $c_m(t)$ and $c_g(t)$,

$$\alpha c_m(t)\left(F(c_g(t)) - F(c_m(t))\right) < \alpha \int_{c_m}^{c_g} c f(c|c_m)\, dc,$$

whenever $c_m(t) < c_g(t)$, and this is true for any t. As $t \to 0$,

$$\begin{aligned}
& \lim_{t \to 0} \frac{\alpha c_m(t)}{1-t}\left(F(c_g(t)) - F(c_m(t))\right) \\
= \quad & \alpha c_m(0)\left(F(c_g(0)) - F(c_m(0))\right).
\end{aligned}$$

The limit of the right-hand side is $\alpha \int_{c_m(0)}^{c_g(0)} c f(c|c_m(0))\, dc$, which is strictly greater than the left-hand side. Hence, for t sufficiently close to zero, the cost of those receiving moderate care *is greater with bundling than with no bundling. A similar computation shows that the cost of those receiving generous care is strictly larger with bundling than with no bundling as long as $c_g(t) < \gamma$. Since the total mass of employees is s, either $c_m(t) < c_g(t)$ or $c_g(t) < \gamma$ must hold, which establishes the result.*[30]

3. **If it is worthwhile to bundle health insurance with employment for $\hat{t}$, it is worthwhile to bundle for all $t \geq \hat{t}$.** *Define*

$$\begin{aligned}\Phi(c_m, c_g, t) &= (1-t)\left(\alpha \int_{c_m}^{c_g^*} cf(c|c_m)\,dc + \int_{c_g^*}^{\gamma} cf(c|c_m)\,dc\right) \\ &\quad - g(c_g^*)\int_{c_g^*}^{\gamma} f(c|c_m)\,dc + w_{res}(c_m, m).\end{aligned}$$

That is, $\Phi(c_m, c_g, t)$ is $(1-t)$ times the employer's compensation cost when the tax rate is t and cut-off values c_m and c_g are chosen. The firm's optimal choices $c_m(t)$ and $c_g(t)$ also minimize $\Phi(c_m, c_g, t)$. Consider $t' > \hat{t}$:

$$\begin{aligned}(1-t')\,\phi(c_m(t'), c_g(t')) &= \Phi(c_m(t'), c_g(t'), t') \\ &< \Phi(c_m(\hat{t}), c_g(\hat{t}), t') \\ &< \Phi(c_m(\hat{t}), c_g(\hat{t}), \hat{t}) = (1-\hat{t})\,\phi(c_m(\hat{t}), c_g(\hat{t})) \\ &\leq w_0.\end{aligned}$$

The second line follows from the fact that $(c_m(\hat{t}), c_g(\hat{t}))$ is feasible but not optimal in the employer's problem, the third line follows from the fact that, holding fixed the employer's compensation plan, decreasing t increases Φ, and the fourth line follows from it being optimal to bundle at $\hat{t}$. Hence, $\phi(c_m(t'), c_g(t')) \leq \frac{w_0}{1-t'}$, and it is optimal to bundle at t'.■

Proof of Proposition 3: *For $c_m = 0$ not to be optimal, it is sufficient that $\frac{\partial \phi(0, c_g^*)}{\partial c_m} < 0$. When*

[30] If it weren't for the fact that the employer must hire a non-negligible mass of workers, the solution to the employer's problem would be to hire only type-γ workers, give them generous coverage, and pay them wage $\frac{w_0-\gamma}{1-t}$. The cost of such a worker would be $\frac{w_0-\gamma}{1-t} + \gamma = \frac{w_0-tg}{1-t}$, which is always less than the no-bundling cost, $\frac{w_0}{1-t}$. Hence it is always worthwhile to bundle in this case.

$c_m = 0$,

$$\frac{\partial \phi\left(0, c_g^*\right)}{\partial c_m} = f(0)\,\tilde{K} - \frac{1}{(1-t)}\alpha, \tag{22}$$

where

$$\tilde{K} = \alpha \int_0^{c_g^*} c f(c)\,dc + \int_{c_g^*}^{\gamma} c f(c)\,dc - \frac{1}{(1-t)} g\left(c_g^*\right)\left(1 - F\left(c_g^*\right)\right).$$

When $f(0) = 0$, *(22) becomes* $-\frac{1}{(1-t)}\alpha$, *which is necessarily negative. Hence if* $f(0) = 0$, $c_m^* > 0$. *By continuity, for* $f(0)$ *sufficiently small, the employer will choose* $c_m > 0$. *Increasing* t *decreases* $\frac{\partial \phi\left(0, c_g^*\right)}{\partial c_m}$, *from which part ii) follows.*■