THE TAX SUBSIDY TO EMPLOYMENT-BASED HEALTH INSURANCE AND THE DISTRIBUTION OF WELL-BEING

MARK PAULY*

I

INTRODUCTION

Health and public-finance economists agree, virtually unanimously, that excluding compensation paid in the form of health-insurance premiums from workers' income taxes causes workers to end up with more health insurance than they would have absent the exclusion. Because the value of the exclusion increases in proportion to a worker's taxable income, this stimulus causes higher-wage workers to be more likely not only to have insurance but to have more generous insurance than lower-wage workers.[1] Despite dissent from some journalists and a few health economists, there is also general agreement that this higher level of coverage leads to increased consumption of medical care, both in quantity and quality.[2] Finally, with the possible exceptions of some very high-risk households, of some very expensive but rare types of care, and of some families at the lower end of the income distribution who still get a modest benefit from the subsidy, there is general agreement that the *additional* coverage induced by the tax subsidy leads to inefficiently high levels of health care spending and administrative costs for insurers.[3]

Although the argument that the tax treatment of employment-based health insurance causes inefficiency is generally accepted, there has been less discussion and less consensus on its distributional implications for high-wage versus low-wage workers. In this article, I will consider some recent contrasting views on the distributional effects of the tax subsidy within employment groups.

1. Charles E. Phelps, *Large Scale Tax Reform: The Example of Employer-paid Health Insurance Premiums* (University of Rochester, Working Paper No. 35, March 1986); Mark Pauly, *Taxation, Health Insurance, and Market Failure in the Medical Economy*, 24 J. ECON. LITERATURE 629, 629–75 (1986).

2. Malcolm Gladwell, *The Moral-Hazard Myth*, NEW YORKER, Aug. 29, 2005, at 44–49.

3. Martin S. Feldstein & Bernard Friedman, *Tax Subsidies, the Rational Demand for Insurance and the Health Care Crisis*, 7 J. PUB. ECON. 155, 155–78 (1977).; I know of no one who argues that it is desirable for some people (with the possible exception of plutocrats like Bill Gates) to have zero coverage; the welfare argument, as far as I know, wholly relates to the incremental coverage and incremental use associated with the exclusion.

I will show that the correct answer depends on what one assumes to be the within-group incidence of the reduction in money wages that occurs when group insurance is part of a competitive compensation package. To do so, I will review the theoretical and empirical evidence on the within-firm distributional effects, in groups with a single and multiple insurance offerings. Finally, I will consider the distributional implications across as well as within groups.

Although the evidence is still incomplete, my tentative conclusion is that the subsidy tends to harm low-wage workers, not by making explicit transfers of money income, but by forcing those low-wage workers to take their income in a form biased toward overly generous health insurance rather than money wages and by restricting their ability to obtain insurance tailored to their preferences. That is, low-wage workers suffer not because income is taken from them and given to others, but rather because they are compelled to use the income they have in less (net) beneficial ways than they would prefer. But this tax-induced incentive to uniformity and costly coverage does not yield a major gain to high-wage workers. Rather, because the subsidy causes substantial overall inefficiency, in aggregate, workers lose, with lower-wage workers suffering more loss than others.

Some terminology requires clarification. Workers as workers differ in terms of their productivity or the value of their skills, which in competitive labor markets will translate approximately into total compensation cost. The relevant redistribution here is that across workers at different productivity levels (measured in money); generally, the terms "high productivity" and "low productivity" distinguish classes of workers. However, almost all the literature discusses the same redistributional concepts by labeling workers by income or wage. Since a low-productivity worker can be in a high-income household (if the household has other sources of income) and since money wages can be affected by the costs of benefits, neither of these terms is as precise as total productivity; income level (money wages) rather than productivity level will nevertheless be used here when needed to be consistent with the literature. And since marginal tax rates are related to a household's total income—not to an individual worker's productivity or compensation—income rather than productivity will be used in discussing tax laws.

One way to phrase the question of equity applied to workers or households with different income or productivities is to ask whether—compared to some alternative tax treatment—the current institutional structure creates an environment under which some subgroups of the population, deserving of a higher welfare level than what would flow from their productivity alone, fall short of achieving such a level. In order to answer that question, one must identify such potential alternative tax treatments and identify those subgroups deserving higher welfare.

The welfare judgment is relatively uncontroversial—there should not be redistribution of resources from low-productivity people to those who are better off. The task of identifying possible taxation alternatives is more

complicated. Two possibilities bear consideration, and these differ in their fundamental assumptions regarding the benchmark or baseline value judgment of the equitable distribution of payment for collectively chosen expenditures.

The first approach assumes that society desires all its members to pay taxes in an amount calculated by applying current statutory tax rates to their true or full income.[4] Applying this assumption, it is clear that laws that permit a portion of one's income to be excluded from taxation effectively under-tax those who make the fullest use of those legal exclusions. If the increased tax revenues achieved by eliminating such exclusions could be used to lower the tax burden on lower-income persons, then the current legal exclusions effectively generate a zero-sum redistribution of wealth from the poor to the rich.[5]

The second approach assumes that society chooses a net or effective marginal-tax burden for each income stratum that collects the amount of taxes it deems appropriate, explicitly taking into account the deductions and exclusions available under present tax law. In this case, the withdrawal of the exclusion would simply result in a lower nominal marginal tax rate for higher-income individuals, leaving the total amount of tax revenue collected from this stratum unchanged.[6] If this distributional pattern is at a political equilibrium, then removal of the exclusion can, by virtue of the differing use of the exclusion, at most induce horizontal redistribution of wealth within the same social stratum.[7] The key point is that the removal does not generate redistribution across income classes.[8] This article generally takes this second view of tax system equity, but it refers to discussions associated with the first view as well.

II

A BENCHMARK CASE

As a benchmark, consider the case of a relatively high-productivity, high-income subpopulation, all members of which currently obtain the same level of tax-excluded health-insurance coverage chosen in some fashion to be their expected utility-maximizing policy, given the value of the exclusion. All

---

4. "True" or "full" income, according to the classic Haig-Simons definition, is annual "income" consisting of all consumption during the year plus the change in net worth. Since health insurance premiums equal covered medical care consumed plus insurance administrative services consumed, they are clearly part of the total consumption that affects a person's well-being. *See* Robert M. Haig, *The Concept of Income—Economic and Legal Aspects*, *in* THE FEDERAL INCOME TAX (Robert M. Haig ed., 1921); HENRY C. SIMONS, PERSONAL INCOME TAXATION (1938).

5. If, however, the increased revenue were to be used to finance additional public spending, the redistributive effect would depend on the nature of that spending.

6. This approach was explicitly taken in the recent report of the President's Advisory Panel on Federal Tax Reform. *See* PRESIDENT'S ADVISORY PANEL ON FEDERAL TAX REFORM, SIMPLE, FAIR AND PRO-GROWTH: PROPOSALS TO FIX AMERICA'S TAX SYSTEM (2005), http://www.tax reformpanel.gov/final-report.

7. James Buchanan and I made this point several years ago. *See* James M. Buchanan & Mark V. Pauly, *On the Incidence of Tax Deductibility*, 23 NAT'L TAX J. 157 (1970).

8. *Id.* at 157.

workers are assumed to have the same risk levels, so that they have the same expected values for benefits under this policy, and the same expected or average out-of-pocket spending.[9] Assume that there are no tax-shielded spending accounts, but that the full premium for the insurance is tax shielded, either because it is treated as fully employer-paid or because any explicit employee premium is tax shielded as part of a cafeteria plan.[10]

It is reasonable to expect that the members of this group would choose a more generous health-insurance plan than they would without the tax exclusion, resulting in increased consumption of medical care.[11] A benchmark case would be one in which this choice has no effect on the insurance coverage or medical-care use of others in the economy, and no effect on the prices others would pay. In other words, there would be no spillovers. This benchmark case would occur only if both the insurance and the health care were supplied by firms experiencing constant returns to scale, with no industry-specific inputs, so that there would be no effects on wages or compensation of workers in other firms.

Although this benchmark case may sometimes be approximated for jobs in many similar-sized firms, there are likely to be spillovers from the behavior of higher-productivity people onto the well-being of lower-productivity people. These spillover effects are in large part responsible for the negative effects of the health-insurance tax exclusion on lower-income households. Broadly speaking, the spillovers discussed in this article come from one of two effects of the tax subsidy: the effect on the cost a household bears for group insurance, or the effect on the determination of the variety of different employment-based insurance products.

III

GROUP INSURANCE

The tax exclusion does not subsidize all health-insurance purchasing. Until the enactment of the Health Savings Account provisions,[12] people could only take advantage of any exclusion if they used it in the context of insurance arranged through their employers. Usually this meant that the exclusion required insurance to be employment-based group insurance. Tax breaks make a difference and encourage people to take their insurance in the group setting. That is not to say that there would be no group insurance without a tax exclusion; rather, the exclusion affects the number of people who choose group insurance and the level of insurance coverage chosen by or for the group. This article deals primarily with the latter case—people who will remain within a group insurance setting but change the level of their coverage depending on the

---

9. The level of tax-excluded health care coverage is chosen to be the expected utility-maximizing policy, given the value of the exclusion to the members of the subgroup.

10. A cafeteria plan, according to Section 125 of the tax code, I.R.C. § 125 (2000).

11. CHARLES E. PHELPS, HEALTH ECONOMICS 344 (3d ed. 2002); Feldstein & Friedman, *supra* note 3; *see also* Pauly, *supra* note 1.

12. These provisions are still small enough to be ignored in most cases.

availability of the exclusion. This article then considers the implications of the exclusion on decisions to drop all insurance coverage or to replace group insurance coverage with individual insurance.

Group insurance makes a distributional difference if there are or can be workers with different productivities or preferences for insurance in the group. In a firm staffed by employees at the same wage level and with the same tastes for insurance and medical care, there is no spillover. With or without a tax exclusion and regardless of any cost advantage over individual insurance, in a homogeneous firm offering a single employer-paid insurance option, all worker wages will fall by the same amount—the pre-tax employer-insurance-premium payment. This is important: Although employers write the checks for their share of the premium, employees pay for that share in the form of lower money wages.[13] In addition, if different firms choose different amounts or types of coverage with different cost-levels, those decisions will only affect wages in each individual firm, without any redistribution across firms.

Although some industries and firms are fairly homogeneous regarding wages, many firms—especially the larger ones—employ workers of varying levels of productivity and, therefore, at varying levels of total compensation; workers in these firms therefore do not necessarily experience the uniform reduction in wages of the benchmark case. Does group health insurance actually redistribute wealth within such heterogeneous firms, and, if so, how? Is there a different net effect on the compensation of some workers compared to others? Speculation about the answer to these questions is wide-ranging.

At one extreme, Len Nichols proposes that the per-worker benefit value and expected cost is the same for all workers, but the wage reduction is greater for higher- than lower-wage workers.[14] He supposes that, in the recent past, until health-insurance costs became a substantial part of total compensation cost, "employers could take enough out of the higher-wage workers' productivity to finance employee contributions for all," so that there need have been no wage reduction for lower-wage workers.[15] This meant that the tax exemption resulted—until recently—in an overall redistribution of wealth from high-wage to low-wage workers. Nichols, however, is concerned that this redistributive effect—though still present—may be eroding under pressure from rising health care costs and spending.[16]

At the other extreme, Clark Havighurst and Barak Richman posit that the bite on wages is uniform across workers at different wage levels but the cost per

---

13. MARK PAULY, HEALTH BENEFITS AT WORK 9 (1997).

14. Len Nichols, *Of Once and Future Kings: The Past and Future of Employment Based Health Insurance* 10 (New Am. Found., Working Paper, 2005).

15. *Id.*

16. *Id.*

worker of providing health care is higher for those at higher wage levels.[17] Thus, they propose that there is redistribution from low-wage to high-wage workers.

Who is right? Probably neither. Strong forces in competitive labor markets produce a result with no net financial redistribution. Two issues here are, one, whether dollar-value health benefits vary with wages, and, two, whether the cost of benefits is distributed in a way that varies (or does not vary) with wages.

On the cost or value-of-benefits side, there has been little direct investigation of the relationship between benefits claimed for a given insurance policy—with some cost sharing—and the income of the claimants. Do higher-wage workers get more (in expected dollars) out of a given nominal benefits policy than low-wage workers?[18] Some economists claim that the effect of income in this cross-section is virtually zero in terms of use, though it may have a positive effect in terms of spending.[19] Estimation of the effect of income per se is difficult to measure precisely and is contaminated by the fact that overall health tends to improve with income or wage level.[20] Similarly, in the Rand Health Insurance Experiment, in which insurance plans had constant coinsurance but involved an income-related deductible, spending (at any level of coinsurance) was found to be approximately equal for people with incomes in the lowest third of the sample and those in the highest income third.[21] Spending by the middle third was lower, though only by about ten percent, and the study estimated that income elasticity—given uniform coverage—was in the neighborhood of 0.2.[22] For example, under policies with the same coinsurance and the same dollar deductible, an increase in household income of ten percent was forecast to be associated with an increase in average total medical spending of two to four percent.

Although the precise magnitude is uncertain, the relationship of productivity, income, or wages to spending, if properly measured, is virtually certain to be positive. This is because, in part, the demand for health presumably rises with income or, more precisely, wage rates,[23] and because higher-wage people will use medical care to improve health and so spend on health care costs rather than take valuable time from work to exercise or

17. Clark C. Havighurst & Barak D. Richman, *Distributive Injustice(s) in American Health Care*, 69 LAW & CONTEMP. PROBS. 7, 42–43 (Autumn 2006).

18. It is possible that, while both high- and low-wage workers value a policy at a level greater than its cost, high-wage workers attach a higher overall value to the policy, thereby benefiting from a greater consumer surplus. That possibility is not considered here.

19. Thomas E. Getzen, *Health Care is an Individual Necessity and a National Luxury: Applying Multilevel Decision Models to an Analysis of Health Care Expenditures*, 19 J. HEALTH ECON. 259, 259–70 (2000).

20. MICHAEL GROSSMAN, THE DEMAND FOR HEALTH: A THEORETICAL AND EMPIRICAL INVESTIGATION (1972); James Smith, *Healthy Bodies and Thick Wallets: The Dual Relationship between Health and Economic Status*, 13 J. ECON. PERSPECTIVE 145, 145–66 (1999).

21. JOSEPH P. NEWHOUSE, FREE FOR ALL? LESSONS FROM THE RAND HEALTH INSURANCE EXPERIMENT 45 (1993).

22. *Id.* at 47.

23. GROSSMAN, *supra* note 20. Grossman finds that health levels are strongly positively related to the wage rate.

improve health. Additionally, the strong positive correlations between income and age and between age and medical expenditure risk are of great importance.[24] These correlations make it virtually certain that higher-wage workers—who are very likely to be older—will as a group claim more benefits than younger, lower-wage workers.[25]

In order to determine the net distributional effect of tax exclusion, one must also determine how insurance premiums are distributed across wage levels. Here, both the Nichols assumption of little or no effect on low-wage workers[26] and the Havighurst–Richman assumption of net redistribution from the lower- to the higher-income workers[27] seem unsupportable and implausible.

The most reasonable conclusion is that if lower-productivity workers get less out of a given group insurance policy than higher-productivity workers, they sacrifice less wages to pay for it. That is, they pay less than the average cost (contrary to Havighurst–Richman) but more than zero (contrary to Nichols). If the wage reduction is tailored to the cost of benefits, there need not be any appreciable net financial redistribution. It is far from obvious theoretically and far from confirmed empirically that important redistribution across wage levels of any type occurs in group insurance.

The theoretical reason why low-productivity workers should sacrifice proportionately less of their money wages in this fashion relates to the constraints in competitive labor markets. A firm offering health insurance cannot compete for workers if it cannot offer a compensation package that compares favorably with packages offered by other employers. Consider the simple case in which the bulk of low-productivity workers work in homogeneous firms providing health insurance.[28] This wage–fringe package sets the market-prevailing level of total compensation and well-being for low-wage workers. This means that an employer with a mixed (non-homogeneous) labor force—offering the same insurance-benefit package—must offer its low-productivity workers cash wages equal to those offered by homogenous firms in order to compete—those wages can then be reduced only by the cost of insurance to low-productivity workers. A heterogeneous firm that tried to shift the cost disproportionately to lower-productivity workers could not attract and retain such workers in competitive labor markets. Furthermore, the employer cannot reduce the wage by an amount greater than the value of the health benefits to low-productivity workers and expect to compete with firms not offering such health benefits. Finally, setting aside the minimum-wage

---

24. INSTITUTE OF MEDICINE, COVERAGE MATTERS: INSURANCE AND HEALTH CARE (2001).

25. U.S. DEPARTMENT OF LABOR, BUREAU OF LABOR STATISTICS, NATIONAL COMPENSATION SURVEY: EMPLOYEE BENEFITS IN PRIVATE INDUSTRY IN THE UNITED STATES, MARCH 2005 6 (2005).

26. Nichols, *supra* note 14, at 10.

27. Havighurst & Richman, *supra* note 17, at 42.

28. The possibility that some low-wage workers may have such low values for health insurance that they would prefer homogeneous firms not providing coverage, as well as the possibility that most firms are not homogeneous, will be dealt with later in the article. *See infra* Part IV.

constraint, there is no reason to reduce the (low) wage by an amount less than the cost of the benefits if the firm is maximizing profits.[29]

Thus, with competition for workers on both the low-productivity and the high-productivity margins, an employer will not be able to reduce low-productivity workers' wages by the average (over all workers) cost of health benefits unless, contrary to the Havighurst–Richman assumption,[30] low-productivity workers actually incur the same benefits costs as high-wage workers. This argument of proportional wage offsets is most plausible for self-insured firms—in which employers explicitly pay whatever benefits costs are incurred.

Most employers will have the ability to set different wages for workers of different productivity levels working in different jobs, making this kind of productivity-level wage discrimination quite feasible. There may be some averaging when workers in a given job category or type vary by risk level, such as redistribution from low-risk to high-risk workers. But there will not be redistribution from low-wage to high-wage workers precisely because different jobs (by definition) can legally and do practically pay different wages.

So in theory, competition sets fairly strict limits on the ways in which wages at different productivity and cost of benefits levels can be reduced. What do we know empirically about how the money wage varies across workers at different wage levels within firms offering health benefits? The empirical evidence on the distribution of benefits-cost within groups has not focused directly on the relationship between benefits and wage levels at different jobs with different productivities. It has focused instead on the relationship between benefits and the wage paid to people at different levels of risk or expected cost for reasons other than the wage level. Studies have examined how money wages are affected for workers of different ages, job tenure, or local market price levels.[31] The idea is that the higher expected expense associated, say, with the older-age (but also usually higher-wage) combination should reduce cash compensation more for older workers in firms that offer health insurance relative to those who do not, or that wages of older workers should be reduced more in high medical cost areas.

Studies[32] reveal strong evidence for a differential incidence of health-benefits costs on wages across jobs or workers characterized by average tenure or age. Whether the differential impacts on wage just track the differential costs of health insurance is less clear (they actually seem to be a little larger).[33] Thus, within-firm incidence is not uniform—and tends to be related to expected

---

29. This is true unless the firm must take into account minimum wage constraints.

30. Havighurst & Richman, *supra* note 17, at 42.

31. *See, e.g.*, MARK V. PAULY & BRADLEY HERRING, POOLING HEALTH INSURANCE RISKS (1999).

32. *See, e.g.*, LOUISE SHEINER, FEDERAL RESERVE BOARD, HEALTH CARE COSTS, WAGES, AND AGING (1999), http://ssrn.com/abstract=165530; PAULY & HERRING, *supra* note 31.

33. SHEINER, *supra* note 32.

benefits cost—but there is no proof that the offset is precise. In fact, some have found that wages do not vary precisely with cost differences associated with chronic conditions in group insurance or in individual insurance.[34] But based on both the strong theoretical arguments and the suggestive empirical evidence, it would be surprising if the cost of benefits was distributed in the close-to-uniform way needed to have had substantial redistribution from low-productivity to high-productivity workers within self-insured firms.

With empirical evidence inconclusive, what do we have to go on and what policy implications for the exclusion should we draw? One way to think about the wage impact for lower-wage workers is to think about the case of firms either adding or dropping a given health-insurance plan. If a firm had a heterogeneous mix of workers of different skills and productivities, what wage would it offer to its lower-productivity or lower-skilled workers when it added a health benefit that (for the sake of argument) provided lower-than-average monetary benefits for those workers because they were less likely than others to use medical care? Would such a firm just reduce wages by the average cost of benefits, or would its management have to be a fully rational, carefully calculating, organized, cost-benefit machine to be perceptive enough to make smaller cuts for lower-wage workers? The firm might start out reducing wages uniformly across the board. But then it would find difficulty in attracting or retaining low-wage but not high-wage workers. It would have to cushion the cut to low-wage workers. Not doing so would imply that low-wage workers were poor job-shoppers—a possibility—but given the higher turnover of low-wage workers and the greater need to squeeze every ounce of benefit out of total compensation, unlikely.[35] So, the firm does not have to be especially sophisticated in benefits policy to produce differential offsets; it only needs to be closely tuned to its labor markets.

Or suppose a firm dropped coverage and then contemplated how much it would need to raise wages to offset the dropped benefits. If it approached this task in a trial-and-error fashion, it would find that it could stop increasing wages at a lower dollar-amount for its low-wage workers than for its high-wage workers. Would anyone believe that it would, nevertheless, offer higher money wages for these workers than it needs to?

The exclusion causes the nominal level of benefits chosen to be greater because it reduces the net tax cost of taking compensation in the form of health benefits.[36] But if there were already (in the absence of the exclusion) little redistribution in group insurance, this increase in coverage would not further affect the amount of redistribution. If, in contrast, redistribution does occur,

---

34. *See* M. Kate Bundorf, Mark Pauly & Bradley Herring, *Health Risk, Income, and the Purchase of Private Health Insurance* (Nat'l Bureau of Econ. Research, Working Paper No. 11677, 2005).

35. JULIA LANE, URBAN INST., THE LOW-WAGE LABOR MARKET: CHALLENGES AND OPPORTUNITIES FOR ECONOMIC SELF-SUFFICIENCY: THE ROLE OF JOB TURNOVER IN THE LOW-WAGE LABOR MARKET (1999), http://aspe.hhs.gov/HSP/lwlm99/lane.htm.

36. Pauly, *supra* note 1.

the real question then is how is it changed when nominal benefits become more generous? Probably any income-related differences in spending become smaller as co-payments fall, which means that generosity tends to diminish the differences based on wages. But higher benefits-cost may make it harder for employers to ignore the need to reduce low-wage worker wages, however much more productive higher-wage workers are. In both the Nichols[37] and Havighurst–Richman[38] models, the tax exclusion would seem to diminish any within-firm redistribution in whatever direction.

Although many workers who obtain health insurance now do so in self-insured settings, among small firms, outside insurance predominates.[39] The premiums charged by community-rated outside insurers to the group generally do not vary depending on the mix of high- and low-wage workers in the group.[40] However, since most external insurance is experience-rated, redistribution may not be the rule.[41]

IV

WAGE VARIATION AND THE DECISION ON GROUP INSURANCE: PRODUCT VARIETY UNDER SUBSIDIZED UNIFORMITY

Subsidized group insurance at a given level of benefits allows little scope for redistribution (or tax-induced worsening of redistribution). But, whatever it does to money flows, it has another impact on the distribution of well-being: the uniformity in nominal benefits that prevails in group settings. In the extreme case, most common among small firms, only one insurance plan is offered.[42] If, as is generally the case, a firm employs more middle- and high-productivity workers than low-productivity workers, the low-productivity workers may be forced to participate in a luxury plan, and this plan may be less cost-effective for them than a more restrictive or lower-cost plan. However, given the assumptions made so far, even though furnishing low-wage workers with the "wrong" plan may be inefficient, the workers should not be expected to pay its full cost if it yields lower net benefits for them than some other preferred, less-costly plan; their wages will fall only by the value of the luxury plan, not by its cost. Thus, rather than result in redistribution, the "'luxury'" benefits package results in inefficiency. And, as insurance has become more costly, the inefficiency effects of uniformity that affect low-productivity workers have probably grown.

---

37. Nichols, *supra* note 14, at 10.
38. Havighurst & Richman, *supra* note 17, at 9–11.
39. Gail A. Jensen & Jon Gabel, *The Erosion of Purchased Health Insurance*, 25 INQUIRY 328, 328–43 (1988).
40. INSTITUTE OF MEDICINE, *supra* note 24.
41. *Id.*
42. KAISER FAMILY FOUNDATION AND HEALTH RESEARCH AND EDUCATIONAL TRUST, EMPLOYER HEALTH BENEFITS 52 (2005).

Uniformity of coverage in the face of variation in preferences may cause inefficiency, but precisely how does uniformity affect the choice of plan and how does that choice affect worker well-being at different productivity or wage levels? To answer these questions, consider a group of heterogeneous firms competing in the labor market to hire two kinds of workers: "high-productivity" and "low-productivity." Assume that firms that hire only one or the other type of labor are not efficient, and, to make the example simple, assume the firms must hire the two labor types in equal proportions. Furthermore, assume that the administrative costs of offering multiple insurance plans are so high that all firms will offer only one plan, if they offer any insurance at all. Also, assume no unions or state laws affecting what benefits can or must be offered. Finally, assume that the higher-wage workers would prefer a more generous insurance plan than lower-wage workers, even if the premiums paid were tied only to the workers' expected expense—what they would choose as an "isolated optimum." This preference for a more generous plan is a result of the large tax subsidy offered to high-wage workers and because generous, or more hassle-free, coverage may be a normal good.

Given these assumptions, what level of coverage will end up being offered? An obvious conclusion is that the final plan is likely to be somewhere between the two isolated optima (including the endpoints). But where will it fall on that spectrum? The answer requires answering yet another question: How are choices made by or for groups? A number of plausible theories address this question.[43] The group might choose the optimum for the average-wage worker. Alternatively, it might choose the optimum for the person with the median wage, which would happen if employees voted and only the evaluation based on the wage mattered. It might instead choose an optimum that maximizes the sum of the net benefits to all workers. Finally, the group might choose the optimum level of some dominant subgroup of workers—whether top management, union executives, or the workers with the loudest voices.

Another theory that seems to have a good theoretical pedigree but that is more complicated than any of the above is worth exploring. It might be plausible to assume that management will set the benefit at a level that minimizes the total compensation cost of whatever number of workers it chooses to hire. Health benefits are an important part of the total compensation package, and an employer offering more such benefits, if those benefits are sufficiently valuable to workers, will allow the employer to continue to attract workers even at lower money wages. If the reduction in wages that can accompany an increase in benefits is larger than the cost of the benefit, the total compensation cost will fall. Both high intrinsic valuation of insurance and the tax exclusion mean that, at least at some point, benefits might be worth more than wages.

---

43. Gerald S. Goldstein & Mark V. Pauly, *Group Health Insurance as a Local Public Good*, *in* THE ROLE OF HEALTH INSURANCE IN THE HEALTH SERVICES SECTOR 73, 73–110 (R.N. Rosett ed., 1976).

One way to think about how an employer might solve this problem is to assume a supply curve for each type of worker that reflects both the worker's wage and his level of benefits, given the wage–benefit opportunities available elsewhere in the labor market. In effect, the employees willing to work for the sample firm will be those for whom that firm's total compensation package is as or more valuable than whatever else they could find in the labor market.

Consider some uniform benefit package B_1. There will be a supply curve of workers of each type at various monetary wages where B_1 is offered. There will also be a per-worker cost (assumed to be uniform) for this benefit; shifting up the money-wage–labor-supply curve for each type by this cost will generate a supply curve in terms of total compensation. Take a simple average of these two supply curves (given the assumed fifty-fifty split between low- and high-wage workers). The resulting curve will show the average compensation cost per worker for any quantity of workers (of either type). Now consider a more generous benefit package B_2. The money-wage–labor-supply curves shift down; however, the effect on the total compensation–cost curve will depend on how the benefit is valued relative to its cost by the marginal worker of each type. Depending on the level of benefits offered, the valuation tradeoff between wages and benefits will yield a curve representing the average compensation cost per worker of each type for a given total number of workers of that type.

Initially, for both high- and low-wage workers, total compensation will fall because benefits are worth more than money. This curve will decline and will reach a minimum at some level of B for each type of worker. That minimum value will be larger for high-wage than for low-wage workers.[44] The employer can then add the curves vertically and choose that level of benefits that yields the lowest value of this summed curve.

At the optimum, moving the package toward more generous coverage will attract higher-productivity workers at a lower net-compensation cost but will require higher net-compensation cost to attract lower-productivity workers. More generally, the distribution of wages of marginal workers in most firms will have a large share of workers who are receiving the most generous tax subsidies. This group is enthusiastic about generous health-insurance coverage. The lower-productivity workers—who prefer less-generous coverage—will usually be in the minority. Bearing this employee pool in mind, most firms will provide generous health care coverage, generating the attendant inefficiency discussed above.

The inefficiency associated with low-productivity workers getting more-generous insurance than they would have chosen in a homogeneous low-wage firm actually has two dimensions. Low-productivity workers are forced toward the more-generous plans desired by high-wage workers by the mechanism of group choice; that shift is then further strengthened by the more-generous tax

---

44. The minimum corresponds to the quantity of coverage at which the marginal benefit equals the marginal cost for that number of workers.

subsidy realized by high-productivity workers. The first shift is inefficient relative to a situation in which this mixed group of workers could have hypothetically provided different insurance policies at the same administrative costs, so it may not be practically relevant. The second shift is a different manifestation of the inefficiency associated with the differential tax subsidy; rather than have the larger increase in coverage for high-productivity workers that would occur in a high-productivity-only firm, the coverage for high-productivity workers shifts less in the mixed firm, but that is offset by a larger shift for low-productivity workers, who benefit little, if at all, from the exclusion.

However, in the model so far, this inefficient spillover does not reduce the welfare of low-wage workers because the value of their total compensation package cannot fall below the value in the homogeneous low-wage firm. Instead, the cost of the inefficiency falls directly on the owners of firms employing mixed labor forces and is typically passed on to their customers.

V

MODIFYING THE MODEL

What if it is assumed, more realistically, that the mixed firm is the rule rather than the exception? If that is the case, the story becomes more complex, and it is possible that the inefficiency costs depress average compensation of all workers, high- and low-wage alike, compared to the setting without the exclusion. Low-wage workers will be among those adversely affected, but the cost probably would be distributed fairly evenly across the board. The other possible assumption is that workers do not all have to accept a single plan. Instead, they may choose among a variety of plans or choose to take no coverage at all. This latter behavior is the most socially concerning, so it is addressed first.

In some segments of the workforce health care coverage is eroding.[45] This occurs primarily in smaller firms, including ones in which many workers are not low-income.[46] These firms have not stopped offering coverage. Instead, they have ceased offering coverage to all workers, excluding many by establishing waiting periods or work-hours requirements. Additionally, these smaller firms have begun charging explicit employee premiums (which can be avoided if coverage is rejected) that apparently are large enough to bring about rejection.

One explanation for this phenomenon, suggested by Len Nichols, is that although the value and cost of health insurance have increased over time, the growth in value for some workers, especially lower-wage workers, has fallen behind the growth in cost.[47] Whereas fifteen years ago a feasible health-

---

45. KAISER FAMILY FOUNDATION, *supra* note 42, at 35.

46. *Id.*

47. Nichols, *supra* note 14, at 10. "Value" here means the amount a person is able and willing to take out of his or her budget to pay for insurance or care.

insurance policy might have had a value for a low-wage worker that exceeded its cost, that is now no longer the case. Instead, some low-wage workers prefer compensation packages that offer no group coverage but pay higher monetary wages. The key point is that, if this is correct, the system does not need to invoke a process of redistribution to high-productivity workers. In fact, the only redistribution low-productivity workers would be trying to avoid is that associated with the transfer of wealth to doctors and hospitals that may arise as a result of overly generous insurance plans.

The other possibility is that a firm may offer workers a choice between multiple policies. In general, this approach would reduce spillovers either on wages or on product types. If the low-productivity workers chose a less generous policy at a lower explicit worker premium, the high-productivity workers chose a more generous policy at a higher explicit worker premium, and the employer contribution were a uniform dollar amount per worker, there would be no need for differential incidence and no reason why the larger tax exclusion would affect the policy choice of other workers. Some employers do try to make fixed-dollar contributions across the board, others try to do so but in a risk-adjusted fashion, and still other firms contribute a fixed proportion of premiums, as a kind of poor-man's risk adjustment.[48] Thus, the availability of cafeteria plans means no tax subsidy is lost.

The reasonable conclusion here is that multiple choices help to reduce spillovers but do not necessarily eliminate them. For example, low-wage workers are more likely than high-wage workers to choose a cost-constraining, aggressively managed care plan.[49] Also, spillovers may be seen in terms of out-of-pocket payments, though the impact of wage levels on the size of such payments is a little more difficult to determine in theory and to observe in practice. While low-productivity workers would value low premiums, they would also be more affected by a given deductible. Evidence suggests that these workers do not tend to choose permissive plans with low out-of-pocket payments; rather, they tend to select aggressive managed-care plans or indemnity plans with higher cost-sharing.[50]

One possible outcome is that low-wage workers will choose "mini-plans" that have modest deductibles but low upper limits, especially on inpatient benefits.[51] They might expect to receive bad-debt or charity care for the bulk of their hospital bills and seek only to ensure coverage for expensive drugs and devices for which charity care is much less common.

---

48. This approach will usually mean a larger contribution to the more generous plan chosen by the less healthy but wealthier older workers.; *See* Alain Enthoven, *The Fortune 500 Model for Health Care: Is Now the Time to Change?*, 27 J. HEALTH POL. POL'Y & L. 37, 40 (2002).

49. Anne B. Royalty & Neil Solomon, *Health Plan Choice: Price Elasticities in a Managed Competition Setting*, 34 J. HUMAN RES. 1, 1–41 (1999).

50. Melissa W. Barringer & Olivia S. Mitchell, *Workers' Preferences Among Company Provided Health Insurance Plans*, 48 INDUS. & LABOR RELATIONS REV. 141, 141–52 (1994).

51. Vanessa Fuhrmans, *More Employers Try Limited Health Plans*, WALL ST. J., Jan. 17, 2006, at D1.

Finally, removal of the tax subsidy probably would appropriately reduce the availability of group insurance benefits at some, though by no means all, firms. Group insurance would disappear for firms for which the advantage of group insurance and lower money wages are outweighed by higher money wages and freely chosen individual insurance. These employers would generally be small firms, low-wage firms, and heterogeneous firms. If the workers then turn to individual insurance, would there be less of a chance of redistribution (from either low to high or high to low)? Individual insurance is risk rated in states that do not forbid such rating, but a buyer's income or wage rate has not been used in the underwriting process.[52] Instead, age and health status are by far the most important criteria considered by underwriters when issuing new policies.[53] Thus, should there be redistribution in the Havighurst–Richman sense in group insurance, such redistribution would also occur to some extent in individual insurance. So, again, the effect of removing the health care tax exclusion on purely financial redistribution might be negligible.

VI

THE MARKET FOR INSURANCE AND CARE

Thus far this article has discussed the effect of the health care tax exclusion on the kinds of insurance that low-wage workers receive and the price they pay for it. Since the additional insurance high-productivity workers receive as part of the more generous benefit package affects the amount and type of medical care they will consume, there might potentially be a spillover onto the medical care available to low-productivity workers. Such medical spillovers would be economically significant even if there are no spillovers within insurance markets per se. How might this work?

With virtually all products, higher-quality goods cost more. The quality level high-income people find attractive tends to be higher than that which low-income people prefer. This is true for restaurant meals, housing, entertainment, college choice, et cetera. In most cases, however, a tax-fueled increase in demand by higher-income people would not affect the well-being of low-income people. If higher-income demand for fine wines and food surges, that surge need not affect the welfare of people on a beer budget at a fast-food restaurant. That is, in the usual case, it is possible to buy in separate markets with different levels of quality and price. As long as the services or products are supplied by competitive firms at constant marginal cost, changes in purchases in one market will not affect other markets.

However, this may not be the case for medical care for several reasons. First, explicit or implicit legal restrictions and incentives often discourage the

---

52. SCOTT E. HARRINGTON & GREGORY NIEHAUS, RISK MANAGEMENT AND INSURANCE 446 (1999).

53. Bradley Herring & Mark V. Pauly, *Incentive Compatible Guaranteed Renewable Health Insurance*, 25 J. HEALTH ECON. (forthcoming 2006).

provision of low-quality (even if low-cost) care or discourage variety more generally. The need to meet a community standard of care—generally as protection against malpractice actions—and the presence of regulatory rules prohibiting low-quality care may mean that an increase in demand for high-quality care caused by too-generous insurance for high-income workers may harm low-income workers by depriving them of the ability to obtain low-cost, decent, but lower-quality care. That is, the lower quality level desired (given the cost savings) by low-income workers may become less available as the tax subsidy differentially increases the demand of high-income people. In addition, to the extent that each medical provider treats people at different income levels, there may be an explicit or implicit cost to offering different quality levels of care. An individual doctor in private practice may find it hard to offer high-quality, high-cost care to well-insured, high-income patients while providing a different style of care to lower-income patients. Even if the high-income patients have generous PPO or indemnity coverage and the low-income patients have restrictive HMO or high-deductible coverage, both the practical difficulties of managing two styles of care and the ethical discomfort inherent in offering varying levels of care may, paradoxically, make lower-income people worse off overall even as they consume very high-quality medical care. The reason is that the large reductions in other aspects of their consumption required by the need to pay for high-quality care may make them worse off in terms of total welfare, even as their medical-care consumption improves.

Thus, an inefficient outcome is possible. Unfortunately, economic theory does not tell us much that is definitive about welfare effects in a world with a limited number of differentiated products. With competitive free entry, there may be too many products or too few, relative to an ideal world in which the number and variety of products just matches the distribution of preferences. However, it does seem likely that a differential subsidy to consumers of high-quality products will have negative spillover effects on consumers of low-quality products in this type of interdependent model.

One solution to the problem of product differentiation is for sellers to specialize in one specific quality level. This surely happens to some extent in medical markets, for providers of care in poor neighborhoods tend to be lower-quality, concierge medicine is offered to the rich, and hospitals seem to stint on quality for the uninsured.[54] Although the closed-panel HMO model like Kaiser has had difficulty spreading in the United States, it solves the individual physician's problem by, in effect, sending the physician only patients who are seeking the same quality level of health care.

Finally, some medical goods and services such as patent-protected drugs and physician consultations may not be supplied competitively at constant marginal costs. In either case, an increase in demand by heavily subsidized higher-

---

54. Jean M. Mitchell & Jack Hadley, *The Effect of Insurance Coverage on Breast Cancer Patients' Treatment and Hospital Choices*, 87 AM. ECON. REV. 448, 452 (1997).

income people may drive up prices that others will have to pay. This increase is inefficient only to the extent that the initial subsidy is inefficient, but it definitely means that lower-income persons are made worse off as higher-income persons are made better off. These effects are likely small in the aggregate, with hospital care being highly competitive, but in smaller market areas the effect on the cost of drugs and specialized physician services may be more significant.

Based on this analysis, it appears that the main negative effect on lower-wage consumers of the high-tax subsidies for the well-off occurs through this type of quality spillover. Is there any way to get an idea of the magnitude of this spillover? First, assume that increases in the quality and price of health care do not cause lower-income workers to change their insurance coverage. Instead, they respond primarily to higher prices, which translate into higher co-payments, by using a smaller quantity of care. The main adverse impact, then, is that low-income workers pay more for health care but get less benefit as compared to their own optimum quantity and quality level of care. Assuming that the tax exclusion causes high-wage people to choose more generous subsidized insurance, which in turn increases the cost of such care ten to twenty percent,[55] this sets an upper bound on the maximum reduction in well-being experienced by lower-income people.[56] The static welfare loss therefore is not negligible.

VII

SPILLOVER EFFECTS, NEW TECHNOLOGY, AND SPENDING GROWTH

Any substantial negative effect of the tax exclusion on lower-income people is most likely to arise in the dynamic context of changes in quality and cost of health care over time. The primary empirical facts here are that (1) the bulk of the reason for spending growth in the United States is the development of beneficial, but costly, new technology and (2) the valuation of this new technology is proportional to income level. The extent to which the tax exclusion has affected the form and amount of new technology is as yet unknown.

Here is a story which, if true, would make the connection most obvious: Suppose that consumers can choose among insurance plans that adopt different policies toward costly, new technology. Suppose also that lower-wage workers, in an effort to limit the amount of increase in their real income going to medical care, would choose policies that limit cost by providing smaller amounts of new technology; on the other hand, higher-wage workers would be more able and willing to devote more of their income to new medical care, preferring to invest more in higher-quality care. Now, if the higher-wage workers receive a larger health-care tax exclusion, they will presumably demand a higher rate of

---

55. PHELPS, *supra* note 11, at 354.
56. The actual amount is probably less, since the additional quality will be worth something.

improvement and innovation in health care, resulting in a higher rate of growth in health care spending. Will this adversely affect lower-wage workers?

The answer to this question is in the spirit of the discussion of the static case. There will be a spillover if providers are unable easily to supply disparate levels of care to the two groups, providing access to new technology to some but not others. The reasons why this might occur are the same as those discussed before—legal restrictions and professional standards of behavior.

Thus the health-insurance tax subsidy harms low-income people in a dynamic context by fueling the already-present eagerness on the part of higher-income people to have access to the latest, expensive technology. The cost of this new technology, when incorporated into insurance premiums, makes health insurance so expensive as to be harmful to the overall well-being of lower-income people. Some low-income workers may be a little healthier, but they will also be much poorer. In principle, the health care market could offer an alternative insurance plan that would be less permissive in its coverage of new technology; such a plan would allow lower-income people to avoid the "medical poverty" phenomenon. Unfortunately, so far, such an option has not emerged in the market.

VIII
CONCLUSION

The tax subsidy to employment-based group insurance does harm lower-income, lower-productivity workers, even though those workers have incomes low enough that the direct effect of the presence or absence of the subsidy is minimal. It does so not by redistributing the financial cost of care or insurance, but rather by imposing higher and more rapidly growing costs for both on lower-income people. Even if abolition or limitation of the subsidy did not affect the distribution of the tax burden across income strata, lower-income people would benefit because they would experience lower-cost health care, lower-cost insurance (paid explicitly out of after-tax income and implicitly out of wages), and a lower rate of long-term growth in both. The net gain would be smaller than the reduction in financial burden, because the more attractive care and the more lavish insurance would be worth something, but the savings would more than offset this loss.

Indeed, though lower-income people would gain from limits on the tax subsidy, under the assumption of an unchanged distribution of the tax burden, higher-income people would gain even more. They too would be induced to be more frugal and could capture the financial savings from their more economizing choices directly. That is, they would be induced to choose less-generous insurance and would therefore have higher money wages and more to spend on other things they value more highly than the insurance they forego.

The situation therefore is not one in which the rich now gain at the expense of the poor. It is one in which both groups now lose, and both could gain from changes in policy. The possibility for mutually beneficial improvement is, after

all, what is meant by "improved efficiency," and in this case it seems that the gains would be shared naturally across income groups.

However, there is a potentially strong impediment to this happy outcome: a reduction in equity as applied to the consumption of medical care and the attainment of health. Removal of the tax subsidy would worsen access to highest-quality and latest-technology medical care. It might lead to greater disparities by income level: although the incentives to economize would increase more for the higher-income people who benefited more from the subsidy, the removal of subsidies to more uniform group insurance might lead to more differentiation in coverage based on income. That overall welfare is improved would be positive for true believers in welfare economics, but it may be small consolation for health-policy analysts who often overly focus on health to the exclusion of other dimensions of welfare. There can be more important things than your health, but this may remain a hard message to spread and therefore a difficult guide to public policy.