Limiting the Tax Exclusion for Employment-based Health Insurance: Are Improved Equity and Efficiency Enough?

Abstract - Removal or limitation of the current exclusion of employment-related health benefit payments from taxable income would improve economic efficiency and equity, but is this sufficient to make such a policy change likely? This paper uses some concepts from public choice theory, based on earlier work by Buchanan and Pauly on the incidence of tax deductibility, to suggest that such a change is more likely in the current context of broader tax reform involving a desire to reform the functioning of the health insurance market and to increase the tax share of upper middle income households.

INTRODUCTION

Health insurance premiums arranged through employment are excluded from taxable income at all levels of government in the United States. The value of this exclusion rises with income, is uneven within income categories, and causes both insurer administrative costs and medical care spending to be higher than would occur without the exclusion. Despite the inequity and inefficiency associated with this tax provision, it has proven politically difficult to have serious consideration of its limitation or elimination. However, in part because of the need to raise revenue to provide subsidies to lower income people at high risk for being uninsured, and because of the federal government's overall need to close the budget gap, the possibility of changing the tax exclusion has again entered the policy debate.

There are two other ingredients in the current discussion that are new. The first is that the Administration has embarked on a broad plan of tax reform that would increase the tax share of higher income households. The second is that, in addition to subsidizing insurance, the Administration's health reform program needs to consider ways to slow the growth in medical spending (sometimes imprecisely called "cost containment"). In this paper, I will argue that these two new considerations might make some modification of the exclusion politically feasible for the first time. I will develop some simple public choice models to show why individual provisions of tax deductibility and exclusion are generally hard to

Mark V. Pauly

Department of Health Care Management, Wharton School, University of Pennsylvania, Philadelphia, PA 19104

change, but claim that the combination of tax reform and health reform may create an opportunity to do so. I will then propose a way of limiting the exclusion that links its revenue raising potential with the goals of tax reform and cost containment, while simultaneously improving equity. I will argue that this combination may have more bipartisan appeal than any of the individual components alone.

THE TAX EXCLUSION, TAX EQUITY, TAX EFFICIENCY, AND HEALTH SYSTEM EFFICIENCY

The tax exclusion arises because employer payments toward employee health insurance or (in the case of self-funded firms) employee health care are excluded from the federal income and payroll tax bases. In addition, explicit employee premium charges can be shielded from federal taxes if the employer puts a so-called "cafeteria" benefits plan in place. Employer contributions are also excluded from state and local tax bases, and in many cases so are the employee payments in cafeteria plans.

The distribution of excluded income by household income level is skewed toward higher income households. Such households are more likely to have employment-based insurance, to have more generous coverage, and to live in areas with high medical care spending (measured either by spending per capita or medical input price indexes). The progressivity of the income tax also means that the tax avoided by the exclusion generally rises with income. The upper limit to the Social Security tax alters this conclusion somewhat but the effect is small. The net effect is that the great bulk of hypothetical taxes avoided by the exclusion are skewed toward higher income households. While the benefit measured in this way definitely increases with income, because insured spending and premiums increase less than proportionally with income, the

value of the benefit is not "progressively" distributed; it does not increase so rapidly with income that the benefit increases as a proportion of income as income rises.

Because the tax exclusion applies only to medical spending covered by insurance payments, it provides a financial incentive to use insurance to cover the cost of care. The average administrative expense proportion in employment-based insurance (the "administrative loading") is about 13 percent of premiums. For a household with a marginal tax rate greater than 13 percent (virtually all working households, given the combined employee-employer federal payroll tax for Social Security and Medicare), this means that the tax advantages more than offset insurer administrative costs, so it is less costly to the household to pay for care through insurance than directly out of pocket (even ignoring the discounts from list price that insurers often receive). However, there are real resource costs to running payments through the insurance system, administrative costs that the tax exclusion hides. In addition, it is generally the case that more generous insurance coverage will cause patients to increase their use of costly health care services more than if they were less well insured; this "moral hazard" can lead to inefficiency if additional coverage causes the use of care to be worth less than its cost without sufficiently offsetting risk reduction benefits. Thus the tax exclusion is probably a major contributor both to high medical spending levels and to high private insurer administrative costs, two much-decried aspects of the U.S. medical care system.

Though the tax exclusion is a loophole that fails to provide a net benefit to the general population, it does benefit a small portion of the population, mainly middle and upper income households. While encouraging more generous coverage for middle and upper middle income households in general does not contribute to any socially defined equity or efficiency goal,

the exclusion does cause more people to have insurance through the workplace than would otherwise be the case, and may cause a few more middle-class people to have at least some insurance; thus, it may reduce the number of uninsured (compared to the no-exclusion scenario). However, the likelihood of being uninsured is strongly related to income. The concept of ideal insurance coverage has never been well defined, but it is surely the case that, among those with some insurance, the likelihood that cost sharing discourages use of high benefit care is going to be greater among lower income households than among higher income households. Tax credits that are more generous for low income populations would obviously be a better subsidy mechanism than continuation or even elaboration of the tax exclusion.

TAX EXCLUSION OR DEDUCTIBILITY AND TAX REFORM

Up to this point I have treated the tax that would have been collected on compensation for a given set of taxpayers as a measure of the value of the exclusion. However, as James Buchanan and I pointed out many years ago, this measure of benefit may not be appropriate in a realistic model of political determination of tax policy (Buchanan and Pauly, 1970). Specifically, the system may need to observe a political equilibrium with regard to broad distributional parameters. If removal of the exclusion would substantially increase the tax collections from a population group, that group may be able to block the removal, or may demand some other tax concession as a substitute for this sacrifice. In formal welfare economics, this should not be a problem. Suppose the set of upper middle income taxpayers (say, households with incomes over $100,000) now all obtain about the same levels of employment-related insurance coverage, and suppose that, because

of the exclusion, the generosity of coverage is pushed beyond the efficient level. By the definition of Pareto optimality, there should be a way to change things that would benefit all in this group and thus get unanimous approval. But that change would consist of cutting or eliminating the exclusion, and then reducing marginal tax rates to absorb the higher tax collections. It may well be that the broader political distaste for cutting taxes on "the rich" would prevent this ideal change, and yet still leave the inefficient and inequitable initial state intact.

To move away from this situation, we need something to disturb the political equilibrium. Suppose that there is a change in the distributional equilibrium that involves raising the total tax burden on the rich. One way to do so would be to raise marginal income tax rates, but that involves distortion and excess burden. A better strategy is to look for tax changes that remove existing distortions—changes with negative excess burden. Here doing something about the exclusion would seem to be a prime candidate—and probably more preferable (both politically and normatively) than changing the tax deductibility of charitable contributions, home mortgage interest, or even small firm depreciation.

WHAT TO DO WHEN

Taking as given the desirability of "doing something" about the exclusion, I now offer some thoughts on how to do so in ways that combine improvements in equity and efficiency with political feasibility. To begin, it seems desirable to limit such changes to better-off households. Not only is this apparently more equitable, it is also where the money is, where the major distortions are, and where the modest potential good effects of the exclusion, like encouraging the otherwise-uninsured to become insured, are likely to be small. That is, it makes limiting the exclusion

progressive in its (potential) net tax impact, it does so in a way that generates negative excess burden, and it potentially helps finance insurance for the lower income uninsured.

Second, curtailing the exclusion for only part of the working population means that the potential negative impact on group insurance is mitigated. There are relatively few firms with labor forces dominated by workers with six-figure incomes. While there is almost always some entity on any margin, the set of firms that would be pushed to drop insurance because management paid more taxes on compensation as insurance is surely a thin margin. Given the substantial administrative cost savings for group insurance, adding taxes for some higher wage workers is not likely to lead the group to want to drop coverage. The high wage workers are unlikely to be the marginal workers when it comes to the presence or absence of some group coverage.

Third, these high wage workers may be those whose preferences are disproportionately driving the generosity of coverage. That is what some commentators strongly suggest (Havighurst and Richman, 2006), and there surely are some product quality spillovers especially in small firms where economies of scale in insurance buying require all to share the same coverage if they take any coverage at all. Hence curtailing the subsidy to the well-off may lead to a preference by them for lower cost health plans, something that can increase the "affordability" of coverage within existing groups, and because of the apparent importance of spillover effects on the quality and intensity of coverage actually lower insurance costs across the board. (I would not argue that getting well-off workers to want less coverage—which I suspect largely will take the form of agreeing to more restricted managed care than excising some supposed "gold plated" features—will in itself make coverage affordable or attrac-

tive for lower income workers.) A truly low priced plan in the absence of subsidies will require giving up much more than even lower income people have been willing to do, and will make "underinsurance" look worse. The conversation about what level of insurance people might be encouraged to have when they should have at least some insurance, as opposed to none, is a conversation we have yet to have. We do not know what affordable insurance is best, but we do know that some insurance is better than no insurance, and that only affordable insurance is worth considering. So we should seek the affordable insurance that is better than nothing, rather than holding out for the ideal insurance that many are unable or unwilling to buy.

A graceful mechanism for limiting the exclusion already exists in the way group life insurance is treated. People who obtain excess group life insurance simply see higher taxable compensation on their W-2 forms at the end of the year. No revolutionary changes in the way wages are paid or pay is reported need occur.

WHAT TO LIMIT

It seems clear that limits on exclusion should be linked in some way to the value or desirability of spending; insurance spending should be taxed if it is of low social value and left untaxed if it is of high social value. While in principle the value of health care or the social desirability of spending might be specified analytically and estimated with empirical information (for example, cost effectiveness calculations), the final decision on what is socially worthwhile will still necessarily be a political one. Therefore we might look for political consensus on what can be cut as our first approach to specifying limits on the exclusion. That is, we at least want to exclude from a tax subsidy spending that we think ought not to occur. Since we think spending is growing too rapidly, we

must think there is some new spending that should not be incurred.

An example of how a process to incorporate this thought into tax policy might work is as follows. A large number of trade groups recently agreed with the administration to take actions to pare 1.5 percent off the rate of growth of medical spending, lowering projected spending growth from 7 to 5.5 percent per year. They agreed to "bend the curve" to what is presumably a more appropriate level. A possible model for limiting the exclusion would then be one in which the value of the exclusion for high income households was permitted to grow (per worker) in a tax-excluded fashion only up to 5.5 percent per year, with any excess growth over that level subject to tax. Informally, if "we" agree that spending in excess of 5.5 percent is violating our social goals, we should at least be able to agree not to subsidize it.

A second example illustrates the same point. The value of the exclusion per worker for a given nominal insurance policy varies across geography in part because premiums per worker vary across geography. High cost states and cities have opposed a flat dollar limit on the exclusion on the grounds that they should not be penalized for being areas where medical costs are high, for example, because the cost of living is high. (There is surely a positive correlation between medical prices and consumer prices for other types of products in a given area, though the correlation is far from perfect; controlling for consumer price levels, some areas spend much more on medical care than others.) A reasonable social decision then would adjust the value of the exclusion by some index of medical input prices (such as the indexes used to adjust Medicare payments across geographic areas). Thus a worker in a high price area would have the amount of spending attributed to his income adjusted downward, while a person living in a low price area would have

it adjusted upward. We know from studies by the Dartmouth group, however, that there is still substantial variation in input-price-adjusted spending across areas (Wennberg, Fisher, and Skinner, 2002).

There are two possible reasons for the remaining high spending. One is that the higher spending represents higher quality of care in some areas than others. There surely are parts of the country where the availability of doctors and high quality hospitals is low; these mostly rural areas do have low average levels of spending. While higher quality in the abstract is a good thing, it probably is not good social policy to subsidize it more for some than for others. So if an individual lives in an area where high medical costs are associated with easier and more convenient access to high quality but costly care—especially new technology—that good fortune should not result in lower taxes on a given level of total compensation than people elsewhere pay.

The other possibility, which the Dartmouth group asserts to be much more common, is that the higher costs represent spending on inefficient services that do not improve outcomes. In this case, it also seems desirable to terminate the tax subsidy, as we do not want to subsidize inefficient spending. Taxing benefits in high cost areas is a less draconian approach to the problem of curtailing area variations in spending than the idea of refusing reimbursement for such spending, recently considered by the Senate Finance Committee.

In all cases, of course, simple tax changes are no substitute for good and nuanced policy judgment if that judgment can be obtained at low administrative cost. Concretely, one would want to have in place a process to exempt some increases in spending over the target rate from the exclusion limit if it was determined that particular year's new technology had a very high health value relative to its cost, and a geographic area might get a pass

on a limit if its high spending arose from larger proportions of truly sick people (who nevertheless are generally uncommon among the well-off). But the concept here is that the benefit of the doubt would be given to the hypothesis that high spending growth for high income people is not of great social value and so is not deserving of subsidy at the margin. Moreover, if we tie the limit primarily to the growth of spending and not its level, we can avoid much of the need to make cross-sectional adjustments in the dollar amount of excludable expenditures.

THE PARADOX OF REVENUE PRODUCTIVITY

James Buchanan and I also considered in our paper that tax funds are collected to pay for public goods, and offered some thoughts on the link between deductibility and the level and financing of some hypothetical (but unrelated) public good. One of the main motivations for bringing discussion of the health insurance exclusion back to life in the current setting is not concerned directly with the equity or efficiency arguments I have reviewed here, but rather with the need to raise money to pay for the Administration's goal of more heavily subsidizing health insurance for those lower income households who now benefit neither from the tax subsidy nor from existing public programs.

In the pure theory of public finance, earmarking (explicit or implicit) has no comfortable role to play. Taxes should be raised from the most efficient and equitable sources and spent on items of highest net value. (Possible interactions between the value of public goods and the taxed good make the theory of optimal taxation more complex than this simple statement, but those interactions will be ignored.) If taxing health benefits at the margin for the better-off is a way to raise revenue with negative excess burden and also is

felt to be equitable, those funds should in principle be used for whatever category of public spending generates the most net benefit. A worthy public spending project might be subsidizing the uninsured, but it might be investing in infrastructure or any other social goals. There are some earmarked taxes that function as marginal-benefit-based user charges, like gasoline taxes to pay for highway maintenance, but that consideration would not obviously apply to taxing the health benefits of the well-off—they get nothing directly from covering more of the uninsured (other than, perhaps, a less troubled conscience). A clever economist can always think of something. It turns out to be true that having a large proportion of uninsured in your community adversely affects your access to high quality care even if you are a well-insured high income person (Pauly and Pagan, 2007). But more realistically, the reason for targeting the tax subsidy to health insurance is that it falls more within the purview of the Congressional committees charged to develop plans to subsidize the uninsured—and not because of some deep theory in public economics.

It is worth noting that taxing the exclusion as a way of raising the revenue that is crucially needed to finance the programs for the uninsured raises some issues that are similar to those that arise in connection with sumptuary (sin) taxes. The more the tax improves efficiency and reduces excess burden, the less money it collects. For example, Paul Fronstin has offered the opinion that a tax on benefits limited in ways that have been discussed might not collect as much revenue as a tax on benefits (Abelson, 2009). He does not say why, but one possibility is that a tax that caps the exclusion might drive almost all insurance premiums back close to the level of the cap. However, health insurance premiums and, by assumption, insurance coverage and health spending, would be

cut. Whether benefits are cut or remain the same, more income would be subject to taxation and thus more taxes would be raised. If that reduction in spending by the well-off lowered premiums the currently uninsured face, it is even possible that a change in taxes that collects little revenue would do as much to reduce the number of uninsured as a tax that paid for a generous subsidy. Lowering premiums by lowering real resource costs would generally be more beneficial to welfare than lowering premiums by subsidizing high priced insurance. Assuming that capping the tax exclusion would lower premiums for everyone is an overly optimistic view, but the general point is that, in this case, limiting the exclusion will collect more taxes with at least a zero excess burden (if benefit spending does not change) or a negative excess burden (if it does). In principle taxing benefits might lower labor supply, but generally this tax change will not change the marginal tax rate on earned income.

The other possibility that Fronstin may have had in mind is that limiting the addition to the tax base to health insurance spending in excess of some target for only a fraction of workers may not add much to the revenue base. However, with $200–400 billion of potential tax revenue, there would still seem to be enough to matter, though perhaps not enough to pay for covering lower income uninsured households. If the cap is limited to growing at a lower rate than the value of the currently excluded premiums, tax collections (relative to the counterfactual) will grow over time.

More generally, is there any reason in political economy to suggest that taxpayers would be more willing to finance subsidies for lower income uninsured households if they were raised by limiting the exclusion than if they were raised in other ways? One possible hypothesis is that taxpayers may consider both the

marginal taxes they will have to pay to raise funds and the excess burden associated with the type of tax financing chosen. That both considerations ought to be taken into account is the burden of optimal tax theory, but that theory is rooted in welfare economics that is not linked to voter choice. There is some literature suggesting the hypothesis that excess burden should matter in public choice models (Johnson and Pauly, 1969) but little conclusive evidence.

The idea that tax financing to pay for medical care generates a burden on the economy has been endorsed by the recent Council of Economic Advisors (2009) report on why the country would gain from reducing medical care spending (if it could do so without too large of a reduction in health benefits) because lower Medicare spending would lower Medicare's tax burden thus imposing less distortion on the economy. The same issue would arise in connection with financing subsidies for the under-65 uninsured. The negative excess burden associated with taxing benefits may mean that, among all the necessarily distasteful ways of raising funding for health care reform from the upper middle class, this method may not only be preferable as a financing choice but may also make those taxpayers not receiving direct benefit more willing to pay to help others (compared to their willingness if they had to pay higher general taxes). They may have to pay for the uninsured (because someone has to pay for the uninsured) but at least they do so in a way that reduces their own insurance premiums and gives them more money wages to be used for other things.

CONCLUSION

Based on normative economic theories of efficiency and equity, the case for curtailing the tax subsidy to employment based health insurance is, in my opinion,

overwhelming. Whether that case will be enough to overwhelm politics and special interests remains to be seen, but the argument based on principle is valid. We may want to allow necessarily imprecise and somewhat amateur judgments of political feasibility to temper our policy advice to some extent, but we should be clear about the principles involved, lest we run the risk of generating more confusion on a topic (and an insurance financing mechanism) that generates confusion enough on its own.

REFERENCES

Abelson, Reed.
"Taxing Those with Health Insurance to Pay for Those Without." *The New York Times* (May 8, 2009): A1.

Buchanan, James M., and Mark Pauly.
"On the Incidence of Tax Deductibility." *National Tax Journal* 23 No. 2 (June, 1970): 157–67.

Council of Economic Advisors, Executive Office of the President.
The Economic Case for Health Reform. Report posted June 2, 2009. http://www.whitehouse.gov/assets/documents/CEA_Health_Care_Report.pdf.

Havighurst, Clark C., and Barack D. Richman.
"Distributive Injustice(s) in American Health Care." *Law and Contemporary Problems* 69 No. 3 (Autumn, 2006): 10–82.

Johnson, David, and Mark Pauly.
"Excess Burden and the Voluntary Theory of Public Finance." *Economica* 36 No. 143 (August, 1969): 269–76.

Pauly, Mark, and Jose Pagan.
"Spillovers and Vulnerability: The Case of Community Uninsurance." *Health Affairs* 26 No. 5 (September/October, 2007): 1304–14.

Wennberg, John E., Elliott Fisher, and John Skinner.
"Geography and the Debate on Medicare Reform." *Health Affairs Supplement. Web Exclusive* 21 No. 1 (February 13, 2002): W96–W114.