

SCALING AND SPLITTING: NEW APPROACHES TO HEALTH INSURANCE

How can we cut health care costs while expanding access to care?

◆ BY CHRISTOPHER T. ROBERTSON AND KEITH A. JOINER

In the United States, cost-sharing in health insurance coverage has become the primary mechanism for reducing insurance expenditures and, by extension, maintaining affordable coverage. Cost-sharing involves patients making various out-of-pocket (OOP) payments for their own health care aside from whatever the insurer pays.

As a patient's spending on health care grows month by month in any given year of coverage, she moves through three different "zones" of insurance, from no insurance, to partial insurance, and finally to full insurance. Looking at these zones individually:

Zone 1—No Insurance: An annual deductible gives the patient complete responsibility for the first health expenditures in a year. In other words, the patient has 100 percent "skin in the game" for those expenditures. In common plans, the average annual deductible is less than \$1,000.

About one-quarter of American workers are in "high-deductible health plans with a savings option," meaning that the insurance is paired with a tax-advantaged savings plan for medical expenditures. The average deductible for these plans is a bit over \$2,000, though federal law allows them to be as high as \$12,000 for families. High-deductible plans, as a percentage of plans offered and of employees enrolled, have increased more than five-fold over the last decade. Many employers are eliminating all other options for their employees.

CHRISTOPHER T. ROBERTSON is professor and associate dean for research and innovation at the University of Arizona's James E. Rogers College of Law. KEITH A. JOINER is professor of medicine, economics, and health promotion sciences and director of the Center for Management Innovations in Health Care at the University of Arizona's Eller College of Management.

This article summarizes and incorporates some text from three prior papers by the authors.

Zone 2—Some Insurance: In this zone, patients share the burden of health care consumption through copayments, coinsurance, or reference pricing. A copayment is a flat fee paid at a doctor's office, hospital, or pharmacy, while coinsurance is a percentage (typically about 18 percent) of the cost of the service or drug. Under a "reference price," an insurer pays a fixed amount for a service and the beneficiary pays all charges above that amount.

These forms of cost-sharing are intended to decrease "moral hazard," in which patients consume more health care than is efficient because they bear little of the cost. But these forms of coverage can lead to "behavioral hazard," in which services or drugs of clear medical benefit are under-utilized because patients shy away from the co-pay. Most consumers do not (or are not able to) distinguish between care that is considered highly beneficial and care that is of minimal benefit or even potentially harmful.

Zone 3—Full Insurance: Finally, there is a zone in which patients have no "skin in the game" at all and insurance covers the full cost of care. Most individual workers have cost-sharing burdens capped at some amount less than \$3,000 per year (or less than \$5,500 for family-coverage), after which all expenses become the responsibility of the insurer. Although most plans included such a cap already, it is now required by the 2010 Patient Protection and Affordable Care Act (ACA).

Imagine an insured, Ms. Mildred Median, whose plan has a \$1,000 deductible, an 18 percent coinsurance burden, and a \$3,000 cost-sharing maximum. Suppose that this year, she will spend tens of thousands of dollars on a heart stent, chemotherapy drug, or other high-cost care. After spending her \$1,000 deductible, she will be exposed to up to \$2,000 more in costs (\$3,000 cap minus the \$1,000 deductible). Given her 18 percent coinsurance rate,

that \$2,000 will be consumed after the next \$11,111 in health care expenses. Thus, Ms. Median has reached Zone 3, where she has full insurance with no more skin in the game, after consuming \$12,111 in health expenses, \$3,000 of which she paid out of pocket.

For Ms. Median, the cap is a good thing. If she also earns a median family income of about \$51,000, her health care expenses will have consumed 6 percent of her pre-tax income. Depending on her other obligations and the amount she has put into savings, she may not have been able to bear more risk. In this sense, the Zone 3 cap is doing exactly what it was designed to do.

Such a cap is anathema for cost control, however. The vast majority of health spending is consumed by a minority of individuals who each account for tens of thousands of dollars of health care consumption in a given year. (This is the familiar “80-20 rule” for health care costs—20 percent of the population consumes 80 percent of the care.) Thus, the cap on out-of-pocket spending deprives the insurer of its primary cost-control mechanism at precisely the point where the expenditure decisions are most effective in controlling aggregate costs.

The two goals for a rational health insurance policy are in inherent tension: too high a cap hinders risk protection; too low a cap hinders cost-control.

SCALING

Individuals have radically different abilities to bear risk. For instance, those in the top quintile of the U.S. household income distribution are paid five times more than those in the bottom quintile. Yet, within a given health plan, they all have the same cost-sharing burdens, given the prototypic process for pricing insurance premiums and cost-sharing obligations. Accordingly, the cap will be too high for some beneficiaries, making them “underinsured” because it undermines the risk-protecting

and access-guaranteeing purposes of health insurance. The same cap will be too low for other beneficiaries, removing the price signal sooner than necessary and making them “overinsured.” These individuals no longer have any skin in the game, and thus they can spend without any market discipline. Prior empirical research shows that wealthy individuals can accept greater risk without adverse consequences for their health.



HEALTH & MEDICINE

The language of taxation may provide normative traction as to the fair distribution of health care expenses between individuals and their collective insurance pools. A “progressive” tax is typically paid as a percentage of adjusted income, with several tiers of increasingly higher percentages. Such a policy is sensible to the extent that income past a certain level is more disposable than income at the lower levels, which must be allocated to basic human needs. A “flat rate” or “proportional” tax is one where everyone pays the same percentage of income. Although rare today in overt forms, a “head” tax is one where each individual pays the same dollar-amount regardless of wealth or income. The head tax is thought to be objectionably regressive.

Our current system, in which each employee faces the same amount of health care costs in order to get the full insurance of Zone 3, is analogous to a head tax. It is regressive in the sense that lower-paid workers must pay a larger percentage of their incomes than higher-paid workers. Coupled with the fact that health insurance premiums are paid with pre-tax dollars, lower-income workers (in lower tax brackets) are effectively subsidizing insurance costs for those with higher incomes. For lower-income workers, choosing high-deductible plans with lower premiums in order to afford insurance does not solve this dilemma, and in many ways exacerbates it.

In contrast, the optimal insurance policy will be one that scales the cost-sharing burdens for each beneficiary to his or her ability to bear that uninsured risk. Scaled cost-sharing (SCS) is feasible in a world where employers pay wages and provide health insurance; the two data points need only be linked together. Even outside the employer-sponsored health insurance market, it is relatively easy to get wage information. Those same data are used to calculate premium subsidies in the individual market. Although scholars have periodically recognized the concept of scaling, it has not yet achieved the sustained attention and prominence it deserves in scholarly and policy debates.

One imperfect but feasible mechanism for such tailoring would look like the following: Instead of exposing all beneficiaries to the same fixed-dollar *amount* of uninsured risk per year, the insurer would expose all beneficiaries to the same *percent* of their wages as the uninsured risk. For example, we could use as a baseline a common cost-sharing profile that includes a \$3,000 cost-sharing cap, and apply it to a median American worker with a \$50,000 income. This median worker with an average plan faces a 6 percent cost-sharing ratio. We could then apply that same 6 percent ratio across the board, scaling the absolute cost-sharing amount upward or downward with each person’s income. In this arrangement, each beneficiary gets roughly the protection from risk that she needs, while also continuing to have as much skin in the game as she can handle. Such an implementation would be analogous to a flat-rate tax, but more progressive options may be worthwhile to consider.

This is not a zero-sum reform that simply redistributes risk from the poor to the wealthy. Instead, because the wealthy are better able to bear that risk, SCS should also appeal to rational employers

who seek to maximize shareholder value while providing employee insurance. Because the distribution of American workers’ incomes is concentrated in favor of the highly compensated, proportional income-scaling will also be asymmetric. One of us, Christopher Robertson, has estimated that a 6 percent scaling would add four times more cost-sharing than it removes, thereby significantly reducing the aggregate burden of insurance premiums.

SCS is also unique among “consumer-directed health insurance” reforms in that its increase of the catastrophic caps for half of the population will affect high-cost health care, which accounts for the bulk of overall spending. Thus, SCS could significantly reduce insurance outlays in this second respect. Moreover, SCS promises to deliver better health outcomes and enhanced worker productivity for each dollar spent on health insurance. For these reasons, SCS may allow a greater bargaining surplus between workers and shareholders.

There are intellectual precedents for SCS. In the market for goods and services, sellers often use differential pricing (e.g., Acura and Honda) to sell to consumers with a range of willingness to pay. This strategy can increase access to goods and improve profits for the seller, enhancing overall welfare. Several European countries have adopted income scaling for speeding tickets in an effort to make the deterrent effect proportionate across heterogeneous drivers.

For health insurance in particular, it is important to distinguish the proposal for scaling cost-sharing (out-of-pocket payments by insured individuals for their medical care) from the scaling of insurance premiums (monthly payments required in order to maintain an active insurance plan). While scaling of costs has received only minor consideration by economists, many scholars have recommended the scaling of premiums, which already appears in some federal programs and in benefit plans for about 10 percent of large employers. While premium scaling may serve fairness goals, it does almost nothing on its own to solve the problems of underinsurance and overinsurance.

Why isn’t scaling already happening? / One may ask why employers have not widely adopted wage-scaled cost-sharing in the insurance plans they offer. There are five potential sources for this apparent market failure.

First, one might wonder why a choice-based system, in which workers would select their own cost-sharing profiles and pay insurance premiums accordingly, has not emerged. Car and home insurance often already operate in this fashion. The long-standing practice of employers contributing to health insurance premiums (even though the costs are ultimately borne by the employee) and the legal limits on individual-rating of premiums present obstacles to this approach. In this domain, choice also presents problems of adverse selection by allowing beneficiaries to exploit their private information, which undermines the risk-pooling function of insurance. More generally, a choice-based mechanism could be compromised by the severe cognitive limits that individuals face when making such complex decisions about

risk. As a result, greater intervention by employers and regulators may be sensible in this domain; they could move toward scaling while holding constant the lack of choice that exists in employer-sponsored plans today. Nonetheless, as an intermediate step, employers could implement scaling as a default rule.

A second significant obstacle is that well-intentioned caps on cost-sharing in the ACA limit the potential application of SCS. This is an unfortunate limitation in the ACA, but the executive branch could facilitate reforms through thoughtful interpretations of the law and the exercise of discretion in enforcement. Congress could also fix this flaw. Even without such reforms, employers have considerable discretion to use scaling beneath those statutory caps.

Third, there may be an agency problem. Managers responsible for designing the firm's insurance policy are paid more than the median worker, which may cause them to reject SCS even if it would improve profits for the shareholders. There is a related collective action problem in the market for talented workers. Employers are hesitant to be the first to impose higher cost-sharing burdens on their highest-paid workers, lest they lose them. Fundamentally, these problems are caused by the tax code, which creates a distortion in favor of health spending through insurance. This distortion is especially pronounced for high-paid workers.

Fourth, there will be difficult questions in the transition, as changes in the actuarial value of insurance for any particular worker could be perceived as changing the value of the worker's overall compensation package, which thus may call for untenable adjustments on the salary side (the compensating differential). However, the overall cost of health insurance should decrease with SCS—assuming that premiums would decline in conjunction with higher cost-sharing obligations for higher-paid workers and the overall higher amount of cost-sharing because of the skew in wages. This efficiency creates a bargaining surplus that allows a win-win redistribution between classes of workers and shareholders. Additionally, for higher-paid workers with health savings accounts and similar tax-preferred vehicles, significantly higher OOP unreimbursed expenses would be partially offset by the income tax deductions associated with such expenses.

Fifth, there may be misperceptions that current federal law prohibits this sort of “discrimination” by salary levels. One of us, Christopher Robertson, has shown elsewhere that the law is actually permissive. Indeed, it arguably obliges employers to use scaling in order to allow lower-paid workers to get the full benefit of their health insurance without hiding tax-free compensation for top workers in the form of unnecessary insurance. Thanks to an expansion of this unenforced rule in the ACA, the IRS now has a legal mechanism to counteract the distortion of the tax preference. It need only clarify its regulations to require

scaled cost-sharing. Lower-paid workers may also have options for litigation. Accordingly, congressional action is unnecessary for this landmark reform. The Internal Revenue Service, or individual litigants, could nudge employers toward scaling, helping to achieve health insurance's goals and reduce its distortive effects on consumption. Scaled cost-sharing is smarter insurance.

SPLITTING

As we have seen, the fundamental problem for cost-sharing is that the cost of treatment can be disproportionate to patients' ability to pay even a portion thereof. That is why we limit cost-sharing exposure. But our proposal, so far, also defeats the cost-containment virtue of cost-sharing by giving patients carte blanche when they make the most costly consumption decisions. Even with scaling, caps on OOP will remain. But we have a way to address this problem while retaining consumer choice in treatment.

It is remarkable that insurance currently provides only an “in-kind” benefit, paid to the provider rather than the beneficiary.

The “split benefit” would divide the health insurance payment between the provider and the patient, rather than simply paying it to the provider. This would guarantee health care access and patient choice.

Alternatively, for expensive treatments (costing, say, \$100,000), the insurer could consider satisfying its coverage obligation by paying a portion (say, \$10,000) directly to the patient. The patient then decides whether to spend that portion on the treatment. If so, the insurer pays the balance (\$90,000) to the provider, thereby insuring access. If the patient instead declines the care, he or she can save or spend the money on anything else. The insurer saves the balance (\$90,000). We call this a “split benefit” because it divides the health insurance benefit payment between the provider and the patient, rather than simply paying it to the provider as is done now.

Strikingly, the split benefit guarantees health care access and patient choice precisely to the same extent as current in-kind insurance benefits. For this reason, the split benefit is consistent with current insurance contracts and regulations because it does not change coverage or the size of the benefit. That feature makes the split benefit practicable, unlike many other theoretical solutions. The split benefit can be used alongside other insurance mechanisms, including traditional cost-sharing (scaled or not), exclusions, and fail-first policies.

The split benefit creates new options for the patient. Because it is fungible, the split benefit creates an opportunity cost, causing some patients to decline the expensive treatment in lieu of medical

HEALTH & MEDICINE

and nonmedical alternatives that they value more highly. In this way, once the split is paid, it functions exactly the same way as a traditional cost-sharing burden. It is an out-of-pocket payment.

The insurer can exercise the split benefit as a unilateral option whenever it is most likely to save money. Insurers will find the split benefit most useful for procedures (whether drugs, devices, surgeries, or diagnostics) that meet four criteria:

- The insurer must cover the procedure.
- The price of the procedure is disproportionate to the patient's wealth.
- The patient would otherwise be likely to consume the treatment.
- The procedure is not proven to reduce health care expenditures on net.

The fourth criterion reflects the rational interests of any insurer. It is folly to use the split benefit or any other mechanism to try to dissuade patients from consuming health care that is actually efficient from a financial point of view.

Even with those limitations, the split benefit proposal covers a broad swath of American health care. A 2006 study found that 21 percent of all prescriptions written in the United States are for off-label uses and that most of those had little or no scientific support. Many of the most expensive treatments (such as heart stents) are proven effective for some patients, but are used widely for others. Estimates from many sources and using a variety of approaches have concluded that approximately one-third of all medical expenditures in the United States are either wasteful or harmful.

Recent initiatives to identify and highlight a list of such interventions in various medical specialties are a step in the right direction, but are unlikely to have a substantial effect on costs in the short or even intermediate term. First, a report from the Institute of Medicine suggested that 17 years are needed before a large majority of physicians adopt new practice recommendations. Second, the interventions identified are few in number. Third, those recommendations come from the provider community, which (at least in the current fee-for-service payment model) is incentivized to maintain the options for reimbursement at nearly a steady state. This much larger zone of unproven, low-value health care would be a prime target for the split benefit, although it would be unlikely to be driven from the provider side.

Whether the split benefit should be extended to other health care that is costly on net, but is proven to have high value for at least some patients, is an open question. One concern is that, given the churn in the health insurance market, insurers may be biased against health care consumption and thus may use the split benefit and other cost-control devices too often on the margins. We tentatively suggest that the split benefit is warranted wherever and to the same extent that traditional cost-sharing is appropriate. In fact, it is a form of cost-sharing that simply solves a wealth problem.

The insurer also has the discretion to select what level of split

is optimal for each patient. The optimal size of the split-benefit payment is an empirical question, one that is likely context-dependent. There will presumably be diminishing marginal returns, such that moving from a 1 percent to a 10 percent split may yield a very large reduction in the rate of consumption, but the equally costly step of moving from a 10 percent split to a 19 percent split may yield little additional benefit.

Testing the concept / To investigate those empirical questions, we fielded a randomized vignette-based experiment in which we exposed a diverse group of 1,800 respondents to stories asking them to imagine a future decision about whether to consume a particular expensive, unproven treatment. We manipulated whether a split-benefit payment was offered, the amount of the split (\$5,000 or \$15,000), and the type of split offered. We conceived three such types:

- payment via bank check that the patient receives in advance of the health care decision, and simply keeps if he declines care
- a rebate offering payment after the decision is made, conditional on declining care
- another prepayment, but paid into an account with fungibility limited to other health expenses, similar to a health savings account

We found that offering a split benefit substantially reduced the intent to consume. It did not much matter how large the payment was or its form. We also asked respondents whether they viewed the insurer as treating them fairly, and they responded very positively. They also supported the idea of a split benefit being used to save costs in Medicare, and amazingly we detected no political polarization on that point.

Our study suggests that the split benefit could be useful as a way for public and private insurers to reduce consumption of the high-cost, low-value care that drives much of health spending. And they can do so without reducing access or infringing on patient choice. Indeed, the data suggest that insurers have flexibility as to how they pay a split, in the amount and type. Those paying a split will be perceived fairly; after all, the only effect of a split is to increase the patient's wealth and options.

Weighing objections / Admittedly, the split benefit may create a perverse incentive for individuals to seek prescriptions for treatments they otherwise may not have consumed anyway (a problem of "false demand"). To some degree, insurers can observe and police these dynamics. Insurers should not use the split benefit for ailments that are easy to fake or where the diagnosis is highly subjective. For example, nonspecific back pain may be one such diagnosis that could be opportunistic for a higher percentage of patients than other diagnoses, such as breast cancer. Still, if small split-benefit payments significantly reduce consumption, there will be some tolerance for the stimulation of false demand. If a

\$5,000 payment prevents consumption of a \$100,000 treatment, as our experiment suggests, the insurer could tolerate a dozen or more false demand payments while still saving money on net.

One might object that the split-benefit payment seems coercive, potentially pushing patients away from consuming health care. To the contrary, the entire insurance relationship on net, even including a split-benefit payment, still induces patients toward consuming health care. The insurer tells the patient that he or she can spend the money on anything, but if the patient wants to consume health care, then the insurer will provide a nine-fold match of the patient's money (assuming a 10 percent split). Merely allowing the patient to decide how she wants to spend one-tenth of her insurance benefit does not constitute coercion.

Still, there may be situations where the patient cannot decide whether to consume a treatment, and where someone else—such as a parent or next of kin—decides on the patient's behalf. The split-benefit proposal may create a conflict of interest if that substituted decisionmaker receives the benefit of the cash payment (either through expropriation or inheritance), while the patient receives the benefit (if any) of the treatment. It may be best to limit the split-benefit program to only those situations where the patient is competent to make treatment decisions. Still, traditional cost-sharing obligations already impose these sorts of dilemmas on substituted decisionmakers, who would rather keep the money. Even worse, traditional cost-sharing obligations, unlike the split benefit, may be so onerous that they deny access to the expensive care altogether.

The poorest patients also present a particular concern. The poor patient's alternative consumption choices for food or housing are more pressing than those of the median patient. This makes the poor patient more likely to decline care in order to pursue those alternatives. Insurers could scale split-benefit payments according to patient wealth (giving the poor smaller splits), but that may appear unfair. With same-size splits, the benefits of a fungible payment will be greatest for the poor precisely because those alternative consumption options—such as food or housing—will actually be better for the poor patient than expensive health care. Even if we focused exclusively on health outcomes, evidence suggests that investments in housing, diet, and education may prove more effective than investments in expensive medical interventions. If the choice is hard, it is because the alternatives are attractive.

A related objection would invoke the “specialness” of health. One might argue, in the words of economist and philosopher Amartya Sen, that “health is among the most important conditions of human life and a critically significant constituent of human capabilities which we have reason to value.” Thus, it may seem perverse for the split-benefit policy to facilitate patients trading health care in favor of other goods such as housing or even jewelry.

This critique does not bear scrutiny. People buy insurance to ensure future access to care that they otherwise could not afford. Ex ante, consumers need not know whether they will actually want to consume a given drug for a given disease that they may

someday suffer. But rational consumers want the option to consume such treatment when they become better informed by their actual experience of the situation. After all, patients want health, not health care. Interventions are not valuable per se, but only because they may increase well-being. Ex post, having secured the option, a rational consumer may nonetheless prefer to spend that benefit on other goods that they judge to have a more substantial effect on their well-being. The split benefit is perfectly congruent with the option-buying purpose of insurance. It has the side benefits of increasing patient wealth and reducing insurance costs along the way. On the aggregate, a split-benefit reform may correct the systematic bias in our economy toward low-value, high-cost health care.

Arguably, the split benefit is a better solution than traditional cost-sharing or rationing by insurers or physicians, which reduce access to care. If a physician refuses to write a prescription because of cost, if an insurer refuses to cover such a prescription, or if an onerous cost-share exceeds a patient's ability to pay, the health care system has denied the patient the choice. In contrast, the split benefit keeps the decision in the hands of the patient. The proposal serves patients' autonomy by giving them additional options.

CONCLUSION

We have presented two reforms to health insurance, which could be applied in the employer-sponsored insurance market, the individual exchanges, or in Medicare. There are, of course, other ways to control health care costs—including the centralized rationing policies that are used more overtly abroad and the move toward partial or full risk-adjusted capitation. However, cost-sharing is attractive in the current fee-for-service payment environment because it leaves the choices in the hands of patients themselves. Scaling and splitting are simply ways of making cost-sharing work better.

We suggest that policymakers and insurance designers use a combination of scaling and splitting. First, adjust cost-sharing burdens according to wealth or income, so that patients will be required to pay as much as they can bear from their own pockets. Then, when patients' health care consumption crosses beyond the scaled cap, begin using the split benefit—paying them the wealth necessary to keep their skin in the game. In this way, we can achieve the purposes of insurance—protection from risk and guaranteed access to care—while also maintaining cost-control through the market-mechanism of consumer choice. R

READING

- “A Randomized Experiment of the Split Benefit Health Insurance Reform to Reduce High-Cost, Low-Value Consumption,” by Christopher T. Robertson, David Yokum, Nimish Sheth, and Keith Joiner. *Innovation and Entrepreneurship in Health*, Vol. 1 (2014).
- “Scaling Cost-Sharing to Wages,” by

Christopher T. Robertson. *Yale Journal of Health Policy, Law, and Ethics*, Vol. 14 (2014).

■ “The Split Benefit: The Painless Way to Put Skin Back in the Health Care Game,” by Christopher T. Robertson. *Cornell Law Review*, Vol. 98 (2013).