Journal of Health Economics 57 (2018) 168–178

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Decomposition of moral hazard

John A. Nyman[a,*], Cagatay Koc[b], Bryan E. Dowd[a], Ellen McCreedy[c], Helen Markelova Trenz[d]

[a] Division of Health Policy and Management, School of Public Health, University of Minnesota, 420 Delaware St. SE, Box 729, Minneapolis, MN, 55455-0392, United States
[b] Cornerstone Research, 1919 Pennsylvania Avenue, N.W., Suite 600, Washington, D.C., 20006-3420, United States
[c] Center for Gerontology and Healthcare Research, Brown University, School of Public Health, 121 South Main Street, Suite 6, Providence, RI, 02903, United States
[d] Optum Labs, 11000 Optum Circle, Eden Prairie, MN, 55344, United States

ARTICLE INFO

Article history:
Received 23 January 2017
Received in revised form
25 September 2017
Accepted 7 December 2017
Available online 11 December 2017

Keywords:
Moral hazard
Health insurance income effects
Efficient moral hazard

ABSTRACT

This study seeks to simulate the portion of moral hazard that is due to the income transfer contained in the coinsurance price reduction. Healthcare spending of uninsured individuals from the MEPS with a priority health condition is compared with the predicted counterfactual spending of those same individuals if they were insured with either (1) a conventional policy that paid off with a coinsurance rate or (2) a contingent claims policy that paid off by a lump sum payment upon becoming ill. The lump sum payment is set to be equal to the insurer's predicted spending under the coinsurance policy. The proportion of moral hazard that is efficient is calculated as the proportion of total moral hazard that is generated by this lump sum payment. We find that the efficient proportion of moral hazard varies from disease to disease, but is the highest for those with diabetes and cancer.

Introduction

The concept of moral hazard has played a central role in U.S. health policy since Pauly introduced it into the health economics literature 50 years ago (Pauly, 1968). It debuted as a counter-argument to Arrow's famous 1963 evaluation of the welfare effects of health insurance derived from Friedman and Savage's risk avoidance model (Arrow, 1963; Friedman and Savage, 1948). Instead of health insurance being overwhelmingly welfare-increasing (under the right circumstances) as Arrow had suggested, Pauly warned that the welfare loss from moral hazard could be so large as to render health insurance welfare-reducing (Pauly, 1968). Analysts who adopted Pauly's model concluded that moral hazard generated so much inefficiency that coinsurance rates should be raised dramatically (to 67% or greater according to Feldstein, 1973, or to about 45% across-the-board according to Manning and Marquis, 1996) in order to ensure that insurance increases welfare. Apparent empirical support of this theory and policy came from the influential RAND Health Insurance Experiment that found that large increases in cost-sharing could be implemented causing substantial decreases in healthcare spending and utilization, but with no important impact on health (Newhouse, 1993).[1]

In Pauly's model, the additional health care consumed when insured was viewed solely as the product of a price distortion (Pauly, 1968). His model did not recognize that the price reduction (that is, the change from the market price to the coinsurance rate stipulated in the health insurance contract) alternatively could represent a vehicle for transferring income from those who purchase insurance and remain healthy, to those who purchase insurance and become ill. It also did not recognize that this income transfer could cause the purchase of additional health care, resulting in a welfare gain. While Marshall (1976) and de Meza (1983) observed that such income effects were possible if insurance paid off with a direct income transfer, neither recognized that an insurance price reduction effectively represented a similar income transfer and could produce a similar income effect.

[1] RAND findings have been challenged by those who point out that attrition rates were 16 times larger in the cost-sharing plans than in the free plan (Nyman, 2007; Aron-Dine et al., 2013). Differential attrition rates suggest that participants who became ill dropped out of the cost-sharing arms in order to revert to their more complete pre-RAND insurance coverage. A lack of ill patients in the cost-sharing arms would help explain the finding that the remaining participants in those arms could receive substantially less health care, especially fewer hospitalizations, but with no important effects on health.

* Corresponding author.
E-mail address: Nyman001@umn.edu (J.A. Nyman).

https://doi.org/10.1016/j.jhealeco.2017.12.003
0167-6296/

In this paper, the theory is adopted that, rather than to avoid risk, at least a portion of health insurance is purchased in order to receive an income transfer when ill, and that the insurance price reduction (represented by the coinsurance rate) is the vehicle by which income is transferred from the healthy to the ill (Nyman, 1999a,b, 2003). Because the literature on the welfare effects of moral hazard did not originally recognize an income effect and has only belatedly attempted to redefine moral hazard in healthcare as the pure price effect (Cutler and Zeckhauser, 2000), we refer in this paper to the entire effect of insurance on spending as moral hazard. Accordingly, this paper seeks to decompose ex post moral hazard—the additional healthcare expenditures caused by becoming insured—into an efficient income-generated portion and an inefficient portion that arises because a price reduction is used to transfer income.

Chetty (2008) uses the term "liquidity" to refer to the subject of this transfer. In some respects, "liquidity" is more accurate than "income" because it captures the fact that the transfer in insurance typically does not have a periodic time dimension, but instead represents the transfer of a stock of resources or wealth that can be captured at an instant of time. On the other hand, if the transfer occurs by means of a price reduction (as it does with health insurance), the transfer is far from "liquid" in the sense that it would be difficult, if not impossible, to resell most health care services (e.g., an appendectomy) on the market and thereby convert them into cash. Accordingly, we have chosen to continue to refer to this notion as an "income transfer" because it matches the simple theoretical model better and because this language was used in the analyses that originally presented this theory (Nyman 1999a,b, 2003).

In the next section, the theory is summarized. It is shown how an income transfer is contained in the structure of a standard coinsurance contract. Then, a 7-step empirical approach is presented. In the following section, the results of the decomposition of moral hazard are presented for the various health conditions. In general, we find evidence of a substantial efficient moral hazard effect, but one that varies in size depending on the type of illness and assumptions of the empirical model. In the final section, the implications of the results are discussed and the limitations acknowledged.

Theory

The theoretical model has been presented elsewhere (Nyman, 1999a,b, 2003), so it is summarized here. First, consider the *ex post* case of a consumer who has become ill, with and without insurance. The consumer maximizes utility when ill, U_s, derived from medical care, M, and other goods and services, Y. If uninsured, the consumerspends his or her income (or wealth), Y_o, on M and Y, assuming that the price of a unit of medical care is unity. As a result, the consumer's problem is written:

$$\max U_s(M, Y)$$

$$s.t. Y_o = M + Y$$

The consumer solves the problem at (M_u, Y_u), where the marginal rate of substitution equals 1, the (negative of the) price ratio.

If the consumer is insured, he or she has purchased insurance with an actuarially fair premium payment, R, out of income, but in return, the price of medical care has dropped from 1 to the coinsurance rate, c. The consumer's problem is now:

$$\max U_s(M, Y)$$

$$s.t. Y_o - R = cM + Y$$

The consumer solves the problem at (M_i, Y_i) where the marginal rate of substitution equals c, the new effective price ratio.

Because the price of M has dropped from 1 to c, the consumer purchases $(M_i - M_u)$ more medical care, which constitutes moral hazard spending using the traditional definition. Assume that an actuarial study has previously estimated M_i and the insurer sets the premium, R, to be actuarially fair: $R = \pi(1-c)M_i$, where π is the probability of illness. Thus, the insurer spends $(1-c)M_i$ on this consumer, of which $\pi(1-c)M_i$ is financed by the consumer's own premium contribution and the rest by $(1-\pi)(1-c)M_i$ worth of income transferred from the insurance pool.

Alternatively, the consumer could have spent the same premium payment, R, on a contingent claims insurance contract that would pay out the same amount, $(1-c)M_i$, as was spent by the insurer under the coinsurance contract, but in this case as a lump-sum payoff, P, paid directly to the insured consumer upon becoming ill. The ill consumer's problem under this contract would be written:

$$\max U_s(M, Y)$$

$$s.t. Y_o - R + P = M + Y$$

The payoff, P, would consist of the consumer's own contribution to the insurance pool, $R = \pi(1-c)M_i$, plus an amount contributed by others and transferred to the ill consumer, $T = (1-\pi)(1-c)M_i$. Thus, the ill consumer with this insurance both pays out R *ex ante* and receives $(T+R)$ *ex post* as a result of this contract, so T, the income transfers, is the net increase in income. The consumer solves this problem at (M_c, Y_c), where the marginal rate of substitution equals 1, the original price ratio without distortion.

This model can be used to decompose the moral hazard generated by the coinsurance contract. Assuming that healthcare is a normal good and that $M_u < M_c < M_i$, then $(M_c - M_u)$ would represent the income transfer effect produced by a coinsurance contract and $(M_i - M_c)$ would represent the price effect. Thus, $(M_c - M_u)/(M_i - M_u)$ is the percentage of moral hazard that is income-generated, and so, efficient.

The *ex ante* portion of the model captures the motivation for purchasing insurance. In the *ex ante* period, the (healthy) consumer does not know whether she will become ill during the contract period. If the consumer is uninsured and becomes ill, the consumer is assumed to spend M_u on medical care and the rest of income on other goods and services. If the uninsured consumer remains healthy, she is assumed to derive no utility from medical care, so $U_h(0, Y_o)$. Thus, this model applies best to the type of healthcare that only those who are seriously ill would consume, the type of healthcare that is likely to represent a large portion, if not the majority, of total healthcare spending. Accordingly, if uninsured, the consumer's *ex ante* expected utility is:

$$EU_u = \pi U_s(M_u, Y_o - M_u) + (1-\pi)U_h(0, Y_o).$$

If the consumer is insured with a coinsurance policy, she would have paid the fair premium, R (worth $\pi(1-c)M_i$ of income), even if healthy. The coinsurance rate payoff would include a net transfer, T (worth $(1-\pi)(1-c)M_i$ of income), if ill, which could be spent on other goods and services after the M_i healthcare spending is accounted for. Again, it is assumed that the healthy consumer derives no utility from, and therefore does not purchase, any of this type of health care. For example, what *healthy* consumer would choose to receive a course of chemotherapy, an amputation, or an organ transplant, even if insured? So, an insured consumer has *ex ante* expected utility of:

$$EU_i = \pi U_s(M_i, Y_o + T - M_i) + (1-\pi)U_h(0, Y_o - R).$$

Thus, even though the insurance is actuarially fair, in that the expected premium payment if healthy, $(1-\pi)R = (1-\pi)\pi(1-c)M_i$, equals the expected transfer if ill,