Moral Hazard and the Production of Health Care

Professor Jeff Richardson
Foundation Director, Centre for Health Economics
Monash University, Australia

Professor Jan Abel Olsen
Professor of Economics, Institute of Community Medicine,
University of Trømsø, Norway

August 2011

Centre for Health Economics
ISSN 1833-1173
ISBN 1 921187 68 9
Correspondence:

Jeff Richardson
Centre for Health Economics
Faculty of Business and Economics
Monash University  Vic  3800
Australia

Ph: +61 3 9905 0754  Fax: +61 3 9905 8344
Email: jeffrey.richardson@monash.edu
Increasing obesity and less healthy lifestyle are commonly seen as a reflection of consumer preferences and the relative prices of alternative sources of satisfaction. From this perspective public health policies and attempts to alter people’s choices are paternalistic. This article introduces a simple model which demonstrates that as a result of comprehensive health insurance, lifestyle induced ill health is at least in part attributable to market failure: like the moral hazard of insurance induced excess health care consumption there is a moral hazard of insurance induced self neglect. It is potentially much more significant than the effect of ‘ex ante’ moral hazard discussed in the literature. This suggests that for reasons of efficiency, not paternalism, comprehensive insurance should be accompanied by policies to encourage people to adopt healthy lifestyles.
TABLE OF CONTENTS

1 Introduction .................................................................................................................................... 1
2 Ex ante moral hazard .................................................................................................................... 2
3 The capacity to determine own health .......................................................................................... 3
   3.1 The ‘macro’ determinants of population health ...................................................................... 3
   3.2 Individual health behaviour ..................................................................................................... 3
4. Trade-off between health care and lifestyle efforts ...................................................................... 4
5. Discussion .................................................................................................................................... 7
6. Policy Implications ........................................................................................................................ 9
7. Conclusions .................................................................................................................................. 9
References ..................................................................................................................................... 10

List of Figures
Figure 1 Effective health care, less: ‘own effort’ and better health.................................................. 6
Figure 2 Effective health care, less: ‘own effort’ and poorer health.................................................. 7

Moral hazard and the production of health care
1 Introduction

Two trends in health and health care can be observed in most rich countries. First, aggregate health care expenditures are increasing at a faster rate than the increase in GDP. This can be explained by new medical technologies that have made health care more effective in the prevention and cure of chronic diseases, at the same time as increasing wealth has made these technologies affordable. The second trend is the increasingly unhealthy lifestyle of a large proportion of the population, resulting, *inter alia*, in an increasing prevalence of obesity and so called ‘lifestyle diseases’. For the OECD as a whole, the proportion of GDP spent on health care has risen from 6.9% to 9.0% between 1990 and 2008. Over a similar period the proportion of the adult population classified as overweight or obese increased from 41% to 50% (OECD 2009). The main exception to the trend towards unhealthy lifestyle has been smoking which has fallen as a result of vigorous interventions in terms of both direct and indirect regulations.

Despite the increasing availability of evidence based information about the benefits of a healthier lifestyle (ie healthy diet, more physical activity, and less substance use), most people fail to meet the recommended standards. The aim of this paper is to introduce a simple micro-economic model that helps to explain why people may *reduce* their own healthy efforts during times when improved medical technologies become freely available. It demonstrates that there is a form of ex ante moral hazard which suggests the need for policy intervention.

Commencing with Ehrlich and Becker (1972) there has been a small literature on the possibility of ex ante moral hazard due to the effect of increasing voluntary health insurance. In this literature the moral hazard is mediated by personal income. ‘Moral hazard’ is the indirect and, in terms of social goals, adverse effect of insurance upon individual behaviour and, specifically, a reduction in self-protection. As most countries have limited voluntary health insurance, and health care is dominated by some form of a national health service (NHS) or public insurance, discussion of this has been limited outside the US and Switzerland (for a UK exception see Courbage and Coulon (2004); Klick and Stratmann (2007). The omission is surprising as the literature on moral hazard is large and there is no reason why it should be limited to financial measures for protecting individuals against health related expenditures. This note demonstrates that insuring population health through the increased provision of effective health care may create an ex ante moral hazard in a publicly financed NHS which is independent of income effects and which may be quantitatively more significant than reported in the literature.
2 Ex ante moral hazard

In their seminal contribution Ehrlich and Becker (1972) argue that increased health insurance creates two conflicting incentives depending upon the response of the private insurer. First, insurance reduces the effects of medical expenditures upon private income. Individuals will therefore be less concerned with illness and the need to prevent it. Self-care and self-protection through healthy lifestyle would be reduced. Gravelle (1986) notes that this inflicts an external cost on others who participate in the insurance scheme as premiums (or taxes) will rise.

Secondly, however, if expenditure on prevention reduced the price of health insurance there would be a financial incentive to increase preventive activities implying that health insurance and prevention might be complements not substitutes. The significance of the second incentive is doubtful. If prevention reduced cost for the insurer then it is also likely to reduce costs for the uninsured individual, providing at least the same incentive. Subsequent writers have suggested a complementary relationship might exist if insurance increased doctor visits and doctors successfully encouraged patients to reduce lifestyle risk factors (Dave and Kaestner 2009). Cook and Graham (Cook and Graham 1977) also extend Ehrlich and Becker, with the concept of an irreplaceable commodity (health) and the suggestion that a dollar has different value in poor than in good health. In both cases the focus is upon a variable level of insurance.

The empirical literature does not suggest that this form of ex ante moral hazard is quantitatively large. Newhouse (Newhouse 1993) found no relationship between the levels of health insurance in the RAND Health Insurance Experiment and lifestyle. Courbage and Coulon (2004) found no relationship between smoking, exercise and private health insurance in the UK. In contrast, using data from US citizens transitioning to Medicare cover, Dave and Kaestner (2009) found an increase in ‘unhealthy activities’ of about one quarter of a standard deviation amongst elderly men, but not women, when these activities were measured by smoking, alcohol consumption and vigorous physical exercise.

The important feature of the existing theory of ex ante moral hazard – and the justification for the present paper – is that the negative behavioural effects of insurance are mediated entirely through their impact upon income. In his review of the field Kenkel (2000) concludes that market failure occurs because the price of market insurance does not reflect individual spending on self-protection. Without market failure the incentive for self-protection would be the increase in income as the premium falls with self-protection. In the most comprehensive theoretical treatment of the subject to date, Zweifel et al. (2009) simply assume that ‘prevention of illness costs money but has no other direct impact upon the individual’s utility’ (p224).

In the alternative model presented here, the primary focus – as in extra Welfarism – is upon health, and variation in this (not income) has the relevant impact upon utility. The next section describes the extent to which individuals determine their own health. A distinction is made between exogenous health circumstances, and those factors which individuals can influence through their health related choices which may result in a problem of moral hazard. The subsequent section presents the new model which follows a household production function approach to the demand for health. The institutional context is one of tax funded health care (or social insurance) based on community rating, ie there is no link between an individual’s contributions to the financing of health care and their expected use of health care. This setting is then compared to one in which there is such a link, through risk adjusted individual rating. The final section draws some policy implications.
3 The capacity to determine own health

The literature on the determinants of health generally takes a macro level epidemiological perspective, studying the causes of observed variations, or inequalities, in population health. In contrast micro-economic models study the health related behaviour of individuals – how they choose within their given choice set.

3.1 The ‘macro’ determinants of population health

There are three classes of determinants of population health (Evans, Barer et al. 1994; Olsen, Richardson et al. 2003). The first, genetics, explains inherited diseases through natural variations in human biology. Second, the physical and social environment includes working conditions, pollution, cultural norms and position in the social hierarchy. Third, health-related lifestyle refers to people’s behaviour regarding diet, exercise and substance use. The three classes of determinants differ in the extent to which individuals have control over them. Genetic endowments are health preconditions and reflect a ‘biological lottery’ over which we have no control. The environment in which people happen to live represents their circumstances or opportunities that – at least for children – reflect a ‘social lottery’ over which they have little control. Lifestyle is the determinant over which we have most discretion, but precisely how much of that reflects autonomous sovereign consumer preferences, and how much reflects adaptation to cultural norms (‘social conditioning’) in the environment in which people happen to live is a contentious issue.

Our genetic inheritance cannot be affected by policy – at least not for those of us already born. The physical and social determinants, however, are of policy concern because these create health inequalities that are considered ‘avoidable’ (Whitehead 1992; CSDH 2008; Wilkinson and Pickett 2009). However, the seminal work on social determinants of health, (Marmot 2005; Marmot and Wilkinson 2006) suggests that as long as people are driven by status, there will always be some degree of hierarchical differences in health.

3.2 Individual health behaviour

The latter literature leaves little scope for the individual to influence their health. But however individuals fare in the biological and social lotteries, these represent constraints under which they have discretion to make decisions. The health related social choices facing the individual have been widely discussed (Evans, Barer et al. 1994; Chaloupka and Warner 2000; Cook and Moore 2000; Kenkel 2000; AIHW 2010). Meta analyses leave unresolved the relative importance of behavioural variables but leave little doubt that, minimally, they have significant marginal effects and, judged by national willingness to pay for marginal treatments, these effects are of great value to people (Peel, McClure et al. 2005).

The uptake of health related behaviours depends upon individual utility functions. In its simplest form in health economics this includes only two goods; wealth and health. Following the seminal work of Grossman (1972), it is widely agreed that health care has value to the individual only insofar as it impacts on health and health care is included in the utility function through a production function for health. Utility functions can, however, be extended beyond wealth and health to include leisure, status, social relations, physical appearance (vanity), concerns for the health or well-being of our fellow citizens, and essentially any attribute that is important to individuals. In the current context, there is an important distinction between those factors that are
exogenous to the individual versus those factors over which the individual may exercise some control. In a subsequent discussion the discretionary behavioural elements motivated directly by the benefit of health gain are referred to as ‘healthy efforts’. Non-discretionary elements attributable to genetics and the social environment are embodied in the production function.

4. Trade-off between health care and lifestyle efforts

Our model consists of a production function for health, H, measured by some notion of health adjusted life expectancy which includes both quality and quantity (such as expected Quality Adjusted Life Years – QALYs). As shown in equation 1, health is a function, \( h \), of health care, HC, and healthy efforts, E. Health care, HC, encompass all forms of preventative or curative interventions provided by a health service. By efforts, E, we refer to any changes in health related lifestyle specifically undertaken for the purpose of improving one’s health (diet, exercise, substance use). Both inputs have a positive effect upon health as health approaches a person’s health potential.

\[
H = h(HC, E)
\] ...

\[
\frac{\partial H}{\partial HC} > 0, \quad \frac{\partial H}{\partial E} > 0
\]

Healthy efforts may have both short term and long term effects. In the long term the positive effects relate to postponing the deterioration of one’s health capital, so that the need for future health care is deferred or reduced. The short term effects would commonly be a lower incidence of complaints such as respiratory and muscular pain – problems that would otherwise have required cure. Some lifestyle efforts may be a direct substitute for preventative medications, eg reduced salt intake versus anti-hypertension drugs to reduce the risk of cardiovascular diseases.

However the pivotal behavioural assumption of the model is that both the marginal benefit of HC and E decline with rising health and therefore with rising values of the other input; that is

\[
\frac{\partial^2 H}{\partial E \partial HC} < 0
\]

In the limit the assumption is necessarily true. If HC has returned a person to full health, then the marginal effect of efforts will be zero. Possibly of greater empirical importance if HC is perceived as capable of maintaining full health the perceived benefits of own efforts will be zero. The cross partial is symmetrical and rising values of E will have a similar negative effect upon \( \frac{\partial H}{\partial HC} \).

Individuals have a utility function, equation (2) that includes wealth, W, health, and efforts. While wealth and health are normal in the sense that ‘more is better’, efforts involve disutility, ie ‘less is better’. This is because they refer to instrumental activities, ie they are undertaken to improve health and not for pleasure. The disutility arises from the fact that there are barriers involved in spending time and energy on physical exercises, as well as in sacrificing current pleasures from unhealthy food and substances. These are perceived disutilities prior to attempted healthy lifestyle changes. While it is possible that after breaking such barriers and new habits have been formed, people may experience positive utility from the same efforts, the current model is driven by ex ante preferences, ie by ‘decision utility’.

\[
U = u[W, H(HC, E), E]
\] ...

\[
\]
\[ \frac{\partial U}{\partial H} > 0; \frac{\partial U}{\partial E} < 0 \]

Wealth, W, is constant, when health care is free at the point of delivery (publicly funded), and there are negligible monetary costs involved in making own healthy efforts. Health care is assumed to have few or no side effects, i.e., there is no direct effect from HC on utility. In order to maximize utility individuals will adjust \( E \) as shown in equation 3 which is derived from the first order condition from the Lagrangian for the constrained maximisation of utility:

\[ \frac{\partial U}{\partial E} = -\frac{\partial U}{\partial H} \cdot \frac{\partial h}{\partial E} \]  

Rearranged, equation 3 is equal to equation 4 which states that the slope of the indifference curve \( \frac{\partial U}{\partial E} \cdot \frac{\partial U}{\partial H} \) between \( E \) and \( H \) on the right hand side of the equation must equal \( \frac{\partial H}{\partial E} \) which is the slope of the production function, equation 1. In other words for maximum utility under the production constraint the marginal rate of substitution and marginal rate of transformation between \( H \) and \( E \) must be equal. This is shown at \( P_1 \) in Figure 1.

\[ \frac{\partial H}{\partial E} = \frac{\partial U}{\partial E} \cdot \frac{\partial U}{\partial H} \]  

The effect of increasing the use of health care due to an autonomous shift in technology or the supply of doctors can be found by differentiating (1) with respect to \( HC \), which results in equation 5:

\[ \frac{dH}{dHC} = \frac{\partial H}{\partial E} \cdot \frac{\partial E}{\partial HC} + \frac{\partial H}{\partial E} \cdot \frac{\partial E}{\partial HC} \]  

The impact of the increase has an initial positive effect, \( \frac{\partial H}{\partial HC} \) and then an indirect effect of \( HC \) on \( E \) which reduces health.

As \( \frac{\partial H}{\partial E} > 0 \), the sign of the indirect effect depends upon the sign of \( \frac{\partial E}{\partial HC} \).

From the initial assumptions this will be negative. Since \( \frac{\partial^2 H}{\partial E \partial HC} < 0 \) an increase in \( HC \) implies a smaller value of \( \frac{\partial H}{\partial E} \). From the utility maximising condition (4) this implies a corresponding decrease in \( \left( \frac{\partial U}{\partial E} \right) \left( \frac{\partial U}{\partial H} \right) \) the slope of the indifference curves in Figure 1. With diminishing marginal utility this can only occur if people reduce \( E \). In sum, normal preference and production relationships imply a substitution of health care for own efforts when individuals maximise utility.

This is illustrated in Figure 1. With less or less effective health care, the optimal level of efforts, \( E_1 \), results in the health level \( H_1 \). The introduction of effective health care involves a vertical shift.
in the production function. The individual will substitute free health care for costly (ie unpleasant) own efforts, resulting in efforts level, $E_2$, and health level $H_2$, ie better health with less efforts.

**Figure 1 Effective health care, less: ‘own effort’ and better health**

The difference between the two production functions illustrates the expected change in productivity from the introduction of new effective health care. An exception might occur with a technology which increased the marginal productivity of efforts, however examples are hard to envisage, with the notable exception of an exogenous change in the doctor-patient encounter such that doctors encouraged greater patient efforts at self-care and patients adopted the doctor’s advice.

An interesting and a special case of this model is when the net effect of increased health care on health becomes negative. In equation 5:

$$\frac{\partial H}{\partial HC} < -\frac{\partial H}{\partial E} \cdot \frac{\partial E}{\partial HC}$$  \hspace{1cm} \ldots (5b)

This occurs when there is a sufficiently strong substitution of health care for own efforts, ie when $\frac{\partial^2 H}{\partial E \cdot \partial HC}$ is quantitatively large as shown in Figure 2. In this, the increase from $HC_1$ to $HC_2$ significantly reduces $\frac{\partial H}{\partial E}$ the slope of the production function depicted in the figure. It remains less than the slope of the utility function over a sufficient range that personal efforts are scaled back until the net health effect is negative. With improved medical care marginal efforts have a lesser health pay-off but have the same cost to utility, inducing a sufficient adjustment of lifestyle that overall health is reduced although, again, utility rises.
5. Discussion

A ‘free’, (at point of delivery) National Health Service creates a form of moral hazard or market failure which is analogous to, but distinct from, the ex ante moral hazard discussed in the literature. In the latter case increased insurance against financial losses increases income when a person suffers avoidable illness and the anticipated increase in income encourages self neglect and the occurrence of the illness. The individual’s concern relates to income. In our model increasing health care resources to insure populations against the adverse effects of illness directly increases health but reduces the consequences of self neglect and thereby encourages it. The individual’s concern relates to health and the reduced health benefit from incremental prevention. There are no behavioural consequences of an income effect as the institutional context is an NHS which fully protects income.
If a link was created between own efforts and the direct personal cost of health insurance, there would be a financial incentive to increase efforts – the greater the effort, the lower the expected health care costs, and thus, the lower the individual health insurance premium. This reintroduces the indirect effect of efforts on wealth which exists in the previous literature. Own efforts are rewarded and mitigates the substitution effect in the model. Typically, however, this nexus between efforts and personal insurance costs do not exist in an NHS – or, for that matter, in most private insurance schemes.

In contrast, there are situations where the link between efforts and wealth is negative. A healthy diet may be more costly than an unhealthy diet, and gym memberships and other exercise initiatives may involve increased costs. Returning to Equation (2), this implies that increased efforts, E, reduces wealth, W. This is equivalent to a change in the relative prices of own efforts and health care in such a way that individuals would substitute even more efforts for free health care.

This form of moral hazard postulated in our model is similar to the iatrogenic effects of health care famously postulated by Ivan Illich (1975). In ‘Medical Nemesis’, Illich argues that the increased medicalisation of illness leads to a cultural change and loss of individual will and autonomy. In our model the problem is that health care is effective and that it is rational – utility maximising – for people to reduce their own health efforts. Rather than changes in culture and preferences, incremental adjustment occurs rationally because of the relatively lower marginal productivity of efforts and the substitution of effective medical care. Even in the extreme case where final health falls, utility rises.

Analytically the key driver of the new model is the assumption that the cross partial between health care and healthy efforts is negative: that as health is increased by one input in the health production function, the marginal product of the other falls. The two inputs are perceived as substitutes. Systematic review of the epidemiological literature indicates the existence of strong links between modifiable risk factors and population health. Health care is widely seen as efficacious when it is available. The benefits of negative effort – self-indulgence – are less obvious. People may choose to diet for reasons of appearance, and exercise for enjoyment rather than for health gain. But the assumption that is needed for the model is that marginal, not total, health related activities are affected by a concern for health; that because of a concern for health these activities will be increased as compared with the level selected for other reasons.

The importance of the nexus between the societal provision of health care and the negative effect upon individual health behaviours depends ultimately upon its quantitative significance, about which we have no direct evidence. The literature cited earlier indicates that the effects of financial insurance upon behaviour are not very important. This does not imply that effects mediated by health per se are insignificant unless the assumption is made that health and income are fully interchangeable. Even in the presence of flexible private insurance and labour markets this is implausible. Rather, the empirical significance of the model depends upon the individual belief in the efficacy of health care and their willingness to adopt unhealthy lifestyles. The epidemic of obesity in western countries suggests that there is this willingness. Whether causal or not, this has coincided historically with a substantial increase in the provision of health services.

With increasing health in most developed countries it appears unlikely that there are currently examples of aggregate national health status deteriorating for the reason portrayed in Figure 2. However, the obesity epidemic may change this as there is evidence that childhood obesity in the US may lead to reduced life expectancy (Olsansky 2005). Consistent with the argument here, the US leads the world in obesity and has the largest and most aggressively marketed health
sector. It would be unsurprising to find that this has influenced the perception of the efficacy of medical care and that the relative return through health related efforts is perceived as being less.

6. Policy Implications

Some people clearly prefer an unhealthy lifestyle to a healthy one. Others would like a healthy lifestyle and better health but the NHS has tipped the balance between rewards and efforts in a way that partially subverts their efforts. Their utility may rise with better health care but their resolve to help themselves is partially undermined. There is increasing evidence that many people wish they had the will to change their admittedly unhealthy habits – most smokers wish they did not smoke, most overweight people would prefer to lose weight (Thaler and Sunstein 2008). But the effect of their failure is cushioned by an effective health service.

As the institutional setting for the model presented here is a publicly funded health system there is no link between an individual’s expected health care use, and the level of their direct contribution to the financing of health care. This leads to market failure as desirable incentives which would exist without an NHS will be suppressed and this justifies compensatory policy.

Following the principle of don’t subsidise substitutes, one approach to this problem would be to increase the out of pocket price of health care. This would, however, violate the widely stated principle of ‘equal access for equal need’ found in publicly funded health services. The alternative approach, therefore, is to lower the relative cost of own efforts: to introduce financial or other incentives for these activities but not via insurance premiums. These include subsidies for healthy foods and for activities which increase physical exercise. The fiscal costs by such subsidies might be matched against the revenues raised by indirect taxation of unhealthy substitutes: tobacco, obesogenic food and drink and products encouraging sedentary behaviour. These suggestions are not new. However, the policies are justified here in terms of efficiency and the need to offset the adverse influence of another arm of social policy (ie free health care). Consequently, the policies should be supported by those sceptical of paternalism but concerned about market failure.

7. Conclusions

The equity argument behind publicly funded health care is founded on the view that how people fare in the biological and social lotteries should not affect their access to health care. However, differences in health also stem from personal health related choices and these may be changed by increased access to or the quality of health services, introducing a form of ‘market failure’: less self-protection than would occur without public policies to increase access. With the social goals underpinning an NHS it is not possible to make health care entitlements depend on past behaviour. Policy interventions can, however, be in the form of ex ante incentives – subsidies to activities which promote healthy efforts, and taxes upon unhealthy behaviours. These suggestions are not new, but the argument presented here is that they may be justified for reasons of market failure rather than as the result of paternalism.
References


