The potential role of risk-equalization mechanisms in health insurance: the case of South Africa

NEIL SÖDERLUND and SOLANI KHOSA

Centre for Health Policy, University of the Witwatersrand, and Anglo American Corporation of South Africa, and Centre for Health Policy, University of the Witwatersrand, Johannesburg, South Africa

International agencies such as the World Bank have widely advocated the use of health insurance as a way of improving health sector efficiency and equity in developing countries. However, in developing countries with well-established, multiple-player health insurance markets, such as South Africa, extension of insurance coverage is now inhibited by problems of moral hazard, and associated cost escalation and fragmentation of insurer risk-pools. Virtually no research has been done on the problem of risk selection in health insurance outside developed countries. This paper provides a brief overview of the problem of risk fragmentation as it has been studied in developed countries, and attempts to apply this to middle-income country settings, particularly that of South Africa. A number of possible remedial measures are discussed, with risk-equalization funds being given the most attention. An overview is given of the risk-equalization approach, common misconceptions regarding its working and the processes that might be required to assess its suitability in different national settings. Where there is widespread public support for social risk pooling in health care, and government is willing and able to assume a regulatory role to achieve this, risk-equalization approaches may achieve significant efficiency and equity gains without destroying the positive features of private health care financing, such as revenue generation, competition and free choice of insurer.

Introduction

Health insurance has evolved to pool the risk of incurring high health care costs across individuals. Typically, serious illness is a rare event but can involve very high treatment costs, often far above the savings reserves of individuals. For most poor and middle-income persons, it is simply not feasible to save sufficient amounts to meet these costs, should they occur. Consequently, insurance, or the pooling of risks across individuals, is an efficient solution to providing cover for high health care costs for most individuals. In most developed countries, formal insurance arrangements, linked largely to employment, arose spontaneously initially, and were only later taken over, mandated or at least actively regulated by the state. Today, few industrialized countries have substantial proportions of the population without any health cover (Abel-Smith 1992).

The evolution of insurance arrangements in many post-colonial countries, particularly in Africa and Asia, has been slightly different. Upon gaining independence, many governments intervened directly in the financing and provision of health care. Particularly in poorer countries, however, state initiatives have floundered because of low levels of funding, inequitable distribution and inappropriate services. Some of the resulting gap in health care access was addressed by foreign donors. In addition, governments have encouraged insurance-based funding in order to relieve the burden on public finances, and to improve the quality of care provided. This approach has been supported by organizations such as the World Bank (Shaw and Griffin 1995). Consequently, in many developing countries, particularly middle-income countries, mutual and state-sponsored social health insurance arrangements have arisen to supplement the basic cover provided by public institutions.

It is convenient to divide types of risk sharing into what we will call ‘insurance’ arrangements, where
a defined benefit accrues to an individual associated with his or her contribution to a health insurance fund, and public systems where general taxation is used to fund a health care system to which all residents have access. This paper deals specifically with insurance arrangements in developing country settings where multiple insurers are reasonably well established, and fund a significant share of health care spending. Examples of such countries include South Africa, Thailand, Korea, Morocco (World Health Organization 1995), many Latin American Countries and some formerly socialist East European countries (Cichon 1990; McCarthy et al. 1995). For the most part, the discussion applies to middle-income countries, since formal health insurance coverage is very limited in most low-income countries (Griffin and Shaw 1995).

As in developed countries, policy-makers in developing countries with long-established insurance arrangements are increasingly having to turn their attention to two problems associated with insurance that hamper the further extension of coverage. The first, moral hazard, refers to the propensity for insured persons to buy more of a good (in this case health care) than they would if they had to pay out-of-pocket for it. Moral hazard in the context of health insurance in developing countries has received some attention in the literature (Peabody et al. 1995; Barros et al. 1986) and is not dealt with here. This paper concentrates instead on the problems of risk distribution between insurers, and the consequent effects that this has on coverage, efficiency and equity in health care.

After giving a brief overview of insurance-based health care financing in South Africa, the paper describes the economic theory relating to risk selection, and possible solutions that could be applied. Particular emphasis is given to the concept of a risk-equalization fund, which has been suggested by the South African Committee of Enquiry into National Health Insurance (Committee of Inquiry into a National Health Insurance System 1995) and implemented in a number of developed countries (McCarthy et al. 1995). Common misconceptions regarding risk-equalization mechanisms are dealt with before describing how such schemes might work. We then discuss practical issues around assessing the suitability of risk equalization and how to implement risk-equalization mechanisms. Throughout the paper, the term ‘health insurance’ is used to refer to all forms of subscription-funded (as opposed to tax-funded) third party cover for health care costs, and should not be interpreted, as is often the case in South Africa, to include only policies offered by for-profit insurance companies.

**Health insurance in South Africa**

Three broad types of risk sharing are commonly found in the health sector: mutual (or non-profit) insurance, where the enrollees are also the sole shareholders of the organisation; for-profit insurance, where private companies sell underwritten policies, from comprehensive medical benefit cover to insurance against a single disease; and social, or national health insurance, where the state (or an organization on behalf of the state) acts as insurer. Currently, in South Africa, the first two varieties exist. Non-profit mutual insurers are the oldest form, and are called ‘Medical Aid Societies’, or ‘Medical Schemes’. They are generally employment-based, did not risk rate contributions until recently, and often involve income-related cross-subsidies from richer to poorer members. They are thus similar to Germany’s ‘Sickness Funds’ (Wysong and Abel 1990). For-profit, open-enrolment health insurance arrangements are a relatively new phenomenon. These are generally offered by large insurance companies, for whom health insurance is a relatively small component of their business. They charge risk-related premiums, offer a wide variety of benefit options, and are aggressively marketed through brokers. Membership of Medical Aid Societies has been shrinking over the last ten years or so, whereas that of for-profit insurers has been growing, although they have a small share of the market overall. In 1992/3, 40% of health care spending was channelled through medical schemes, compared to an estimated 3% through risk-rated products sold by insurance companies (McIntyre 1995).

In addition to these two options, the South African Department of Health is considering the introduction of mandatory state-administered Social Health Insurance (SHI) for low-income formal sector employees currently unable to afford either of the above forms of insurance. This implies that in the future, all three types of health insurance may exist in South Africa. In addition to the above-mentioned financing mechanisms, a system of primary health care clinics and hospitals, both financed and provided by the state, acts as a provider of last resort when insurance funds run out, as well as a sole provider for the majority of the population unable to afford any health insurance. Approximately 39% of health care spending is by the state sector, four-fifths of
which goes on hospital care provision (McIntyre 1995).

Since 1989, medical schemes have been deregulated in South Africa, allowing them to risk-rate premiums and removing stipulations that required them to accept all applicants. A second wave of deregulation proposed by the outgoing Nationalist Party Government in 1994 was not implemented, and it is the intention of the current government to re-regulate aspects of the industry, especially with regard to risk selection and dumping. It is recognized, however, that the same regulations will have to apply to both medical schemes and for-profit insurers, as failure to do this will result in low-risk individuals being attracted out of the regulated medical schemes industry into the largely unregulated commercial health insurance industry.

Health insurance, risk and adverse selection

As with all forms of insurance, health insurees have differing probabilities of claiming from their insurer. If insurers charge all enrollees the same premium, and consumers have some information on their likely probability of claiming as a result of illness, then low-risk individuals will tend to opt out of the scheme, with consequent requirement to increase premiums for remaining enrollees, causing the remaining lower risks to exit, and so on. In their classic paper on the topic, Rothschild and Stiglitz suggest that such a system is inherently unstable, with no equilibrium being possible (Rothschild and Stiglitz 1976) or if equilibrium does occur, it is at sub-optimal levels of cover. In this situation, the potential enrollee’s preference for high or low levels of cover itself acts as a signal to the insurer of their likely risk of requiring health care, with those choosing high levels of cover being more likely to be high risk. This phenomenon is known as ‘adverse selection’, which may be defined as the propensity of low-risk individuals to underinsure.

Attempts to reduce adverse selection may cause problems of their own, however, particularly with regard to equity. If the health characteristics of an individual are known to insurers, then, in a competitive market without information costs, insurers can charge perfectly risk-rated premiums to all insurees, avoiding the need for separating contracts. This will eliminate adverse selection, since low-risk people will be charged appropriately lower premiums. High-risk enrollees, however, will have to pay much higher premiums, and in many cases may not be able to afford insurance at all. Alternatively, insurers may exclude persons from cover altogether when it becomes evident that they are likely to have high health care costs (‘dumping’), or make special efforts to attract low-risk, healthy enrollees (‘cream-skimming’) (van de Ven and van Vliet 1992). In fact, as Newhouse points out, in the real world, the existence of information differences between insurers and insurees may reverse the Rothschild Stiglitz equilibrium result, with high-risk persons being underinsured, and low-risk ones obtaining desired levels of cover (Newhouse 1996). Both the forced exclusion of high-risks from cover (dumping), and the self-exclusion of low-risks (adverse selection) are evident in the health insurance market in South Africa (van den Heever 1994).

The excess of information that enrollees have relative to insurers has been termed ‘consumer information surplus’ (van de Ven and van Vliet 1995). This means simply that an individual probably has a better idea of their own risk of illness than their insurer. The same notion of information differentials can be applied to the relationships between insurers and state regulators, and between providers and insurers (Newhouse 1996).

The concept of information asymmetry is crucial both to understanding the way health insurance markets operate, and to designing possible solutions. The information that insurers and insurees use may be understood as a set of risk ‘signals’. A signal is a piece of information that indicates, in the case of health insurance, the probability of requiring medical care. The most common examples include age, gender, previous medical history and health service utilization, social class, place of residence, occupation and ethnic origin. In the first instance, signals are used by insurers to determine risk-rated, or so called ‘actuarially fair’ premiums. For example, insurers would charge more for health insurance for a 70-year-old than for a 30-year-old. Similarly, amongst 30-year-olds, one with a strong family history of heart disease would be charged a higher premium than one without. At best, however, all such signals combined rarely predict more than 5-10% or so of the variation in health care costs incurred over a year (van Vliet and van de Ven 1993; van de Ven and van Vliet 1995). Insurers are thus still left with considerable uncertainty regarding the likelihood of claims at the individual enrollee level.
Table 1. Possible solutions to risk fragmentation in health insurance

<table>
<thead>
<tr>
<th>Policy reform</th>
<th>Risk selection¹</th>
<th>Risk stratification</th>
<th>Political feasibility²</th>
<th>Consumer choice³</th>
<th>Enforceability</th>
<th>Insurer competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Single national insurer</td>
<td>Eliminated</td>
<td>Eliminated</td>
<td>Poor</td>
<td>Nil</td>
<td>Good</td>
<td>Lost altogether</td>
</tr>
<tr>
<td>2. Mandating open enrolment</td>
<td>Decreased</td>
<td>Unaffected</td>
<td>Good</td>
<td>Full</td>
<td>Poor</td>
<td>Maintained</td>
</tr>
<tr>
<td>3. Restricted choice models</td>
<td>Decreased</td>
<td>Unaffected¹</td>
<td>Fair</td>
<td>Limited</td>
<td>Fair</td>
<td>Partially maintained with franchise type</td>
</tr>
<tr>
<td>4. High-risk pool (HRP)</td>
<td>Decreased</td>
<td>Unaffected</td>
<td>Fair</td>
<td>Limited for bad risks</td>
<td>Fair to poor⁵</td>
<td>Only possible for good risks</td>
</tr>
<tr>
<td>5. Risk-equalization methods</td>
<td>Decreased</td>
<td>Decreased</td>
<td>Fair</td>
<td>Full</td>
<td>Fair</td>
<td>Enhanced</td>
</tr>
</tbody>
</table>

¹ i.e. cream-skimming or dumping.
² Political feasibility obviously depends heavily on the prevailing political environment. This refers specifically to political feasibility in South Africa.
³ This refers to consumer choice of insurer, not provider.
⁴ Risk stratification could still occur unless accompanied by a risk-equalization approach.
⁵ Depends on criteria for entry into HRP.

Active selection of low-risks is evident predominantly in competitive, for-profit insurance environments. In mutual non-profit insurance arrangements, however, such as the traditional medical scheme in South Africa, there is also the potential for stratification of risks without selection. This is because members of mutual insurers are not selected at random from the population. In general, such schemes are employment-based, and enrollees mirror the employee mix of the industry in which they work. A common phenomenon is that, where an industry has undergone significant contraction, medical schemes to which its employees belong are left with an excess of elderly people, since new recruitment of young employees has not occurred. The main medical scheme covering miners in South Africa, the Mines Benefit Society, has gone bankrupt because the significant labour force contraction in the mining industry led to it being left mainly with retired enrollees, who had high health care requirements.

The potential for any form of active risk selection (i.e. adverse selection, cream-skimming or dumping) is offset by the costs of gathering information to facilitate selection, and inertia in both consumers and insurers (Pauly 1974; Newhouse 1996). It is also unusual to find pure competition and open enrolment for health insurance, with most schemes or choice of schemes being linked in some way to employment. Empirical studies from the US show definite evidence of the existence of cream-skimming and adverse selection, but significant risk-pooling still occurs spontaneously, suggesting that information and friction costs, as well as spontaneous risk-pooling resulting from large group purchasing, offsets pure selection phenomena in the real world (Schuttinga et al. 1985; Dowd and Feldman 1992; Thorpe 1992).

Possible solutions to adverse selection, cream-skimming and dumping

We will go on to outline a number of possible health care financing arrangements that ameliorate selection phenomena in health insurance. All assume the public acceptability of risk pooling, and by implication, the importance of social solidarity in health care financing. There is an implicit social contract in social risk-pooling arrangements, i.e. that in exchange for participation while healthy, individuals will be guaranteed future membership of the pool when unhealthy. Table 1 summarizes the potential impact of each approach.

Single insurer models

Mandatory membership of a single national health insurer ensures that all selection problems disappear.
Premiums can be set in any way that is considered equitable, so long as total costs are covered. While the large size of the National Fund may give it signiﬁcant monopsony purchasing power, it also implies that there is no competition between insurers, and this might be associated with inefficiency and non-responsiveness to consumer preferences. Where there is already extensive private insurance provision, its forced replacement by a National Health Insurance fund would also arouse considerable opposition from existing insurers. Finally, in less-developed countries, the ability of the state to administer such a large operation might be called into question.

Mandating acceptance of all enrollees

Selection may also be reduced by mandating that insurers accept all applicants, and by forbidding insurers from dumping existing members who become unhealthy. This approach has already been attempted in parts of the US and a number of other countries, but is very difficult to enforce as long as consumers are allowed to choose which plan they join (Luft 1995). If enrollees can choose which insurance plan to join, plans can encourage the enrolment of certain groups by offering benefits preferred by the groups. For example, an insurer may offer full coverage for sports injuries and maternity care, but limit coverage for long-stay institutions, thus appealing primarily to the young, fit and healthy. Even if the package of benefits is ﬁxed by regulation, as suggested by Enthoven (1988), subtle differences, like the number of paediatricians versus geriatricians on the panel, might swing the marginal potential enrollee (Gauthier et al. 1995).

Restricting choice of insurer

If choice is dispensed with, and enrollees are randomly, or otherwise, allocated to plans, equitable risk pools with full coverage can be achieved in the presence of multiple insurers. Alternatively, insurers may be allocated responsibility for a geographically deﬁned population, as is the case for District Health Authority purchasers in the UK NHS internal market. While there is still the potential for risk stratification between geographically deﬁned populations, this will be relatively static over time, and can be easily compensated for by risk-adjusted capitation payments (Smith et al. 1994). However, the loss of competition between insurers in this arrangement may have a negative effect on efficiency (Enthoven 1988).

Some competition and consumer choice may be maintained through franchising arrangements, as is the case with some Medicaid and Medicare purchasing agreements in the US (De Lew et al. 1992). Under this arrangement, an insurer is contracted to cover a given population for a set period of time. At the end of this period, the covered population can decide to keep their current insurer, or award the franchise to another insurer.

Another variant of the ‘restricted choice’ model that maintains some competition is the Health Insurance Purchasing Co-operative (HIPC), whereby an external agency, not linked to employment, purchases insurance from competing insurers on behalf of a large group of potential enrollees with mixed risks (Ellwood et al. 1992). Irrespective of the effects of competition on efﬁciency, some see the loss of consumer choice as a negative feature in itself.

High-risk pools (HRPs)

Under this arrangement, high-risk individuals, deﬁned either as those unable to purchase commercial cover, or those with some objectively deﬁned ill-health characteristics, are guaranteed access to one high-risk insurer at an affordable rate. This rate is subsidized from tax revenues or a levy on other insurers. Risk rating may or may not be allowed in the rest of the market, where multiple insurers continue to compete for clientele. The classic example of a public HRP is the Medicare system for the elderly and disabled in the United States. Medicare was instituted because of the inability of the elderly and disabled to purchase affordable cover in a risk-rated commercial insurance environment.

While an HRP may provide some cover for bad risks, it has a number of associated problems. Firstly, unlike risk-equalization methods, this approach eliminates the potential for insurer competition for a sizeable chunk of health care activity and expenditure. Secondly, if objective criteria are not set on who may gain entry to the HRP, a secondary moral hazard phenomenon may operate, whereby community rated insurers, once they have paid their HRP levy, have incentive to dump all enrollees with above average expenditure into it. If the primary insurance market is allowed to risk-rate, then enrollees paying above the HRP average premium will have incentives to try to enter the HRP and beneﬁt from the subsidy. The potential for HRP growth, and consequent health care cost growth, under this policy approach is obviously signiﬁcant.

The HRP approach is essentially a partial solution, where regulators are unable or unwilling to influence
the mainstream business of insurers, and thus adopt all of the insurance system ‘outcasts’ themselves. The private health insurance industry in South Africa has recently put forward a proposal for an HRP to counter government plans of mandatory community rating or risk-equalization mechanisms (Gore A 1977, unpublished paper). They suggest, however, that it could be paid for by a 1% levy on existing insurers, which seems implausibly low when compared with the US Medicare experience.

Risk-equalization funds

Risk-equalization approaches have appeared relatively recently in the health insurance literature. This is despite the fact that they have existed for some time in geographically organized, publicly funded systems, such as the UK National Health Service, where they are referred to as ‘weighted capitation’ (Smith et al. 1994; Carr Hill et al. 1994). In principle, a risk-equalization fund operates as follows. Multiple insurers, of both the open enrolment and mutual society types, are allowed to operate and to charge premiums on whatever basis they choose. At the same time, a central fund is constituted from a levy on insurance premiums, general taxation, or earmarked wage deductions. This fund then compensates insurers according to the expected health care costs of their enrollees. This compensation may be paid directly to the insurer, or to enrollees themselves in the form of health care vouchers. The fund aims to cover, for each enrollee, the difference between a standard premium per person and the full expected costs of care for that person, given their age, sex, chronic conditions, etc. Consequently, insurers no longer have a disincentive to recruit high-risk enrollees, since they are compensated from the fund for such persons’ expected higher health care costs. Risk equalization has the advantage that it also eliminates inequities due to non-intentional risk stratification between insurers.

Bovbjerg defines reinsurance as ‘. . . an assumption of risk between or among insurers’ (Bovbjerg 1992). Health insurers usually reinsurance in an effort to share the risk of extraordinarily high-cost illness episodes. The aim is to protect against insolvency brought about by unexpectedly high claims. Unlike reinsurance, a risk equalization system pays out primary insurers on the basis of their ex ante probability of incurring high costs, given their enrolee risk profile. Whether they actually incur high costs or not does not affect their payment from the risk-equalization fund, nor does risk equalization protect from insolvency due to bad management, fraud or unexpected catastrophes.

Reinsurance per se does nothing to ameliorate risk selection, since reinsurers face similar concerns as primary insurers and will thus charge risk-rated premiums to primary insurers. This means that primary insurers with bad-risk pools will still have to charge higher premiums or exclude high risks (Bovbjerg 1992; Pauly 1992). Furthermore, a second cycle of moral hazard is introduced, with reinsured primary insurers having less incentives to practice good cost management, and limit risky business expansion programmes. The Belgian Government operated a mandatory reinsurance pool for sickness funds whereby insurers with surpluses were required to transfer funds to those with deficits. Unsurprisingly, all funds soon ensured that they ran deficits (Nonneman and van Doorslaer 1994). Risk-equalization mechanisms, on the other hand, offer no greater payment to inefficient insurers, and thus do not deplete incentives to control costs.

Finally, reinsurance introduces another cycle of transaction costs. In South Africa, brokerage fees on reinsurance alone constitute a significant inefficiency in the health insurance market (Registrar of Medical Schemes 1996, personal communication).

Income cross-subsidies and risk equalization

There is also potential for confusing the concepts of risk equalization and income-related cross-subsidization. Income-related contributions, although they might have merit for other reasons, have no necessary implications for risk selection. In most instances, we would expect poorer people to have higher levels of need for health care. This may well be offset by lower utilization levels for equivalent need, however, because of poorer access to facilities, the proportionately greater influence of co-payments, and a lower sense of entitlement to health care benefits.

Common misconceptions regarding risk equalization

Reinsurance and risk equalization

Superficially, reinsurance for high cost events, and risk equalization may seem like the same concept.
The implementation of risk equalization, however, may initially have effects opposite to those of income cross-subsidies. In South Africa, the average age of people currently insured is considerably higher than that of low-income potential enrollees, who are all of working age. Since age is probably the single most important predictor of health care need, risk equalization would imply a transfer of funds from the new, low-income members to existing scheme members. The members of existing schemes, however, have significantly higher incomes, and an income-based cross-subsidy scheme would thus involve transfer in the opposite direction. With time, the health care needs of low-income enrollee groups are likely to increase as their age structure becomes more similar to that of the general enrollee population.

From a policy point of view, it is important that these two transfer arrangements are kept separate, and based on independent evidence of income and health care risks for cross-subsidy and risk-equalization pools respectively.

**Design of risk-equalization mechanisms**

**Measurement of risk**

This is the area of risk equalization that has been most researched. Virtually all of this research has been done in developed countries. Ideally, what the risk-equalizer seeks is a set of signals that will perfectly predict the likelihood of an enrollee needing health care over the insurance period (say one year). An ideal set of signals should be valid (i.e. they explain a large proportion of the variation in health care costs), generalizable to all enrollees, easy to elicit, and objectively verifiable. Commonly used signals include age, gender, previous medical history, behavioural factors (such as smoking) and socioeconomic status. Age, gender and other demographic factors are generalizable, and easy to elicit and verify. Unfortunately, they explain a very low proportion of the variation in health care costs, typically less than 3% at the level of individuals (van Vliet and van de Ven 1992). Nevertheless, they have been adopted as the basis for risk-equalization mechanisms in a number of countries including Israel (Chinitz 1994), Medicare in the US, Switzerland, Germany (Graf von der Schulenburg 1994), and the Czech Republic (World Health Organization 1995).

More sophisticated approaches essentially make use of information on past health care usage or costs (van Vliet and van de Ven 1993; van de Ven and van Vliet 1995; Hornbrook et al. 1991; Newhouse et al. 1989; Matsaganis and Glennerster 1994). The problem with risk-adjustment based on past costs is the potential for over-servicing because past costs are not entirely exogenous. Insurers that purchase care in excess of need (either in quantity or quality) in year one receive higher payments in year two than other insurers with a similar enrollee mix but who were more cost-conscious in the past. A compromise position is to use the occurrence of externally verifiable indicator diagnoses as risk markers. An example of this approach is the Diagnostic Condition Groups system (DCGs) developed by Ash and others (Ash et al. 1989), which uses diagnostic information from previous hospitalizations to allocate enrollees to one of nine risk groups. This allows for more sensitive adjustment than would be possible just with age, but is less easily influenced by insurers than crude claims history. Even with use of past utilization information, however, typically no more than 7-10% of health care cost variation in the following year is explained (van de Ven and van Vliet 1995).

As van de Ven (van de Ven and van Vliet 1995) points out, however, the poor absolute predictive ability of commonly available signals should not result in their rejection as instruments for risk equalization. The key question is whether they predict a significant proportion of the ‘theoretically predictable’ variation in future health care costs. This point is illustrated by Figure 1.

Of the total variance in future health care costs, only a small proportion (Newhouse suggests about 13%) are non-random in nature (Newhouse et al. 1989). If we assume that one player, say the insurer, has all the information required to predict the theoretical maximum, then the difference between the variation predicted by the insurer’s information, and that predicted by the information of a less-informed regulator (predicted variation in Fig. 1), may be termed the insurer information surplus. In devising an ideal risk-equalization formula, the regulator need only seek to reduce the insurer’s information surplus over that of the regulator to zero, not predict all variance in future health care costs. Put another way, all the signals that the insurer has at his disposal should be available to the regulator if residual risk selection by the insurer is to be prevented (Schwartz 1995). Costs of information gathering for both players are not zero, however, and it is unlikely that either will try to obtain the theoretical maximum, but will instead opt for a more cost-effective, intermediate level of information.

---

Equalizing risks in health insurance 347
The above discussions all apply to predicting future health care costs for an individual. When predicting costs for groups of individuals, for example employees of a single employer, a considerable proportion of unexplained cost variation disappears because of the dissipation of random errors in large samples. Robinson et al. (1991) show that using a standard risk-adjustment formula, there is a median prediction error for costs in the next year of 2.6% for groups of around 100 in size, but this drops to 0.5% for groups of around 1000 members, and to less than 0.1% for groups over 5000 members.

In summary, therefore, an optimal risk-equalization formula will depend on the signals available to insurers, the degree to which insurers use those signals to risk select, and the costs to the regulator of acquiring similar information.

Signals such as age, sex, area of residence and occupational group could be easily elicited for the currently insured population in South Africa. Poorer countries with less sophisticated health information systems might find it difficult to generate risk estimates based on past diagnoses since they often do not record diagnostic data. Health insurers in South Africa, have, until recently only collected data on procedures performed on enrollees, not data on the underlying illnesses for which the procedures were performed.

In practice, most risk-equalization funds have chosen to compensate insurers at the level of the expected cost of enrollees predicted by the risk signals being used. That is, if 60-year-olds consume 200% more health care resources than the population average, the risk-adjusted premium an insurer receives for 60-year-olds will be 200% more than the average. Since there is almost always information asymmetry between insurer and regulator, however, some residual selection is likely to occur.

Evidence from the US suggests higher than expected costs for those who already have high predicted costs, and lower costs for those who are already rated as low risk (Schwartz 1995). In other words, despite the risk-equalization mechanism, insurers covering mainly high-risk cases are still being undercompensated. Consequently, an appropriate risk-equalization mechanism may require payment for high-risks in excess of the signal-predicted cost.

In addition to the empirical observation noted above, a number of theoretical and statistical arguments may be put forward to support overpayment on risk signals, which are not described here (Glazer and McGuire 1996; Chernikovsky and Ninio 1995; Pettengill and Vertrees 1982; Thorpe et al. 1988).

**Financial flows and administration**

There are a number of alternatives for administering a risk-equalization scheme. One option is to have a
central risk-pooling agency which collects a set fee per enrollee from all participating insurers, and pays them back a risk-adjusted fee per enrollee. Alternatively, a central agency may give to each member of the population a risk-adjusted insurance voucher, with which they can purchase risk-rated insurance from competing providers (Pauly 1992). The central risk-equation agency may be funded from general taxation, earmarked taxes, or a per enrollee tax on health insurers.

While the conceptual fundamentals of each approach are the same, one may be more feasible than another. In South Africa, and we would suggest, many of these settings, a system whereby the risk-equation pool deals only with insureds might be more suitable because of possible difficulties related to language and literacy in using vouchers, as well as no history of voucher use in other areas of daily life.

Compulsory versus voluntary membership

The issue of compulsory membership needs to be dealt with at two levels: compulsory membership of individuals within insurance schemes, and compulsory membership of schemes within a risk-equation pool.

With regard to the former, failure to mandate a basic level of insurance coverage for all who could afford it would be expected to result in adverse selection if the risk-equation pool drew its funds directly from insurers, rather than taxation. Under this situation, low-risk individuals would effectively be paying an average-risk premium, and might thus decide to self-insure more cheaply.* Consequently, it would seem that, at least for formal sector employees, mandatory membership, or strong inducements such as tax exemptions and penalties for late-age joining, would be needed to maintain a broad mix of enrollees in insurance generally, and hence in the risk-equation pool.

Mandatory membership of the risk-equation pool by insurers would also be essential. If some schemes were allowed to opt out, these would attract an imbalance of low-risk enrollees, thus loading the risk-profile of participating schemes. All medical insurance providers, including mutual insurers, for-profit open enrolment insurers and any public insurer, should be treated equally by the risk-equation process.

Risk equalization and differential benefit packages

In many multiple health insurer environments, schemes aggressively differentiate their products in order to avoid direct price-related competition, and to enable implicit enrollee signalling about their perceived risk level. While some countries have tried to regulate the benefit packages that insurers offer, wide income differentials, particularly in middle-income countries, make this difficult. The fact that different insurers offer a range of different benefits creates a number of complexities for any risk-equation process. If the risk-equation process were to compensate insurers on the basis of the risk of incurring costs given their own unique benefit package, then schemes offering few benefits would end up cross-subsidizing comprehensive schemes with the same risk profile simply because of their higher benefit levels. It is consequently essential that risk equalization applies to a defined package of benefits that all schemes are mandated to cover. As a corollary, risk selection and/or risk-rated premiums would be permitted for any benefits not in the core package.

Implementing risk equalization in developing countries

The South African Committee of Inquiry into a National Health Insurance System (Committee of Inquiry into a National Health Insurance System 1995) has acknowledged the need for more detailed research and consultation prior to implementation of risk-equation mechanisms. What processes may be required to assess the feasibility of risk equalization in developing countries? We believe the following are some of the issues that need to be addressed in the South African context, both in terms of suitability of the approach and implementation, should risk equalization be deemed suitable.

Involving stakeholders

Where government’s main obligation is to directly provide health care for the poor, issues around private medical care financing are often neglected, or approached with considerable circumspection. This appears to be the case, until recently, in South Africa. Government thus needs to develop expertise in the area of private financing, and use this to formulate consensus on broad objectives before further advance in this area is advisable. In South Africa, it has been useful to point out the state’s role as default insurer for those excluded from commercial insurance, a role
that the Department of Health is increasingly unwilling to undertake.

It is thus essential for both technical and political reasons that a significant proportion of the private health care industry is involved in discussing proposals. There can be no better source of information on appropriate signals to include in a risk-equalization formula than the insurers themselves. By the same token, it is unlikely that any would show their full hand if they believed they had information that other insurers did not. The more players that buy into the process, the more open discussion is likely to be, since the potential gains from pooled information would exceed the losses from disclosure of own information. Nevertheless, an independent research process, together with information from other countries, will be needed to supplement the input from insurers.

In South Africa, two particular lobby groups would appear to be particularly opposed to the risk-equalization approach. The first is the subgroup of the actuarial profession engaged in assessing enrollee risk for the health insurance industry. Since most other areas of insurance have limited market expansion potential, and long-established actuarial input, health insurance is one of the few areas of potential job growth for this group. Since risk equalization would not cover non-core benefits, the role of health actuaries would not disappear entirely, however.

The second group opposed to risk equalization is that part of the long-term insurance and pension fund industry interested in selling ‘health pension’ type products. In extreme cases, they are advocating the abandonment of pay-as-you-go, insurance-type approaches for personal medical savings accounts that carry savings over from young to old age. The financial instability of many health insurers in South Africa encourages individuals to attempt to cover themselves through savings vehicles. Since risk-equalization mechanisms would go some way towards ensuring the solvency of insurers and the insurability of individuals, they would be expected to decrease demand for ‘health pension’ type products. In addition, even if insurers do not have specific objections to the concept of risk equalization, many appear to favour the status quo simply out of resistance to change in the regulatory environment which they are currently adapted to.

It needs to be emphasized that risk-equalization processes pose no real threat to the health insurance industry as a whole. Those who stand to lose are insurers who have neglected cost control in favour of risk selection, while those who have got to grips with cost control, and still provide an acceptable product, should benefit. Overall, risk equalization should improve the efficiency of the health care industry by concentrating attention on cost-effective purchasing, thus making care more affordable and extending coverage to those excluded by selection phenomena (McCarthy et al. 1995). Consequently, risk-equalization mechanisms have the potential to increase the total market for health insurance. Final decisions on a risk-equalization formula and administration process would probably need significant, but not necessarily majority, support from the private health insurance industry, especially in situations where the state is the largest single insurer.

**Measuring the importance of risk selection and stratification**

While most parties in South Africa seem to acknowledge differences between the risk pools of various health insurers, there is little evidence on the magnitude of the problem. It is essential for two reasons that this be estimated before the introduction of risk-equalization mechanisms. Firstly, since the introduction of a risk-equalization scheme will incur substantial costs, as well as opposition from insurers that participate in the open-enrolment market, evidence is needed that risk selection or stratification affect a significant proportion of the population seeking insurance, or are increasing over time. Secondly, baseline information on risk distribution is needed in order to judge the success of any remedial measures.

The extent of risk stratification can be measured in the first instance by looking at the age, sex and income distribution of members of existing insurers, and the extent to which these differ. If diagnostic data were available, examining the distribution of chronic conditions would also be essential, since these are generally good predictors of expected health care utilization. Examining differences in costs alone would not be useful, since differences in risk profile would be mixed in with differences in levels of benefits, as well as differing degrees of cost control. It is also useful to study ‘joiners’ and ‘leavers’ from insurance schemes. If joiners tend to be sicker than average, or leavers healthier, then there are probably adverse selection problems. If leavers are more sickly, or joiners healthier, then dumping and cream-skimming are likely to be happening. Lastly, the
changes in risk profiles of schemes over time might give an indication of the urgency of the problem.

To study these issues, however, individually linkable data would need to be obtained from all insurance market players. In South Africa, statute does not require that data on individual enrollees be provided to the regulator, and in the past, insurers have proved reluctant to release these data voluntarily.

**Determining a risk-equalization formula**

As discussed above, an ideal set of signals to be included in a formula should be valid, generalizable, easy to elicit, and objectively verifiable. These can quite easily be identified from existing utilization data collected by insurers. Suitable candidates for evaluation in the South African setting might include age, gender, occupation, and objectively verifiable chronic diseases. While the first three could be relatively easily obtained when enrolling members, the gathering of chronic disease data would require recording of diagnostic information in claims, and a considered approach to diagnostic screening. Specific risk categorisation tools, such as Diagnostic Condition Groupings (DCGs), have not been evaluated in developing country conditions, and their suitability will depend both on the incidence of different conditions, and prevailing treatment practices.

Suitable signals are likely to change with improvements in medical technology, better hospital information, and accumulated research in this area. There will also be a need to monitor residual risk-selection based on signals not included in the risk-equalization formula. Any formula will thus need to be reviewed over time.

**Establishing an administration structure**

In most instances internationally, a single national body has been appointed to administer a risk-equalization fund. Where the participants of the fund include a mix of public and private insurers, it would seem sensible for the risk-equalization fund to be independent of government. While an existing private sector insurer or scheme administrator could be appointed to manage the fund, there would be significant potential for conflict of interest if they continued to sell health insurance as well. Consequently, a semi-autonomous state body, independent of existing private sector interests and government, seems optimal in settings such as South Africa.

While external commentators have generally supported the idea of risk-equalization mechanisms in South Africa, there appears to be a perception that they would be too complex for existing administrative capacity (National Economics Research Associates 1996). This appears to ignore the fact that insurers, by charging risk-rated premiums, have already mastered the application of risk differentials in a given setting. By extension, if risk selection is widespread, then capacity for risk adjustment must also exist (although attracting the capacity into a public sector institution may be more difficult).

**Evaluating risk equalization**

Before the introduction of any risk-equalization scheme, it is essential that good baseline data on coverage and costs exists, and that there is a framework for collecting and comparing future data to establish the success of the scheme. Furthermore, as with all regulation, regulated bodies can be expected to adapt to the new environment, at least in part, by exploiting loopholes in the rules. Consequently, risk-equalization mechanisms will themselves be required to adapt. This might involve use of additional risk signals to determine adjustment payments, changes to the total size of the risk pool, or changes to the basic package of services covered by the scheme.

**Conclusions**

While a significant number of developed countries have introduced risk-equalization mechanisms, or are considering doing so (McCarthy et al. 1995), their use remains infrequent in developing countries, even where competing insurers are a major source of health care finance. Experience in South Africa suggests, however, that the number of people covered by medical schemes is contracting, at least in part because insurance risk pools are fragmenting. This is the main immediate reason why a risk-equalization mechanism has been suggested for South Africa.

More generally, however, the suitability of a risk-equalization approach to improve equity and efficiency of health insurance in countries like South Africa would seem to depend on several factors:

- The current seriousness of risk stratification and selection in health insurance, and the rate of change of these phenomena.
• The ability and willingness of the state to act as a ‘last resort’ insurer for those excluded from insurance.

• Public and political support for cross-generational, and healthy-to-sick risk pooling arrangements.

• Some degree of acceptance of the need for social risk-pooling amongst insurers themselves.

• The definition of a minimum package of essential benefits to be provided by all insurers.

• The identification of a set of signals that can reasonably predict costs of providing core package care to potential enrollees.

• The capacity for administering, enforcing and monitoring a risk-equalization scheme.

Research in each of these areas is required before decisions can be made about the suitability of risk-equalization mechanisms. Furthermore, extensive debate involving key stakeholders will be necessary in parallel with the research process if results are to have any impact on policy. Potentially, however, risk-equalization mechanisms offer a way of improving equity and efficiency in private health care financing in South Africa without the extensive use of public funds or measures that reduce competition between insurers. They are increasingly becoming the preferred policy option in developed countries faced with balancing risk-pools of competing insurers (McCarthy et al. 1995), and many middle-income countries, including South Africa, might usefully consider their introduction.

Endnote

* If the risk-equalization pool drew its funds from general taxation, however, then at least the risk-equalizing portion of premia would effectively be mandatory, and low-risk individuals would be charged an actuarially fair premium when purchasing insurance, thus eliminating the potential for adverse selection. It is unlikely, however, that the South African Government, or that of many other developing countries, would be prepared to raise the additional tax revenue required.

References


**Acknowledgements**

Thanks to the Anglo American corporation and the Africa Groups of Sweden for funding this overview, and to Lucy Gilson, Roland Bright, Jane Doherty and Alex van den Heever for useful comments on an earlier draft. Any errors remain the responsibility of the authors.

**Biographies**

Neil Söderlund, MBBCh, MSc, DPhil, is a senior researcher at the Centre for Health Policy, University of the Witwatersrand, Johannesburg, on secondment from the Anglo American Corporation of South Africa. His current research is into regulation of private health care financing in South Africa, and the design of low-cost health insurance schemes. After completing a medical degree at the University of the Witwatersrand in 1989, he worked as a clinician for two years, then obtained a masters degree from the London School of Economics and Political Science (1992), and a doctorate from the University of Oxford (1995) on the role of case mix adjustment in hospital cost analysis in the United Kingdom National Health Service. Previous research activities include work on the economics of HIV prevention strategies in developing countries (for WHO Global Programme on AIDS), and research examining the impact of transport policies on health, both in developing countries and the United Kingdom.

Solani Khosa, BA (Hons), is a researcher at the Centre for Health Policy at the University of the Witwatersrand. He obtained a bachelor's degree in Economics from Grinnell University, Iowa, in 1992 following which he worked in primary care management for three years. Prior to his current work on the design of low-cost health insurance packages, Mr Khosa was attached to the Health Economics Unit at the University of Cape Town, where he undertook research on user fees in the South African public sector.

*Correspondence:* Dr R Neil Söderlund, Centre for Health Policy, University of the Witwatersrand, SAIMR, PO Box 1038, Johannesburg, 2000, South Africa.