One hundred and eighteen years of the German health insurance system: are there any lessons for middle- and low-income countries?

Till Bärnighausen*, Rainer Sauerborn

Department of Tropical Hygiene and Public Health, Medical School, University of Heidelberg, INF 324, 69120 Heidelberg, Germany

Abstract

A number of low and middle income countries (LMICs) are considering social health insurance (SHI) for adoption into their social and economic environment or striving to sustain and improve already existing SHI schemes. SHI was first introduced in Germany in 1883. An analysis of the German system from its inception up to today may yield lessons relevant to other countries. Such an analysis, however, is largely lacking, especially with regard to LMICs. This paper attempts to fill this gap. For each of the following lessons, it considers if and under which conditions they may be of relevance to LMICs. First, small, informal, voluntary health insurance schemes may serve as learning models for fund administration and solidarity, but in order to achieve universal coverage government action is needed to formalise these schemes and to introduce a principle of compulsion. Once compulsory health insurance exists for some people, incremental expansion of coverage to other regions and social groups may be feasible to achieve universality. Second, in order to ensure sustainability of SHI, the mandated benefit package should be adapted incrementally in accordance with changing needs, values and economic circumstances. Third, in a pluralistic SHI system equity, as well as risk pooling and spreading, can be enhanced if funds merge. The optimal number of funds, however, will depend on the stage of development of the SHI system as well as on other objectives of the system, including choice and competition. A risk equalisation scheme may prevent the adverse effects of risk selection, if competition between insurance funds is introduced into the system. Fourth, as an alternative to both state and market regulation, self-governance may serve as a source of stability and sustainability as well as a means of decentralising and democratising a health care system. Finally, costs can be successfully contained in a fee-for-service system, if cost-escalating provider behaviour is constrained by either political pressure or technical means. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Health care reform; Health care financing; Social health insurance; Universal coverage; Benefit package; Risk equalisation; Self-governance; Cost containment; Low and middle income countries; Germany

Introduction

A number of low and middle income countries (LMICs) are considering social health insurance (SHI) for adoption into their political and economic environment or striving to sustain and improve already existing SHI schemes, e.g. China (World Bank, 1997), Thailand (Tangcharoensathien, Supachutikul, & Lertiendumrong, 1999; Nitayarumphong & Pannarunothai, 1998; Kho- man, 1997), Viet Nam (Ensor, 1999), Indonesia, Philippines, Bangladesh (Tan, 1998; DSE; SHINE/ GTZ, 1998), South Korea (Yang, 1995; Shin, 1996), Kazakhstan (Ensor, 1999), Russia (Sheiman, 1995), Bosnia, Romania (The InterHealth Institute, 1998), Hungary (Donaldson & Gerard, 1993; Deppe & Oreskovic, 1996), the Czech Republic (Deppe & Oreskovic, 1996).

The main reasons for choosing SHI as the method of health care financing are that SHI can provide a stable
source of revenues, a visible flow of funds into the health sector, and a combination of risk pooling with mutual support. In addition, a SHI scheme can be established independently from the government, while it can nevertheless operate in line with government health policy. Disadvantages of SHI, on the other hand, comprise problems with insuring informal sector workers and a lack of cost control (Normand & Weber, 1994).

While the experience of Latin American countries in developing SHI systems has been analysed with respect to transitional economies (Ensor, 1999), it is striking that literature analysing the lessons learnt from the evolution of the German system and their potential relevance to LMICs is lacking. With this paper, we intend to fill this gap. We will do so by analysing the long-term evolution of the German health insurance system and examining what, if any, lessons can be derived for the design of SHI in LMICs.

As Germany has the world’s oldest SHI system, it naturally lends itself to historical analyses. Yet, although the value of diachronic analyses that trace the paths of an insurance system’s evolution over time is increasingly recognised (Field, 1999), most English language articles about the German health care system take a synchronic perspective, providing a snapshot, rather than a longitudinal, overview of the system as it exists at moment (see, for instance, Brown & Amelung, 1999; Jost, 1998; Wahner-Roedler, Knuth, & Juchems, 1997; Lassey, Lassey, & Jinks, 1997; Roy, 1993; Von der Schulenburg, 1992; Reinhardt, 1990). The exceptions to this rule tend to concentrate on one—often technical—aspect of the system (for instance, Busse & Schwartz, 1997; Henke, Murray, & Ade, 1994; Kirkmann-Liff, 1990) or confine themselves to a background chronology without working out the more general implications flowing from history (for example, Altenstetter, 1999; Iglehart, 1991). If one leaves the more short-term political and statistical vagaries behind and chooses instead to look—where meaningful—with the widest-angle lens that history permits, some sequential dynamics and cumulative effects otherwise hidden become apparent which are of potential relevance to the context of LMICs.

In the following, we will examine how during this history up to date (i) universal coverage was achieved, (ii) equal access to a comprehensive benefit package was established, (iii) equity in financing was improved, (iv) consumer choice and competition were introduced into the system, (v) sustainability was ensured and (vi) costs were contained. We will focus our analysis on the mode of development and the institutional arrangements. For each question analysed, we will consider whether the experiences from the German case may be of use to the contexts of LMICs.

Incremental achievement of universal coverage

Developments in health care systems can be categorised according to their scale and the pace at which they occur on a continuum between transformational and incremental change. Unlike many other health care systems (such as the British NHS), the formation of the German system has been characterised by incremental changes and adjustments during both its nascent and its more mature stages.

Small, voluntary, informal risk-sharing schemes as the starting point

Statutory sickness funds evolved out of the relief funds that had originated as solidarity-based support systems within the medieval guilds. Since the end of the 17th century, five types of relief funds had developed in different regions of Germany: relief funds for journeymen, relief funds for craftsmen modelled after the mutual support systems of the guilds, factory relief funds founded by socially-oriented entrepreneurs, relief funds founded by local authorities for workers or trades people and community relief funds for people who could not otherwise find insurance (Zorn, 1912; Peters, 1978).

When Bismarck reformed the German health care system in 1883, the policies he needed to implement to establish a comprehensive, social insurance system were—compared to their far reaching consequences—of relatively minor immediate impact. The Bismarckian ‘work of a century’ was, in fact, an incremental, rather than a transformational change of structures already in existence. The law of 1883 built on, first, experiences gained in the administration regional relief funds and, second, social change brought about by membership in the funds (Caron, 1882; Peters, 1978; Abel-Smith, 1992; Herder-Dorneich, 1994). Administratively, the voluntary relief funds had served as an apprenticeship stage for the development of skills in insurance administration and actuarial science at the level of the fund as well as in insurance regulation at the level of government. More specifically, the basic principles under which Bismarck’s system was to operate had already been tried and proven to work in its numerous, regional predecessors.¹ (1) The support funds were largely self-governed. (2) Both employers and employees were represented in the bodies of self-governance in most of the company-based funds. (3) Company-based funds were financed in part by employers, in part by employees. (4) Compulsory insurance had already been introduced in many municipalities (Alber, 1992).

¹ In 1876, 869,204 people were insured in 5239 officially recognised regional sickness funds (Peters, 1978).
Socially, the voluntary relief funds served at least two important functions: first, as an opportunity to build trust between the management of a scheme and its participants, and second, as learning models for solidarity, the functioning of which could be experienced first hand among people with similar social identities. Both a basic trust in risk-sharing schemes and an understanding of solidarity eventually translated into increased willingness to participate in larger schemes. The German experience during this phase of the evolution of the health insurance system suggests the following lesson:

Small, informal, voluntary health insurance schemes may serve as learning models for fund administration and solidarity, both of which will make introduction of larger, more formal, compulsory schemes an easier task.

Whether such a development can be repeated in LMICs may depend on how far the experiences with voluntary schemes in other countries mirror those in Germany. From the study of the rural cooperative health insurance schemes in rural China since the 1960s, evidence exists that the development of good management practice as well as of a trust relationship between administrators and beneficiaries are crucial elements for the long-term survival of schemes (Feng, Tang, Bloom, Segall, & Gu, 1995; Liu, Hsiao, Li, Liu, & Ren, 1995; Khan, Zhy, & Ling, 1996; World Bank, 1997). While the existence of voluntary, non-profit health insurance schemes in itself does not guarantee that such developments take place, it offers opportunities for learning, which may not be otherwise provided. The implicit danger is that, if the scheme fails, the patch may be spoilt for more promising efforts at a later time (World Bank, 1997).

In many African countries traditional risk-sharing schemes exist. For instance, anthropological studies from Burkina Faso (to be published elsewhere) have revealed a dense network of traditional mutual aid organisations based on profession or on risks (such as funeral funds). It is intuitively tempting to compare these schemes to the guild-based relief funds of the pre-Bismarckian era and to consider initiatives to integrate health risk into the cover of these funds. It has, however, been argued that the development of SHI in 19th century Germany (and Europe) is unlikely to repeat itself in today's Africa. The first argument is that traditional informal risk-sharing schemes differ in logic and function from insurance (Criel, Van Dormael, Lefevre, Menase, & Van Lebergh, 1998).

The different logics underlying informal risk-sharing and insurance (as understood in developed market economies) have been analysed from different perspectives, notably anthropological and economic (see for instance, Platteau, 1997; Besley, 1995; Coate & Ravallion, 1993; Lespès, 1990; Cashdan, 1985). As a rule, these analyses start with a dichotomy between 'real' insurance and informal risk sharing, but often end up emphasising that many similarities exist. In a study of the traditional *tontine* systems in Africa, Lespès (1990) found that as the *tontines* grow they take on many of the formal characteristics of insurance. Platteau (1997), in a review of the concepts underlying traditional risk sharing, starts by pointing out that traditional mutual aid schemes are based on balanced or generalised reciprocity while insurance is based on conditional reciprocity. Under both balanced and generalised reciprocity people expect, over time, to receive as much from a scheme as they contributed; under balanced reciprocity there are tight rules on how and when a return will be paid, under generalised reciprocity these rules are much looser. Under conditional reciprocity, on the other hand, the giver will only receive a return if she herself falls victim to the event she insured against by enrolling in the scheme, i.e., income is redistributed between the lucky and the unlucky. Platteau ends his analysis by arguing that traditional risk-sharing schemes may be able to serve an insurance function if either some standard of balanced reciprocity is upheld or redistribution does take place, but is not visible to enrolees. Similarly, Cashdan (1985) argued that generalised reciprocity could act in the same way as conventional insurance. In addition, under generalised reciprocity, those who have gained most during a certain time period are expected to give most to those who lose most during the same period—even if at some unspecified point in the future reciprocal action is expected. As such, a higher degree of solidarity is realised in the short term than in both risk- and community-rated insurance, in which contributions are independent of the financial situation of the individual. In Thailand, for instance, a number of voluntary community-based funds originally founded as non-insurance schemes to provide loans to members—have for some years successfully provided health care insurance coverage as well (Nitayarumphong & Pannarunothai, 1998).

Going back to the origins of the German funds for mutual aid as they developed within the medieval miners' associations and guilds, especially before the 16th century, generalised rather than conditional reciprocity stands out as the main principle. What is more, it was under this principle that an understanding of community self-help, social justice and solidarity was developed, which later formed the conceptual basis for the evolution of more formal insurance funds (Herder-Dorneich, 1994). Thus, the European and African histories of risk sharing, though separated by time, appear to be related in concept.

The second argument points to the fact that where voluntary health insurance schemes exist, they are mostly initiated by agencies external to African society, namely foreign NGOs, and thus lack the dynamic of an
endogenous social movement. As such, they usually lack some components, which, it could be argued, are important for scheme performance, notably participation, accountability and social control (Atim, 1999). In addition, non-movement-based schemes normally do not pose a threat to government and will thus fail to prompt government regulation or take over of those schemes, as in Germany. Without government involvement, it has been argued, broad social protection will likely remain elusive for a majority of the population (Criel & Van Dormael, 1999).

Atim (1999), in a comparison of voluntary, non-profit health insurance schemes in Ghana and Cameroon, found that the character of social movement was not a main determinant of scheme performance. In addition, he argues that, where circumstances suggest that social movement characteristics would improve performance, non-movement-based schemes could over time incorporate elements of a social movement.

Government involvement as well does not depend on a social movement character of voluntary schemes. While health insurance legislation based on existing voluntary schemes may, in fact, be intended to gain support of industrial workers, as in Germany (see above) and in many Latin American countries (Abel-Smith, 1976), the schemes in themselves do not need to be an expression of social dissatisfaction in order for that goal to be reached. The current Chinese government, for instance, views well-functioning community-based health insurance as a means of ensuring ‘social stability in rural areas’. It therefore supports still existing schemes (which were originally established by the central government through legislation in the 1960s), while promoting the re-establishment of schemes in communities where currently none exist (Gwatkin, 1999). What is more, it has been argued that non-movement-schemes may, over time, incorporate elements of a social movement and thereby enhance their success (Atim, 1999).

To sum up potential lessons learnt with respect to LMICs, the German case, among others, demonstrates that small, informal, voluntary, community-based health insurance schemes may serve as crystallisation points from which larger, more formal, compulsory schemes can be developed. Countries should investigate how to promote such schemes, especially if alternative insurances currently do not seem feasible. As an alternative to the de novo creation of health insurance schemes, consideration should be given to including health in the cover of pre-existing non-health risk-sharing schemes.

Before 1883: incremental legislative changes to achieve supraregional compulsory insurance

Six major laws led up to the reform of 1883, attempting either to regulate more closely existing structures, to establish new structures of social support, or to expand coverage. On a more abstract level, three lines of incremental developments in these laws can be distinguished.

First, the content of the laws moved from general principles to more and more concrete rules and regulations. While, for instance, the Prussian Common Land law of 1794\(^2\) established basic, general tenets of public welfare and officially sanctioned the existing chain of subsidiarity (individual–family–guilds and relief funds–communities–the state), the laws that followed up on those tenets laid out detailed rules on how sickness funds should be organised (including provisions about contributions, the benefit package, entry conditions and the management of the funds).

Second, the character of the laws gradually changed from permissive to obligatory. In 1843, the Common Law of Trade\(^3\) conceded municipal authorities the right to recognise existing voluntary funds and make insurance in these funds compulsory—a first, albeit hesitant, step away from the liberal principle of the early relief funds towards compulsion. The right was extended in 1849, when local governments were given permission to make insurance compulsory for certain employment groups. In 1854, local governments were allowed to pressure all uninsured into creating insurance funds for mutual support (Hirsch, 1875; Gladen, 1974; Peters, 1978; Herder-Dorneich, 1994).

Third, the laws moved from regional to supraregional competence. In 1854, compulsory insurance was for the first time established on a supraregional level covering the entire territory of Germany for one employment group: all miners were required to join one of the many regional miners’ insurance funds. A number of non-Prussian states had, at this time, already established a compulsory health insurance for workers. In some states, compulsion was tied to a specific fund (such as in Hannover); in other states, workers had a choice between different sickness funds (such as in Hamburg) (Herder-Dorneich, 1994).

In 1883, these three lines of incremental development were brought together in Bismarck’s workers’ insurance. It laid out detailed rules for the provision of health insurance including a minimum benefit package, the types of sickness funds, management of the funds and the extension of coverage to family members (Vogel, 1951; Gladen, 1974; Herder-Dorneich, 1994). It made health insurance coverage a legal obligation for most workers and people employed in trade and crafts. And, above all, it replaced the existing regional principle of compulsory insurance by a supraregional principle—a epoch-making, but nonetheless incremental step. In addition to laying the groundwork for universal

\(^2\) Allgemeines Preussisches Landrecht.
\(^3\) Allgemeines Handelsrecht.
coverage, the supraregional principle solved the problem of providing insurance coverage for an increasingly mobile population, for whom—as social ties were severed and regional insurance was lost—social security had become a more and more elusive concept.

In introducing the workers’ insurance Bismarck did not primarily intend to further social justice, but to fortify the state against the threat from a proletariat, which had become both well organised in the trade unions and politically powerful, as its interests were represented by the Social Democratic Party. By incorporating formerly self-administered insurance into the state Bismarck hoped to counter the increasing politicisation of the working class. Indirect evidence supports this view. First, compulsory insurance at the inception of the system was limited to workers. Although blue-collar workers were employed in the formal sector economy, they were harder to insure than many other groups, as they had low incomes and high risks of work-related illness, accidents, and disability. Better risks in terms of income and health care needs such as civil servants were not included under compulsory health insurance cover until 30 years later. The likely explanation is that Bismarck expected civil servants to be naturally loyal to the state and interested in preserving the status quo. Second, a number of provisions in Bismarck’s ‘Socialist Law’, passed in 1878, were intended to obstruct the functioning of those sickness and relief funds that had been founded by workers—the stick preceded the carrot. Third, both trade unions and the Social Democratic Party were openly opposed to Bismarck’s social insurance, as they—correctly—viewed the programme as a means to tie the workers to the existing state structures (Herder-Dorneich, 1994).

If informal risk-sharing schemes exist, the creation of legal frameworks formalising these schemes and eventually making them compulsory, can be an important step towards establishing universal social health insurance.

Introducing compulsory health insurance has been part of health care reform in many countries of South East Asia, Central and Eastern Europe, and has come under consideration in a number of African countries (Zwi & Mills, 1995; Bennet & Ngalane-Banda, 1994). Countries where successful informal voluntary risk-sharing schemes for health exist should consider making them more formal, as without such government action risk-sharing social protection will remain limited and contingent on local circumstances (Criel & Van Dormael, 1999). A legal framework may define a minimum benefit package and regulate contributions, provider payment mechanisms and scheme administration.

Whether the three transitions that took place at this phase in the development of the German health insurance system—from informal to more formal, from voluntary to compulsory, and from small to larger schemes—can and should be emulated in LMIC may be highly contingent on the context: the power structure, trust and legitimacy between the different actors. Will informal schemes be opposed to government regulation? Will the participants and the current non-participants in the scheme resist being compelled to join? Will solidarity (or generalised reciprocity) suffer as the average distance (physical and social) between members increase? Will trust in the scheme and its management remain intact as community participation and social control become increasingly difficult?

As in the German case, it has been suggested elsewhere for the context of LMICs that incremental changes may be easier to implement than transformational changes in developing efficient and equitable health insurance (Carrin, De Graeve, & Deville, 1999a). In many of the formerly socialist Eastern European countries a development towards SHI similar to that in Germany had been taking place, starting in the 18th century until restructuring of the health care sector by the communist governments (see, for instance, Observatory, 1999a, b). In Hungary, voluntary self-help funds for industrial workers were legally legitimised in 1840; a voluntary General Fund for sick and disabled workers was established in 1870; a national compulsory insurance for industrial workers, similar to Bismarck’s workers’ insurance, was established; and, finally, at the turn of the century a national insurance fund for agricultural workers was set up (Observatory, 1999b).

It is safe to say that a number of Eastern European countries picked up threads of development that had been abandoned during the communist era when they (re-)established SHI. The mode of reforms, however, did not follow the earlier incremental pattern. A crisis in health care financing accompanied by a fall in life expectancy in an environment characterised by rapid political and economic changes had opened a ‘window of opportunity’, in which fast and drastic, rather than slow and steady, action was felt to be required.

Of the three transitions described above the move to compulsion may be the most difficult to achieve, even if it is only for one segment of the population and in one region of the country. In many contexts, the establishment of any form of compulsory insurance may be deemed not to be politically feasible. In such situations, voluntary schemes may remain a second-best option. The recommendations regarding rural community-based (or cooperative) health insurance schemes in the Peoples’ Republic of China in 1998 to promote voluntary community-based health insurance in the countryside exemplify this. Although policy makers were aware that problems with adverse selection could arise and that willingness-to-join could be generally low because of negative past experiences with community-based
insurance, the Chinese State Council shied away from compulsion. Another compulsory payment to a state organisation, it was feared, would further increase anti-government sentiments among farmers (Gwatkin, 1999, Hsiao, 1984). Similarly, in Nigeria a social health insurance scheme was established on a voluntary basis, as a compulsory scheme was judged not to be politically feasible (Bennet & Ngalande-Banda, 1994).

In addition, compulsion may prove not to be enforceable for administrative or economic reasons. In China, under national law all urban workers in state-owned companies and their family members fall under the compulsory cover of the Labour Insurance Scheme, a company-based Bismarckian health insurance. As more and more state-owned enterprises near bankruptcy, the company-based funds become increasingly insolvent. In theory, workers still receive full reimbursement for a comprehensive benefit package, their family members are covered with 50% of their eligible health care expenditures. Increasingly, however, deficit-running enterprises have been unable to pay contributions to the funds. As a result, in 1998, about one-third of workers in state enterprises had no health insurance coverage at all, many others received only marginal health care benefits from their insurance (Grogan, 1995, Yip & Hsiao, 1997; Hu, Ong, Lin, & Li, 1999; Center for statistical information of the Chinese ministry of health, 1999).

In many Eastern European countries, the implementation of a compulsory SHI after 1989 has faced serious problems (such as in Bulgaria and Hungary) or even failed (such as in Kazakhstan). In these cases, both employees and employers were unable to pay social health contributions in the time of economic crisis. In addition, the state lacked the capacity to collect contributions from those companies and workers that were able to pay (Observatory, 1999a–c).

In sum, the German case suggests that the central government plays a crucial role in establishing a SHI, as it is the institution best placed to create a legal framework for SHI. Legislation in Germany incrementally formalised and expanded insurance as well as made it compulsory. While many LMICs already have successfully introduced compulsory health insurance for some segments of the population, other cases from LMICs suggest that introducing compulsion—even if only in the formal sector—may be difficult to enact or to enforce, if the government is politically or administratively weak or the economy is flagging.

After 1883: incremental expansion of coverage to achieve universal coverage

The incremental approach taken to develop the system after the introduction of the workers’ insurance manifested itself mainly in the expansion of population coverage, the size of the risk pools and the benefits covered. By some estimates, Bismarck’s law doubled sickness insurance coverage among workers from around 5% to 10% of the total population. Thereafter, coverage in the statutory health insurance grew steadily from 11% in 1885 to 37% in 1910. By 1930 about 50% of the total population were covered and by 1950 about 70%. Since 1975 more than 90% of the population are enrolled in the statutory health insurance; the remaining
10% are nearly completely covered under private or other health insurance. Today, less than 0.5% of all people living in Germany do not have health insurance (Tennstedt, 1977; Peters, 1978; Neubauer, 1988; BMG, 1999a, b) Fig. 1.

The extension of population coverage in an insurance system can be classified based on three principles: the regional, the personal, and the place of work principle (Zeöllner, 1963). According to the regional principle, an insurance scheme is first established in some regions of a country (usually the most industrialised ones), and then gradually extended to cover other (usually less developed) regions. An extension of coverage according to the personal principle can be either oriented at horizontal criteria such as occupation or vertical criteria such as income. Similarly, extension according to the place of work principle can be along horizontal lines (e.g. economic sector) or vertical lines (e.g. size of company).

Before the introduction of the workers’ health insurance, access to insurance had depended either on the region of residence or on the place of work. Correspondingly, coverage was extended according to the regional or the place of work principle. By contrast, once the principle of supraregional compulsory insurance was introduced for different occupational groups in 1883, coverage was extended according to the personal principle. Along horizontal lines coverage was expanded incrementally to cover more and more occupational groups and—in three major shifts of expansion—to cover the unemployed, all primary dependents and retirees (see Table 1) (Wasserrab, 1889; Lang, 1925; Peters, 1978; Alber, 1989; Manow, 1999). Vertically, coverage was expanded by increasing the income ceiling above which health insurance is no longer compulsory, as was done, for example, in 1918 when the monthly income limit was doubled from RM 2500 to RM 5000. For people who fell under the law of compulsory insurance, but who did not have access to a sickness fund through their work and could not insure in a town-based fund, every municipality had to provide insurance through a municipal sickness fund.

This second phase in the development of the German health insurance system suggests the following lesson:

If compulsory insurance already exists for some people, extending it incrementally to other regions and social groups will—if a number of conditions are met—be a feasible way to achieve universal coverage.

Expanding compulsory insurance coverage is a task many LMICs face today. In Viet Nam, a compulsory SHI scheme was introduced in 1993, which covers civil servants and workers in larger enterprises, but—in spite of attempts to expand cover to family members, farmers, and urban informal sector workers—more than 90% of the population remain uncovered (Ensor, 1997). South America has enjoyed a long tradition of social insurance. But population coverage is highly variable. It ranges from less than 10% in the Dominican Republic to more than 80% in Costa Rica, although the proportion of the population covered in most countries is rising. The population group least likely to be covered is the growing number of urban informal sector workers (Donaldson & Gerard, 1993). In Africa, a number of countries have established social health insurance schemes, such as Cameroon, Ghana and Kenya. Again, coverage is mostly confined to the formal workforce (Bennet & Ngalande-Banda, 1994).

In Chinese cities, up to the market reforms in the 1980s, SHI schemes covered more than 90% of the urban population, although only 50% of health care costs incurred by spouses and children of the insured were covered. Today, the number of urban residents covered has dwindled to less than 50%, as more and more state enterprises declare insolvency and people increasingly find work in private enterprises or in the informal urban sector (Center for statistical information of the Chinese ministry of health, 1999; World Bank, 1997; Hsiao, 1995). The major stumbling block to universal coverage in these cities is the growing number of informal sector employees and migrants from the countryside (both legal and illegal). The central Chinese government currently attempts to promote a stepwise expansion of coverage in all cities from state to non-state enterprises to the self-employed and—and eventually and perhaps with the help of subsidies—to the urban poor (Hu, 1999; Zhu, Zhou, Zhang, Ma, & Gao, 1999; Bloom, 1998). In accordance with a national policy recommendation, some city governments consider offering voluntary enrolment in the city-wide funds to anybody not yet under mandatory cover, if they are able to contribute as much to the fund as do workers earning 60% of the city’s average annual salary (see, for instance, Social health insurance administration office of Shenzhen city, 1999; Labour office of Yichang city, 1998).

The German case suggests that compulsory coverage can be extended incrementally to achieve universality. This ‘lesson’, however, cannot be drawn without some general qualifications and without considering, whether the specific methods used in Germany can and should be transferred to other countries and times. From an ethical point of view, it has to be kept in mind that the German government in adopting an incremental approach towards universality was motivated by arguments of power rather than social justice (Rimplinger, 1971; Observatory, 2000b). If one accepts a utilitarian ethic that preservation of power may be a legitimate goal of social policy as long as the ultimate outcome serves social justice, it has to be kept in mind that an incremental approach to establishing SHI may, in fact, lead to more inequity.
<table>
<thead>
<tr>
<th>Year</th>
<th>Creation of components of social security</th>
<th>Population coverage of social health insurance</th>
<th>Scale and scope of the mandated benefit package of social health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1854</td>
<td>Miners</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1883</td>
<td>Statutory health insurance</td>
<td>Blue collar workers (in saltworks, processing plants, factories, metallurgical plants, railway companies, shipping companies, shipyards, building companies, trade companies, power plants)</td>
<td>Minimum benefit package: Sickpay (63% of all benefits)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Craftsmen</td>
<td>Restricted in- and outpatient care</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Persons employed by lawyers, notaries, bailiffs, industrial cooperatives, insurance funds</td>
<td>free pharmaceuticals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medical aid devices</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>deathpay</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>maternity support</td>
</tr>
<tr>
<td>1884</td>
<td>Statutory accident insurance</td>
<td>Transport workers</td>
<td></td>
</tr>
<tr>
<td>1885</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1889</td>
<td>Statutory pension insurance</td>
<td>Commercial office workers</td>
<td></td>
</tr>
<tr>
<td>1892</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1901</td>
<td></td>
<td></td>
<td>Increase in the duration of sickpay from 13 to 26 weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sickpay extended to cases of sexually-transmitted diseases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Increase in allowances to family members in case of a hospitalisation of a relative</td>
</tr>
<tr>
<td>1911</td>
<td>Agricultural and forestry workers</td>
<td>Domestic servants</td>
<td>Increase in maternity support</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Itinerant workers</td>
<td>Increase in sickpay of high-wage workers</td>
</tr>
<tr>
<td>1914</td>
<td></td>
<td>Civil servants</td>
<td>Earlier start of sickpay</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Family support for spouses and children</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Midwife services</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obstetric services</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pregnancy allowance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nursing mother’s allowance</td>
</tr>
<tr>
<td>1919</td>
<td>Persons employed in public cooperatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Persons employed in private cooperatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Persons who are only partially capable</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>of gainful employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wives and daughters without own income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927</td>
<td>Statutory unemployment insurance</td>
<td>Seamen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Persons employed in the educational and</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>social welfare sectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td>All primary dependents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1935</td>
<td>Midwives</td>
<td></td>
<td>Increase in the duration of maternity support</td>
</tr>
<tr>
<td>1938</td>
<td>Self-employed workers in nursing and</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>child care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>Retirees</td>
<td></td>
<td>Full cover of the treatment of all notifiable diseases</td>
</tr>
<tr>
<td>1953</td>
<td>Refugees and expellees</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The seriously disabled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td>The physically disabled</td>
<td></td>
<td>Increase in sickpay for workers</td>
</tr>
<tr>
<td>1970</td>
<td></td>
<td></td>
<td>Prevention</td>
</tr>
<tr>
<td>1972</td>
<td>Self-employed agricultural workers</td>
<td></td>
<td>Pediatric screening</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salary of a temporary replacement workers for sick farmers</td>
</tr>
</tbody>
</table>
For one, revenues in a SHI system traditionally flow from formal wages and salaries. As a result, the population groups likely to be covered last are the most vulnerable segments of the society: those without incomes (the unemployed, retirees, and family dependents) or those with incomes that are variable and hard to assess (urban informal workers and farmers). This has an important implication. Current members of social health insurance schemes may be opposed to including other groups in the insurance cover. On the one hand, since those who are as yet without insurance are likely to be low income and high risk people, those who are currently insured would likely pay part of the price of including these groups in the form of higher insurance contributions. On the other hand, the incremental approach to include people along employment or regional lines implies that the social proximity and thus solidarity between current members is higher than between members and non-members. Unemployment and informal sector employment are increasing in many LMICs (ILO, 1999). Thus countrywide solidarity across employment lines may be increasingly hard to establish in LMICs. But continuing commitment to solidarity among all people living in a country—as it still exists in Germany (Hinrichs, 1995)—is a basic condition for establishing universal SHI.

Moreover, a stepwise passage to universality may result in decreased access to health care for the uninsured in the interim periods of partial coverage (which may be quite long, if political will is lacking or socio-economic conditions are unfavourable). Resources may be drained away from the uninsured to provide health care for the insured (Normand & Weber, 1994; Abel-Smith, 1992). A case in point is a compulsory insurance scheme for Indonesian civil servants. Equity concerns have been raised, as the scheme is subsidised from general government revenues. In addition, beneficiaries were found to use public hospitals at a rate that was five times the national average (Prescott, 1991). For these reasons, a fast-track approach to universal health insurance, may be preferable in some circumstances, although it may require a much larger effort.

In the formerly socialist countries of Eastern Europe establishing social insurance step by step, starting with partial coverage, would have meant reducing equity in comparison to the universal access guaranteed before under the command-and-control Soviet model of health care (Observatory, 1999a,b, 2000a,c; Twigg, 1999).

Whether or not the specific means by which groups without formal wages or salaries were integrated under the SHI cover in Germany can be replicated in LMICs depends on a series of factors. First, the unemployed and the retired are covered through the wider system of social insurance. The statutory pension insurance and the statutory unemployment insurance provide the two groups with regular, taxable incomes from which mandatory health insurance contributions are automatically deducted. Obviously, this is only possible in countries where comprehensive social insurance exists or is established at the same time as the health insurance.

Second, children and spouses are included under the cover of the breadwinner. Since contributions are independent of family size, a re-distribution from singles to families and from families with fewer to families with more children results. How far a re-distribution is feasible within a SHI system depends on the dominant hierarchy of values within a society.

Third, self-employed farmers were not covered until 90 years after the introduction of Bismarck’s workers’
insurance (Holler, 1977). This reflects the difficulty in assessing and taxing farmers’ incomes. For many countries, especially those where farmers constitute a large proportion of the total population, such as China, Vietnam and Thailand, such a delay may not be practicable. The difference in context between 19th century Germany and today’s LMICs is even more prominent in the case of the informal sector. In Germany, the proportion of informal sector workers declined during the years of the system’s existence. In many LMIC, on the other hand, the informal sector is expected to continue to grow in the next years (ILO, 1992, 1996, 1999).

Self-employed farmers and informal sector workers are hard to cover in a SHI system since their incomes fluctuate and are hard to assess objectively. Still, many systems have been devised for assessing the incomes of the self-employed and charging contributions (Normand & Weber, 1994). These systems, however, may be administratively cumbersome and costly or may be extremely crude, such as flat-rate contributions. Consequently, countries that have large or growing informal sectors should consider alternatives to SHI, unless some means of including the informal sector in the social insurance cover has been shown to work well.

Yet, even in a context of high formal sector employment payroll deductions may prove problematic as the (sole) revenue base for insurance. In Germany, as in other developed countries, wages and salaries constitute a decreasing proportion of total GDP as the contribution of business profits and capital investment to GDP is growing (Statistisches Bundesamt, 1997; OECD, 1998). As a result, payroll deduction rates to SHI in Germany increased, even in times when sickness fund expenditures as a proportion of GDP remained fairly constant (Bärnighausen, 2000; Bärnighausen et al., 1999; Braun, Kühn, & Reiners, 1998).

A Bismarckian health insurance system can, in fact, be implemented in a MIC following an incremental process very similar to that in Germany. In 1965, the first voluntary health insurance fund was organised in Korea. By 1977, when compulsory insurance was first introduced, there were 11 voluntary funds, which covered about 0.2% of the population. Compulsory insurance was expanded vertically step by step to companies with 500, 300, 100 and finally 16 employees over the following six years. Similarly, coverage was expanded horizontally to government officials and private school teachers (1979) and families of military servicemen and employees of private school foundations (1980). Universal compulsory coverage was achieved 26 years after the establishment of the first voluntary fund through schemes covering the rural and the urban self-employed (in 1988 and 1989) (Moon, 1998; Peabody, Lee, & Bickel, 1995; Anderson, 1989). The achievement of universality in South Korea shows that a Bismarckian health insurance can be established in a country with a social, political and cultural history which is very different from that in Germany. It also shows that the pace of incremental development can be much accelerated. It has to be kept in mind, however, that this happened against a backdrop of fast and sustained economic growth and a shrinking informal sector—two conditions that, while neither necessary nor sufficient, are conducive to establishing a SHI.

In sum, Germany succeeded in achieving universal coverage following an incremental pattern of expanding compulsory insurance. This success has been contingent on a number of social, economic and institutional circumstances. Yet, as the above cases suggests, a similar approach holds promise for countries that have already established partial coverage, such as many countries in transitional Asia and South America.

**Incremental extension of the benefit package to attain comprehensive coverage**

The approach to extend the mandated benefit package was incremental as well. It occurred along three dimensions.

First, the largest changes in the scope of the benefit package were brought about by the introduction of new types of statutory social insurance system, such as accident, pension and unemployment insurance. Each new type extended the benefit package to an area of social need, which the social net had not covered before. The principle of compulsion applied to the same groups of the population as before. Today, all types of statutory insurance cover health-related benefits. The latest addition to the statutory insurance system was long-term nursing care insurance. Introduced in 1995, it pays for ambulatory as well as in-patient nursing care (Bloch, Hillebrandt, & Wolf, 1997).

Second, the benefit packages of already existing types of the statutory insurance were gradually extended to additional disease groups and services. Examples encompass occupational diseases (which were added to the coverage under the statutory accident insurance in 1925 and 1929), the treatment of sexually transmitted diseases and a broad spectrum of preventive measures (which were added to the benefit package of the statutory health insurance in 1952 and 1955, respectively).

Third, already existing benefits were more or less gradually increased in amount or duration. For instance, amount and duration of sick pay were increased in 1957; the time limit on coverage of in-patient care was eliminated in 1974 (Lang, 1925; Peters, 1978; Winterstein, 1980a, b).

The expansion of both coverage and benefits has lead to a gradual transformation of the statutory health insurance system. On the one hand, as more and more groups of society fell under the laws of compulsory
health insurance, the original intention of the system—to prevent social unrest among workers—was replaced by the political will to include nearly all strata of society in the comprehensive safety net of statutory insurance. On the other hand, the gradual extension of the benefit package shifted the focus of the system from protection against loss of income to the coverage of medical care. This development is mirrored in the change over time of the ratio of cash benefits (sick pay) to benefits in kind (medical care). Livelihood protection rather than access to health care was the initial objective of health insurance. While at its inception, the system provided 1.7 times more cash benefits than benefits in kind, by 1955 this ratio had reversed to 1:4, by 1977 to 1:10 and by 1984 to 1:16 (Neubauer, 1988; Alber, 1992; Manow, 1999).

The very generous benefit package has recently been curtailed in response to cost containment requirements (Manow, 1999). The reductions in scale and scope of benefits have again been incremental. Since the economic recession following the first oil crisis in the early 1970s, coverage of some services has been reduced through the gradual increase in co-payments or through the introduction of eligibility conditions, notably for pharmaceuticals (in 1977, 1981, 1983, 1988, 1992, 1996 and 1997), glasses (in 1977, 1981, 1996, 1997), dentures (1981, 1988, 1996, 1997), medical cures (1981, 1983, 1996) and hospital stays (1981, 1983, 1988, 1997) (Bandelow, 1999; SGB V, 1998; OECD, 1997; Herder-Dorneich, 1994; Reimers, 1993; BMG, 1993, 1996). Some benefits were entirely excluded from the benefit package such as certain dental services, medical aid devices, death pay for those insured after 1989 and pharmaceuticals for the so-called petty diseases, common colds and diseases acquired during tourist travel as well as for pharmaceuticals which are either cheap or of unproven medical benefit (Braun et al., 1998).

The benefit package in the German statutory health insurance system was incrementally expanded from a limited to a very comprehensive set of benefits. In the course of time, the original intention—to secure workers’ income in times of illness—was first taken over by the new political imperatives to provide free medical care to all sick people and, later, to do so while constraining spending. These transformations appear dramatic only from the telescopic perspective of a one-hundred-year-plus history. Because they were brought about by incremental modifications rather than abrupt shifts, the process did not undermine the basic organizational principles of the original system.

In that, the German case suggests the following lesson:

Incrementally adapting the mandated benefit package in accordance with changing needs, values and economic circumstances will contribute to the sustainability of a social health insurance system.

When mandatory SHI schemes were introduced in Eastern European countries after 1989, the financing and decision-making structures changed more or less fundamentally in comparison to the former state-owned and centrally planned health care system of the Soviet era. Some of the basic objectives of the system, however, did not change (McKee, Figueras, & Chenet, 1998; Observatory, 1999a–c, 2000a–c). Universality of population coverage and comprehensiveness of services covered were intended to retain or regain a principle of free access to a full range of health services for all, as had existed under the communist model. While establishing a mandatory SHI system with universality was clearly desirable to preserve equity and feasible given the starting point, many of the formerly socialist countries were soon forced to rethink their stance towards comprehensiveness. Much like Germany, Hungary incrementally redefined the mandated benefit package through exclusions (and some new inclusions), copayments and decreases in the scale of services covered (Observatory, 1999b). For the Russian Federation, some analysts recommended to accept deteriorating quality and co-payments and hope for future economic growth and gains in efficiency rather than to reduce the scope of entitlements in place before 1989 (Chernichovsky & Potapchik, 1997; World Bank, 1996). Yet, in 1998 a law was passed that made the scope of benefits contingent on the anticipated level of revenues mobilised for use in the health care system in any given year (Twigg, 1999).

The German case suggests that a middle road between the approach recommended and the one enacted in Russia will be advantageous for both sustainability and public acceptance of a SHI system: neither was the benefit package left completely unchanged over time, nor was it changed fast and unpredictably in accordance with revenues (contribution rates and income ceilings were slowly adapted as well). Slow adaptations of scope and scale of benefits guaranteed a balance between the financial capacity, on the one hand, and need and social expectations, on the other. In that, the current benefit package is only one fleeting endpoint soon to be replaced in a dynamic of more or less incremental change. This is exemplified by the more recent, but ongoing discussion to ‘slim down’ the still very comprehensive mandated benefit package and leave services above the new statutory minimum to be covered by voluntary insurance (see, for instance, Flintrop, 2000)

In countries that have established compulsory health insurance systems, policy-makers have some latitude in defining and incrementally redefining the benefit package based on need and financial capacity. Countries, on the other hand, that consider introducing a principle of compulsion for the first time or intend to broaden
coverage in voluntary schemes need to consider the demand for the services covered. If bottom–up risk-sharing schemes exist, such as in pre-Bismarckian Germany, the cover they offer for a certain level of contributions can be assumed to represent demand (at the level of the current premium or contribution rate) and be used as a starting point, from which to expand depth and breadth of a health insurance scheme. In the absence of such bottom–up arrangements, research needs to be undertaken that elicits how attractive different benefit packages are to potential enrollees and how much they would be willing to pay for them. We will report the results of such studies from China and Burkina Faso in due course. To sum, the German case with regard to LMICs, countries embarking on a social health insurance system should start with a package limited enough to be financially viable, yet with services that are both relevant to need as well as attractive to enrollees. LMICs, which have already established a social health insurance system should aim to extend the benefit package to comprehensiveness, but should eschew maintaining too luxurious a package, if the contribution rates rise to too high levels.

**Evolution of the system of sickness funds**

The German social health insurance system is split into a number of separate sickness funds. Within one sickness fund, vertical equity is achieved through contributions according to ability to pay (a fixed proportion of members’ pay checks is deducted each month), while horizontal equity is realised by guaranteeing access to care on the basis of need. In effect, income is redistributed from the rich to the poor, the healthy to the sick, the young to the old, men to women, singles to families and the employed to the unemployed as well as to the retired (Henke, 1988). This interpersonal redistribution constitutes the basic expression of the principle of solidarity in the German health insurance system (Hinrichs, 1995). But—while horizontal equity in terms of access to services was built in the system since its genesis through a legally guaranteed benefit package for all people enrolled in a statutory fund—vertical equity in financing was (and, in part, still is) compromised as the regional funds and the different types of company—and occupation-based funds constituted separate risk pools.

The sickness funds as non-profit, statutory organisations are obliged by law to translate any profit or losses that have occurred in their running accounts into decreases or increases of the contribution rate, so that the accounts will be balanced to zero (Iglehart, 1991; SGB V, 1998). The expenditures of a sickness fund depend on the morbidity of its member pool; contributions on the other hand are the only source of a fund’s revenue. Since contributions are proportional to income across all incomes up to a ceiling, a fund can charge a lower contribution rate if its members are healthier and earn more. The disparity in risk—and income structures manifests itself in contribution rate differentials. For instance, in 1993, the highest and the lowest contribution rates varied by nine percentage points between 7.8% and 16.8% (Minn & Pfeiffer, 1995; Jacobs, 1998).

**Fund mergers**

At the outset, the sickness funds were rather small in size with membership on average ranging in the hundreds. Pool sizes grew as, firstly, the total number of insured increased and, secondly, sickness funds merged (Abel-Smith, 1992). The increase in the total number of insured resulted from population growth and the expansion of coverage. Except for a momentary large increase in 1989 due to the reunification, the number of insured increased at an almost constant rate. The decrease in the number of sickness funds, on the other hand, displays less of a pattern. The number of funds fell precipitously following the First World War and as a result of two major reform laws. The first wave of mergers was caused by the Reich Insurance Ordinance of 1911, which disbanded the community funds and stipulated a minimum size for the membership of the sickness funds (Manes, Mentzel, & Schulz, 1912; Manow, 1999). Within one year the number of funds was more than halved, dropping from 21,238 in 1913 to 10,004 in 1914 (Statistisches Reichsamt, 1914, 1915). Through a slow process of mergers the number of the funds was halved a second time between 1919 (9145 funds) and 1938 (4524 funds). In the wake of the Second World War, the rate of consolidation gained again in speed; between 1940 and 1948 the number of funds was reduced from 4456 to 1760. The most recent (and still ongoing) surge in sickness fund mergers was brought about by the infusion of competition into the payer system in 1996 through the Health Care Structure Reform Act of 1992, which forced funds to realise economies of scale. In addition, the increase of some high risks during the 1990s (e.g. AIDS and costly technological procedures such as MRI) are likely to have—albeit more indirectly—further stimulated fund consolidation, as their coverage needed larger pools of insured. As a consequence of these forces, the number of funds decreased again by more than 50% from 1015 in 1994 to 483 in 1998.

The concomitant expansion of the number of insured and the consolidation of sickness funds, increased the risk pool size incrementally, but exponentially, from an average of 229 insured per fund in 1885 to about 300 in

---

4 Reichsversicherungsordnung.

5 Gesundheitsstrukturgesetz.
1890, 3000 in 1930, more than 10,000 in 1950, and
16,000 in 1970. Since then, the rate of growth has further
accelerated. In 1991, an average of about 40,000 persons
were insured in each statutory sickness fund; in 1997
that number amounted to more than 91,000 (Kaiserlich
Statistisches Reichsamt, 1885–1918; Statistisches Reichsamt,
1919–1942; Statistisches Bundesamt 1952–1999)

Countries considering establishing or reforming a SHI
system need to consider the number of sickness funds.
Regarding the exact number of funds and fund size in a
pluralistic system one experience from Germany is that
the optimal number and size of health insurance funds
may be contingent on the stage in the evolution of a
system. At an early stage, the advantages of small funds
(social proximity among enrollees and enrollees’ direct
ability to monitor the fund management) may outweigh
the disadvantages. As the system becomes more formal,
universal and compulsory, however, it may be increas-
ingly desirable to merge existing funds to improve risk
pooling and spreading as well as efficiency and equity.
On the other hand, if choice, competition and decen-
tralisation remain or emerge as important objectives of
the health care financing system, the number of funds
should not be reduced too far. Thus, the German case
suggests:

In a pluralistic insurance system equity, efficiency,
risk pooling and spreading can be enhanced, if funds
merge. The optimal number of funds will depend on
the stage in the development of a SHI system as well
as the objectives of the system, including choice and
competition.

In Estonia, the Soviet health care system was replaced
by a SHI in 1992. In an effort to decentralise the system,
the administration of the health insurance for the
population of roughly 1.5 million was devolved to 15
county funds, six city funds and one fund for seamen.
After concerns were raised that this had created
insurance pools too small to be scale efficient and to
effectively spread the risk of catastrophic care, the
number of funds was reduced from 22 to 17, and it has
been recommended to further reduce that number
(Observatory, 2000a).

In rural China, the cooperative medical schemes,
where they still exist, currently keep separate accounts
for workers and farmers. It has been recommended to
combine these two accounts to increase risk pooling and
spreading (Carrin et al., 1999b). In urban China, health
insurance is traditionally organised through separate
funds based on work unit, within two compulsory
insurance schemes, the Labour Insurance Scheme
(which covers state-enterprise employees) and the
Government Insurance Scheme (which covers civil
servants). Currently, attempts to revive the urban SHI
schemes are under way, since population coverage has
decreased drastically since the market reforms in the
1980s (see above). One of the key elements of reform is
to combine the many company-based funds into one
city-wide fund, in order to increase risk pooling,
administrative efficiency and purchasing power with
respect to providers. In two demonstration projects in
the cities of Jiujiang and Zhenjiang city-wide risk
pooling in one single fund has been successfully
established (Hu et al., 1999; Ma, 2000). In Shanghai,
the separate health insurance funds have been united into one city-wide fund as well and cover has been gradually expanded to employees of private enterprises of different sizes (Sun, 2000). In the city of Shenzhen, a city with one of the highest proportions of private companies and joint ventures, one health insurance fund took over the function of the separate company-based funds and the funds for government officials in 1992. Since then, coverage of state-owned as well as non-state-owned companies has been gradually expanded to cover 87% of formal sector workers (Fu, 1998; Ou, 1999).

In sum, as the above cases exemplify in LMICs in which achieving or preserving universality is a main objective of health sector reform, the advantages of a smaller number of funds may outweigh the disadvantages and the merging of risk pools may be advisable.

**Risk equalisation**

In addition to the increase in the size of the risk pools through fund mergers, in Germany the separation between the funds in terms of revenues and risks was torn down by two ‘risk equalisation schemes’. In 1977 a ‘risk equalisation scheme’ for retirees was set up. This scheme equalised costs incurred for the treatment of retirees—in 1989 about 40% of total expenditures in the system of statutory sickness funds were financed across fund borders (Wasem, 1986; Leber & Wasem, 1989). Since real expenditures were equalised *ex post*, the equalisation scheme was in reality a form of reinsurance. In 1993, another equalisation scheme, the so-called risk structure equalisation (RSE), came into effect. It is based on an *ex ante* assessment and equalisation of the ‘risk structure’ of a fund in comparison to its ‘financial strength’ (Wasem, 1993). ‘Financial strength’ is a function of the total contribution income of one specific fund in relation to the total contribution income of all funds. ‘Risk structure’, on the other hand, is defined as the expenditures an *average* sickness fund would incur, if the composition of its members was the same as that of the *specific* sickness fund under consideration. As a result, fund revenue is redistributed from financially strong to financially weak funds and from funds insuring low to funds insuring high risks. The criteria according to which risk is equalised comprise age, gender, the number of family members covered by the policy of the family head, and the number of disabled (OECD, 1997; Giehler, 1995; Möller, 1995; Giehler & König, 1993; Glanz & Rogalski, 1997; Die Ersatzkasse, 1998).  

The RSE challenged the prevailing interpretation that reforms in the German health care system could not be transformational because of structural impediments (Altenstetter, 1998). The resulting changes as well—while neither rapid nor drastic—were, indeed, significant. In 1999, the RSE amounted to 23.5 billion German Marks. The local sickness funds received most of the equalisation payments (19.4 billion in 1999), while the substitute funds for employees paid the lion’s share (13.6 billion) (BKK; 2000; Die Ersatzkasse, 2000; Der Hausarzt, 2000). Vertical equity in financing between the sickness funds improved as a result. While in 1993 71% of all people insured in the statutory health insurance paid a contribution rate that fell within plus/minus 1% of the average contribution rate, that number has risen to 93% in 1999 (BMG, 2000; Bandelow, 1999).

The German case demonstrates that:

A risk equalisation scheme based on a few and easily obtainable, verifiable and universally applicable criteria may increase vertical equity in financing in a pluralistic health insurance system.

For LMICs in which (i) multiple funds exist, (ii) each fund insures a significant part of the population and (iii) overall risk is fragmented along fund lines, a risk equalisation scheme holds much promise. Such countries include South Korea, many Latin American countries and some formerly socialist Eastern European countries.

A case in point is Romania, where strong interregional differences in revenues and expenditures give rise to considerable inequity in a SHI system based on several regional funds. Currently, the regional funds are required to pass on 7% of their revenues to a National Insurance Fund to be redistributed among regions as needed to meet national goals of equity. There are two problems with this approach. Firstly, 7% of the individual regions’ revenues is not sufficient to achieve interregional equity. Secondly, there is no specification of the criteria by which monies are to be redistributed among the regions, leaving the redistribution subject to political discretion (InterHealth Institute, 1998). In such a situation, a risk equalisation scheme comparable to the one in Germany is likely to be more effective in increasing equity in health care financing.

In South Africa, the predominant model of insurance is employment-based non-profit insurance funds, while a small market share of health insurance is held by for-profit private insurance companies. In 1989, the employment-based funds were deregulated. They were allowed to risk-rate premiums and did no longer have to accept all applicants. Predictably, risk selection and dumping resulted. Population coverage decreased, at least in part because of the ensuing fragmentation of risk pools. In this situation, it has been suggested that the introduction

---

6Risikostrukturausgleich.

7The current Minister of Health, Andrea Fischer, plans to bring before parliament a proposal to extend risk equalisation to a number of clusters of chronic diseases (Jessen, 2000).
of a risk equalisation scheme would ameliorate the adverse effects of risk selection and thus reverse the trend of contracting coverage (Söderlund & Khosa, 1997).

Whether such a scheme can be successfully implemented will, however, depend on a number of conditions, including high public acceptance of financial redistribution between social groups (as redistribution will increase due to the scheme), the definition of a minimum benefit package, information available on valid equalisation criteria, and the capacity for administering and monitoring the scheme.

In sum, risk equalisation could improve vertical equity of financing in LMICs with universal compulsory coverage and mandatory assignment of enrollees to funds. In countries which have not introduced compulsory and allow risk selection, risk equalisation could help increase population coverage.

**Competition**

On the theoretical continuum from one fund for all and one fund for each, one would expect scale efficiency in a one-fund system, if there are economies of scale in administrative costs and other overheads over the entire range of the population. A one-fund system, however, lacks the discipline of competition and may thus be less efficient. Adverse selection is not possible in a one-fund system. Innovation, on the other hand, is likely to be more rapid in a multi-fund system. In addition, if fund membership is based on geographical region or place of work, there is greater potential for social control of fund management and direct solidarity among enrollees. Another trade-off between the advantages of one fund vs. many funds exists with regard to the purchasing function of insurance. One fund may be able to wield monopsony power to control provider prices, but only multiple funds competing with each other on the basis of price have incentives to control providers.

German policy makers did not face a ‘Stunde Null’ decision between a monopolist and a pluralistic insurance system, as insurance had historically developed along pluralistic lines. What is more, while the German system clearly had some of the disadvantages of a multi-fund system, it had so far failed to realise its advantages. Compounding the equity and efficiency problems inherent in a system made up of separate risk pools was a lack of market competition through consumer choice. While free choice of providers and, ensuing, competition between providers were instituted gradually in Germany since the end of the 19th century and culminated in a ruling by the Supreme Court in 1960 that indirectly extended patient choice to any physician who chose to set up a practice (Behaghel, 1994; Huercamp & Spree, 1982), the freedom to choose a sickness fund was very limited. Traditionally, people were assigned to a sickness fund as compulsory members based on occupation or region of residence. The funds had a guaranteed membership and could, in general, neither gain nor lose members. Therefore, they had no incentive to lower contribution rates (or increase quality) by reducing inefficiencies in their operation. Unleashing market forces through free choice of funds without a risk equalisation scheme, however, would have forced the funds to compete for ‘good risks’ while trying to be less attractive for ‘bad risks’ (‘cream-skimming’) (Reiners, 1993). This would have undermined the principle of solidarity. Moreover, since sickness funds differ widely in the make-up of their membership, funds with healthier and wealthier members would have enjoyed a head start in a competitive market, while funds with sick and poor members would have been punished for the mandatory assignment of insured. Two safeguards largely prevented these adverse effects of competition. First, the RSE—even if imperfect—took the profit out of risk selection and created (more) equal competitive chances for all funds. Second, a supporting law made it illegal for the sickness funds to dump enrollees or to refuse to enrol anyone.

Theoretically, the introduction of competition in 1996 should be expected to have increased administrative efficiency through a profit incentive and to a change in the memberships of the funds, if people perceived significant differences between funds.

Administrative costs did not fall but rose slightly between 1996 and 2000: in former West Germany from 5.27% to 5.57% and in former East Germany from 6.12% to 6.18%. Possibly, sickness funds had already exhausted their potential for efficiency gains before 1996, in order to be prepared, when competition started. Administrative costs had fallen steadily between 1993 and 1996 from 5.28% to 5.08% in the West and from 6.12% to 6.18% in the East. 

---

9The resulting monopoly power of the funds was somewhat limited, however, as some people could (collectively) choose to insure with a company-based or guild-based fund and others could (individually) choose an occupation-based substitute fund (Wasem, 1996). Still, in 1989 only about 50% of individuals had some freedom of choice between different sickness funds (Henke, 1989).

10Critique of the RSE centres around the question whether the criteria for which risk is currently equalised really suffice to make risk selection unattractive a competitive strategy and whether the RSE should be have organised regionally, so that differences in supply structures will not be cemented by cross-subsidies between regions (Moeck, 1995). Moreover, it has been pointed out that more important than the actuarial facts may be whether fund managers perceive different insured to be equally attractive customers (Brown & Amelung, 1999)
6.87% to 6.12% in the East (BMG, 1999a, b). Furthermore, competition may increase administrative overhead, if administrative intensity is perceived as a decisive competitive advantage, e.g., in marketing and customer service. All in all, in Germany administrative overhead of health insurance (within the statutory system) is astonishingly low in international comparison. But, comparison of overheads may not be a good measure of efficiency, as high overhead costs do not automatically mean less overall efficiency. Measures financed out of overheads may be expended to control total system costs (for instance through technology assessment) and thus may incur a benefit to society (Folland, Goodman, & Stano, 1997). Choice, on the other hand, was indeed practiced. Between 1996, when enrollees could for the first time choose their sickness fund, and the year 2000 the membership of the largest type of funds, the local sickness funds, was reduced by 9.9%, the membership of the substitute funds for employees was reduced by 1.1%. By the same token, membership in the company-based funds, the substitute funds for workers and the guild-based funds increased by 41.3%, 24.9%, and 7.7%, respectively (Der Hausarzt, 2000). The potential adverse effects of choice have been effectively prevented by the RSE. The contribution rate differential decreased, although the differences of fund memberships composition increased with regard to the criteria equalized by the RSE, as evidenced by the steady rise of RSE transfer payments by approximately 1 billion DM/year since 1993 (BKK 2000). The transfer payments due to the RSE should be expected to fall over time, if the risk pools become more similar. By contrast, the transfer payments have steadily risen since 1993 by approximately 1 billion German Marks per year (BKK, 2000). This is evidence that risk selection has not been eliminated by the RSE. Yet, as the closure of the gap between contribution rates suggests, it has prevented the adverse effects of risk selection. Analysing consumer choice in the German health care system, again a stepwise approach becomes apparent. At the inception of the system neither free choice of providers nor free choice of funds existed. Both were successively added. This is especially important, since choice and competition may be detrimental to equity in health care, if safeguards are missing that ward off the adverse effects of choice. In the German system, the following safeguards have been established: On the provider side, physicians cannot compete on the basis of price as the fees for services provided are negotiated and are legally binding. This prevents price discrimination, which physicians could use to exploit any monopoly power they have, to charge higher prices to smaller sickness funds and—if coverage is only partial—to the uninsured (Reinhardt, 1999). Competition is thus focused on quality (as perceived by the patient), leading to higher consumer satisfaction (but not necessarily to higher clinical effectiveness).

On the payer side, risk selection is theoretically avoided by a number of measures. First, risks are equalised to make risk selection a meaningless competitive strategy to pursue (see above). Second, two measures prevent funds from exploiting any residual risk differences remaining after equalisation. Funds should not be able to ‘dump’ enrollees, as any fund must insure anybody who wishes to be insured. Funds should only be able to marginally ‘cream skim’, as they are required to cover a comprehensive ‘minimum’ benefit package for all enrollees. Since the benefit packages in all funds are very nearly equal, it could be argued that the German system is not different from either a contribution-based system with one national sickness fund or a system based on an earmarked tax with a needs-based allocation formula, except for imperfections in the RSE and where the law of large numbers does not apply. There are a number of important differences, however. First, incentives to be administratively efficient do not exist in a one-fund system. Second, except for the benefit package decision-making is more decentralised in a multi-fund system. Third, a multi-fund system leaves open future developments towards competition based on price and on scope or scale of benefits. Recent reform proposals to reduce the mandated benefit package exemplify this. The benefits excluded from compulsory cover could—in different combinations—either be offered in different combinations by private insurance or be tied to the remaining mandated benefits in the statutory system (see, for instance, Flintrop, 2000). Under the latter scenario, by offering benefit packages more attractive to good risks funds again would have the capacity to cream skim, which they would likely use, if the RSE left open incentives to do so. Funds that cream skimmed less effectively would attract a pool of worse risks which, in turn, would drive up contribution rates and drive away enrollees to other funds. As a final result, bad risks would finance part of a benefit package, which to them had less value than to good risks. The reforms would thus lower solidarity in health care financing.

The German case suggests the following lesson:

In a Bismarckian health insurance system, choice and competition do not have to be built-in from the start, but can be added on at later stages of development when administrative and legal capacity allow to install safeguards preventing adverse effects. These safeguards, on the other hand, will likely prevent some of the benefits from competition to be realized.

Many of the formerly socialist Eastern European countries, such Kazakhstan, Hungary and the Russian
Federation have emphasised consumer choice in the formulation of health care reform (Observatory, 1999b, c; Sheiman, 1994). The German lesson may be one of caution. While choice and competition may serve to stimulate high quality service provision, patient and consumer satisfaction, the negative side effects may be hard to control if administrative skills and actuarial information are lacking. In addition, the health insurance system needs to be embedded in an effective legal system to deter potential offenders, such as physicians charging under-the-table payments, as is common practice, for instance, in the Russian Federation and China (Ensor & Savelyeva, 1998; Delcheva, Balabanova, & McKee, 1997). Moreover, as the German case demonstrates administrative efficiency may not automatically result from the introduction of a profit incentive. Finally, alternatives must exist for choice to be meaningful. Such alternatives may be absent in hospital-based health care systems, where many services are exclusively available at one hospital in a region.

How and when choice and competition are introduced into a health care system will likely depend on the institutional history of the system. The health care financing reform in the Russian Federation offers some interesting contrasts to the German reforms. Policy makers chose to introduce a purchaser–provider split by establishing obligatory regional insurance funds to contract with providers. Competition in this system is confined to the provider side, while funds hold the monopoly for health insurance in each region. Forces of tradition likely influenced this choice, as monopsonistic purchasing constitutes a less radical transformation of the former integrated system than pluralistic purchasing (Twigg, 1999; Sheiman, 1994). Still, in comparison to the former Soviet health care system, the introduction of SHI has increased total administrative overhead considerably (see, for instance, Curtis, Petukhova, & Taket, 1995).

In sum, the German case suggests that policy makers in LMICs should not introduce choice and competition until the health care and insurance system has reached a more mature stage. Legislation can prevent the negative effects as well as some of the positive effects of competition.

Beyond the public/private dichotomy: self-governance as steering structure

While the mode of decision-making in the German health care financing system can be described as largely incremental, the structure in which decisions are reached is characterised by a self-governance of corporatist organisation, operating within a legal and political framework set by the federal government. Self-governance has been imbedded in the German statutory health insurance system since its inception in 1883 (Tennstedt, 1977); it was modelled after the self-administration structures under which most support funds and guild-based sickness funds had already operated (Schmidt, 1977).

The relative independence of the self-governance from both state and market is the principal reason why—in contrast to the health care financing systems in most Western European countries, the German system has not only survived but, at least along some lines, been strengthened over time. The health care financing system has provided institutional continuity through a phase in German history characterised by extraordinary tumult: two World Wars, four regimes (Imperial, Weimar, Nazi) and the post-war division into two separate states (Moran, 1999; Alber & Bernadi-Schenklun, 1992) and two currency collapses (1923 and 1948). Furthermore, after the German reunification in 1989, the structures of self-governance speedily and completely replaced the command-and-control system in the former German Democratic Republic.

The imposition of West German structures on East Germany did not happen without opposition. Resistance among Eastern German stakeholders was low, since they lacked personnel with political experience and organisation as well legitimacy to demand the retention of former East German structures. But resistance existed in the West. The main opposition party, the Social Democrats, proposed a uniform social insurance in the East, as opposed to the separation of social insurance into (at that time) different types. The association of the local sickness funds, on the other hand, argued that the SHI should—in an interim period—be organised by the local funds alone, as the Western pluralist set-up would be too complicated to be implemented right away (Wasem, 1998; Lüschen, Niemann, & Apelt, 1997; Henke & Leber, 1993). These proposals were not without clout as the West German health insurance was, at the time, being strongly criticised as inequitable because of its pluralist structure (Scharf, 1999)—a criticism that, among others, lead to the transformational reforms in 1992. Still, in the final outcome the West German self-governance was imposed unchanged on East Germany. It was successfully implemented within only eight months’ time. This transition was made easier by the political support from both parties in the coalition government and the Chancellor’s office as well as by financial support in form of subsidies from the West (Wasem, 1998). Above all, however, it is testimony to the capacity of the self-governance—both of the sickness funds as well as of the physicians’ organisations—to forcefully react to the challenge of an alternative institutional arrangement and quickly expand into new territory (Döhler & Manow-Borgwardt, 1992).
The organisations of self-governance

The self-governance is enacted by non-profit payer and provider organisations. These organisations perform a triple role between interest representation of their membership, self-administration and a steering function within the health care financing system.

On the payer side, the statutory sickness funds are self-governed corporations under public law. They perform their legal functions organisationally and financially independently, but remain under the supervision of the government. The funds are regionally based, but are represented by independent umbrella organisations at the federal level. On the provider side, ambulatory physicians are compulsory members in two regional corporatist associations: the Medical Councils (Ärztekammern) and the Panel Physicians' Associations (Kassenärztliche Vereinigungen). The Medical Councils are responsible for residency programmes and professional licensure; the Panel Physicians' Associations contract with the sickness funds' associations for ambulatory care for their members. In contrast to the ambulatory physicians' organisations, the hospital associations are not incorporated under public law, although their status has increased in recent years (Bandelow, 1998).

Intraorganisationally, the payer and provider associations are self-governed by a democratically elected governing board and an executive committee. Interorganisationally, the system is managed through mesolevel negotiations at the Länder level between the sickness funds' associations and the Panel physicians' associations or the individual hospitals. As most decisions belonging to the realm of "low" politics (Walt, 1994) are formulated and implemented at the Länder level or beneath, decision-making is effectively decentralised.

At the federal level, interorganisationally self-governance manifests itself in two main ways. Firstly, within the confines of the in—or the outpatient sector, the umbrella organisations of the sickness funds negotiate with their provider counterparts to work out terms of contract that serve as a framework for all Land level negotiations. In the ambulatory sector, these collective framework agreements are binding. They define the rights and duties of Panel Physicians the mandated benefit package as well as a relative value scale for the services within the package. The relative values are expressed in points. They are converted into monetary values through multiplication with conversion factors. The conversion factor differs according to region. It is intended to mirror regional differences in input prices. In the hospital sector, the umbrella organisations of the sickness funds and the German Hospital Association jointly issue non-binding recommendations concerning standards of care, ground rules for the diffusion of hospital technologies, nurse education, training, staffing and quality assessment. Secondly, across sector lines, a mixed private and public body, the Concerted Action in Health Care, brings together representatives of all major institutional components in the health care system to issue non-binding annual recommendations (SGB V, 1998; Beske & Hallauer, 1999).

Evolution of the self-governance

The self-governance of the German health care system can neither be described as a market system (because important market elements such as the steering function of the price on the provider side are lacking) nor as a state bureaucracy (because health policy is neither decreed centrally nor implemented by public organisations, but materialises in negotiations between independent associations). A number of alternative socio-political concepts have been proposed to analyse self-governance. They comprise interest representation by pressure groups, neocorporatism and policy networks (see, for example, Gäfgen, 1988; Döhler, 1990; Perschke-Hartmann, 1994; Lamping, 1994). These concepts have explanatory power for different phases of policy making in the development of the German health insurance system since the Second World War. First, the fact that during the 1950s and 1960s a number of attempts at health sector reform failed, has been attributed to the power of various pressure groups, such as the Association of Panel Physicians, to block any reform that ran counter to their members' interest (Webber, 1988, 1989; Rosewitz & Webber, 1990; Mayntz, 1990). Second, during the 1970s and 1980s the government integrated a more limited number of interest groups into the policy making process. The most prominent among these neocorporatist arrangements is the Concerted Action in Health Care (Bandelow, 1998). In this phase, a consensual mode of decision-making prevailed and a number of incremental reform steps were formulated and implemented (see, for instance, Gäfgen, 1988). Last, the Health Care Structure Reform Act of 1992 marked a turning point in health policy making in Germany, as corporatist arrangements partially ceased to be effective. A new mode of policy making emerged which is centred around working groups and informal policy networks (see, for example, Döhler & Manow, 1997).

Throughout these three phases, the system of self-governance has been able to functionally adapt to changing external constellations (Döhler & Manow, 1995a). Over the course of time, these adaptations have

---

11 Konzertierte Aktion im Gesundheitswesen.
12 The analysis of 'reform blockages' has been criticised as neglecting the complexity of interactions between the different factors as well as their historical variability (Döhler & Manow, 1995a).
suggested that: In sum, the German experience with self-governance system from the government and political turbulences. It also partly insulated the health insurance administering the system and decentralised decision-making. It also partly relieved the government of the burden of German health care financing system. This has effectively relieved the government of the burden of administrating the system and decentralised decision-making. It also partly insulated the health insurance system from the government and political turbulence. In sum, the German experience with self-governance suggests that:

A self-governance may serve both as a source of stability and sustainability as well as a means of decentralising and democratising the health care system. As such it is an interesting alternative to both state and market regulation of a health care financing system.

Whether lessons from the German case regarding self-governance can be applied to current health care reforms in Eastern European countries can be doubted on two counts. First, the starting points are different. In Germany, the self-governance evolved in a countervailing build-up of corporatist associations based on pre-existing structures of interest representation. In the civil societies in formerly socialist Eastern European countries, third-party payers and professional associations of physicians had been largely absent before 1989, except for the powerful, compulsory trade unions of health workers. Physician associations started to develop in the 1990s but—as of yet—do not have much influence as they have neither gained statutory standing nor formal representation in policy-making bodies (Observatory, 1999a–c, 2000a, c, d, e, f). In addition, while in Germany health insurance and self-governance created for the first time a system of comprehensive and universal coverage, in the formerly socialist Eastern European countries one comprehensive system replaced another (see, for instance, Curtis et al., 1995). Of greater interest to Eastern European countries may be the establishment of a self-governance in the former German Democratic Republic, as described above. It is however, unlikely that the process would have been similar to the one described above, had West Germany not provided financial resources, technical expertise and personnel.

Second, the direction of development is different. While in Germany, the self-governance—as well as health policy making as a whole—have increasingly centralised; in the former Eastern bloc countries one of the main aims of health care reform has been to decentralise decision-making. Inflexible decision-making under the centralised Soviet model has been blamed, in part, for the deteriorating quality of services, the low efficiency of service provision and low patient and physician satisfaction, especially since the 1980s (for instance, Observatory, 1999a, 2000d). Instituting a self-governance is viewed as one way to decentralise the health care system (for instance Observatory, 1999b, 2000a). In this context, centralisation through a growth in the role of the central government, national bodies of interest representation and national-level payer and provider associations in the formulation of health policy would be controversial.

Nevertheless, in Eastern Europe centralisation tendencies appear against the backdrop of overall decentralisation. In Hungary, the governance of health insurance was transferred from the state to several self-governed health insurers—an important decentralisation,—but at the same time the financing function in health care was centralised into one common health insurance fund. In a similar vein, centralisation and decentralisation were simultaneously realised in Croatia when the responsibility to finance health care was delegated from local government authorities to a single
national health insurance fund (Saltman & Figueras, 1997).

Governance policies need the support of a country’s civil society and cannot be applied uniformly in different contexts (Zwi & Mills, 1995). How relevant some of the experiences from the German case are to the contexts of other countries, must thus be judged in light of the structures currently prevailing in the health care systems and the civil societies of these countries as well as the objectives driving the change towards a self-governance.

But a self-governance may be a means to yet another end. Talking about the statutory accident insurance Bismarck is reported to have remarked that ‘the insurance itself is for me a secondary consideration. My chief consideration is to use this opportunity to attain corporate associations … In this way we will establish the basis for a future representative body that will become an important participant in the legislative process’ (Vogel, 1951; Pflanze, 1990). In that respect, a de novo creation of a self-governance is desirable, if increased interest representation and participation in policy making is a policy goal considered important.

In sum, compared to a central state-controlled system a self-governance of regional provider and payer associations decentralizes decision-making and increases lower level accountability and transparency. It may also be a source of democracy, if representatives in the associations are elected.

Cost containment

From the mid 1970s throughout the 1980s, West Germany was more successful than most other OECD countries in curbing the growth of health care expenditure: Total expenditures for health care as a proportion of GDP rose only slightly from 8.0% in 1975 to 8.2% in 1990. Sickness funds’ expenditures show a parallel development, rising only 0.1% during the same period (from 5.2% to 5.3%) (OECD, 1998; BMG, 1999a; Statistisches Bundesamt, 1999).

The fact that total health care expenditures rose from 5.7% of GDP in 1970 to 8% in 1975 (sickness funds’ expenditures: 3.3% and 5.2%, respectively) was mainly a result of, firstly, cost-shifting out of the statutory pension insurance system into the health insurance system (Wiesenthal, 1981); secondly, the financing of back-log demand for capital investment in the hospital sector that had accrued due to chronic shortage in financing during the preceding periods (Kühn, 1980) and, thirdly, the improvements in the benefit package described above.

In order to save the statutory pension funds from financial difficulties caused by demographic and economic changes in the early 1970s, the contributions which the pension funds made to the sickness funds for the provision of health care to the retired were reduced, effectively shifting costs out of the pension insurance into the health insurance system (Wiesenthal, 1981; Braun et al., 1998).
The rise from 8.2% to 10.4% of GDP between 1990 and 1997 (sickness funds: 6.4–7.5%), on the other hand, was principally caused by the reunification and by the introduction of the statutory long-term nursing insurance in 1995 (BMG, 1999a; Reinhardt, 1999; OECD, 1997). Thus, overall cost containment in the German health care system can be considered reasonably successful. But costs have not been controlled equally well in all sectors. While expenditures in the ambulatory sector have remained stable in comparison to contributory income, expenditures in the hospital sector have been rising steeply since 1975 (see Fig. 3) (BMG, 1997, 1999a, b; Statistisches Bundesamt, 1999).

The success of cost control in the ambulatory sector is economically counterintuitive. Ambulatory physicians work in private, for profit practices, whereas hospitals are largely non-profit. It is even more counterintuitive, if one considers that ambulatory physicians successfully controlled spending even in times, when they were reimbursed under a fee-for-service system without an expenditure cap, and that they did so against the backdrop of an increasing physician density (Rosewitz & Webber, 1990). To large parts, the ‘miracle of German cost containment policy’ (Von der Schulenburg, 1994) should be amended to read ‘in the ambulatory sector’.

Costs in the ambulatory sector were controlled well during four periods (before 1932, from 1933 to 1955, from 1975 to 1997, and after 1997).

From 1883 until 1931, the sickness funds acted as an oligopoly or (near) monopsony that was able to dictate the terms of contract to the provider polipoly (the individual physicians), because the sickness funds were well coordinated in spite of their large number. Individual physicians working in their own practices contracted to provide care for sickness fund members a contracting system that resembles the arrangement in independent practice association HMOs (see, for instance, Caughey & Sabin, 1995). The two most common methods of remuneration under these contracts were a fee-for-service system with a case-based expenditure ceiling (about 60% of practices) and capitation (about one-third of practices) (Seitz, König, & Jelastopulu, 1998; Herder-Dorneich, 1994; Kirkmann-Liff, 1990).

From 1933 until 1955, the Great Depression, the second World War and the post-war restoration period constrained ambulatory physicians from using their newly gained organisational power (with the legal establishment of a ‘countervailing power’ (Galbraith, 1952), the Panel Physicians’ Association, in 1932) and their monopoly status in the provision of ambulatory medical services (granted in 1933) to change the capitated payment system, —which they had agreed to in return for the corporatist privileges,— into a reimbursement system more favourable to physicians (Behagel, 1994).

From 1975 to 1997, a fee-for-service system limited by either expenditures caps or targets replaced a cost escalating, ‘pure’ fee-for-service system. Under an expenditure cap, a global budget is negotiated by the self-governance. If the physicians’ bills collectively exceed the negotiated budget, then the fee per claim is reduced in order to stay within the budget.

After 1997, the competence of the self-governance to negotiate provider reimbursement was partially restored in 1997, when government mandated, global budgets were replaced by specialty-specific, ‘flexible’ practice budgets, which are targeted at growth rates negotiated by the self-governance at Länder level (SGB V, 1998; Galas, 1997).

The relative success of cost-containment in the outpatient as opposed to the inpatient sector has been attributed to the stronger structures of interorganisational self-governance in the former (Alber, 1992). The development of expenditures in the outpatient sector has been contingent on the reimbursement mechanisms, which, in turn, were negotiated or mandated within various power constellation of payers, providers and the federal government. Analysing these constellations and political and environmental forces acting on them it becomes apparent that costs were controlled well, whenever the cost-escalating nature of the fee-for-service system was constrained. Such constraints originated either within the self-governance, under the influence of outside forces, or from direct action by the federal government.

14 In the German hospital sector up to 1993 ‘not-for-profit’ meant that hospitals could neither make a profit nor incur a loss, as their running costs were reimbursed in full, while investment finance was provided directly from Länder tax revenue. Since 1993, hospitals can both make a profit and incur losses, if they under- or overspend a negotiated target budget (Bärnighausen, 2000).

15 Although the statutory health insurance system is not steered by market forces, but by self-governance structures, the economic termini poli-/mono- poly/psony are used here, because they serve as a useful description of different levels of aggregation on the buyer and supplier-side of services and because the power of payers and providers in the collective negotiations depends, in part, on their (virtual) power as buyers and suppliers in the health care market.

16 On the other hand, it can be argued that the Panel Physicians’ Associations have hampered the success of cost containment in the system as a whole. Since 1933, the associations have successfully fought off any threat to their collective monopoly for ambulatory care in the form of clinics ran by the sickness funds (Zweifel, 1998). This has upheld the strict division between ambulatory and hospital sector, which is the cause of much inefficiency as it leads to duplications of services. Recent reform proposal by the new coalition government of the Green and the Social Democratic Party seriously challenge the associations’ monopoly for outpatient care.
government. Within the self-governance, cost control was achieved, when the sickness funds were in a more powerful position than the physicians, because the physicians were not organised (such as between 1883 and 1931) or weakened by interprofessional allocation conflicts (between 1992 and 1996) (Brenner, Heuer, & Pfeiffer, 1994; Busse & Schwartz, 1996). Outside forces were instrumental, when the physicians practiced self-constraint either in response to environmental factors (such as between 1933 and 1955) or under political pressure (such as from 1975 to 1977, from 1986 to 1988, and again after 1997). Political pressure has been exerted by two means: a threat by the federal government to strip outpatient physicians off some of their professional autonomy and by ‘moral suasion’ via expenditure targets issued by the Concerted Action in Health Care. In the first case, the physicians’ associations have repeatedly chosen to give up income positions rather than professional status (Döhler, 1987; Webber, 1988). In the latter cases, the targets were non-binding and usually not observed. Still, cost containment often sustained successfully under the impression of the targets, as the self-governance fell back upon creative means to control costs (Rosewitz & Webber, 1990; Powell, 1993). Examples include case-base reimbursement in 1982 and a regional experiment with outpatient gate-keeping in the state of Bavaria (Herder-Dorneich, 1994; Kirkmann-Liff, 1990).

The federal government acted directly to contain costs, whenever it took away the power to negotiate from the self-governance. Both in 1977 and from 1989 to 1991 caps on the collective budget for ambulatory physician became a legal obligation, although nominally total expenditure was still to be negotiated by the self-governance. From 1992 to 1996, the federal government interfered even more directly in the steering function of the self-governance by mandating that the growth in physician expenditures be capped at the increase in contribution income (Schwartz & Busse, 1996).

By contrast, when none of these three constellations were present, costs increased dramatically. During those periods when the provider organisations were relatively centralised and homogenous in comparison to the payer side and the central government did not interfere in the negotiations, the Panel Physicians’ Association used their unchecked bargaining power to press for reimbursement mechanisms, which would result in higher physician incomes (from 1955 to 1966 and from 1967 to 1974). In 1955, the Association succeeded in adding yearly retrospective adjustments to the existing capitation fees according to type and amount of services actually rendered. After 1966 all physicians were reimbursed according to a ‘pure’ fee-for-service system. As could be expected, under both systems costs escalated (Rosewitz & Webber, 1990; Herder-Dorneich, 1994).

With respect to cost containment the German case suggest that:

Costs can be successfully contained in a fee-for-service system, if institutional constellations constrain provider behaviour. The means by which cost control is achieved can be either technical (negotiated or mandated expenditure caps) or political (threat and ‘moral suasion’).

Fee-for-service remuneration has been established or considered in many of the newly established SHI systems in Eastern Europe and some of the older SHI systems in Asia (such as in the Philippines and in South Korea). There are a number of reasons for using a fee-for-service system as the means to reimburse providers: first, a desire to improve satisfaction among patients through freedom of choice; second, to improve quality of services in a system where money follows the patient; third, to increase utilisation of priority services; fourth, to stimulate the provision of good data on health care utilisation, and, last, to increase satisfaction among physicians through more flexible working conditions and potentially higher incomes (Sheiman, 1995; Observatory, 1999b, c, 2000c).

There is ample evidence from LMICs (e.g. South Africa, South Korea, China) that fee-for-service remuneration leads to an expansion in overall service volume and rising health care expenditures because of supplier-induced demand (Kutzin, 1997; Moon, 1998; Tan, 1998). Consequently, not to introduce unregulated fee-for-service reimbursement is one of the few unequivocal lessons of health care financing (Kutzin, 1997; Barnum, Kutzin, & Saxenian, 1995). There is no clear-cut answer, however, to the question which type of regulation works best, if a country should decide to introduce fee-for-service reimbursement.

Even if a fee-for-service system is regulated, costs may be hard to contain. In Hungary, the national Health Insurance Fund incurred a growing deficit as service volume expanded under a mixed payment system with FFS, capitation and DRG remuneration in the outpatient, primary care, and inpatient sector, respectively. Ceiling caps introduced on budgets of different expenditure items improved the situation somewhat, but could not effectively control costs as caps could not be imposed on some benefits and the capitated sector had incentives to shift patients into the two non-capitated sectors (Observatory, 1999b).

The German case emphasises once more that policy makers considering introducing FFS systems should from the start think about how they will curb the increases in service provision likely to result (unless, of course, such an increase is desired). Expenditure caps on global budgets as a technical means to control costs may be applied in other countries. More contingent but also
more interesting is the German experience that political pressure on providers through ‘moral suasion’ or threat may suffice to induce cost-containing self-constraint.

Conclusion

A number of positive as well as negative experiences during the evolution of the German social health insurance system may be of relevance to the contexts of LMICs. Spelt out as lessons learnt, they represent a highly aggregated gist of what is more often than not the result of many individual, often diverging effects. As a consequence, many qualifications apply and trade-offs are commonly discovered at lower levels of analyses. The transferability of the lessons to other contexts needs to be carefully considered in the light of these qualifications and trade-offs.

The lessons, we believe, may contribute to the debate on health sector reform in LMICs on two levels. On a more conceptional level, they may suggest policy action or provide evidence to argue for or against certain policies: the creation of small voluntary insurance schemes, the incremental enlargement and formalisation of existing voluntary schemes, the gradual expansion of compulsory health insurance, the incremental adaptation of the benefit package in compulsory schemes, the merging of funds, risk equalisation and competition in a pluralistic insurance system, self-governance and mechanisms to control costs in a fee-for-service system. If policy makers find some of the experiences drawn out above potentially of practical interest, a closer investigation of the underlying technical details may be warranted. The literature listed below may provide a starting point for such study. While for some situations the technicalities of point change in the German case may indeed provide a blue-print for (or against) action, in general they are likely to be highly contingent, at times they are certainly idiosyncratic, such as the means by which population groups without formal wages were included under compulsory cover, the services included in the mandated benefit package or the criteria used for risk equalisation. Yet, if the concepts seem of interest, the technical particulars may prove amenable to adaptation to other realities and constraints.

Acknowledgements

This work was supported by the collaborative research grant ‘SFB 544’ of the German Research Foundation (DFG). Till Bärnighausen received a scholarship of the DAAD (Deutscher Akademischer Auslandsdienst, Bund-Länder-Hochschulsonderprogramm III).

The views expressed are those of the authors. We want to thank Michael Borowitz and David Dror for their helpful comments on an earlier draft of the paper. We also want to thank Christopher Murray, Sanguan Nitayarumphong, Willy de Geyndt, Jürgen Hohmann, Hermann Schulte-Sasse, Veronica Vargas, Armadeep Thind, Zhang Xinping, Chen Maoshen and Luo Wujin for their questions and ideas.

References


Further reading.


