Hospitals’ response to increasing autonomy and market forces: evidence from 4 countries

Barbara McPake and Kara Hanson

Corresponding author, Barbara McPake, Professor and Director, Institute for International Health and Development, Queen Margaret University College, Corstorphine Campus, Clerwood Terrace, Edinburgh EH12 8TS, Tel +44 131 317 2491, Fax +44 131 317 3494, e-mail: bmcpake@qmuc.ac.uk

Kara Hanson, Senior Lecturer in Health Economics, Heath Policy Unit, Health Economics and Financing Programme, London School of Hygiene and Tropical Medicine.

Keywords: Hospital autonomy, hospital reform, Zambia, Colombia, UK, Indonesia.

\[\text{Keywords: Hospital autonomy, hospital reform, Zambia, Colombia, UK, Indonesia.}\]
Abstract

The policy of hospital autonomy is contingent on the idea that market forces can discipline hospitals and so improve their performance. That idea is contested from two perspectives. First, under most forms of autonomy implemented, hospitals are not the residual claimant to surpluses generated by their activity and may pursue other objectives with unclear implications for their behaviour. Second, hospital contracts are inevitably incomplete, perverse incentives may arise and hospitals may seek to increase surplus by excluding or under treating particular patient groups.

Evidence to assess the validity of the market forces model, or alternatively of either conflicting model is scarce. The operation of health sector reform is multi-faceted, time lagged and beset by inter-relationships between multiple variables. Research funding is usually small scale and haphazard. This paper seeks to draw together evidence across four countries (the UK, Zambia, Indonesia and Colombia) where evidence of hospital performance in the wake of reform is relatively robust, in an attempt to identify patterns of response to reform.

Across these four countries, the evidence suggests that hospital reform may be associated with productivity improvement, although in all cases it is difficult to link robustly the reforms with the measured improvements. In three of the four countries: the UK, Zambia and Indonesia, it is easier to link reform with the development or further development of two-tier or multi-tier service provision strategies, raising questions regarding the equity
impact of the policy. In Colombia, there is no similar evidence and certain features of the Colombian programme seem to protect against negative equity effects. These are an absence of out-of-pocket payments, and measures which mandate cross-subsidies among different groups. Such measures have been difficult to implement in Colombia, a middle income country, and may not be feasible in lower income countries.
Introduction

Problems in the functioning of developing country hospitals have been both legion and legendary (especially at tertiary level) throughout recent decades. Compared to the importance of their services in relation to the profile of disease in the population, they have consumed a disproportionate share of resources in many countries (Hanson, Atuyambe et al. 2002; Ssengooba, Atuyambe et al. 2002). Yet, given the higher unit costs of services, this has not implied that services have been adequately funded. Failure to invest in hospital infrastructure and consequent reliance on outdated technologies are well documented in a range of countries (for example, (Purvis 1997; Uganda Health Planning Department 1998))

Problems of irregular supply of drugs and other consumables; and insufficient and poorly paid and under-motivated human resources have also been commonplace, as in lower levels of the health system. Over-crowding, staff attrition, waiting and waiting lists variously characterise hospitals in different settings.

Hospitals in higher income countries have undergone considerable change following from technological revolutions. Less invasive techniques have enabled declines in length of stay and a reorganisation of the hospital workforce, reducing the role of long term carers and increasing the role of technical specialists. Trends such as the increasing professionalisation of nursing staff can partly be explained from this perspective. Little of these developments can be discerned in hospitals in low income countries. Whether this reflects a difference in efficient technological mix arising from relative price differences, or the failure to seize important opportunities is less clear.
These problems have been compounded by hospitals’ failure to play their designated role in the health system. Gatekeeping is an underdeveloped function in most health systems of low income countries and even national hospitals often do not have systems in place to enable them to turn away self-referred patients likely to be presenting with trivial conditions, for example in Zimbabwe (Hongoro and Musonza 1995). This implies that such hospitals have been dealing with a larger share of the need pyramid than envisaged.

Hospitals are important symbols of the health system as a whole. In some countries, national hospitals were constructed in the first years of independence as monuments to the new government’s commitment to health care. Prominent politicians may ensure hospitals are constructed in their home areas of core political support. In this context, hospitals’ physical presence and size become more important objectives than their contribution to people’s health, impeding technical solutions. Within the hospital, the medical profession dominates, posing constraints to the imposition of efficiency from a managerial perspective. While doctors’ training in management is usually peremptory at best, introducing specialist managers to run hospitals has met with little success. Without the support of the clinical team, managers are powerless to impose strategies that, for example, interfere with clinical freedom by trying to economise on resource use.

Against this background, hospital reform arrived late in the programme of health sector reform which began from the mid-1980s. There is little mention of hospitals in the World Bank’s influential 1987 document ‘Agenda for Reform’ (World Bank 1987) or in the follow
up World Development Report ‘Investing in Health’ in 1993 (World Bank 1993). Since Alma Ata, the focus of health sector development efforts has been on primary health care (PHC), and the role of hospitals has been relatively neglected.

Since the mid-1990s, the issue of hospital reform has been much more prominent in the health system development debate. In particular, a large number of countries have set about changing the relationship between hospitals and the health system as a whole. The long list of problems, and their varied incidence in different low and middle income countries, implies that the objectives of reform, and presumably then appropriate strategies, are likely to be different in different places.

Nevertheless, in low income countries, the main discourse around hospital reform has concerned hospital autonomy, indicating that hospitals are singled out for devolved authority. In higher income countries, similar changes have been enacted through system level reforms – purchaser-provider splits, for example, which suggest that hospital autonomy arises in a systems framework where the roles of other actors have been equally transformed. Underpinning at least some of the enthusiasm for autonomy has been a model based on the idea that market forces can discipline hospitals in ways that will resolve these many different issues surrounding their operation. The next section examines the arguments related to this model.

The market forces model of hospital reform
In a World Bank document, (Harding and Preker 2003) present the options for hospital arrangements as ranging from ‘privatization’ to ‘corporatization’, ‘autonomization’, and ‘budgetary organizations’ (labelled P, C, A and B on figure 1 below). They place these within a spectrum of incentive environments (Manning, 1998) and argue: ‘The incentives for efficient production are higher moving outward, where service delivery is often better than at the core’ (p42). Harding and Preker explain that incentive environments are the product of governance, incentives embedded in funding and payment mechanisms and the market environment. It may be deduced from this that the rationale for moving from B to A (for example) is to enable responsiveness (through mechanisms of responsive governance) to market forces strengthened either through heightened competitive pressures and/or through funding and payment mechanisms that reward activity. We term this the market forces model of hospital reform.

The relationship between market competition and hospital performance implied in this model has been the subject of considerable theoretical and empirical investigation in the literature dealing with the US and the UK. Both the level of competition in a particular market, and the precise form that competition takes, have been argued to mediate the relationship between market competition, on the one hand, and price and quality outcomes on the other (Chalkley and Malcolmson 1996 (OXREP paper); Propper 1996). For instance, in markets characterised by imperfect competition, arising for example from product differentiation owing to the importance of reputation in assessing quality, increased competition could lead to increased price (Satterthwaite 1979). Lower price
associated with higher competition may drive down quality; or alternatively, as in arguments about the “Medical Arms Race”, competition may encourage providers to increase quality and therefore prices (C&M 1996 OXREP).

Payment mechanisms and contracting have also been examined in the context of hospital competition, with the literature identifying the problems of contracting when quality is multi-dimensional and difficult to measure, as well as some potential contractual solutions (Chalkley and Malcolmson 1998). Interestingly, few of these issues have been considered in the context of hospital competition in low and middle-income countries, though the potential for both lower absolute levels of competition, and greater problems of quality monitoring is greater in these settings.

**INSERT FIGURE 1**

Two additional problems can be identified with the market forces model. First, there are potential problems of incomplete contracts. Public hospitals have multiple objectives and missions, and there is a risk that the sharpening of incentives through the application of market forces will lead to differential responses to the needs of different patient groups. For example, Harding and Preker (2003) recognise that ‘[I]f serving a larger number of healthier or otherwise low cost patients is the easiest way to increase income, then these efforts will ... figure into hospital behaviour’ (p46). This raises immediate concern for the access of patients who do not offer surplus generating opportunities for the hospital. Who these patients are depends on the nature of the arrangements through which hospitals
provide services. But even where there is a funded contract intended to provide universal coverage, patients presenting with particular conditions, with additional social needs, or in some circumstances (more fully discussed below), no ability to augment the contract based payment from their own pocket may offer unattractive business to surplus focused hospitals. This is probably the most important respect in which hospital contracts are inevitably incomplete (Macneil 1978), or cannot provide for every eventuality, leaving scope for opportunism (Williamson 1975).

Second, and perhaps more fundamentally, it is not clear that hospitals will seek out opportunities to generate surplus (Propper 1995), resulting in attenuated hospital incentives to pursue quality and consumer responsiveness goals. For example, a public autonomous hospital may not be the residual claimant to the surplus generated, or its decision making authority over the use of any surplus may be highly circumscribed by the regulatory regime. In such circumstances (which approximate in all the contexts the authors are aware of), resource allocation procedures leave a gap for political rather than market forces to determine outcome.

As suggested above, the market forces model of hospital reform appears to have been highly influential in the design of reform programmes in diverse corners of the world, and yet evidence that can discriminate between the alternative views of its effects has been slight as is the case for evidence in health sector reform debates in general. Evidence is not always a valued commodity in highly politicised reform debates – the deliberate neglect of health systems research in the UK in the wake of the ‘Working for Patients’ reform
programme at the end of the 1980s is a case in point (Hamblin 1998). Reforms are usually
implemented in packages with many of the same target outcomes, but differential timing
of implementation, partial implementation, and time lags before impact would be expected
together reduce the opportunities to link policy and impact.

The purpose of this paper is to review evidence of the impact of the market forces model
of hospital reform in diverse settings contrasting the insights that incentives strengthen as
more market forces apply; contracts are inevitably incomplete with implications for quality
and equity; and interest in the residual on the part of hospital decision makers is
attenuated. These insights are not mutually incompatible but the reform effects they imply
may play themselves out in consistent or contrasting ways across contexts.

There have been two major efforts to conduct and collect case studies of hospital reform,
the first in the early to mid-1990s by the Data for Decision Making Project of Harvard
University for the United States Agency for International Development (Govindaraj and
Chawla, 1996); and the second was conducted in the late 1990s by the World Bank and
published in an edited collection (Preker and Harding, 2003). These have described reform
measures, compiled data from routine sources and from any evaluations of the reforms that
have been available, and hence sought to provide an overview. Depending on timing and
availability of data, some of these studies have considered trends in hospital performance
following reform (sometimes in comparison with pre-reform) in some cases bringing to
bear other trends and external factors that may be influential.
Besides these cross-country efforts, there have been a small number of more isolated research projects that have been able to interrogate single case studies in more depth and in particular to collect primary data driven by hypotheses of reform effect. We have selected four case studies which we believe to provide more robust evidence of the effect of hospital autonomy and market forces. Two of these (UK, Indonesia) use secondary data. The other two (Zambia, Colombia) have collected primary data. Three (Indonesia, Colombia and the UK) have been the subject of at least two different reviews, providing a minimal check on the possibility of distortion from a single perspective. All used a research design that sought to control for bias and to explicitly consider potential alternative hypotheses.

On the basis of these four case studies, some patterns of hospital response to autonomy and market reform may be recognised.

Response to hospital reform in the UK

In the United Kingdom, health service provision is dominated by the National Health Service (NHS) through which care is mainly funded through national taxation and delivered through public providers, although primary level doctors are self-employed. A purchaser provider split was introduced in 1991 to the National Health Service. On the purchasing side, district health authorities (DHAs) were to contract with providers for services and some general practitioners were granted ‘fundholder’ status enabling them also to purchase services from hospitals on their patients’ behalf. On the provider side,
hospitals were able to acquire trust status, or circumscribed autonomy, if they met specified pre-conditions.

The new arrangements implied that a trust would be run by a board of directors and be free to determine its management structure and the profile of services offered (with some provisos). Trusts were to be directly accountable to the centre, employ their own staff and set employment terms and conditions. Financially, a trust’s income would be determined by its contracts with health authorities, GP fundholders and the private sector, and a trust could retain any surplus generated for use the following year. For non-trust hospitals, the DHA would retain most of these controls, although all hospitals became dependent on the sale of services through contracts. The main constraints on trusts pertained to pricing (a formula based on recurrent costs and use of capital was to be applied), and to borrowing on capital markets which was centrally controlled (Ham 2003).

Despite lack of much public investment in research, a number of studies are available which contribute to an understanding of the UK experience. With respect to efficiency, there is some agreement that there were hospital productivity increases following the reforms. The evidence of Soderlund et al. (Soderlund, Csaba et al. 1997) is the more convincing on this point. Their regression analysis showed that gaining independent trust status was associated with significant productivity gains, although they could not rule out that hospitals may have chosen to become trusts because they anticipated being able to increase productivity, and found that some productivity gain was associated with ‘episode inflation’ by which hospitals may break up episodes of care to claim higher numbers.
Figure 1 shows the relative trends for four groups of hospitals: those that became trusts in the first wave (1992), the second wave (1993), the third wave (1994), and those that remained directly managed throughout the period studied. It suggests that efficiency gains were greatest in those hospitals that acquired trust status and occurred later in those trust hospitals that acquired trust status later.

An earlier study had suggested that those hospitals that acquired trust status in the first wave were already more efficient (Bartlett and LeGrand 1994). According to Soderlund et al., this result arose from failure to control for case mix. (Soderlund, Csaba et al. 1997) data suggest first wave hospitals were least efficient at the outset, perhaps because of artificial inflation of costs to allow them to show later efficiency gains, or perhaps because of the costs of preparing for trust status (although this last explanation was not supported by evidence from third wave hospitals). However (Maniadakis, Hollingsworth et al. 1999) further control for quality change and argue that a proportion of the productivity gains may have arisen at the expense of quality. One might conclude that although productivity gains seem to have been made over the period, factors other than reform might at least partially explain the trend observed.

INSERT FIGURE 2
With respect to the other goals of hospital reform (quality and choice) there has been very little research (Hamblin 1998). The review of evidence in (Ham 2003) suggests there was no evidence that trusts increased patient choice, or became more accountable to local populations. Notable public quality failures such as those in cancer screening programmes might accord with the view of (Maniadakis, Hollingsworth et al. 1999), that quality declined over the period, although Ham also points to improvements in the care environment and service organisation predicated on management freedoms. A recent study by Propper et al. found that competition may have actually reduced clinical quality, with more competitive hospital markets associated with higher rates of mortality from acute myocardial infarction (Propper et al. 2004).

As the reforms bedded down, it became apparent that competitive pressures would not be allowed much rein in the internal market. (Ham 2003) provides a detailed account of the extent to which the market was manipulated by political interests, to ensure that concerns for stability and the protection of hospitals’ survival dominated those of competition.

This environment implied that trusts could make little headway by competing for the main contracts of DHAs through cost and quality improvements but could look for sources of additional income in more marginal areas of activity. They are argued to have competed more energetically for ‘extra-contractual referrals’ (mainly patients referred across district boundaries) and for the contracts of general practitioner purchasers – both of which were considered to be exchanged in more competitive markets. As a result, it was alleged that those patients covered by a district health authority contract waited longer for treatment.
Although evidence relating to this is anecdotal, it was sufficient to instigate a Ministry of Health instruction to hospitals to pool waiting lists of different purchasers in July 1997 (Harrison 1998).

Although it has received much less attention in the UK literature, implicit in the new freedoms afforded to trust hospitals was the freedom to develop pay-beds, privately (out-of-pocket or private insurance) financed beds which have been present in hospitals since the inception of the NHS. These also provided a source of marginal income for trust hospitals which was energetically seized upon, most notably in the form of the development of pay-bed units (Keene, Light et al. 2001; Ham 2003). Data on NHS pay-beds are not centrally collected but in the first year of trusts (1990-1), (Williams and Nicholl 1994) report an estimate that there were 81,366 patients treated in pay-beds. For 1994-5, (Keene, Light et al. 2001) provide an estimate of 99,399 for this number, a growth of 22% over 4 years. According to (Keene, Light et al. 2001), policy guidance states that pay-beds should not make financial losses and evidence suggests that this guidance has been followed (National Economic Research Associates (NERA) 1995). There are also a series of instructions that should prevent users of standard services from negative impact from pay beds. These include an explicit requirement for a common waiting list and clinical criteria for categorising paying and non-paying patients. However, (Keene, Light et al. 2001) found that: ‘The ... requirement to maintain common waiting lists for non-elective patients was, as far as we could establish, not being met. This is hardly surprising given that the main point of private treatment in the UK is to secure treatment more
quickly than in the NHS.’ Table 1 shows some sample waiting times for elective admissions which they derived from an earlier study (Williams 1997).

(Keene, Light et al. 2001) conclude that while the share of pay bed activity in the scale of total NHS provision is too small for there to be a significant effect of pay bed users on the rest, the data testify to two-tier clinical service provision within the NHS and a violation of its equity principles.

Zambia

In Zambia, the public health system is intended to offer universal coverage through a network of publicly owned facilities structured to offer basic care in small peripheral facilities and increasingly complex care through a referral system at health centres, district hospitals, regional hospitals and national referral hospitals. However, problems of inadequate resources and poor management of resources in the public sector have ensured that a significant proportion of health expenditure (62% according to the World Health Report, 2000) is channelled through the private sector which includes non-governmental facilities (mainly Christian missions) that are based in rural areas, formal for-profit health facilities that are based in the main towns of the country and informal for-profit providers that sell drugs and traditional medicine throughout the country.
(Kamwanga, Hanson et al. 1999) provide a review of the background to hospital reform in Zambia. The reform programme in Zambia was largely put in place in the first years of the government of the Movement for Democratic Change, which replaced the United National Independence Party’s 30 year rule in 1991. The blueprint for the reforms was influenced (some would say driven) by the views of donor agencies which were eager to support the new government. The reforms created a ‘purchaser-provider split’ structure by which a Central Board of Health acted as national purchasing agency, contracting with district health boards for services up to the level of the district hospital, and with referral hospital boards (public and private not-for-profit) for services at that level. This created hospital autonomy at secondary and tertiary level in the sense that the management relationship between the Ministry of Health and the hospitals was intended to be mediated through the contract rather than operate directly.

However, implementation of the reforms was interrupted. There were 3 changes in leadership of the Ministry of Health between 1994 and 2000 and each Minister took a different view of the reforms, creating a stop-go cycle of implementation over the period. Structures were put in place – a considerable achievement – but they were never effectively used. Contracts between the CBoH and hospitals never developed beyond the specification that the CBoH would make a given payment (which in any case was often not fully paid), while hospitals would deliver services according to demand. Attempts to shift the locus of hospital staff contracts from the Public Service Commission to individual hospitals, in order to achieve effective management of staff at hospital level, were frustrated by union
discontent and an absence of political will to overcome that. However, some key staff, especially those in new roles associated with the reforms such as accountants, were recruited directly by Boards. Direct interference in hospitals’ day to day management by the Minister of Health continued, despite the designated roles implied by the purchaser-provider split.

A number of other developments concurrent with the reforms need to be considered in evaluating the Zambian experience of hospital autonomy. Three tertiary hospitals in the country were the object of efforts to support management strengthening in the Copperbelt region of the country under the KANDO project. These efforts included the development of management information systems, investment in financial management through recruitment and training and technical assistance in relation to debt management and revenue generation. Second, after relative neglect of the hospital level within the reforms, in 1998 a hospital steering committee was created and a hospital policy developed in 2000. This emphasized the ‘right sizing’ of hospitals – the redistribution of hospital beds on a more equitable basis through the country. Third, the background to reform implementation was one of economic decline and shrinking resource availability for the health and other sectors.

Financing policy was also under reform throughout the health sector. A policy by which ‘cost sharing fees’ (at levels intended to be token) were to be applied to services within a defined essential package of services, and ‘cost recovery fees’ (intended to represent the full cost of the service) were to be applied to services outside the package, had been negotiated and agreed within the health sector early in the reform period. Although the
policy was never ratified, lacking a political backer who would support its prioritisation on the cabinet’s agenda, it was widely understood by health sector decision makers to be the \textit{de facto} official policy and the guiding financial document at primary care level.

The problem at tertiary level was that the relevant essential package was only defined in 2001. Until then, the interpretation at that level was to consider cost sharing fees as appropriate for all services delivered in ordinary wards and clinics, and cost recovery fees applicable to ‘high-cost’, ‘fast-track’ and ‘private’ wards and clinics which hospitals increasingly developed during the reform period. Similar to UK NHS pay-beds, these offered patients the choice of a higher priced service with some characteristics superior to those on ordinary wards or clinics, for example shorter waiting time or access to a consultant of choice. The development of such services was a particular emphasis of the KANDO project, and seems to have been the main use made of the degree of greater autonomy which secondary and tertiary hospital managers achieved. Like UK hospitals, Zambian hospitals could not gain by competing for ordinary patients. The lump sum budget for these patients was unresponsive to hospital performance. High cost patients offered the only opportunity to increase revenues.

In practice, therefore, the implication of greater hospital autonomy in Zambia and the form which greater exposure to market forces took was most clearly witnessed in the development of two-tier charging. This was encouraged by the KANDO project, and presumably embraced enthusiastically by hospitals outside the project because of its revenue generating capacity. This is understandable in any context but particularly one in
which hospitals had accumulated large debts (Haddon 1998) and might be described as in financial crisis. This suggests that revenue or surplus maximisation provide reasonable alternative motivations for hospitals in this context.

Research aiming to measure the distribution of resources within the hospital to ordinary and superior service users was conducted in two of Zambia’s tertiary hospitals in 2000 (McPake, Nakamba et al. 2004). This aimed to compare costs between “high-cost” and “low-cost” service users, to identify the extent of cross-subsidy in either direction, and to measure the allocation of specific resources which were (a) within the control of hospitals; (b) relevant for clinical quality of care; and (c) important to users and therefore likely to affect the relative demand for high-cost and low-cost care.

The study used step-down costing methods to measure costs in one of the hospitals, and made specific measurements of resource availability in 8 wards in each hospital including staffing (observing ward round duration and staff involved, presence of staff on wards at other times and checking the staff allocation), cleanliness (checking toilets and the presence of cleaning staff in wards), drug use (auditing prescription, receipt and the purchase of drugs outside the hospital), use of diagnostic tests (lab and X-ray) and use of the operating theatre for major and minor operations.

Table 2 shows cost by the 8 wards in the costed hospital. It shows that cost differences ranged from 34% to 145% higher in “high-cost” than “low-cost” wards, and non-staff costs ranged from 43% to 229% higher.
Analysis of use of diagnostic services and of theatre suggested that numbers of users of laboratory, X-ray and theatre for major and minor surgery were broadly in line with numbers of high and low-cost service users in the hospital population as a whole. There was no control for case mix in this comparison. Drug use patterns showed marked differences, however. High cost patients were prescribed higher numbers of drugs, were less likely to be asked to buy the drug outside the hospital, and consequently received a markedly higher number of drugs per patient (figure 3). This might reflect demand or supply side bias. For example, low-cost patients might be more likely to refuse diagnostic services given the charges levied for those, making active management of their condition more difficult, resulting in fewer numbers of drugs being prescribed.

Using a staff index which weighted staff of different categories, there were also marked differences between assigned staff, time spent on ward rounds, and staff presence on wards (figure 4), all favouring high-cost patients. Presence of cleaning staff on wards also suggested bias in favour of high cost patients and this was reflected in the observed cleanliness of toilets.
The Zambian data have considerable limitations, most notably their inability to indicate whether resource allocation patterns reflect differences in case mix between high and low cost patients, or preferential treatment. Nevertheless, the cost differences are marked and both analyses indicate that low-cost patients secure less access to inputs, particularly non-staff inputs.

INSERT FIGURE 4
Indonesia

The health care system in Indonesia includes public facilities ranging from health centres to hospitals, and a large private sector including non-governmental providers (religious missions), and private clinics and hospitals. Compared to other South-East Asian countries, public health expenditure is low (Int$21 per capita, compared with Thailand, $108; Malaysia, $116 and Singapore $268), accounts for 37% of total expenditure on health (WHO, 2000) and has been decreasing in the context of a contracting economy and declining oil revenues. Approximately 14% of the population is covered by any kind of health insurance (Thabrany, 2000). Indonesia’s ratio of hospital beds of 0.7 per 1,000 population in 2002 compares to an average of 2.5 in East Asian and Pacific countries. Decentralisation starting in 1987, delegated responsibilities to provincial and district level governments.

(Bossert, Kosen et al. 1997) and (Lieberman and Alkatiri 2003) provide accounts of Indonesia’s 1991 hospital reform programme by which hospitals were granted autonomy in cohorts, on the basis of meeting a set of criteria in relation to cost recovery, bed occupancy and length of stay, and the prosperity of the surrounding economy. The first hospitals to be granted autonomy were largely ‘vertical’ (or centrally managed by the Ministry of Health) but increasingly, those managed by provincial and district levels have been included. According to (Bossert, Kosen et al. 1997), the major objective of hospital autonomy was to

---

encourage hospitals to recover costs, providing a somewhat more limited focus than in Zambia. Like in Zambia, autonomy in Indonesia’s case is described as highly circumscribed, characterised by controls from central, provincial and district levels. However, (Bossert, Kosen et al. 1997) list the following respects in which autonomy increased for Swadana (autonomous) hospitals: the hiring and firing of hospital staff; setting fees; distribution of beds among fee levels, except class III (set aside for the poor and charged at the lowest rates), purchasing drugs and supplies, selecting incentive systems for use of a specified portion of own source revenue; and contracting with the private sector. This suggests that Indonesian autonomous hospital managers had considerably greater room for manoeuvre than their Zambian counterparts. Perhaps the most important source of greater autonomy was in relation to their own generated revenues, which constituted a high share of total hospital revenue – between 30 and 80% (Bossert, Kosen et al. 1997), and 40% was set as the minimum to achieve Swadana status (Lieberman and Alkatiri 2003). Indonesia’s hospitals have a graduated fee structure for successive levels of amenity, ranging from class III to VIP (for the very rich). Swadana hospitals were free to set their own fees, except for class III beds, and could determine service patterns across classes of beds, subject to providing a minimum of 50% class III beds.

(Bossert, Kosen et al. 1997) conducted research in 10 hospitals, 5 of them Swadana, 3 public, managed by either provincial or district authorities, and 2 private (one small, one large) in 1996, assessing trends in relation to levels of fees, pattern of service provision (bed availability by class), unit costs, staffing patterns, length of stay, bed occupancy and
management systems. (Lieberman and Alkatiri 2003) conducted analysis of 6 Swadana hospitals and one traditional district hospital in 1998.

Both studies found that hospitals operated in a favourable financial environment, with increasing government subsidy throughout the periods covered. This was surprising given the apparent objective of the reform programme being to increase the private financing of hospitals. According to both evaluations, fees increased considerably in the Swadana hospitals after autonomy was introduced. In (Bossert, Kosen et al. 1997) sample, the average percentage increases from the year before to the year after the transition to Swadana status, ranged from 34% for Class I beds, to 144% for VIP beds (although only one of the five Swadana hospitals provided VIP beds both before and after the introduction of autonomy), or 117% for class IIIb beds. However, fees increased in the non-Swadana public hospitals to similar degrees – a finding that might have been related to the requirement to increase cost recovery rates in order to qualify for Swadana status. Fees increased in private hospitals too, although gradually and throughout the period analysed. (Lieberman and Alkatiri 2003) also found fee increases for class III beds across the board, averaging 51% over the period 1994/5-1998/9. This figure was highest (100%) for the non-Swadana hospital included in their survey.

Bossert et al. found a trend toward decreasing availability of beds for the poor related to Swadana status(Bossert, Kosen et al. 1997). Although the pattern was variable (figure 5), there was a combined reduction of 365 out of 1559 class III beds in 3 Swadana hospitals, with only one hospital reporting a significant increase. In the other public and one private
hospital which reported this data, the number of beds available to the poor was unchanged. Lieberman and Alkatiri also found that the number and share of class III beds fell in some Swadana hospitals.

As in Zambia, Bossert et al.’s comparison of unit costs and revenues in two of the Swadana hospitals, suggests that the higher classes of beds do not generate net additional resources to subsidise the lower classes (table 3). In one of the two Swadana hospitals, there was a regressive relationship between bed class and fee recovery level by which the poor effectively subsidised the rich. Unfortunately, there were no data available to compare with the situation in these hospitals prior to autonomy. However, in the small private hospital, the relationship between bed class and fee recovery was progressive (table 3).
The evidence from this study on the direction of subsidy is rather tentative, given the small sample of hospitals from which data were available. It is supported by data from another study (Suwandono, Gani et al. 2001) which costed services in a further three hospitals and found that commercial beds (those above class III) failed to cover even their recurrent costs. An important reason for this was suggested to be the high proportion of revenue used for staff (mainly doctor) incentives. In this study, no specific link to Swadana status is alleged.

However, Lieberman and Alkatiri’s data suggest the opposite result at least in two of their sample hospitals. Subsidy appears progressively targeted towards the poor in one of the sample hospitals whereas in the other it is middle level beds (classes I and II) that seem to provide the subsidy for class III beds rather than VIP and super-VIP beds (a further category which seems to have emerged since Bossert et al.’s study). The authors do not comment on the divergence of their results from others’ here (although Bossert et al.’s study is reviewed within their paper), and they do not indicate what weight they would place on their findings, given the limited sample. There is little explanation of methods applied in this study, which would allow understanding of whether the costs reported included incentive payments to doctors and others, for example.
Both studies suggest that alongside these potential problems in relation to the equity effects of the hospital reform programme in Indonesia, some progress has been made in relation to hospital efficiency. In Bossert et al.’s study, quantitative measures of efficiency, for example length of stay and bed occupancy rate did not change markedly, nor were there significant differences between their levels in Swadana and non-Swadana hospitals. However, management systems were deemed to have improved, and in particular, the incentive payments made possible by the combination of increased own-generated revenue, and greater autonomy to decide incentive systems were deemed to have reduced physician absenteeism. Lieberman and Alkatiri reached similar conclusions, and also found increases in bed occupancy rates in the Swadana hospitals. However, the non-Swadana hospital in their sample had a better than average growth in its bed occupancy rate.

Colombia

Prior to reform, the Colombian health system was ‘segmented’ into the three sub-systems public, private and social security, that are typical of Latin American countries (IABD, 1996). In 1990, private expenditure, made up of contributions to private insurance funds and out-of-pocket expenditure on treatment was a high proportion (64%) of the total. Expenditure on social security made up of compulsory contributions by formal sector workers and their employers, and of public subsidy accounted for 22% of total expenditure. Expenditure on services provided by publicly owned facilities, the resort of the majority of the population without formal employment or the ability to pay for private sector services, made up the remaining 13%. These proportions are comparable to the group of Latin
American countries with ‘segmented’ health systems (data from IABD, 1996 and Suarez, 1994). The case of Colombia provides a contrasting example of hospital reform to those of Zambia and Indonesia in that the move towards greater hospital autonomy took place within a reform programme that aimed to reduce the extent of out-of-pocket payment for health care and hospital services, rather than to increase it.

Reforms to hospitals occurred in the context of a sector-wide reform programme. Law 100 was enacted in 1993 and intended to come into effect in 1995. Its provisions are explained by (Londono 1996) and (Yepes 2000). It aimed to universalise insurance coverage by transforming the existing social insurance programme into a contributory regime and creating alongside it, a subsidised regime to which there would be means tested entitlement. A screening instrument (SISBEN) was applied by municipal governments and categorised households as able to pay 12% of their income and enrol in the contributory regime, or able to pay smaller percentages, on a sliding scale, and enrol in the subsidised regime. Packages of care for the two regimes were differently specified but intended to be aligned by 2010. The insurance market was liberalised. Both contributory and subsidised regime packages could be offered by a new kind of institution, the EPS (Health Promotion Enterprise), private organisations which were to compete for households’ subscriptions.

A series of cross-subsidies were designed within the reform programme. One twelfth of the revenue from the contributory regime was to be transferred to help finance the subsidised regime. A risk equalisation fund was designed to both provide for cross-subsidy and protect against cream skimming of richer households. The relevant subscription paid by each
family would be forwarded to the fund and an age and sex adjusted capitation rate returned to the EPS with which to ensure provision of the relevant package of care.

The implication of the reform programme for public hospitals was that, in principal at least, they moved from being directly managed by the relevant health authority and became ESEs (Autonomous State Enterprises), run by a board of trustees and contracting with EPSs for business.

There have been difficulties in implementing the above measures which have left gaps between the system as it was designed and as it has emerged (Plaza, Barona et al. 2001; McPake, Yepes et al. 2003). 33% of the population, approximately those who were previously covered by social security, were covered by the contributory regime by 2000. A further 22% were covered by the subsidised regime. These people would previously have relied on public services or have had little access to health services. However, this still falls short of the goal of universal coverage. Second, there were multiple abuses by insurers at the start of the programme, including the failure to provide registration cards (allowing claim of premium but avoiding provision of services) and exclusion of the poorest. These abuses were considered to be in decline by 2000. Third, abuses of the SISBEN process were also alleged including the exclusion of some rural areas and urban squatter communities, and the channelling of subsidy to politically favoured populations. Fourth, it proves difficult in practice to discriminate between those covered by the subsidised and contributory regimes in terms of the package provided, or even in terms of who should be billed for treatment received – the relevant EPS of the subsidised regime or the territorial
health secretariat who in theory should cover any deficit. Finally, at first competition among EPSs within the contributory regime was slow to develop as those covered tended to maintain their affiliation to the Colombian Institute of Social Security (CISS) which had previously managed the social security programme. However, this started to change after 1999. 1.2 million shifted their affiliation in the period 1999-2000 – approximately 8.5% of all affiliates.

A noticeable phenomenon in the immediate aftermath of the reform programme has been dramatically increased expenditure to a point that makes the sustainability of the reforms open to question (Sanchez and Yepes 1997; Jaramillo, Olano et al. 1998). (Jaramillo 2002) reports that hospitals’ resources have increased 1.5, 1.9 and 2.6 fold for level 1 (smaller and less specialised), level 2 and level 3 hospitals respectively.

One study sought to track performance of hospitals in Bogotá in the period before and after the implementation of the reforms (McPake, Yepes et al. 2003). The study found increases in overall activity (figure 6: total admissions), bed occupancy (figure 7) and productivity indicators (outpatient visits and emergency outpatient visits per contracted staff hour). These changes could not be easily accounted for by case mix data which, though inconclusive, suggested that complexity of case mix increased over the period. Similarly, quality of care data failed to provide evidence that productivity indicators grew at the expense of quality. The availability of inputs increased, to offer a wider range of services and measures of process quality of maternal care and patient satisfaction with services in general suggested no significant changes.
Qualitative data focused on how hospital workers reacted to and perceived the changes that had taken place. It concluded that the change process itself was perceived as threatening by health workers but that as change bedded down, positive features of reform – such as clearer roles for staff and responsiveness to patients - were increasingly identified, though negative views – dominated by the view that money rather than human values now ran the system - persisted. A transformation in administrative procedures, necessary to enable hospitals to deal with the greater complexity of billing multiple insurance bodies, was widely recognised, and its corollary in the increased importance of the administration and greater prioritisation of management training, in some cases resented.

This appears to be the only published attempt to track performance in hospitals over the relevant period and is hampered by its small, Bogota-based sample. Other studies provide limited information relevant to hospitals’ experience of reform more widely. (Plaza, Barona et al. 2001) suggest that the difficulties that hospitals in Bogotá faced in transforming their administrative systems to those capable of billing multiple payers were reflected in other parts of the country. Overall, system accessibility was judged by the same authors to have improved, with a reduction in those reporting non-use of services for financial reasons from 51% to 41% between 1992 and 1997. (Jaramillo 2002) report that at national level, public
hospitals became more accessible, increasing their coverage levels from 35% of the population in 1990 to 63% in 1997. The basis for this estimate, derived from Ministry of Health data, is not reported however.

Conclusions

The market forces model of hospital reform suggests that efficiency gains can be achieved through strengthening incentives as hospitals transform from budgetary towards private units by gaining autonomous or corporate status. Critiques of that model suggest that incomplete contracts leave the potential for perverse incentives to arise, which might have negative effects for equity or quality of care. Further, where autonomous or corporate hospitals are not the residual claimant on surplus generated, they might focus on a different and unknown objective, with unpredictable consequences for efficiency, equity and quality of care. Table 4 shows the different ways in which each country’s reform produced market forces, contractual incompleteness and attenuation of interest in the residual.

Evidence in this debate, as in other health sector reform debates, is scarce. What exists in four slightly better researched countries gives some limited scope for optimism with respect to efficiency gains. In the UK, productivity increases have been documented, but these may have been at the expense of quality diminution, or reflect selection bias in those hospitals accorded trust status. In contrast, in Bogotá, improvements in productivity measures did not
diminish after admittedly limited attempts to control for case mix and quality change. Whether the Bogotá experience reflected broader Colombian experience could not be ascertained from published literature. In Zambia, an environment of economic crisis and a context of severe under-funding make quantitative measures of efficiency suspect. Concurrent activities may have supported management improvement but are not inherently related to the market forces model of reform. In Indonesia, there was also evidence of management improvement which could more plausibly be linked to the reform programme, and productivity improvement, which could not.

Evidence that incomplete contracts leads to disadvantage for particular patients whose net impact on hospital objectives is small or negative would seem more consistent, at least across three of the country experiences reviewed. In the UK, the development of pay-beds in response to the lack of market opportunities in the mainstream has received little attention, presumably because these still constitute a small share of total NHS business. Nevertheless, the increased volume of activity and evidence of preferential treatment in relation to waiting times is a clear departure from principles of equity presumed to dominate in the NHS market. The development of two-tier services in Zambia, and the apparently inequitable trends in multiple tier service provision in Indonesia are a more manifest feature of hospitals’ responses to market forces. In these two countries too, the implications seem more severe. Evidence supports the contention that at least in some circumstances, users of lower cost services subsidise users of higher cost services and that the development of high cost services may have a negative impact on low cost users.
In all countries, there is a gap for political rather than market forces to determine market outcomes. Whether desirable or not, this would appear unavoidable given the high political profile of health services and hospitals in particular in almost all settings. It is probably least the case in Colombia where both purchaser and provider sides are both decentralised and competitive, rendering political intervention more difficult. Ironically, at least for those who assume that market forces and equity are inevitably in conflict, in the UK and Zambia it is likely that constraints to competition have exacerbated the equity effects discussed in the previous paragraph. In the UK, the lack of scope for competition in DHA-trust contracts focused energy and activity on marginal areas of hospital business: extra-contractual referrals, general practitioner fundholders contracts and pay-beds. In Zambia, the maintenance of a low-powered global budget for the main contract between the CBoH and autonomous hospitals has similarly focused business development on “high cost” wards and patients – extra resources cannot be obtained by seeking (for example) quality improvements in low cost wards. In Indonesia, the situation appears different. Regulation of the market at least appears to protect space made available to the poorest patients, although is probably unable to protect against intra-hospital non-bed resource allocation which may disadvantage poor users. This contrasts with the apparent success with which UK regulation avoids loss making pay beds.

There was no evidence reported in Colombia’s case suggesting that particular groups of patients were preferentially or adversely served by hospitals’ greater exposure to market forces. There are a number of design features of the reforms that would seem to protect against this problem to some extent. There are in principle, no out-of-pocket payments and
those using hospitals do so under the cover of the contributory or subsidised insurance regime. Thus, if there are groups who represent a poor return to hospitals, they are not easy to identify and thereby to compare with other groups. For example they may suffer from a particular condition for whose treatment hospitals have not been able to negotiate profitable prices, or they may be covered by particular insurers whose contracts offer more marginal returns to the hospital. Whether or not particular diagnostic groups face discrimination in hospital use is equally unresearched in the other three countries. The services received by those covered by different insurers and the cross-subsidies between contracts in Colombian hospitals would be a fruitful area of research. Another irony within the evidence available is that where discriminatory treatment of different groups has been specifically included in reform design (in Colombia between contributory and subsidised regime members) it has proved difficult to implement. It is probable that treatment does differ between these two groups but there seems to be agreement that it does not do so to the extent envisaged.

The evidence supports the view that a reform model cannot be assumed to operate similarly in highly different contexts. It also supports the view that the specific features of the reform design are perhaps of greater importance than the model which provides the initial impetus for reform: the devil is in the detail. These two points are not wholly separable as the feasibility of specific design features is context dependent. For example, it would seem that the avoidance of out-of-pocket payment in Colombia’s reform design has been a key feature in protecting against certain kinds of inequitable outcomes. However, avoidance of out-of-pocket payment may not be considered a feasible option in all settings. In Zambia, public resources for funding hospitals are considered absolutely insufficient and remain so,
partly owing to macro-economic realities and partly owing to the weight of hospitals and health services in government priorities which are unlikely to change. Resources from out-of-pocket payments are a crucial marginal resource that keep hospitals running. In Indonesia, there was an explicit policy priority to withdraw public resources from hospitals, one that may be justified in the light of the assessment of the role of hospitals in health systems outlined in the introduction to this paper. Other features which enforce cross-subsidy by applying a universal means test and requiring compulsory insurance membership by the self-employed for example, have proved difficult in Colombia and may be considered impossible in lower income settings.

Nevertheless, there seem to be some insights from this review of experience that can inform the design of hospital reform programmes. First, the review highlights the need to consider the role expected to be played by competition in hospital markets. The operation of market forces is not wholly dependent on the existence of competition on the supply side, as seems often to be assumed. Monopolists are motivated to minimise cost, and profits are dependent on ability to attract demand which constrains prices and offers incentives for quality improvement, even if less so than competitive firms. However, it would seem helpful in designing reforms to seek clarity as to exactly what competitive forces are expected to apply and how feasible it will be for such forces to operate given realistic assessments of the political environment. Second, recognition of contract incompleteness and of the potential for perverse incentives requires consideration of the most likely forms in which these might arise, and the costs and benefits of three alternative responses: increasing the
completeness of contracts, applying regulation external to the contracting parties, and allowing perverse incentives to prevail.

The evidence suggests that there are a range of possible outcomes associated with the application of the market forces model of hospital reform. Designing reform so that the balance struck between benefits and adverse consequences is as good as it can be is the challenge for policy makers.
References


Sanchez LH, Yepes FJ. 1997. La Ley 1000: Dos anos de implementacion. Santa Fe de Bogota, ASSALUD, FESCOL, Fundacion FES, GTZ. Informes Tecnicos No. 1.


Table 4: Summary of 4 countries reforms

<table>
<thead>
<tr>
<th></th>
<th>Structure of reform</th>
<th>Exposure to market forces</th>
<th>Contractual incompleteness</th>
<th>Attenuation of interest in residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Part of purchaser:provider split. Hospital trusts run by Board of Directors.</td>
<td>Income dependent on sale of services through contract to own DHA, other DHAs through extra-contractual referrals and GP purchasers. Small share from private patients. Own DHA revenue effectively guaranteed, competition for extra-contractual and GP referrals.</td>
<td>Weak provision against separation of waiting lists for extra-contractual referrals, GP referrals or private patients.</td>
<td>Surplus to be retained for following year. Apparently effective incentive.</td>
</tr>
<tr>
<td>Zambia</td>
<td>Part of purchaser:provider split. Hospitals governed by Autonomous Hospital Board.</td>
<td>Income dependent on block contract with CBoH and fees charged to patients. Block contract invariant to any activity, variance only through fees charged.</td>
<td>No effective monitoring or minimum standards for clinical quality levels of public patients.</td>
<td>Revenues retained, reducing deficits in hospital finance and increasing credit-worthiness and supplies of basic inputs.</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Independent reform following separate decentralisation reform. No new accountability mechanisms; new authorities for hospital management</td>
<td>Income dependent on direct government subsidy and fees charged to patients. Policy intention to reduce subsidy and increase importance of fees charged.</td>
<td>Limited controls over minimum provision for lowest service level.</td>
<td>Revenues raised retained but only allowed to be used for certain purposes. Apparently effective incentive.</td>
</tr>
<tr>
<td>Colombia</td>
<td>Followed decentralisation reform; part of purchaser:provider split and insurance reform. Public hospitals became autonomous state enterprises, run by Board of trustees</td>
<td>Income dependent on contracts with insurance agents of contributory and subsidised programmes; and on state governments to cover uninsured services. Public hospitals continued significant direct financing by state government</td>
<td>Failure to enforce discriminatory treatment of patients with different entitlements.</td>
<td>Surpluses may be retained but in practice unpaid bills by local government means real (not paper) surplus very rare.</td>
</tr>
</tbody>
</table>