Private wards in public hospitals: two-tier charging and the allocation of resources in tertiary hospitals in Zambia.

Barbara McPake [1,3], Pamela Nakamba [2], Kara Hanson [3] and Bernard McLoughlin [4]

HEFP working paper, 05/04, LSHTM, 2004

[1] Health Systems Development Programme, LSHTM
[2] Department of Economics, University of Zambia
[3] Health Economics and Financing Programme, LSHTM
[4] Independent consultant

ABSTRACT

Many public hospitals in Zambia are developing two tier charging structures by which private or 'high-cost' services are offered alongside standard or 'low-cost' ones. These are expected to offer higher levels of amenity but comparable levels of clinical quality of care.

The common budgetary base of low-cost and high-cost services has given rise to the concern that the public subsidy may be diverted to supporting high-cost services. In other words, instead of the apparently intended consequence that high-cost services subsidise low-cost ones, the reverse might apply. Evidence from other countries suggests that this concern is not confined to Zambia.

The research reported in this paper consists of two parts. The first uses a costing exercise that was conducted in Kitwe Central Hospital (KCH) in 1997-8 to derive a comparison of costs and cost structures in high and low-cost wards and to assess the ratio of revenues to costs in order to measure whether high-cost wards generate a net surplus which can be used to subsidise low-costs wards. However, the likelihood that resources allocation patterns reflected in cost structures have clinical significance cannot be judged using cost data alone. The second part of the study examines the extent to which resources of clinical significance are preferentially directed towards high cost patients in two of Zambia's public tertiary hospitals, the University Teaching Hospital (UTH) and KCH.

The comparison of costs suggests that high cost patients have better access to hospital inputs. There are problems with the comparison, for example wards are not exactly analogous, but consideration of the sources of bias suggests that the comparison understates rather than overstates the services high-cost patients receive relative to low-cost patients for similar conditions.

Comparison of allocation of resources of clinical significance found disproportionate use of resources by high-cost patients in a number of areas of hospital activity, consistently favouring high-cost patients. We have not compared case mix and need across the two groups of patients, and it is possible that, for example, disproportionate use of theatre by high cost patients is explained by a high-cost case mix biased towards surgical cases. However, since instances of disproportionate use arise in each of the areas addressed, a general explanation

that high-cost patients are more severely ill or requiring greater treatment intensity in all respects, rather than disproportionately represented in particular conditions would be necessary to support this alternative interpretation. This is questionable on a number of bases.

A number of explanations of preferential allocation to high-cost patients can be hypothesised. Whichever is dominant, it would seem that the financing policy requires amendment to enable an incentive environment more conducive to the encouragement of high standards in low-cost services.

ACKNOWLEDGEMENTS

The Health Economics and Financing and Health Systems Development Programmes are funded by the UK Department for International Development (DFID). This research was funded under a further project grant from DFID's HPACORD programme and a second project grant from the Swedish International Development Agency.

INTRODUCTION

In the wake of the primary health care movement there has been relatively little attention paid to the problems of tertiary hospitals in developing countries. These institutions seem to have been regarded as, at best, irrelevant and at worst, the devourers of resources better spent elsewhere.

Most of those concerned with developing country health systems are nevertheless aware of a series of typical problems confronting tertiary hospitals. Despite consuming a disproportionate share of national resources when assessed in relation to the proportion of the burden of disease appropriately treated at that level, hospitals are 'under-funded' in relation to the absolute level of resources needed to perform their intended role. Many tertiary level hospitals lack staff and equipment for their specialist role. Failures of referral systems are also commonplace.

Tertiary hospitals in Zambia have long been characterised by this complex of problems. Secondary and tertiary hospitals together consumed 42% of public health expenditure in 1997 [1, 2]. The University Teaching Hospital (UTH) in Lusaka contains 1544 beds, but fails to provide minimally adequate tertiary level services. For example, there are shortages of basic supplies, critical medical and technical skills and mismanagement of these critically short resources [3]. The public tertiary hospitals in Kitwe (628 beds) and Ndola (948 beds), towns in the Copperbelt regions of the country which was once its powerful industrial centre, have benefited from support to management systems through the KANDO project [4]. However, quality problems are pervasive in the hospitals of the country as a whole. A hospital accreditation programme aiming to set very basic quality standards was able to accredit only one of eight hospitals surveyed in April 2000 [5].

Hospital policy has now been given greater prominence in Zambia's reform programme. The National Health Services Act (1995) restructured the Zambian health system to create an internal market within which the Central Board of Health (CBoH) purchases services from districts (primary health care and up to the level of the services provided by the district hospital), and newly 'autonomous' secondary and tertiary hospitals. In practice, the autonomy of hospitals is heavily circumscribed and they are subject to considerable intervention on even minor matters by the Ministry of Health [6, 7]. In 1999, explicit recognition that the reform

programme had neglected to support secondary and tertiary hospitals in the transition to more autonomous status led to the creation of the Hospital Steering Committee (HSC).

Changes in the financing policy have also taken place, with implications for incentives at hospital level. Hospitals receive two sources of funds. A block grant, determined on the basis of the number of approved beds, is received from the CBoH, though often not fully and there is a lack of transparency in the degree to which hospitals receive their calculated entitlement.

Hospitals also receive transfers from districts, to cover first and second level services provided, though the basis of their calculation and payment differ widely, and there is an allocation for each hospital in the Medical Stores budget. In addition, hospitals charge user fees. To varying degrees, public hospitals have introduced a fee structure with two tiers – 'high cost' and 'low cost'. In principle, patients opting for 'high cost' services receive higher quality hotel services and wait for less time. For example, KCH has a 'fast-track' outpatient clinic intended to enable higher fee paying patients to jump the queue. The rationale for the implementation of this two-tier system is widely understood to be the opportunity to cross-subsidise low-cost patients from a net financial surplus generated by high-cost service provision. This is implicit rather than explicit in policy documents. High-cost services may be paid for on a fee-for-service basis by users, or on a pre-payment basis by which a monthly fee is collected, and fee-for-service bills subtracted from the accumulated amount in the users account. This resembles a medical savings scheme. In 1997, cost sharing of all types accounted for 14% of the income of UTH and 24% of the income of Kitwe Central Hospital (KCH).

This is a dimension of public-private mix which has received very little attention despite not being a peculiar feature of developing countries. In addition to the questions to which other forms of public-private mix give rise, the common budgetary base of public and private (or low-cost and high-cost) services has given rise to the concern that the public subsidy may be directly diverted to supporting high-cost services. In other words, instead of the apparently intended consequence of high-cost service cross-subsidy of low-cost services, the reverse might apply. For example, McPake et al. [8] argue that the incentives associated with this financing structure are likely to imply differences in clinical quality of care, and may

encourage hospitals to subsidise high-cost services in order to generate valuable discretionary resources.

According to the health financing policy, a service package is defined at each level of the health system. Services which form part of the package should be priced on a 'cost-sharing', subsidised basis. Services outside the package should be priced on a 'cost-recovery' basis, fully covering the cost of services received. While this is clear in principle, there are a number of confusions in practice. Particularly relevant to the subject of this paper, it is not clear whether all services provided to patients on 'high cost' wards are classified as outside the package and therefore to be charged on a cost-recovery basis, or only additional services beyond those provided to patients on low-cost wards. In the latter case, 'high-cost' patients would be entitled to subsidised basic clinical services and a net-subsidy on the total package of care they receive. This implies a less progressive policy than if the former principle is supposed to apply, and the opportunity to subsidise low-cost patients by high-cost patients' fees considerably reduced. Under the former understanding of the policy, all the subsidy provided by the CBoH would be used to provide 'low-cost' services.

Evidence from other countries on the effects of two-tier fee structures on quality of care is mixed. There is a surprising absence of research on the issue of 'pay-beds' in NHS hospitals, despite considerable and at times heated debate (for example, [9-13]). A study in 1995 concluded that the NHS' rate of profit on private patients was 17% and that this could be used to provide publicly funded care[14]. However, critics continue to be concerned that the system creates an incentive to increase waiting lists for non-paying patients [15, 16, 23], and that inequities in access are entrenched through the system. Autonomous hospitals in Indonesia, and public hospitals in Ireland were found to charge prices for 'cost-recovery' or 'private' beds that failed to cover their costs[17, 22]. The Irish study assumed that 'private' patients were not preferentially allocated resources relative to 'public' patients in the same diagnostic group.

The aim of this paper is to examine the impact of high-cost ward activity on patients using low-cost wards. The paper consists of two parts. The first uses a costing exercise that was conducted in KCH in 1997-1998 to derive a comparison of costs and cost structures in high and low-cost wards and to assess the ratio of revenues to costs in order to measure whether high-cost wards generate a net surplus which can be used to subsidise low-cost wards.

However, the likelihood that resource allocation patterns reflected in cost structures have clinical significance cannot be judged using cost data alone. The second part of the study examines the extent to which resources of clinical significance are preferentially directed towards high cost patients in two of Zambia's public tertiary hospitals, UTH and KCH. Following from both analyses, an assessment is made of whether high-cost patients are fully funding those elements of their care which are additional to those received by low-cost patients. Although the one and two hospital samples for each part of the study respectively are small, there are only three tertiary hospitals in Zambia. It is at this level of the system, where hospitals are accessible to a large urban population containing the small middle class. Evidence on perceived quality of care from alternative hospital providers indicates that public referral hospitals are preferred to private sources for more serious conditions, given the recognition of superior clinical skills at these locations [19]. Two tier charging is therefore most likely to thrive in these settings and its implications are therefore of most importance there. Although a sample of two appears small, it has already been noted that expenditure at this level of the system represents a significant share of the national public health budget.

Thus, we are not asserting that the experience of these hospitals is generalisable to other hospitals in the country but rather that they are significant in themselves in the Zambian context, and that their experience can provide lessons for central and tertiary hospitals in other countries in the region which are considering or undertaking similar charging strategies.

METHODS

The estimation of cost used a 'step-down' costing methodology which ran for over two years. Its outputs were reconciled to the Hospital's audited financial statements. KCH was the only Zambian hospital which were given an unqualified audit statement in the 1997-8 financial year and we consequently have some trust in the accuracy of these data. The methodology allocated total costs between intermediate (administration, indirect support services such as pharmacy, cleaning and security, and direct support services such as blood back, catering, laundry and laboratories) and final (outpatient clinics and inpatient wards) cost centres; and then from intermediate to final cost centres. Our comparison was based on inpatient data only as high cost and low cost patients were not always separable within outpatient clinics.

For the second part of the study, the approach was to measure the allocation of a specific set of resources which would appear to be within the control of hospital actors, transferable between wards, which are important for clinical quality of care, and which are identified by users of hospital services as important [19], and therefore likely to affect demand for the two levels of care.

The allocation of the following resources was examined: staff, drugs, diagnostic tests and x-rays, and use of the operating theatre. The allocation of cleaning staff was considered as part of the staff assessment, and the cleanliness of ward toilets was examined as an indicator of application of cleaning resources.

Four high-cost wards and four low-cost wards were purposively selected in each hospital with the aim of including one medical, one surgical, one maternity and one paediatric ward on each side. Small numbers in high-cost wards meant that in some cases wards covered more than one of these purposes. For example in KCH, medical and surgical wards were combined. The 16 wards were selected to enable comparison between high-cost and low-cost sides, as far as was possible. Four scheduled ward rounds on different days in a single week were observed on each ward.

The length of the ward round and the staff carrying out the ward round (by level – consultant, senior doctor, junior doctor, clinical officer[i], registered nurse, enrolled nurse, cleaner or porter) were noted. At four separate points of time on different days during the same week, the staff present on the ward were observed and their numbers and levels (as before) noted. The staff allocation – the team assigned to the ward - was also listed by level. In some cases, research assistants did not note whether or not a doctor was junior or senior and as a result, the junior and senior doctor categories were pooled in the analysis. Ward toilets were visited five times during the week and an assessment made of whether or not they had been cleaned in the last 12 hours. This allows for inter-observer variation but the same two research assistants assessed both high-cost and low-cost wards in each hospital in order to remove this bias. It would be inappropriate to compare these data across hospitals, however. These assessments resulted in the compilation of five variables summarising staff availability: staff ward round time per patient; mean clinical staff presence on the ward per patient; clinical staff allocation per patient; cleaning staff presence on the ward per patient; and toilet cleanliness.

Drugs provided and prescribed were assessed by a drug audit. Research assistants, who were medical students in the hospital concerned reviewed patient records and interviewed all patients in each ward or their carers, to determine the list of drugs which had been prescribed and whether or not each had been provided by the hospital or purchased outside. The purchasing of drugs outside the hospital is a common recourse in Zambia when drugs prescribed are not available in the hospital pharmacy. Although high-cost patients effectively purchase drugs whether from the hospital or from elsewhere, the distinction is important for low-cost patients for whom cost differences matter, and whose entitlements to hospital stocked pharmaceuticals might be affected by the presence of high-cost patients. In other settings, the delays, inconvenience and uncertainties associated with having to purchase pharmaceuticals and other supplies outside the hospital would be an important source of disutility to both groups of patients but these are likely to be of only minor relevance in these settings. This resulted in the compilation of three variables related to drug availability: drugs prescribed; drugs provided; and drugs purchased.

The originating ward or clinic (high cost or low cost) was identified for each of the last 200 laboratory tests or X-rays ordered. In the operating theatre all operations conducted in the last three months were classified according to the patient's ward of origin. This information is included in the routine records contained in each department. This produced four further variables: shares of high (low) cost patients in diagnostic tests, X-rays, major and minor operations which were compared to shares of high (low) cost patients in total hospital utilisation.

The number of patients present on each ward on the days of observation was also noted. All data collection in both hospitals took place in October 2001.

DATA ANALYSIS

The costing methodology enabled the direct comparison of high-cost and low-cost inpatient wards in revenue and costing terms without further assumptions; including comparative analysis of cost structure.

Clinical staffing pattern data were analysed by constructing an index of ward round time and observed presence for each ward. The index was constructed using weights on a positive vector in increasing seniority[ii]. Sensitivity to alternative constructions containing a range of weights chosen on the same basis was tested and results were not materially affected. The numbers of cleaning staff were analysed separately as a simple count per patient.

Drug data were analysed using a simple count of the number of items prescribed, provided and purchased elsewhere. The number of days in which the toilets were 'clean' or 'dirty' were added and compared.

Where appropriate, a two-sample Wilcoxon rank-sum (Mann-Whitney) test of difference in medians is used to test for statistical difference between high cost and low cost wards.

RESULTS

A. Cost and revenue analysis: KCH

Table 1 provides unit costs, the share of staff in unit costs, and non-staff costs by ward (high-cost) and ward type (low-cost).

A number of important points emerge. First cost differences ranged from 34% to 145% higher in high-cost than low-cost wards, and there is a significant difference between the proportions of total costs accounted for by staff costs in the two types of ward. In general in Zambia, staff costs have been better protected from cuts in resource allocation to hospitals than other costs with the result that it is complementary inputs which are most pressurised. These data suggest that high-cost patients are particularly advantaged in the allocation of these resources. Further analysis indicates that this is largely explained by a greater share of expenditure being allocated to food in high-cost wards, consistent with the argument that quality of care differences between the two services are associated with differences in hotel attributes. Nevertheless, the overall higher level of expenditure on high-cost services is reflected in higher unit expenditure on all categories of cost. The second component of the study aimed to explore in greater depth the likely clinical significance of differences in resource allocation patterns between high and low-cost services. Non-staff costs ranged from 43% to 229% higher in high-cost than low-cost wards.

Table 1 Cost and revenue by ward (1998, Kwacha)

	Unit cost	% cost = staff	Non-staff costs
High cost wards			
Medical and surgical, male			
Medical and surgical, female	31,469	56	13,846
Paediatrics			
Maternity	32,432	55	14,594
	24,346	49	12,416
	113,885	69	35,304
Low cost wards			
Medical	16,681	71	4,837
Surgical	11,579	65	4,052
Paediatrics	9,921	62	3,770
Maternity	85,162	71	24,697

B. Allocation of clinically relevant resources: KCH

Table 2 shows the proportion of tests, X rays and surgeries in the operating theatre consumed by low cost and high cost patients respectively in KCH. This should be compared with the high-cost share of all patients which, in the year up to September 1999, was 39% of outpatients and 10% of inpatients. Since diagnostic tests and X-rays are consumed both by inpatients and outpatients, levels of use by high-cost payers between 10 and 39% may indicate a fair allocation of resources between high-cost and low-cost patients. High-cost payers disproportionately use microbiology services. Haematology, biochemistry and X-ray services are not disproportionately allocated. The origin of patients using the operating theatre is exclusively inpatient wards. While major surgery appears to be allocated on a proportionate basis between high-cost and low-cost patients, minor surgery appears disproportionately accessed by high-cost patients. To the extent that major and minor surgery map to emergency and elective surgery, these results are anticipated by evidence provided by one hospital director who described the practice of prioritising high-cost patients in the waiting list for elective surgery.

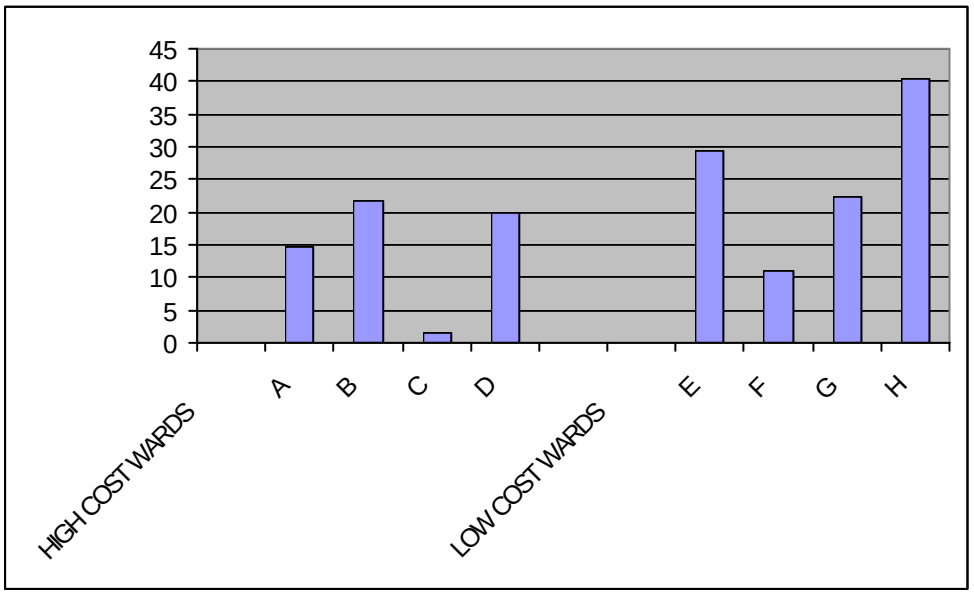
Table 2. Diagnostic tests, X-rays and surgeries by patient payment type, KCH

	LOW COST	HIGH COST	% HIGH COST
Laboratory Audit			
Micro-Biology	120	80	40
Haematology	145	55	27.5
Biochemistry	147	53	26.5
X-Ray	162	38	19
Theatre			
Major	157	17	9.8
Minor	219	39	15.1

The numbers of drugs prescribed and received per patient are higher for high-cost patients in each paired comparison (male and female medical and surgical are jointly compared with the two medical wards; the two paediatric wards and the two maternity wards are compared).

The average patient in high cost wards was prescribed 7 items and received 5 (see Figure 1). In low cost wards the average patient was prescribed 5 items and received 3. These differences are both highly significant ($p<0.01$). These comparisons imply that high-cost patients are prescribed more drugs than low-cost patients, and to an even greater extent, are more likely to be provided with those drugs which are prescribed. While paediatric patients provide an exception as far as rate of purchase from those prescribed is concerned, low cost paediatric patients are still prescribed and receive less drugs than high cost ones.

Figure 1 % Items purchased from the list of drugs prescribed, KCH



High-cost wards are distinctly better staffed on a per-patient basis and ward rounds allocate more time to each individual high-cost patient (see Figures 2 and 3). This is true for each paired comparison but only marginally so for observed staff attendance in the paediatric wards. The indices render the data difficult to interpret directly in terms of absolute staffing levels. As a point of comparison, in the two least served low-cost wards (medical and paediatric) the staff observed ranged from 2-3 nurses on each occasion and on one occasion, a junior doctor, to cope with wards of 20-25 patients. In the same two wards, ward rounds ranged from 1-2 hours (approximately 3-6 minutes per patient) and were conducted by a junior doctor and an enrolled nurse. In contrast, on the high-cost maternity ward, similar staffing levels and ward round lengths coped with only 7-8 patients.

Figure 2 Index of clinical staff presence per patient on wards, KCH

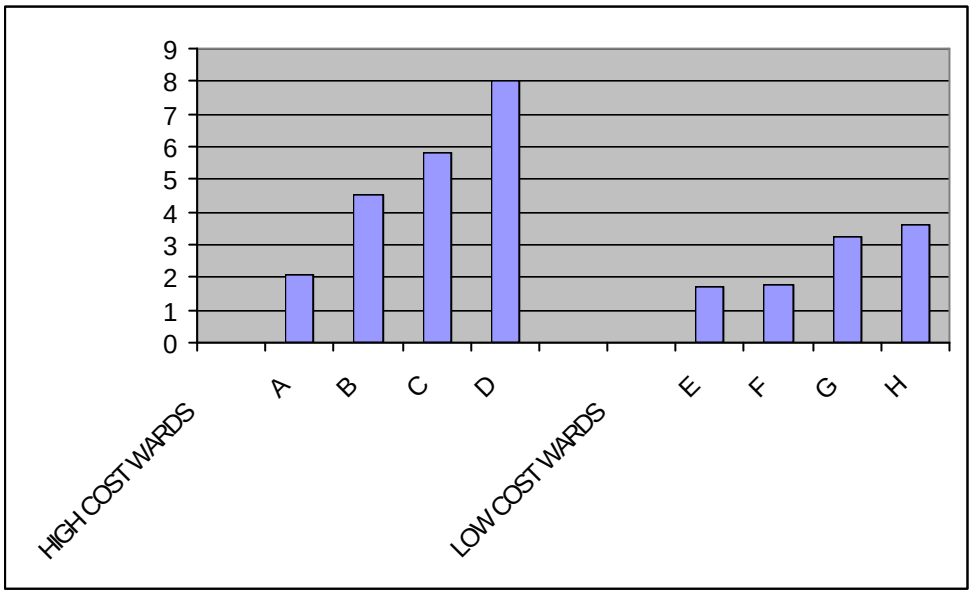
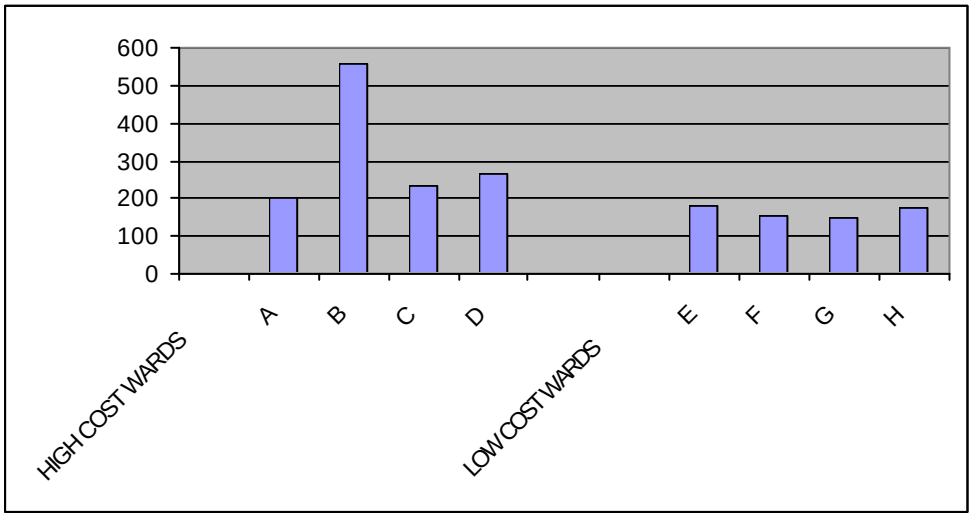


Figure 3 Index of staff time spent on ward rounds per patient, KCH



The index of staff assigned to each ward shows a similar pattern although one low cost medical ward had similar assigned staffing per patient to the two high-cost medical and surgical wards. However, there is a higher ratio of assigned staff to observed staff for the low-cost than high-cost wards (1.4 and 1.1 respectively). This suggests that at the level of the hospital, there is a preferential allocation towards the staffing of high-cost wards. This then appears to be further inflated by the behaviour of the staff, who appear to be more likely to attend to their duties conscientiously if assigned to high cost wards, carrying out more thorough ward rounds and attending on the wards more regularly.

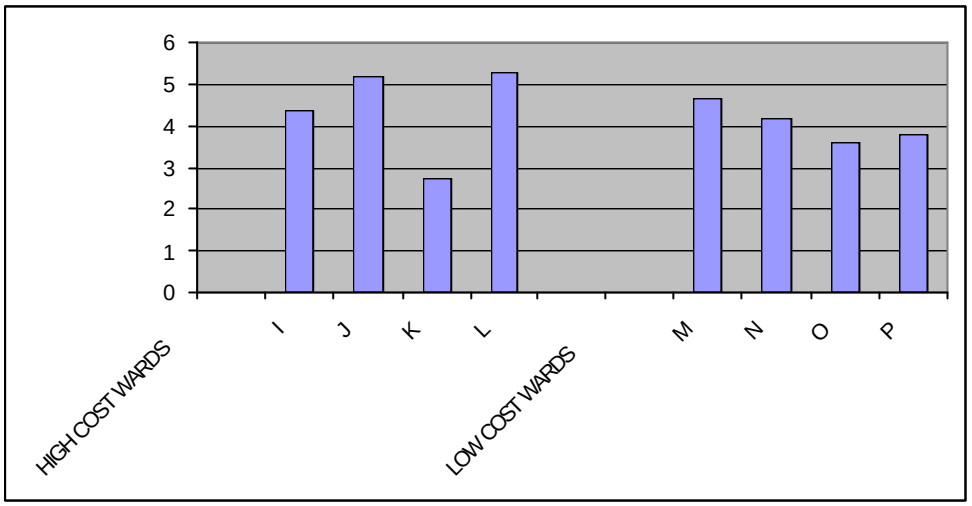
While toilets in the low cost wards were usually dirty, (60% of observations) except in the maternity ward (where 80% of observations were 'clean'), those in the high cost wards were almost always clean (95% of observations). A mean ratio of 5 patients to 1 cleaner was observed across the high-cost wards, and of 13:1 across the low-cost wards (p=0.02). The underlying explanation appears to be that one cleaner is usually on duty on each ward irrespective of its patient load. High-cost wards have fewer patients than low-cost ones.

C. University Teaching Hospital

At UTH, approximately 6% of inpatients were admitted to high-cost wards, and 7% of outpatients to high-cost clinics between January and September 2001. Among the diagnostic services, only use of X-ray and microbiology laboratory services seems to reflect that rate of use, with high-cost patients accounting for only 1.5 to 2.5% of the use of haematology, parasitology and histopathology laboratory services. However, high-cost patients disproportionately occupied surgical theatre time, accounting for 12.3 and 15% of major and minor surgical operations respectively.

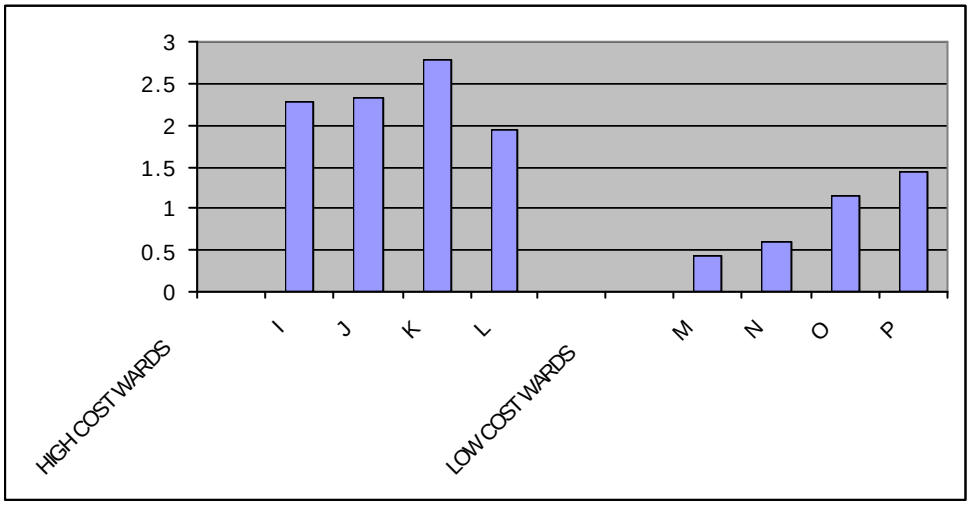
As at KCH, patients at UTH were more likely to be asked to purchase drugs prescribed if they were admitted to low-cost rather than high-cost wards (p=0.09). However, differences in the total number of prescriptions and number of drugs received were neither significant or consistently of the expected direction (see Figure 4). Patients in the high cost maternity ward received the lowest number of prescribed drugs.

Figure 4 Total items received per patient, UTH



Analysis of the presence of staff on wards indicated similar patterns to those of KCH. Both clinical and cleaning staffing levels were higher on a per-patient basis on high-cost wards ($p=0.02$ and $p=0.08$ respectively). As in KCH, there are considerable differentials between staff assigned and staff present, however in UTH these relative differences are similar for high cost and low cost wards (Figure 5).

Figure 5 Index of clinical staff presence per patient on wards, UTH



In other words, there is no suggestion that individual staff behaviour further biases the distribution. This is slightly surprising given that the gap between present and allocated staff is considerably greater at UTH than at KCH (ratios of approximately 2.8 for both types of ward at UTH but of 1.1 and 1.4 for high-cost and low-cost wards at KCH respectively).

In contrast to KCH, more ward round time is allocated per patient in low-cost than high-cost wards, although this difference is not statistically significant. The comparison is disproportionately affected by an outlying mean ward round time per patient in the low-cost maternity ward. Further examination of the data shows this to have been determined by one extremely long ward round involving a senior and junior doctor for 6 hours which occurred because the ward round was disrupted by a difficult labour case. Nevertheless, in the absence of this distortion, ward round time appears to be equitably distributed between high and low cost wards in UTH.

Finally, the pattern of toilet cleanliness at UTH is almost identical to that at KCH with toilets consistently 'clean' in the high-cost wards (100% of observations) and rarely so (25% of observations) in the low-cost wards.

Table 3 summarises the findings from the analysis in the two hospitals. Where there are disproportionate uses of resources identified, these consistently favour high-cost patients.

Table 3 Summary of findings

	KCH	UTH
Diagnostic tests:		
Lab	Excess use of microbiology by high cost	No evidence of inequitable use
X-ray	No evidence of inequitable use	No evidence of inequitable use
Theatre:		
Major	No evidence of inequitable use	Excess use of both major and minor by high cost patients
Minor	Excess use by high cost	
Drugs:		
Number prescribed	High cost prescribed more***	Low cost purchase more*
Percent purchased	High cost receive more***	No difference in number prescribed or percent received.
Percent received	Low cost purchase more**	
Staff:		
Ward presence	More staff in high cost *	More staff in high cost*
Ward round time	Longer ward rounds in high cost**	No difference in ward round time.
Ward allocation	More staff assigned to high cost	Staff assignment not further biasing allocation
Cleaning staff	More cleaning staff in high cost**	More cleaning staff in high cost***
Cleanliness	High cost usually clean.	High cost usually clean.
	Low cost usually dirty.	Low cost usually dirty.

DISCUSSION

Cost comparison

On one interpretation of the current policy, high cost patients are entitled to receive services classified as included in the essential package on a subsidised basis and should then fully cover the costs of services classified as not included. Absence of reliable revenue data prevents analysis of whether this version of the policy is achieved. However, there is a difficulty of defining the 'package' as whatever low-cost patients actually receive and the rest as the additional services received by high-cost patients, as this interpretation of the policy implicitly does. Where there is a clear principle determining what services are core and what services are additional, rights of low-cost patients are clear (though admittedly difficult to enforce). Once clinical service variation between the two sides is admitted, the definition of the package, and hence the allocation of resources, become manipulable according to the incentives facing hospital decision makers. These incentives are more fully explored by McPake et al.[8] . The models developed in that paper suggest that the progressive erosion of standards of care for low-cost patients is a likely outcome of several possible scenarios. The point made here is that such development is unimpeded by this version of the financing policy.

Comparison of costs suggests that high cost patients have better access to hospital inputs. This comparison is subject to some caveats however. Ward types are not exactly analogous. The comparison assumes for example that the case mix in male and female medical and surgical high-cost wards is broadly similar to that in the medical and surgical low-cost wards. If this is not the case, it is conceivable that the essential package component of the care received in high-cost wards is more expensive than the average cost of care in low-cost wards.

However, we believe that this explanation of the gap between unit costs on the two sides is implausible. Comparison of the data for maternity wards is informative. Anecdotally, those able to pay may choose KCH's high-cost service for their normal delivery whereas those arriving at KCH's low-cost maternity ward are far more likely to be referred in the case of complications. This would imply a higher cost of the essential package component of the low-cost delivery service. The data presented in table 1 are consistent with that scenario. The low-cost unit cost is 75% of the high-cost unit cost for the maternity service whereas for other wards the range is between 36 and 53%. This suggests that the unit cost difference between

high and low cost maternity service users under-states rather than over-states the services high cost patients receive in excess of what might be considered their 'essential package'.

The clinical significance of resource allocation patterns

We have not compared case mix and need across the two groups of patients, and it is possible that, for example, disproportionate use of theatre by high cost patients is explained by a high-cost case mix biased towards surgical cases.

However, since instances of disproportionate use arise in each of the areas addressed, a general explanation that high-cost patients are more severely ill or requiring greater treatment intensity in all respects, rather than disproportionately represented in particular conditions would be necessary to support this alternative interpretation. This explanation is questionable on a number of bases. First, a socio-economic gradient in use of high-cost services is highly likely and there is an association between socio-economic status and severity of illness. Second, it is likely that there is a large proportion of the population for whom high-cost services are unaffordable, whatever their condition. Third, high-cost fees increase in illness severity and treatment intensity whereas low-cost fees are relatively invariable [20]. In general, use of higher priced options in health service markets decline as cost increases. Use of the private sector is concentrated in primary care for example. Fourth, evidence of difference in toilet cleanliness and in the proportions of drugs received once prescribed, are independent of need or case mix.

Although in a very different context, Nolan and Wiley [22] compare the case mix of users of public hospitals in Ireland between those entitled to fully subsidised services, a means-tested benefit, and those not, about two thirds of whom are users of 'private' services in public hospitals. Adults of working age account for a larger share of the latter group and this predicts a number of diagnostic group divergences, including a greater share of obstetric and gynaecological conditions, and a higher prevalence of surgical procedures. Overall, the conditions with which those who are not entitled to full subsidy present were associated with a slightly higher resource intensity per bed-day, but lower resource intensity per admission, calculated using a diagnostic related group system used in Ireland. A key difference between the Irish and Zambian contexts is that private services are accessed via insurance which would be expected to reduce or eliminate the extent to which their use would decline as service cost increased.

We conclude from this data that the pattern of allocation of clinically significant resources at ward level to those patients paying high-cost fees is consistent with bias in their favour. The resources implicated in this allocation suggest differences in the clinical quality of care received by patients in the two groups. This is contrary to the intention of the introduction of the two-tier fee system.

Preferential allocation – underlying processes

The most obvious explanation for the patterns of resource allocation and revenue generation observed is an over-enthusiasm among hospital managers for the development of a high-cost service irrespective of its overall financial contribution. This is understandable from a number of standpoints. First, high-cost services allow health professionals to work in an environment where they have greater opportunity to practice their medical skills with the resources they require. Second, high-cost services provide a service option for middle class Zambians who include the health workforce. Hospital staff may even have free access to high-cost services. Third, even if high-cost services do not generate a surplus overall, they decentralise control over resources to the hospital level. Hospital managers have little control over staff for example, but do control high-cost service revenue which staff time can be used to generate. Finally, the availability of additional discretionary cash from high cost fees can make a substantial contribution to the smooth operation of the hospital. Funds available at short notice to enable purchase of essential medical supplies or to repair a water-pump are highly valued, and can benefit patients on both 'sides' of the hospital.

However, there are also other processes which may be influencing the allocation of resources. Low-cost ward patients may refuse certain services on the basis of inability to pay the cost-sharing fees. They might, for example, refuse diagnostic tests or more expensive treatment options. This has direct implications for the distribution of costs, and for clinically relevant resource allocation. It may also have indirect implications. Active disease management involves continuous response to diagnostic test information and to observed patient reactions to treatments initiated. If tests and treatments are rejected, opportunities to improve clinical management will be missed.

Finally, pre-payment is a common means of covering high-cost bills. KCH staff informed us that when faced by a patient who has pre-paid, it is more difficult to tell them to purchase

drugs elsewhere, whereas a patient who has not yet paid is considered to have less entitlement.

All these are understandable processes through which resources are preferentially allocated to high-cost patients. However, they are also processes through which inequalities may be entrenched, and by which low-cost patients can become losers from the development of the high-cost service.

The pattern of bias is less consistent at UTH than at KCH. UTH has been slower to develop its high-cost services, as indicated by its lower proportion of high-cost patients. The data are not helpful in distinguishing the extent to which this is a demand or supply driven contrast. Low demand for high-cost services at UTH may follow from less consistent resource allocation favouring high-cost patients, or less interest in developing the high-cost service on the part of hospital management. Certainly, active pursuit of the high-cost market is an acknowledged strategy of KCH management whereas UTH management is less convinced of the potential of this policy to contribute to the hospital's goals.

The explanation may also be found in exogenous factors. The private hospital market has been developing in Lusaka and some employers who initially registered their employees for high cost services at UTH have shifted to private providers. Elements of instability at UTH, which has been most severely affected by strike activity (see below) may also have played a role. The larger gap between allocated and observed presence of staff on wards at UTH may reflect either greater difficulty in managing staff or greater inefficiency. These factors combine to suggest that UTH may have found it more difficult to develop the high cost service given its local circumstances.

CONCLUSION AND POLICY IMPLICATIONS

Better understanding of a number of factors in the hospital's operating environment would enable the development of policy recommendations. We identify two important areas. First, an understanding of hospital demand patterns and determinants would enable greater clarity as to the policy options within two-tier pricing which would enable more equitable outcomes. For example, the version of the financing policy which allows high-cost users to be subsidised to the extent of the value of the essential package would not harm low cost patients in an absolute sense if these users would, in the absence of the high-cost service use the low-cost one. Better understanding of the demand for hospital services in these markets is another activity of the project from which this paper reports[19], and will enable a more informed assessment of the assumption that high-cost patients would otherwise use low-cost public hospital services. Such understanding would also inform decisions about pricing structures, monitoring and negotiating the use of the CBoH subsidy and regulatory strategies such as requiring minimum quality of care standards for low-cost services.

Second, policy options would be informed by a better understanding of the factors influencing human resource retention, motivation and performance. It is clear that the high-cost service provides motivation for health staff. As has been suggested, it offers a more acceptable working environment, recourse in the event of illness of family members, and also additional remuneration. Given the shortage of human resources, low-cost patients may indirectly be advantaged by the presence of the high-cost service if a larger and better motivated hospital workforce is the outcome. However if opportunities to work in the high-cost service are perceived to be unfairly distributed, the effect may be demotivating for staff providing low-cost services, or if staff absent themselves from low-cost service to 'moonlight' in the high-cost service, the effect may be increased staff shortages in low-cost wards. Such insight would improve understanding of the current situation in addition to informing policy detail such as the process for allocating high-cost employment opportunities.

Two clear conclusions emerge from our analysis. First, a version of the financing policy which defines whatever low-cost patients receive as the 'essential package' and allows high-cost patients to be subsidised to that extent fails to protect standards for low cost patients in an absolute sense. Second, the apparent difference in clinical quality of care between high-cost and low-cost patients contradicts the stated intention of the two-tier fee policy.

REFERENCES

1. Central Board of Health, 1998. National Health Strategic Plan. Lusaka: Government of the Republic of Zambia.
2. Government of the Republic of Zambia. Financial reports 1995-7. Lusaka: Government of the Republic of Zambia, 1996-8.
3. Purvis G, 1997. University Teaching Hospital in Zambia: The Strategic Plan Environment. Bethesda, MD: Partnership for Health Reform.
4. Haddon B, 1998. Kitwe-Ndola Hospital Management Project, Strategic Review of the KANDO Project. Lusaka.
5. Bukonda H, et al. 2000. Setting up a national hospital accreditation program: The Zambian Experience. Bethesda, MD: The Quality Assurance Project (QAP).
6. Kamwanga J, et al. 1999. Autonomous hospitals in Zambia and the equity implications of the market for hospital services, Phase 1 report: A description of the current state of hospital autonomy policy. London: London School of Hygiene and Tropical Medicine, 1999.
7. Hanson K, et al. *Towards improving hospital performance in Uganda and Zambia: Reflections and opportunities for autonomy.* Health Policy, in press.
8. McPake B, K Hanson and C Adam. 2002. *A model of the equity implications of hospital reforms in Zambia*.
9. Shennan J, 1979. *Pay beds.* British Medical Journal; **2**(6182): p.136.
10. Morris P, 1979. *What COHSE should remember about pay beds.* Nursing Mirror, **148**(24): p. 6.
11. Bolt DE, 1979. *Saving pay beds.* British Medical Journal: p. 895-6.
12. Davidson N, 1979. *Long journey to the abolition of pay beds.* Health and Social Service Journal, **89**(4632): p. 256-8.
13. Klein R, 1979. *Ideology, class and the National Health Service.* Journal of Health, Policy and Law; **4**(3): p. 464-90.
14. National Economic Research Associates (NERA), *Healthcare report. Are pay beds profitable?* 1995, Norwich Union Healthcare: Eastleigh.
15. Yates J, 1987. *Why are we waiting? An analysis of hospital waiting lists.* Oxford: Oxford University Press.
16. McAvinchey ID and Yannopoulos A, 1993. *Elasticity estimates from a dynamic model of interrelated demands for private and public acute health care.* Journal of Health Economics; **12**(2): p. 171-86.

17. Suwandono A, et al. 2001. *Cost recovery beds in public hospitals in Indonesia.* Health Policy and Planning, **16**(Supplement): p. 10-18.
18. Carmel S and J Halevy, 1999. *Patient satisfaction and hospital services evaluation: comparison of public and private patients.* Harefuah; **137**(9): p. 363-70.
19. Hanson K, et al. 2002. *Preferences for hospital quality in Zambia: Results from a discrete choice experiment*.
20. Nakamba P, Hanson K, and McPake B. *Markets for hospital services in Zambia.* International Journal of Health Planning and Management, 17(3): 229-247.
21. Barnum H and Kutzin J, 1993. *Public hospitals in developing countries: resource use, cost, financing*: Published for the World Bank, The Johns Hopkins University Press, Baltimore and London.
22. Nolan B and Wiley MM. 2000. Private practice in Irish public hospitals, Economic and Social Research Institute, Oak Tree Press, Dublin.
23. Shirom A. 2001. Private medical services in acute-care hospitals in Israel. *International Journal of Health Planning and Management* **16**: 325-345.

[i] This is a clinical health worker who is trained to an intermediate level of health worker between nurse and doctor, equivalent to a 'medical assistant' in many other developing countries

[ii] The index used in the data presented weighted as follows: Consultant=10, other doctor=7, clinical officer=5, registered nurse=4 and enrolled nurse=2.