The impact of micro health insurance on Rwandan health centre costs

Pia Schneider1,2* and Kara Hanson1

Accepted 3 August 2006

While the implications of payment mechanisms for provider behaviour and cost have been amply explored in industrialized countries, there is little empirical evidence from developing countries. This study exploits the opportunities created by a pilot study of micro health insurance with capitation in Rwanda to address this issue. Using cross-sectional data collected in 52 health centres, the paper employs an econometric cost function with payer-specific outputs to assess the cost impact of two provider payment mechanisms: (1) user fees for care paid by the uninsured, and (2) capitation payment paid by informal insurance schemes for the insured. The cost function allows payer-specific marginal and average costs and scale measures to be calculated. Findings point to significant differences in cost between the two payment forms. These may be due to the incentives embodied in the capitation provider payment or the less severe case-mix among insured patients arising from improved access to care for this group, or both. For both payment types there are important short-run economies of scale, which could be exploited through more intensive use of idle resources in health centres.

Keywords Cost in health centres, provider payment, insurance in low-income areas

INTRODUCTION

While the implications of payment mechanisms for provider behaviour and cost have been amply explored in industrialized countries, there is little empirical evidence from developing countries. This study examines this issue based on information collected during a pilot study of micro health insurance (MHI) in three districts of Rwanda. MHI schemes are informally organized risk-sharing groups with voluntary enrolment, the objective of which is to improve members’ access to care by lowering the out-of-pocket (OOP) price at the time of purchase (Hurley 2000). Research on the performance of MHI is limited (Ekman 2004), and mainly focused on the organizational and financial sustainability of MHI (ILO 2002; Bennett 2004); the determinants of socio-demographic and socio-economic factors in insurance enrolment and care seeking behaviour (Schneider and Diop 2001); horizontal inequity in service use (Schneider and Hanson 2006); and the link between enrolment and quality of care (Criel and Waelkens 2003) and trust (Schneider 2005). To date, there has been no attempt to quantify and compare the impact of insurance with capitation payment versus user fees on cost in health facilities in a low-income context, hence the value of this paper. Using a cross-sectional data set collected in Rwandan health centres, it compares the effect on provider cost of two payment mechanisms: (1) user fees for drugs and services paid by the uninsured, and (2) capitation payment paid by MHI for the insured.

KEY MESSAGES
- Providers involved in a micro-insurance scheme with capitation payment in Rwanda adjusted the resources used to treat patients depending on the patient’s insurance status.
- Lower marginal cost among insured patients arose from a combination of payment mechanism (capitation payment for insured patients) and less severe casemix among the insured (due to improved access to care).
Provider Payment and Cost Impact in Health Centres

Rwanda is amongst the poorest countries, with a per capita GDP of US$250 in 2003. Patients have paid user fees for medical services and drugs to public providers since 1976 (Shepard et al. 1993). From 1996 to 1999, utilization of primary care services dropped from 0.3 to 0.25 visits per capita per year, resulting in idle resources in health centres. Aiming to improve access to care for the poor and raise additional financial resources for health, the Rwandan government, in collaboration with the local population and the technical assistance of the Partnerships for Health Reform (PHR) project, developed and implemented 54 MHI schemes with voluntary enrolment in three rural districts. Each MHI signed a contract with one of the 54 health centres; provider payment is through capitation per member per month together with a low co-payment (Schneider 2005). At the end of the first operational year, 88,303 individuals had enrolled in MHI, representing about 9% of the districts’ 1 million population.

Little evidence exists about the link between insurance, payment mechanisms, provider behaviour and cost structure in health facilities in low-income countries (Barnum and Kutzin 1993; Yip and Eggleton 2001). Three behavioural incentives caused by insurance, namely adverse selection, moral hazard and supply-side induced demand, may increase treatment cost and endanger the financial situation of the insurance fund (Hurley 2000). Of the three, adverse selection was of most concern in Rwanda. Voluntary enrolment may lead to adverse selection resulting in small high-risk insurance pools with predominantly those with greater expected health-care need choosing to insure (Cutler and Zeckhauser 2000). At the end of the first year, enrolment in the 54 MHI schemes varied widely (from 1 to 55% of target population). Health centres partnering with small MHI reported considerably higher use levels (up to 3 visits per member per year), suggesting that adverse selection is a risk with low enrolment and that higher treatment costs for insured patients may be expected in health centres with small pools.

On average, insured individuals reported markedly higher visit rates than the uninsured (about 1.3 versus 0.2 visits per capita per year in health centres, respectively). The higher visit rate for the insured is not indicative of frivolous service use or supply-side induced demand. Rather, MHI seems to improve access to care, with the insured using care based on their need and independent of their socio-economic situation, whereas among the sick uninsured, visits strongly correlate with individuals’ socio-economic background (Schneider and Hanson 2006).

One way the MHI can respond to adverse selection and supply-side induced demand is to shift part of the insurance financial risk to providers through the provider payment mechanism. Capitation payment to providers imposes the full insurance risk on providers and discourages them from oversupplying care (Ellis and McGuire 1993). Capitation sets an incentive to produce efficiently (Cutler et al. 2000) by adjusting treatment intensity within a medically acceptable quality range (Dor and Farley 1996). In the worst case, this treatment adjustment may lead to dumping and skimping of insured patients, resulting in lower cost caused by medically inadequate quality levels (Ellis and McGuire 1996). Studies conducted in industrialized countries on provider behaviour suggest that providers respond to financial incentives embedded in the payment method and adjust treatment intensity within a clinically acceptable range to keep cost low (Dor and Farley 1996).

If providers have several revenue sources, they may try to shift costs across different financers (Newhouse 1996). If MHI with capitation payment does not cover providers’ medical expenditures for the insured, Rwandan health centres may increase prices for the uninsured, aiming to cross-subsidize from uninsured to insured services. While cost shifting will increase user fees and thereby limit access to care for the uninsured, it also sets an incentive to insure.

To assess the cost and efficiency impact of MHI with capitation in health centres, the Rwandan Ministry of Health (MOH) and the PHR project collected monthly routine data on individual patients’ service use, expenditures and finances in all health facilities for insured and uninsured patients in three districts during the first year MHI was available. This cross-sectional provider data permits unique analysis to be performed. The analysis employs an econometric cost function to compare the effect of user fees and MHI with capitation payment plus a small co-payment on provider cost and efficiency in health centres. Fayer-specific marginal and average costs are estimated. Scale measures are derived to identify resource capacity in health centres.

Two hypotheses are tested. First, a higher share of MHI revenue as a result of a larger MHI pool size leads to a decrease in total facility cost (insurance pool effect). This could result either through the payment system effect alone, through the effects of pool size on the strength of the incentives to alter payer-specific treatment intensity, or through the effects of mitigating the financial risk related to adverse selection and moral hazard in smaller risk pools. Secondly, capitation payment for insured patients leads to lower marginal cost (MC) and average cost (AC) per patient than fee-for-service payment for uninsured patients (payment system effect). This would arise through the provider response to the financial incentive created by the different payment systems, resulting in adjustment of treatment intensity. The data do not permit testing a third question, namely whether reduced treatment intensity for insured patients is within the medically acceptable range of quality care; or whether the lower cost brought about by adjustment of treatment leads to medically inadequate quality, arising from dumping and skimping of insured patients (Ellis and McGuire 1996).

Previous research

The goals and decision-making process of public and non-profit health care providers are not well described by the conventional profit maximization model of the competitive market. Generally, it is hypothesized that providers select a resource mix for treatment within the technically feasible set and subject to a budget constraint, which depends on health financing and provider payment (Barnum and Kutzin 1993). Although such profit-maximizing behaviour may affect quality of care, legal liability for medical malpractice and professional ethics may constrain providers from cutting quality (Weisbrod 1991). However, providers might still reduce
those dimensions of quality that are difficult for insurers and patients to observe (Ellis and McGuire 1996).

The few studies that have analyzed the behavioural and cost implications of different provider payment systems used data collected in the USA. Dor and Farley (1996) employed a generalized translog model and data from 331 US hospitals to estimate a multi-product cost function in which case-mix adjusted outputs by provider payment source are treated as distinct outputs. They aim to identify whether providers vary the resource intensity of interventions in response to insured patients' payment sources. Resource intensity is used to proxy quality. A profit-maximizing hospital with fixed prices is expected to select the quality level where the marginal effect on revenue of additional quality equals the marginal costs of producing quality. Results indicate that costs vary widely, which is systematically related to adjustment of treatment intensity given the expected insurance payment. The authors concluded that providers allocate resources on a case-by-case basis and within a clinically acceptable range. Compared with less generous payers, more generous insurance reimbursement serves to increase payer-specific resource intensity in treatment, causing marginal costs to increase (Dor and Farley 1996).

In low-income countries, providers' response to alternative payment systems has been subject to limited study. Using data from the Philippines Social Insurance, Gertler and Solon (2000) showed that hospitals with market power extract insurance benefits through price discrimination instead of passing benefits on to patients in the form of reduced out-of-pocket fees. Providers also price discriminate by using insurance status information to exploit different price elasticities among insured and uninsured patients (Gertler and Solon 2000). Results from a study on provider payment reform in Hainan Province in China indicate that the change from fee-for-service (FFS) to a prospective global budget system in hospitals led to a slower rate of growth of overall expenditures (Yip and Eggleston 2001). Another Chinese study found that provider payment reform needs to be accompanied by measures such as an essential drug list and regulation of average length of stay, readmission rates and quality of care, to limit provider response to undesirable incentives (Meng et al. 2004). Recommendations focus on institutional changes to enhance provider efficiency and reduce costs (Barnum and Kutzin 1993; Yip and Eggleston 2001). These findings suggest that providers respond to financial incentives, indicating that a range of technically feasible treatment styles exists, and pointing to room for cost saving in the provision of care.

In Rwanda, providers can influence their efficiency by improving the management of drug storage; through more flexible use of staff; by changing their treatment style to improve drug prescriptions; and by exploiting scale economics and treating more patients.

**Study context**

The study area covers three Rwandan districts (Byumba, Kabgayi and Kabutare) where the population has the choice of either enrolling in one of 54 MHI schemes or remaining uninsured. Households pay a fixed annual enrolment fee of RWF 2500 (US$7.8 in 2000) to the MHI affiliated with their 'preferred' health centre. The enrolment fee was set to cover all services and drugs provided in health centres and ambulance transport to the district hospital, where few additional services are covered (physician consultation, overnight stay and case-payment for Caesarean section) (Schneider and Hanson 2006). The annual fee was calculated on the basis of health centre recurrent costs, utilization rates, existing user fees, the target population, the desired benefit package and people's ability to pay.

MHI schemes pay a monthly capitation amount to partner health centres, based on the number of insured. Insured patients pay a co-payment of RWF 100 (US$0.30) per episode of illness in health centres and user fees for care not covered at the hospital. Uninsured patients continue to pay user fees for services and drugs in all facilities. On average an uninsured patient pays RWF 1987 per episode of illness, which is considerably more than insured patients (RWF 497).

Rwandan public and church-owned health centres operate formally on a non-profit basis. Their revenue comes from government, donors and the population. Donor and government subsidies contribute about 30% of total health care revenue in the form of salaries for nurses and drugs. Health centres depend on revenue from the population to pay for other operational costs. They retain and manage all revenue sources to reinvest part of it in covering operational costs, and to accrue bank savings. It is, therefore, plausible to assume that health centres select inputs subject to the budget constraint set by the anticipated revenue, and seek to minimize losses or cover costs.

**Data, model specification and estimation procedures**

**Assumptions**

The analysis is based on three assumptions: cost minimization, similar structural quality and that health centres are informed about their cost and revenue.

First, cost-minimizing behaviour implies that health centres choose input factors at their cost-minimizing levels to produce care at the minimum possible cost (Begg et al. 2000). The assumption is that health centres have an incentive to minimize cost under capitation payment (Kass 1987). Deviations from cost-minimizing behaviour are assumed not to influence total cost in a systematic way (Wouters 1993).

Secondly, in the absence of data on quality, it is assumed that health centres provide care to all patients within a similar structural quality range. This assumption follows from household survey results where the same proportion (68%) among insured and uninsured patients said that the health centre always had drugs available (Schneider and Diop 2001).

Thirdly, a major assumption is that providers are aware of the cost implications of different treatment styles, know their structural quality range. This assumption follows from household survey results where the same proportion (68%) among insured and uninsured patients said that the health centre always had drugs available (Schneider and Diop 2001).
The cost model

The analytical framework closely follows Dor and Farley (1996), although the present setting contrasts with the US situation in which insurers, not patients, made fee-for-service payments to providers. The health centre is the unit of analysis. Health centres are multi-product firms with nurses providing basic care to insured and uninsured patients. Total costs (TC) rise with the total volume of services provided on a payer-specific basis (Dor and Farley 1996), which is:

\[
TC = c(X_i M_i Q_i, \ldots) \quad (1)
\]

where \(X_i\) is the number of services provided to patients by payment source \((i)\), \(M_i\) represents the relative case-mix adjustment index for patients’ severity of illness, and \(Q_i\) is a quality index reflecting the payer-specific treatment intensity within a medically acceptable range.

Dor and Farley (1996) had a larger sample size and data on patients’ severity of illness. Consequently, they used a translog cost function with risk-adjusted service use. Due to the small number of observations of 52 health centres and to retain sufficient degrees of freedom, this analysis employs a simpler specification. Total cost enters the equation in its logarithmic form to reduce problems of heteroskedasticity in error terms. The right-hand side includes the logarithm of outputs but no higher order terms, which allows other parameters to be included. To examine the payment-specific effect on total health-centre cost, curative consultations (not severity-adjusted) are distinguished by payment source:

\[
\ln(TC) = \alpha + \beta_1 \ln(CM) + \beta_2 \ln(CNM) \\
+ \beta_3 \ln(PREV) + \beta_4 \ln(PrMrev) \\
+ \beta_5 \ln(PrNMrev) + \beta_6 O + \varepsilon \quad (2)
\]

The variables are described in Table 1. \(\varepsilon\) is a random disturbance term. Factor prices are excluded. They are fixed by the MOH for staff and drugs, and do not vary across facilities.

Parameter estimates are obtained using ordinary least square (OLS) regression. The respective elasticities show the percentage increase in total costs for a 1% increase in an output, ceteris paribus. Modifications of this model were estimated to assess its robustness.

The cost model tests the first hypothesis, that a higher share of MHI revenue leads to a decrease in total facility cost; and it examines the payer-specific cost impact for insured and uninsured curative care consultations. Provider behaviour as predicted by this model depends upon the shape of the cost function. According to Dor and Farley (1996), providers discriminate patients’ treatment intensity by payment source depending on their output level, resulting in different payer-specific marginal costs across output levels. To test the second hypothesis, more detailed cost-structure analysis is needed to examine scale effects, and whether patients receive treatment of the same resource intensity in health centres with different utilization levels (Dor and Farley 1996).

Cost structure analysis

To identify scale effects on payer-specific costs, the health centre cost structure is examined across different patient output levels. Product-specific marginal cost and average cost (AC) values are compared by payment source (Dor and Farley 1996). All health centres are ranked according to their total number of consultations and divided into three groups with low, medium and high consultation output levels. For each group, the following variables are calculated at their respective mean value: the number of insured consultations (CM), the number of uninsured consultations (CNM), MHI pool size, payer-specific average and marginal costs per consultation, and the total number of curative consultations per staff member per year.

Payer-specific AC per curative visit are examined by dividing payer-specific total costs \(C_i\) by the total number of payer-specific curative consultations \(y_i\) (Kass 1987).
Following equation (1), the first order condition for expected cost minimization with respect to risk-adjusted (M) output $X$ implies that providers select a quality of care level $Q_i$ where the marginal cost (MC) of producing quality equals the marginal effect on revenue from additional quality. It results in payer-specific marginal costs at a given quality level $Q_i$ (Dor and Farley 1996):

$$MC_i = \frac{\partial TC_i}{\partial (M_i X_i)} = \frac{c'(Q)}{Q_i}$$

(3)

If providers are able to differentiate treatment across payment source, then they will select payer-specific resource intensity, a proxy for quality that equates payer-specific MC and marginal revenue (MR). Differences in payer-specific quality $Q_i$ would imply payer-specific resource intensities resulting in different MC; while equal quality levels ($Q_{MHI} = Q_{user fees}$) would suggest no treatment discrimination across payment sources and different MC would reflect different case-mix (M) (Dor and Farley 1996).

The cost structure analysis estimates payer-specific MC(i) for consultations as a linear combination of payer-specific elasticity estimates from the log-linear regression shown in equation (2) and their respective AC(i) (Dor and Farley 1996). Formally, this is:

$$MC(i) = \beta(i)(C(i)/y(i))$$

(4)

where (i) = insured or uninsured patients; $\beta(i)$ = elasticities of cost with respect to insured or uninsured consultations; $C(i)$ = payer-specific total cost; and $y(i)$ = payer-specific number of curative consultations in health centres.

Short-run economies of scale measure the effect on costs of a proportional increase in all inputs on the output of the $i^{th}$ product while the level of output of all other products remains constant (Barnum and Kutzin 1993). The analysis examines short-run economies of scale to assess health centre capacity by comparing the percentage changes in payer-specific MC of consultations across health centre output levels (Kass 1987; Dor and Farley 1996). Additional insight is sought on health centre technical efficiency by comparing the total number of curative consultations per staff member per year across the three health centre output groups. F-tests examine the null hypotheses of no significant difference between payer-specific MC of visits and across health centre output levels.

**Data and variables**

Table 1 contains definitions and descriptive statistics for the variables used. The analysis uses data collected monthly from 52 of a total of 54 health centres during the first year in which MHI was operating (August 1999 to July 2000). Two health centres are excluded due to incomplete data. Information was collected from patient registers and health centre accounting reports on facility expenditures and payment-specific revenues, staffing patterns, drugs and services (all consultations, deliveries, laboratory tests) provided to insured and uninsured patients. The total number of services used by each patient was registered and summarized for each group. Drug costs were calculated based on the price paid and quantity bought by the health centre at the district pharmacy, and drug donations to the health centre were valued based on district pharmacy prices. Detailed patient drug prescription records for insured and uninsured patients served to summarize the total value of drugs purchased by the health centre and then sold to insured and to uninsured patients. Overhead costs include staff salaries and other expenses (electricity, water, transport etc.). Overhead costs were allocated to insured and uninsured patients in proportion to their respective total service use. Total annual payer-specific costs of treating insured and uninsured patients consist of treatment and overhead costs (Barnum and Kutzin 1993). Total annual recurrent costs (TC) were calculated using an accounting-based step-down approach. TC is the sum of the two payer-specific total costs $C(i)$.

The inpatient variable (INPAT) was computed by aggregating health centre admissions and deliveries into one inpatient variable. On average, health centres have five beds.

The size of the MHI pool is a proxy to identify whether smaller MHI pools suffer from adverse selection leading to higher health centre costs. As the absolute MHI pool size is correlated with the number of insured consultations, the model includes measures of the percentage of total health centre revenue from insured (PrMrev) and uninsured patients (PrNMrev) (Dranove 1998). PrMrev includes monthly capitation payments and insured patients’ co-payments paid to health centres. It is expected that under capitation the insurance financial risk shifted to providers becomes less of an issue with growing MHI pool size, implying a negative association between TC and PrMrev.

The variable PrNMrev is used to examine whether health centres have fewer incentives to minimize costs with increasing user fee revenue where fees are higher than marginal costs; a positive coefficient would be anticipated (Kass 1987).

The ownership dummy (O) helps to inform whether there are differences between public and church-owned health centres in overhead costs.

**Results**

The analytical strategy employed is based on equations (1) and (3). First, on a health centre level, the factors that affect health centre total recurrent cost are evaluated. Secondly, the cost structure analysis examines the extent to which payer-specific utilization of services at different output levels affects health centre marginal and average cost.

**Cost function**

Table 2 presents results. The model behaves well with an R-squared of 0.79. In support of the first hypothesis, health centre cost seems to be significantly lower as the proportion of revenue from MHI capitation payment (PrMrev) increases (insurance pool effect). The elasticities of cost with respect to insured (lnCM) and uninsured consultations (lnCNM) are likewise positive and significant and both are less than 1, with no significant difference between the two payers ($P = 0.673$). Other explanatory variables, namely preventive care visits (lnPREV), health centre ownership (O) and the share of user fee revenues (PrNMrev), are not significantly related to cost, suggesting that user fee revenues do not appear to set any
incentives to providers that would affect cost. The Breusch-Pagan and the Ramsey RESET test are insignificant, implying that the hypotheses of homoskedasticity and correct model specification cannot be rejected (Kennedy 1997).

These results are relevant in a resource-constrained health system. They suggest that an increase in the number of insured or uninsured visits affects health centre costs similarly, whereas health centres partnering with larger MHI pools report lower total costs. Under capitation payment, this may be a direct result of providers mitigating their financial risk related to insured patients (adverse selection and moral hazard) over a larger MHI membership pool. It may also be related to effects of economies of scale caused by higher utilization by insured patients, which is further examined in the cost structure analysis.

Three alternative specifications of the model were estimated to examine the robustness of the model (results not shown). The first and second versions add the number of inpatient admissions. Version 2 excludes facilities with zero-values for inpatient care. Version 3 excludes the inpatient variable and adds two interaction terms between the ownership dummy and number of insured/uninsured consultations. All three alternative specifications produced similar parameter estimates, and comparable standard errors and significance levels. None of the results were superior to the model in Table 2 with respect to significance levels and test statistics; thus, the model in Table 2 is the most parsimonious model.

Cost structure measures

The cost structure analysis examines payer-specific marginal cost and average cost per visit, and scale economies at different output levels, to evaluate resource intensity of treatment and idle resources in health centres (Dor and Farley 1996). Results are used to test the second hypothesis that insurance with capitation payment leads to lower marginal cost and average cost per insured patient than fee-for service payment by uninsured patients (payment system effect). Cost measures are computed based on above parameter estimates (Table 2) and mean values of respective outputs, following the methodology applied by Kass (1987), Dranove (1998), and Dor and Farley (1996). Table 3 presents results for the three health centre groups ranked on the basis of total consultations.

In all three groups, MC and AC for visits by uninsured patients (CNM) are significantly higher than for insured patients (CM) \((P < 0.001)\). Capitation payment for insured patients leads to lower marginal cost per patient than user fees for uninsured patients (payment effect). The observed pattern of payer-specific MC and AC is consistent with the hypothesis that providers respond to the financial incentives set by the

---

Table 2 Results: log-linear cost function model 1

| \(\ln TC\) | Coef. | Std. err. | \(P > |t|\) |
|---|---|---|---|
| \(\ln CM\) | 0.395 | 0.080 | 0.001 |
| \(\ln CNM\) | 0.327 | 0.101 | 0.002 |
| \(\ln PREV\) | 0.148 | 0.083 | 0.079 |
| \(PrMrev\) | −3.115 | 0.777 | 0.001 |
| \(PrNMrev\) | 0.049 | 0.416 | 0.907 |
| \(O\) | −0.058 | 0.087 | 0.510 |
| Constant | 8.994 | 0.791 | 0.001 |

\(N = 52; \text{R-sq} = 0.79; F(6, 45) = 28.33\)

Breusch-Pagan for heteroskedasticity:

\(\chi^2(1) = 1.24; P > \chi^2 = 0.266\)

RESET:

\(F(3, 42) = 1.81; P > F = 0.161\)

---

Table 3 Cost measures in health centres, by output level

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Low C output</th>
<th>Medium C output</th>
<th>High C output</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (health centres)</td>
<td>17</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>Total consultations per health centre per year (C)</td>
<td>2788</td>
<td>4875</td>
<td>9942</td>
</tr>
<tr>
<td>% of medium C ((=4875))</td>
<td>57%</td>
<td>100%</td>
<td>204%</td>
</tr>
<tr>
<td>Range of C</td>
<td>1806–3604</td>
<td>3654–6052</td>
<td>6170–15 737</td>
</tr>
<tr>
<td>CM: total MHI consultations per health centre per year (mean)</td>
<td>690</td>
<td>1001</td>
<td>2142</td>
</tr>
<tr>
<td>CNM: total uninsured consultations per health centre per year (mean)</td>
<td>2098</td>
<td>3874</td>
<td>7800</td>
</tr>
<tr>
<td>MHI pool size (mean) per MHI, per year</td>
<td>1191</td>
<td>1141</td>
<td>2617</td>
</tr>
<tr>
<td>% of MHI at medium MHI pool</td>
<td>104%</td>
<td>100%</td>
<td>229%</td>
</tr>
<tr>
<td>MC MHI consultation, RWF</td>
<td>244</td>
<td>249</td>
<td>240</td>
</tr>
<tr>
<td>MC uninsured consultations, RWF</td>
<td>402**</td>
<td>322**</td>
<td>311*</td>
</tr>
<tr>
<td>AC per MHI consultation, RWF</td>
<td>617</td>
<td>628</td>
<td>607</td>
</tr>
<tr>
<td>AC per uninsured consultation, RWF</td>
<td>1230***</td>
<td>983***</td>
<td>950**</td>
</tr>
<tr>
<td>Number of total C per staff per year</td>
<td>719</td>
<td>816</td>
<td>987</td>
</tr>
<tr>
<td>Number of total C per staff per day</td>
<td>2.70</td>
<td>3.07</td>
<td>3.71</td>
</tr>
</tbody>
</table>

Note: C = total curative consultations, N = number of health centres included, MC = marginal costs, CM = insured curative consultation, CNM = uninsured curative consultation, AC = average costs. Costs are reported in RWF, exchange rate: US$1 = RWF 390 (June 2000). Difference between insured and uninsured consultations with respect to (a) MC and (b) AC per consultation, and within output level group, two-tailed test: ***Sig \(P < 0.01\); **Sig \(P < 0.05\); *Sig \(P < 0.10\).
payment system resulting in payer-specific treatment intensity and marginal costs, independent of their output level.

For both insured and uninsured consultations, AC is considerably higher than MC, pointing to significant scale economies and idle capacity across facilities. The insignificant difference across different output levels combined with the very low number of consultations per staff per day (see Table 3) indicates first, a payment system threshold effect, meaning that a critical share of insured patients is needed that makes it worthwhile for small health centres to reduce the treatment intensity for capitated patients, and secondly, equally positive scale economies in the three health centre groups, independent of their output levels. However, the payer-specific cost difference across health centres becomes less significant in health centres with the highest output levels, suggesting that health centres could improve their financial situation by increasing the number of patients and using idle resources.

Discussion

This analysis leads to five main findings. First, a larger MHI pool and the related higher proportion from MHI revenue in health centres leads to a decrease in total facility cost. Under MHI with capitation payment, this may be a direct result of providers mitigating their financial risk related to insurance over a larger MHI membership pool.

Secondly, there is no significant difference between the cost elasticities of insured and uninsured consultations, implying that insured and uninsured visits have a similar impact on total facility cost.

Thirdly, marginal and average costs for insured patients under capitation are markedly lower than for uninsured patients who pay user fees. This evidence is consistent with the theory that providers behave as economic agents and adjust treatment intensity according to the expected payer-specific revenue.

Fourthly, for both groups, average costs are considerably above marginal costs, pointing to significant scale economies in health centres.

Fifthly, average revenue per consultation from MHI and from user fees is above the relative marginal cost, implying important short-term profits in health centres. Others found that in a mixed financing system, providers may try shifting costs to other sources (Ellis and McGuire 1993). As Rwandan health centres are paid by different sources, health centres may shift cost. The lower average cost recovery under user fees compared with MHI suggests that providers shift a larger proportion of treatment cost for uninsured patients to the government and donor sources, who subsidize the supply of medical care.

Uninsured patients’ significantly higher marginal cost may be explained by factors other than provider response to different payment mechanisms: severity of illness, adjustment of treatment intensity, and scale economies.

Adverse selection by more severely ill patients into the insurance pool would lead to differences in illness severity between insured and uninsured patients, and result in higher treatment intensity and marginal cost for the insured. However, the opposite was found in this analysis, suggesting that the insured may be less severely ill than uninsured patients. Additional evidence supports our interpretation. First, even in small pools, insured patients report lower marginal costs. It would have been helpful to address the selection questions directly by examining longitudinal data on individuals’ health status among the insured and uninsured; however, such data were not collected.

Secondly, there is qualitative evidence that insured patients were less severely ill than uninsured patients. According to interviews with nurses, insured patients were less severely ill because they sought care at the onset of illness and consequently needed less intensive treatment (ONAPO 2000: p. 58). This is corroborated by patient exit surveys where uninsured patients reported a greater delay in care-seeking (5.5 days) than insured patients (4 days) (p < 0.05) (Schneider et al. 2001). These observations imply that the severity of illness is influenced by the insurance effect and the provider payment mechanism. Testing for the effect of patients’ severity of illness on costs would require additional payer-specific information on health status.

Another possible explanation for the observed marginal cost differences is that the quality of care provided to the two groups differed, with insured patients potentially receiving sub-standard care. In the absence of detailed data on quality of care we have assumed that variations in quality occur within medically acceptable standards, and that the difference in marginal cost reflects higher levels of efficiency in treating insured patients. Some support for this stance is provided from exit survey data, in which insured and uninsured patients judged quality of care to be equal. Quality was proxied by the availability of drugs, waiting time, patient assessments of tests, prescribed drugs, the levels of staff knowledge and overall patient satisfaction. While insured patients reported receiving all prescribed drugs, 10% of uninsured patients did not receive their prescribed drugs, pointing to less intense treatment among the uninsured (Schneider et al. 2001). Nonetheless, insured and uninsured patients may not have been able to judge the technical quality of care received. Assessing the technical quality of care would have required observation of patient consultations.

Finally, findings from the cost structure analysis point to a scale economy effect caused by higher service use by insured patients, leading to lower marginal costs. This improved efficiency would arise through more intensive use of fixed inputs (such as staff) in facilities with higher utilization levels. While the scale effect is caused by the insured patients’ higher utilization, it is expected to affect the marginal costs of insured and uninsured consultations equally where idle resources exist.

An important limitation of this study is the absence of information on individual health status to identify selection problems, and on technical quality, which might help to support or refute one or more of the above alternative explanations.

Conclusion

In this paper we employed an econometric cost function with payer-specific outputs and conducted cost structure analysis to examine the effects of provider payment mechanism on provider behaviour. Findings imply that the lower marginal cost among insured patients compared with uninsured patients is
caused by both adjusted treatment intensity resulting in increased efficiency in treating insured patients under capitation (payment system effect), and the less severe case-mix among the insured, due to their improved access to care (insurance effect). In addition, markedly higher average than marginal cost for insured and uninsured patients point to significant idle capacity for treating both groups. The combined effect may have neutralized any impact of adverse selection or moral hazard that would have caused facility costs to increase under insurance.

This analysis contributes to current research on health financing in developing countries and in particularly to the design of community-based health insurance (CBHI) and the choice of provider payment. Under MHI with capitation payment, providers do not primarily shift cost across payment sources, as is alleged when fee-for-service provider payment is selected for CBHI (ILO 2002; Bennett 2004). Rather, they adjust costs and the resource-intensity of treatment to the expected payment source, which has a cost-containment effect. This finding is particularly relevant for resource-constrained countries that aim to improve access to care through insurance.

The analysis provides arguments to donors and governments to encourage insurance enrolment into schemes with capitation payment. Based on these arguments, and on the grounds of practical feasibility of administering capitation payment for providers and insurers, capitation payment is recommended in the insurance design. Provider performance should be evaluated regularly to prevent providers from skimping on care to insured patients (Ellis and McGuire 1996).

Future research on health insurance and provider payment should collect data to examine insured and uninsured individuals’ health status and care-seeking behaviour before and after insurance, the resulting case-mix and providers’ response with respect to quality of care and resource intensity of treatment.

Acknowledgements

This paper is adapted from the PhD dissertation of Pia Schneider, completed at the University of London, London School of Hygiene and Tropical Medicine (LSHTM) under the supervision of Kara Hanson. The authors are grateful to the USAID-funded Partnerships for Health Reform Project (PHR) administered by Abt Associates Inc., for the use of the data set; and to two anonymous reviewers for helpful comments on an earlier draft. Kara Hanson works with the Health Economics and Financing Programme which is funded by the UK Department for International Development.

References


