

4. THE COLOMBIAN HEALTH SYSTEM

Hospital autonomisation has been adopted in Colombia as one component of a widespread health care reform undertaken in 1993. However, some steps had been previously taken that set the stage for the large reforms of 1993. This chapter describes the years previous to the reforms, and the major changes introduced in 1993. It also describes hospital reforms in more detail, its evolution and the available empirical evidence at the country level. It further describes in even more detail the process of hospital reform in Bogota and reviews three available empirical studies on hospital autonomy at the Bogota hospitals.

4.1. Country profile

Colombia is a middle income country with a population of 41.2 million (DANE, 2006), GDP of US\$ 122 billion dollars in 2005 and a Gross National Income per capita of US\$ 2,290 (World Bank, 2006). Total health care expenditures as a percentage of GDP were 7.8% in 2004 (85% public and 15% private) (WHO, 2006b). Health expenditures per-capita by 2004 were US\$ 170, or 568 in international dollars (WHO, 2006b). The country is divided into 32 subnational administrative units called departments, 1098 municipalities, and the Special District of Bogota, i.e., the capital city.

4.2. The decentralisation reform

The administrative structure of the public sector in Colombia has gone through a two-decade process of decentralisation. Regarding the public health system, Law 10 of 1990 strengthened the decentralisation process by transferring more resources and competencies to the municipalities. Level I hospitals were the responsibility of the municipality, and level II and III hospitals were the responsibility of the department.¹ (Jaramillo, 1999).

A Constitutional Reform in 1991 strengthened the decentralization process and health care for the poor and additional resources from general taxes were definitely transferred to municipalities. As a result, public resources to provide health care for the poor (in

¹ Hospitals were classified, according to their level of complexity, as level I (lowest complexity), level II (intermediate complexity) and level III (tertiary care). This three-level classification still persists although its boundaries have been progressively blurred.

terms of per-capita per-poor person) increased 87% in real terms between 1991 and 1997 (Castano et al, 2001).

4.3. The reform to the health care system

Until 1993, the health system in Colombia was segmented for two separate populations: those in the formal sector of the economy and the rest of the population. The first group was covered by social health insurance based on their employment, and all employees were compulsorily enrolled in one of the 1,040 funds existing at that moment. This segment included 20.6% of the population. These funds had no benefit packages but just open-ended coverage with some exclusions, i.e., a negative list of benefits. Those without enrollment in the social security funds were provided care at the public network of hospitals, which were funded via supply-side subsidies. Services provided by the public network were supposed to be self-targeted by the poor, but secondary and tertiary care facilities were disproportionately used by the non-poor.² Budgeting based on historic costs was the prevailing system to transfer resources to hospitals.

A wide span reform was enacted in 1993 by Law 100, with the purpose of universalising social health insurance coverage, and introducing a market-based system. The reformed system resembled in some respects Einthoven's (1988) proposal of managed competition, as premiums and benefits were regulated, and competition between insurance plans would concentrate on quality rather than on inefficient behaviours such as market segmentation. This was supposed to reduce the failures of health care markets arising from information asymmetries.

The reform aimed at universal coverage by the year 2001, with 70% of the population enrolled in an Employment-Based Scheme (EBS), and 30% in a Subsidised Scheme (SS). These two groups were expected to receive the same benefit package, which would be managed by third-party payers, in a managed-care-like fashion. Health care providers were expected to generate all their revenue from sales to insurers; this was not new for private providers but it was for public providers, because it implied that supply-side subsidies would fade out to be replaced by demand-side subsidies. Consequently, public hospitals would have to compete like any other provider to win contracts with third-party payers. They also had to develop capabilities for billing services and

² This is a common finding for self-targeted public networks in developing countries. See Gwatkin (2005)

collecting accounts receivables.³ The idea was that if they failed to raise revenues by winning contracts in a competitive market, they would face closure. However, as shown below, the expectation of universal coverage with a comprehensive package did not realise.

4.3.1. The Employment-Based Scheme (EBS)

Workers in the formal sector, pensioners, and high-income self-employed workers, have to be compulsorily enrolled in the EBS. The worker chooses any of the competing insurers operating in the EBS. These insurers are called Health Promoting Enterprises (EPS for their Spanish acronym).⁴ Once the worker chooses the EPS, it collects the compulsory contribution, a 12% payroll tax. This tax is shared between employer (8%) and employee (4%). Pensioners or self-employed pay the whole 12% themselves. This contribution is then transferred by the EPS to an equalisation fund called FOSYGA, where contributions are separated into 11 and 1 shares. Eleven shares are put into the equalisation account and one share is taken to help fund the Subsidised Scheme (see below). Once the equalisation account has all the collections of the workers' compulsory contributions, it returns a risk-adjusted premium⁵ to the EPS, not only for each worker but also for his family members.

The per-capita premium that the EPS receives entitles the enrollee and family members to a basic benefit package called the POS (Plan Obligatorio de Salud). Unlike the system before 1993, the POS is an explicit positive list of benefits. EPSs rarely contract with public hospitals, but sometimes the latter provide emergency care for EPS enrollees. In this case public hospitals bill the EPS for the services delivered. In general, public hospitals' share of revenues from EPSs is very small. Although expected to enroll 70% of the population by the year 2001, as of 2004 the EBS enrolled only 35% of total population. See figure 4.1 for a schematic view of the EBS and its relationships with other schemes.

4.3.2. The Subsidised Scheme (SS)

³ Services billed to patients or third parties, that remain unpaid to the hospital.

⁴ EPS were created by Law 100. Among the previously existing public social health insurance funds some were converted into EPS, while the rest disappeared.

⁵ It is risk-adjusted by age, gender and geographic location of the worker.

Informal employment in Colombia has fluctuated around 60% of total employment during the last five years (Herrera, 2005). Enforcing enrollment in the EBS is very difficult among informal workers, and their income is usually very small. Thus, these workers and their family members are eligible for the subsidised scheme (SS). This scheme is aimed at the unemployed, the poor self-employed (mostly informal workers), and the poor vulnerable (the elderly, the chronically ill, people displaced by the armed conflict, etc).

The SS is funded by two main sources. On the one hand, the solidarity share that was mentioned above (one twelfth of payroll tax), and is administered by FOSYGA at the national level. On the other hand, the national budget transfers to the municipalities a given amount for local spending on health, education and basic sanitation. These resources are called the General System of Participations (GSP), and are general tax collections, raised basically through value-added tax and income tax, by the national tax agency. Regarding the health care share of these outlays from the national budget to the municipalities, it has three main destinations: enrolling the poor in the SS, paying for public hospitals, and paying for public health interventions. The SS component of the GSP is sent to each municipality according to its number of current enrollees and the expected increases in enrollment for the year.

Thus, the municipality receives money for the SS from FOSYGA (solidarity share) and SS component of the GSP, which are deposited in a local health fund. With the money deposited at the health fund, the municipality has to decide whom to enroll in the SS. Given that this scheme is aimed at the informal and poor population, the municipality has to identify the poorest in order to select them for the subsidy. Thus, it carries out a census of the poor population applying a proxy-means test called SISBEN, which is a detailed questionnaire that collects information on household characteristics and gives a poverty score according to those characteristics. The score values vary from 0 to 100, for 0 = the poorest, so the municipal authorities have detailed cardinal information that allows them to rank the poor.

Once ranked, the municipal authorities select, in ascending order, the households that will be given the subsidised insurance. These households are notified of their selection and they are allowed to choose the insurer of their preference, among those authorised to operate in the municipality. Insurers in the SS are called ARS, for Administradoras

del Regimen Subsidiado (Subsidised Regime Administrators).⁶ The ARS then goes to the municipal authorities with the list of individuals that have chosen it, and the municipal authorities transfer a premium for each individual for one year. This premium is not risk-adjusted, and is fully subsidised, i.e., neither the individual nor the household has to pay any part of the premium.

Enrolment in the SS entitles the individual to a benefit package that is reduced as compared to the EBS's benefit package. The subsidised package, called the POS-S (for Plan Obligatorio de Salud Subsidiado), includes primary care, emergency care, low-complexity interventions (e.g., hernia repairs, bone fractures, inpatient care at first-level facilities), and high-cost conditions⁷ (cancer, HIV/AIDS, renal dialysis, intensive care unit). It also includes prescription drugs. Enrollees under one year of age have full coverage (the same benefit package as the EBS). Pregnant women have full coverage of any morbidity related to pregnancy. Unlike the POS of the EBS, the POS-S has no waiting periods, and no copayments are charged to users.

The gap between the packages of the EBS and the SS is covered by the local health authorities and is typically provided at the public hospitals. Thus, if an ARS enrollee needs care of a level II complexity, it is not covered by the SS. This creates particularly complex coordination issues when a patient needs care across several levels of complexity. For interventions not included in the POS-S, SS enrollees not infrequently use private providers, paying out of their pockets (O'Meara et al, 2003).

ARSs carry out the purchasing function, within the boundaries of the POS-S. At the beginning of the reform, they were supposed to set up their health care delivery networks as they saw fit, so that public hospitals would have to upgrade their infrastructure and production processes to compete for contracts with ARSs. However, strong resistance on the public hospital side forced the government to enact regulations to protect public hospitals (Jaramillo, 1999). In this sense, Law 715 of 2001 forced ARSs to compulsorily contract 50% of the premium if the ARS operated in an urban

⁶ A single entity can operate both in the EBS and the SS, but these two business units have to be clearly separated to prevent cross-subsidisation as much as possible.

⁷ The benefit package of the EBS and the SS is divided into four levels of complexity. Levels I (low complexity), II (intermediate complexity) and III (tertiary care, high complexity), roughly coincide with the classification of hospitals mentioned above. Level IV are those high-cost conditions that are treated separately in the package, for the purpose of reinsurance and waiting periods. However, these events are treated at level III, or tertiary care, facilities.

center that had tertiary-care public hospitals. This restriction was aimed at 1) avoiding ARS from setting up their own networks, which would have left public hospitals with idle capacity that had to be financed by the government because of high exit costs; and 2) avoiding ARSs from cream-skimming the portfolio of public hospitals by taking the profitable services to their own networks and leaving the ESEs with the loss-making services.

The current situation of the SS is as follows: enrollment by the year 2002 was 11.4 million and it has increased to 15.8 million in 2005⁸. The benefit package has not been widened, except for minor inclusions, but the gap with the EBS remains wide.

4.3.3. The uninsured

The rest of the population, i.e., those who are not enrolled in the EBS or in the SS, are called the uninsured. The uninsured are provided care in the public network of hospitals, although they also use private providers paying out of pocket (O'Meara et al, 2003). When they use public hospital services, they have to pay cost-sharing rates that are set according to their SISBEN classification. Given that most municipalities have run the SISBEN test for all the poor residing in the municipality, most of the times the uninsured have a SISBEN card that includes information about their SISBEN level, and the hospital charges cost-sharing rates accordingly.

Care for the uninsured is the responsibility of the municipality, for which it receives the hospital component of the GSP. The municipality uses a menu of payment mechanisms ranging from line-item budgets to fee-for-service payments, to transfer this money to the hospitals in its area. These monies are also used to pay for interventions to SS enrollees when such interventions are not included in the POS-S.

4.3.4. Other insurance schemes

In parallel with the EBS and the SS, there are a series of minor schemes with diverse degrees of overlap with the EBS and SS. One of them is the Traffic-Accident Insurance

⁸ In addition to the 15.8 enrollees, 3 million beneficiaries were enrolled with a very restricted benefit package. These are called the partial-subsidies plan. These partial subsidies were not in place by 2002 in Bogotá.

(SOAT), which is a mandatory insurance scheme for car owners. Every car has to be covered by this SOAT scheme, and the owner pays a premium to a SOAT insurer, based on the car's value. Individuals injured in traffic accidents are taken to any hospital (public or private), and the provider bills the SOAT insurer on a fee-for-service basis, irrespective of patient's coverage status. There is a pre-established fee schedule that only applies to SOAT events on a national-level basis.

4.3.5. Public health

Previous to Law 100 of 1993, public health interventions were the responsibility of local public hospitals and the social health insurance funds. Personal-level public health interventions were provided by the public hospitals to the uninsured, whereas those enrolled in the 1,040 social health insurance funds received these interventions from their respective funds. On the other hand, collective-level interventions (like environmental protection, water and sanitation, etc.) were arranged and provided by public hospitals. Programs like malaria and tuberculosis control were coordinated from the Ministry of Health as vertical programs with their own hierarchy.

Law 100 divided public health interventions in two separate plans. On the one hand, personal-level interventions (e.g., immunisations, screenings, etc.) called Promotion and Prevention (P&P) activities, are provided by the EPS for the enrollees in the EBS. For the SS, P&P interventions were initially the responsibility of ARSs, but some of them⁹ were transferred to the responsibility of the municipality, by Law 715 of 2001. For the uninsured, these interventions are supposed to be paid by the municipality to the public hospital that provides them.

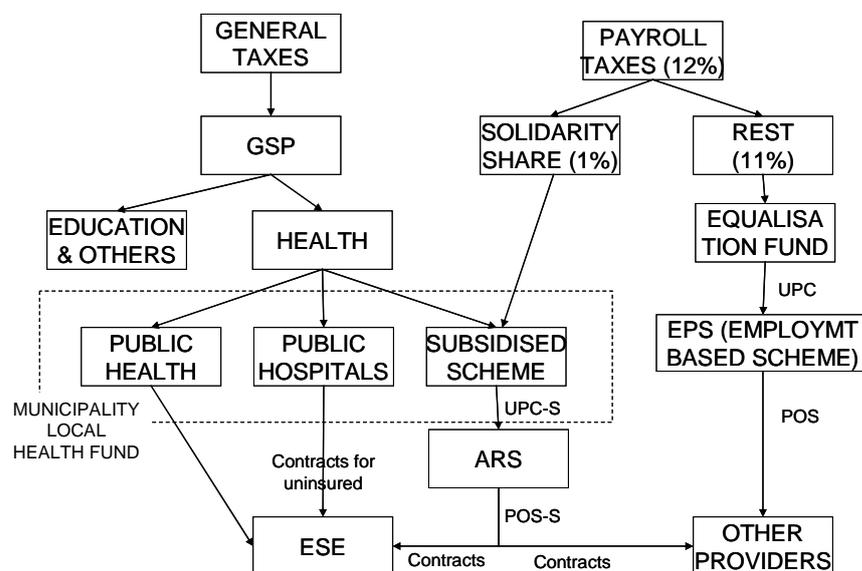
On the other hand, collective-level interventions are the responsibility of the municipality, and are bundled in a "package" called the Basic Plan (PAB, for Plan de Atención Básica). The municipality receives resources from the GSP to finance the provision of these interventions, which the local authority contracts out, very frequently with the local public hospital, but not necessarily. The interventions, their processes and

⁹ Specifically: health promotion, immunisation, family planning, early detection of cervical cancer.

outputs, are defined each year by the Ministry of Social Protection¹⁰. The PAB also includes interventions related to diseases with high externalities, like malaria, TB, etc.

Figure 4.1.

A summarizing sketch of the Colombian health system



4.4. Reform of the public hospital network

4.4.1. The public hospitals before the 1993 reform

One key aspect of the 1993 reform was a radical reform of the network of public hospitals. Before then, public hospitals were integrated into a national network called the National Health System, and it was part of the Ministry of Health organisation. This vertically integrated network was coordinated via the national government and the departments. Hospitals were administrative units of the large national network, and they received input-based budgets that were controlled externally by the departmental and national governments. The incentives embedded in this input-based funding scheme rewarded, as expected, higher-cost and inefficient hospitals, which led to overcapacity and low productivity. Hospital directors were accountable to departmental authorities, and enjoyed no autonomy.

¹⁰ In 2003, the Ministry of Health was transformed into the Ministry of Social Protection by merging it with the Ministry of Labor.

As a result of Law 10 of 1990, many municipalities increased hospital supply and bed supply, but these were still administrative units of the municipal administration, to the extent that they were not given the new legal status of autonomy. In spite of some advance towards autonomy, the lion's share of hospitals' revenues was still administered by municipalities via line-item budgets based on historic costs, to provide services to the uninsured. Therefore a Purchaser-Provider Split did not exist and hospital autonomy was not realised as expected from Law 10.

4.4.2. The reforms introduced by Law 100 of 1993.

Law 100 of 1993 radically changed the decentralised but still largely vertically integrated structure of public hospitals. As mentioned before, the reform introduced a market-based approach to the health care system. While in the EBS and the SS, those eligible were given choice of EPS or ARS respectively, and EPS and ARS were expected to compete on quality (because the benefit package and the premium was regulated), people in the uninsured segment of the population were also given choice of provider, i.e., public hospital.

The market-based scheme was very clear when analysed through the expected context of universal coverage and comprehensive benefit package that Law 100 envisioned for 2001. In that scenario public and private providers would have to compete for insurers' contracts. On the side of public hospitals, it was clear that no supply-side subsidies would be necessary, because they were expected to raise all their revenues from contracts with EPS and ARS. Thus, the "money follows the patient" principle would have come true, and a full transformation of supply-side subsidies (transferred by local health authorities to hospitals) towards demand-side subsidies (administered by ARS) was expected to take place. Put in other words, hospital revenues would no longer be based on inputs, but on outputs.

In order for the market-based approach to work effectively, hospitals had to become autonomous entities with all the necessary flexibility to compete. Accordingly, a Purchaser-Provider Split was necessary. This is a big difference with the process of autonomisation started by Law 10 of 1990. Whereas at that time hospitals still received most of their revenues from supply-side subsidies, and the lack of a market approach

did not create the need for autonomy, after 1993 it was clear that switching subsidies to the demand side within a market environment made it necessary to grant autonomy to hospitals. In consequence, the Law created a new legal entity that was called Social State Enterprises (ESE for Empresa Social del Estado). Every public hospital should be converted into ESE. Hospitals converted into ESEs were characterised by legal and administrative autonomy, and were ruled by the same legal framework as the private sector, except for the use of special clauses to favor the public sector. ESEs had to create their boards of directors on the lines set by Law 10 of 1990, and were enabled to engage in contractual relationships with EPSs, ARSs, and other payers, as well as with upstream input providers.

In terms of Harding and Preker's (2003) analytical framework, ESEs had the following decision rights:

- Strategic management: setting/executing strategic plan
- Finance: budget setting/execution, raising and retaining user-fee revenues. Not allowed to raise capital or borrow from private lenders.
- Marketing: Setting/executing marketing strategy including pricing strategies, but excluding cream skimming.
- Medical management strategy: operating with a network orientation established by health authority.
- Human resources: civil service regulations are kept for current employees. New employees can be contracted on a fixed-term basis, outside civil service regulations and the ESE manager has discretionary power to hire and fire them.
- Procurement: autonomous direct procurement from upstream input providers. Autonomy to join collective procurement strategies.

The other determinants of hospital behaviour were also changed to varying extents. Regarding market exposure, ESEs were supposed to compete vis à vis other ESEs and private providers, to win contracts from EPSs and ARSs. No historic, input-based budgeting would remain after full transformation of subsidies from supply-side to demand-side, when all Colombians would be covered either by the EBS or the SS with equally comprehensive benefit packages.

However, the transition towards universal coverage has been indefinitely extended, because the economy and labor market dynamics have not generated the necessary resources to achieve such universality. Therefore, municipalities are still responsible for a large share of revenues for ESEs, because they receive money from the central government (the GSP) to pay for services provided to the uninsured, and to SS enrollees for services not included in the POS-S.

Regarding residual claimant status, it was largely changed, as ESEs are allowed to retain all revenues, including user fees. However, user fees are not a large source of revenue for ESEs. Instead, most of their revenues still come from contracts with local health authorities, which gives the health authority the levers to compensate deficit-generating ESEs with the surpluses of the other ESEs. In any case, if a given ESE still made a surplus after cross-subsidisation took place, it is allowed to keep the surplus and spend it as it sees fit.

Accountability mechanisms were also strengthened in Law 100. Not only are all the hospitals required to create their boards of directors, but also the regulatory decrees provided explicit guidelines about how to select the directors. These boards have to have three blocks of representatives: the first is the so called “political-administrative” block, with the representation of the Mayor and the local health authority. The second block is that of the scientific area of the ESE, with one representative from the professional staff (physicians, professional nurses, etc) and another from the scientific societies of health professions existing within the area of the ESE. The third block is that of the community, with the participation of ESE’s user organisations and a representative of the businesses in the area of the ESE. Directors from the second and third block are elected by the groups they represent, not by the open population.

These boards must have at least six members, although they can have more members. One of the key roles of the board of directors is to select a short list of three candidates to the position of ESE manager. This short list is sent to the Mayor, who makes a choice among them and issues the appointment. Managers are appointed for a three-year period, and can be reelected by the board, by going through the same selection process.

Other accountability mechanisms are exerted through the tutelage function carried out by the health authority, which, acting as the regulator of health services at the local

level, holds ESEs accountable for issues that it cannot control through its representatives on the board of directors. Contracts between health authorities and ESEs also work as accountability devices.

Regarding unfunded mandates, ESEs are the providers of last resort, and they have to provide care to the uninsured and enrollees in the SS for interventions not included in the POS-S. This generates a source of unfunded mandates, specifically when cost-sharing amounts are much higher than the poor's ability to pay.

Public hospitals in Colombia also showed that autonomy is not a dichotomous state, because although Law 100 created the legal entity of ESE, public hospitals had to fulfill some basic requirements to be converted into ESE. But even having fulfilled these requirements, some ESEs were more advanced than others in terms of the decision rights and the determinants of hospital behaviour described above. Even sometimes an ESE's advance through the different components of autonomy was restricted by higher level authorities, despite the Law's formal transfer of decision rights and residual claimant status.

4.5. Empirical evidence on hospitals after the reform

Although the empirical evidence about the evolution of public hospitals after the reform is not abundant, some studies, particularly run by the government agencies, are available. These studies though, do not isolate the effect of autonomisation, but just analyse aggregate data. A study by Perez (1997) showed that public hospitals increased their revenues by 47% between 1993 and 1995. Their expenses also grew, although at a lower rate (41%). However, they increased their share of revenues coming from supply-side subsidies from 61% of total revenues to 71% during that period.

In a study of 203 level I hospitals using Data Envelopment Analysis, Pinzón (2003) found an association between low levels of relative efficiency and excess of human resources. As has been the experience in other countries, legal restrictions in the public sector make it difficult to downsize overcapacity, human resources being the most difficult to downsize. The study also found an association between high dependency on supply-side subsidies and lower efficiency.

After the enactment of Law 100 of 1993, political turmoil around public-hospital policy did not recede. The most common and persistent complaint relates to hospital accumulated deficits that persistently forced the central government to transfer additional funding. The situation has been compounded by the allegedly high levels of accounts receivables that ESEs have accumulated from ARSs. In fact, there seems to be a chain-reaction whereby the flow of resources from the large funds (solidarity fund and national budget) through the different intermediaries is delayed, so that hospitals, being the last link of the chain, face the sum of all the delays. It has been suggested that delays are stimulated by the possibility to earn interests on bank deposits, whereas no sanctions apply for delays in disbursements to links down the chain (Universidad Nacional, 2002).

The political visibility of public hospitals made it impossible to deliver the promise of a hard budget constraint that would have been expected for ESEs. Nonetheless, the national government issued a technical paper laying out its diagnosis and policy recommendations about public hospital issues (DNP, 2002). This document, called the CONPES 3204¹¹ also summarises the development of government policy regarding public hospitals after the 1993 reform. Although it is not a research paper, it is highly relevant for the issue of hospital autonomy in Colombia.

CONPES 3204 recognises a series of flaws in the implementation of the hospital autonomy policy. The increase in hospital and bed supply mentioned above led to investment initiatives that were not coordinated with other municipalities. This caused an excessive and disorganised growth of bed supply and of hired health workers. It is also shown that converting hospitals into ESEs did not yield the expected benefits in terms of efficiency and competitiveness.

One key aspect regarding budgetary autonomy generated a fiscal problem for the national government. CONPES 3204 shows that autonomous hospitals changed their strategy to set their budgets, using expected revenues (from sales to third party payers) as the basis to calculate expenditures for the following fiscal period. Thus, a hospital was able to increase its expenditures just by creating the expectation that it will generate enough revenues to pay for those expenses. However, at the end of the year they had increased their expenditures but their revenue expectations were not fulfilled. After

¹¹ The CONPES documents are technical papers released by the National Department of Planning to set the reference frameworks and make recommendations on matters of social and economic policy to the National Council of Economic and Social Policy.

several years of this process, which started in 1998, many hospitals ended up with rapidly increasing expenditures but lower than expected revenues, and thus a growing deficit that backfired on the central government. This means that regardless of their autonomous status, ESEs faced a soft budget constraint, because the central government was usually pressed to bail out the hospitals to appease the political antipathy aroused by hospital deficits.

Human resources is the item that shows the largest increases, as it grew from 40% of total hospital expenses to 66% between 1994 and 2001. (Ministerio de Salud, 1998). Surprisingly, the CONPES 3204 points out that increases in labor costs were not paralleled by increases in hospital output.

The soft budget constraint problem, i.e., growth in expenses supported on the expectation of increased revenues that were not realised, caused deficit-running hospitals to increase their accounts payable, accumulate large deficits and stop paying their workers' salaries for several months. As expected, this lowered morale to record lows and frequent strikes altered the public network during the late 90's and early 2000's. As a pilot test of a radical solution, the Ministry of Health started a program to restructure 26 hospitals in the most critical situation (Ministerio de Salud, 2001).

Hospital restructuring is a process led and funded by the national government with the purpose of reducing labor costs at public hospitals (particularly regarding civil servants who enjoy job protection until retirement), to make them more competitive in the marketplace. Key measures of this process include early retirement and severance payments. The results for the 26 pilot test hospitals show that they were able to increase their output and decrease their labor costs and idle capacity. The program also showed that the most successful experiences were those that involved the whole network of public providers in a given geographic area. This suggests that network coordination problems beyond the boundaries of a given hospital were also key determinants of hospital performance.

The paradoxical situation that is highlighted by CONPES 3204 is that even though a sharp reduction in the population that received care at public hospitals took place, and resources available for hospitals increased heavily, they were running large deficits. Whereas in 1994 the uninsured population (who were expected to use public hospitals

only) was 14 million, this number had dropped to 5 million by 2001. Accordingly, the per-capita resources to provide care for that population increased from about CP\$75,000 to about CP\$260,000 in constant CP\$. The drop in population that received care at public hospitals is largely explained by the inclusion of family members in the EBS, and the expansion of enrollment in the SS. However, these figures have to be adjusted down because public hospitals also provide services to enrollees in the SS for those interventions not included in the POS-S. Although this adjustment was not made in the CONPES 3204, it certainly reduces the per capita amounts in 2001 (Cueto, personal communication, 2006).

As said above, the empirical evidence from the studies analysed does not isolate the effect of hospital autonomy on performance. However, CONPES 3204 attributes many of the observed difficulties to the fact that ESEs were using their decision rights in a poor manner. In addition, the lack of a control group to isolate the effects of autonomy reduces the analysis to pre-post comparisons. Given the drastic changes in market environment, payment mechanisms, governance and regulations, it might not be risky to attribute an important part of these changes to autonomy.

4.6. The case of Bogota

Bogota, the capital city of the country and its largest urban center, had a population of 6.7 million in 2002. Enrolment in the EBS was 58.8% of the total population, in the SS was 14.3%, and in the exception schemes was 4.5% (Acosta, 2005). This means that, based on the projected population for year 2002, the uninsured were 22.4% or about 1.52 million people. Bogota's legal status is a Special District, which means that it has both department and municipality characteristics. Given that Bogota is the capital city, there are many public hospitals that belong to other jurisdictions, like departmental-level, national-level and EPSs hospitals. But the hospitals that are under the control of the Bogota Health Authority, the Secretariat of Health (SOH), are the 32 hospitals that are studied in this thesis. After several mergers that took place in 2000, the 32 hospitals were reduced to 22, as described below.

4.6.1. First stage: The evolution of hospitals into ESEs

After the 1993 reform, Bogota started the process to upgrade its 32 hospitals to fulfill the requirements to become ESEs. This process concluded in late 1997, with the District Council enacting the decree that granted them ESE status. Although between 1993 and 1997, the Bogota hospitals did not enjoy much autonomy, the SOH supported a process of institutional strengthening. A study by Giedion et al (1999) shows that although hospital output increased during this period, an increase in labor costs was evident, caused by a 1995 national decree that leveled health worker salaries across the country. The authors also point out that autonomy was still too weak, as hospitals had no control of labor and budgeting. However, hiring workers on a fixed-term basis was emerging as a way to circumvent rigidities in human resources management.

4.6.2. Second stage: the introduction of fee-for-service reimbursement

Once hospitals had reached their ESE status in 1997, the SOH started the contractual relationships with them which required the writing and negotiation of contracts. This first round of contracts was very time consuming, as neither party had experienced such transactions with each other. In order to stimulate increases in output and the setting up of managerial and administrative skills, the SOH implemented a fee-for-service payment mechanism. Given the incentive inherent to fee-for-service payments, it was expected that ESEs would engage in demand inducement. In fact, two years later many of the 32 hospitals had markedly increased their outputs, but many of them also increased their deficits.

An analysis of the Bogota ESEs carried out by Giedion et al (2000) shows that despite increased revenues, public hospitals were not self-sustainable by that time because increases in expenditures were larger than increases in revenues. But the effects of incentives to bill services were readily apparent: whereas in 1997 about half of revenues were supply-side subsidies, by 1999 these have decreased to 4% and the rest was transferred on the basis of billed services.

Bogota showed the same problem pointed out in CONPES 3204 regarding the paradoxical growth of supply-side subsidies in the presence of increased insurance coverage. Although in the case of the SOH it paid ESEs against services billed, that is to say, they must be named demand-side subsidies, hospitals were increasing their

expenditures at a higher rate. In the end, it backfired on the SOH in the form of high pressure to keep cash flowing to ESEs to pay their now higher expenses.

4.6.3. Third stage: from fee-for-service to per-case prospective payment with billing caps

By 1999, the SOH faced the consequences of an economic recession that hit the country in 1998. Reduced tax collections at the national level led to a reduction in the GSP transfers to municipalities. In consequence, the SOH received less money for health care for the poor. In addition, some ESEs showed large increases in billings to the SOH, while others increased their expenditures without concomitant increases in revenues, leading to the accumulation of growing deficits. Given this situation of financial stress, by late 1999 a new Secretary of Health decided to change the payment mechanism towards a sort of per-case prospective payment called Fixed Global Prospective Payment (FGPP). Lopez et al (2002), the designers of the FGPP, stated that the idea with this mechanism was to resemble the DRG payment mechanism, or a per-case payment, although data limitations did not allow implementation of such a sophisticated mechanism. However limited the available data, and keeping in mind that they were extracted from claims data files from the fee-for-service era, Lopez et al used them to define a set of intermediate activities like lab tests, therapies, medications, etc., that were packed into final activities, trying to arrange them so that each package showed the lowest possible variance. The packages (final activities) designed are listed in annex 2

These packages included intermediate activities that were supposed to be provided by the ESE, but given the incentive inherent in the prospective per-case payment, ESEs were expected to reduce the overprovision of intermediate activities, a typical behaviour seen during the fee-for-service era. Each package was differentiated for the three levels of complexity, so that, for example, a hospital discharge at a level I hospital had a lower rate than a discharge at a level III hospital.

However, the FGPP also had perverse incentives, similar to those reported in the literature on DRGs, namely 1) the “quicker and sicker” discharge problem; 2) the cream skimming problem; and 3) code creep. Regarding the “quicker and sicker” problem, the contracts included a no-pay clause for re-admissions within 30 days of discharge.

Regarding the cream skimming problem, it has two different features: cream skimming patients, and cream skimming services. To reduce the incentive to cream skim patients, outlier cases were paid separately on a fee-for-service basis, above a certain threshold. To avoid cream skimming of services (the hospital choosing to provide the most profitable packages and not to provide the less profitable), Lopez et al decided to face that problem by keeping a small number of packages. However, an obvious tradeoff with cream skimming patients (because of increased cost variance) was inevitable.

Another perverse incentive inherent to the FGPP was that ESEs would induce demand for final activities in order to maximise revenues, as pointed out by Marini and Street (2006) for the UK activity-based payment. Thus, ESEs would show higher admission rates, more surgical procedures, more outpatient visits, and so on. Anticipating that event, the SOH set yearly caps for the services that each ESE was allowed to bill to the SOH. If the ESE bills to the SOH exceeded the cap, the SOH would not reimburse these excess bills. These caps were estimated with the explicit purpose to reduce the bloated expenditure patterns that some ESEs developed during the fee-for-service era. But such reduction was not possible overnight, so the billing caps were set based on the previous year Minimum Essential Expenditures (MEE). The MEE was determined by fixed costs and a share of variable costs. The cap was set based on that MEE less a small cut aiming at a convergence between local demand and supply. Two situations were possible for a given ESE: on the one hand, an ESE that increased dramatically its expenditures but not its output during the fee-for-service era would find it difficult to bill enough FGPP to reach the billing cap, because of low productivity. On the other hand, an ESE with high productivity would find it easy to reach the cap if this high productivity was the result of induced demand, because the ESE was expected to quickly reduce output and match the cap; but if high productivity was not associated with demand inducement but real demand, this ESE would find it difficult to stay below the cap.

If a low-productivity ESE were unable to reach the billing cap, the SOH could not stay passive, because that ESE would default on the payment of salaries and accounts payable, and the problem would kick back on the SOH. Therefore, the SOH, acknowledging the impossibility to reduce redundant capacity and personnel overnight, transferred the ESE the required amounts to reach the cap, but conditional on the satisfaction of progressive goals to reduce redundant staff. These conditional transfers used the mechanism of the Performance Agreement. Therefore, the SOH had two

complementary tools, the contract itself and the performance agreement, working as accountability devices. Whereas the former worked as a market accountability device, the latter worked as an ownership accountability device.

Another incentive problem of FGPP, commented on by Giedion et al (2000), concerns information. While during the fee-for-service era, ESEs had the incentive to report detailed information about activities (because payment depended on detailed billing), the shift to a per-case payment gave no incentive for the ESE to provide that detailed information any longer. Thus, the SOH was left with much less information than it previously had.

A last problem of the FGPP was the incentive to skimp on care and reduce quality. In this sense, Marulanda et al (2002) analysed the effects of the change in payment mechanism from fee-for-service to FGPP on financial, output and clinical outcome indicators. Their study shows that hospital output increased, but intermediate activities decreased. Patient satisfaction and clinical outcome indicators did not show relevant changes. However, a large decrease in expenditures was observed, which allowed a reduction in deficit for the year 2000 by 34.5%, a strong impact but still not sufficient to let ESEs accumulate surpluses. A qualitative component in this study shows that when the ESE is approaching the billing cap, the manager starts to apply strategies to restrict demand, mainly raising access barriers.

4.6.4. Fourth stage: hospital mergers

Another big change that took place after the arrival of the new Secretary of Health in 1999, was the decision to merge some ESEs that were unsustainable with other better-performing ESEs in the same geographic area. As a result, the 32 ESEs were reduced to 22. The stated purpose of the mergers was to rationalise and optimise the use of hospital resources, and achieve a better match between supply and demand. Thus, merging ESEs was a strategy to avoid closing hospitals, and instead of that, the resulting merged hospital would make it easier to downsize physical capacity and number of employees, particularly administrative personnel. Giedion et al (2000) report that these mergers allowed a reduction in hospital overall costs by about 10%. An internal report of the SOH (Secretaría Distrital de Salud, 2002) analysed output at the merged hospitals and found that all of them increased the production of tracer activities. However, the report

does not include control data on the non-merged hospitals, or the data of the merged hospitals in the periods prior to the merger.

4.6.5. ESEs and the Subsidised Scheme

The transformation of subsidies from the supply side to the demand side in the jurisdiction of Bogota was more advanced than in any other municipality. However, it was unlikely that universal health insurance coverage and a full transformation of subsidies to the demand side would take place. Two main reasons underlie this failure to achieve universal coverage. On the one hand, although the SOH had enough resources to enrol all the eligible poor in the SS, not all of them took the subsidised insurance, which left people with no insurance at all. On the other hand, expanding the interventions in the POS-S is not likely to occur in the short and medium term, because of difficulty to mobilise additional resources (Castano, 2004). In consequence, these interventions not included in the POS-S have to be paid for by the SOH until they are gradually included in the benefit package.

4.6.6. Overview of ESEs sources of revenues

ESEs in Colombia, and specifically in Bogota, generate the vast majority of their revenues from three large sources:

- 1) services delivered to:
 - a) the uninsured
 - b) to SS enrollees for those items not included in the POS-S.
- 2) services delivered to the enrollees in the SS, for items included in the POS-S.
- 3) services included in the PAB.¹²

All these revenues are mediated by a contract. For sources 1 and 3, contracts are signed with the SOH, whereas for source 2, contracts are signed with ARSs.

Additional sources make up a small share of revenues at Bogota's ESEs, but are still worth mentioning. One of these is the provision of ambulance services. Ambulance

¹² In the case of Bogota, as said above, all PAB services are contracted with ESEs, but this is not necessarily the case for other ESEs in the country.

services are contracted with the SOH separately from services to the uninsured and PAB services. Most ESEs have one or several ambulances that had been previously allocated by the central government. Thus, the national government allocates vehicles to public hospitals all over the country but hospitals are responsible for the recurrent expenditures related to their operation, namely salaries for the crews, equipment, maintenance and fuel costs. In the case of Bogota, the SOH decided in the mid nineties that it would pay for the services of these ambulances but take them to a central location, the Center for the Regulation of Emergencies (CRU) operated by the SOH. Accordingly, when ESEs contract with the SOH, they put their ambulances and their respective crews at the CRU and the CRU operates them. Put in other words, it is like a fleet of ambulances administered by the SOH, whose operating costs are outsourced to ESEs.

Other sources of revenues for ESEs, much smaller than the four listed above, are:

- Services delivered to enrolees in the EBS, mostly for emergency care.
- Contracts with local Mayoralties or Local Administrative Boards, either for the delivery of specific services or for capital investments.
- Services delivered for traffic accidents, which are billed to SOAT insurers.
- Services delivered to uninsured patients coming from other municipalities and departments.
- User fees (paid by the uninsured and the SS enrolees when they receive services outside the POS-S) and copayments (paid by the SS enrolees when they receive services included in the POS-S, and by EBS enrolees).
- Services delivered to patients from other ESEs.

Table 4.1. gives a summary view of services provided to each of the populations, and describes who pays and who provides the services. The smallest sources of revenues are not included for the sake of simplicity.

4.7. Empirical evidence on autonomisation in Bogota

The lack of empirical evidence on the effect of hospital autonomisation on efficiency, quality and equity in developing countries is similar in the Colombia case. Only three studies were found relevant to the Bogota case.

Table 4.1.
Summary of services provided to each of the populations
for an ESE

Population Services		Uninsured	Subsidised Scheme		Employment Based Scheme
			POS-S	non POS-S	
Curative care	Level I	SOH contracts with ESEs on a FGPP basis with a cap for billed services. Services are provided to open population. User fees are based on SISBEN level.	ARSs must contract at least 40% of the premium with local ESE for level I services	Not applicable	EPS contract with private providers, and seldom with public hospitals.
	Level II		ARSs must contract at least 10% of the premium with level II or III ESEs. The remainder can be contracted with private providers; some of them are vertically integrated to the ARS. Comprehensive coverage for the four levels of care is included only for maternal (prenatal, delivery and postpartum) and child care (during first yr of age).	Same as uninsured	
	Level III			Same as uninsured	
	Level IV			Not applicable	
	Promotion and Prevention (P&P)			ARSs contract with ESEs, usually within the 40% compulsory contracting	
Public Health (PAB)	SOH contracts with ESEs for PAB interventions. ESEs provide them directly or subcontract some of them to private providers				
Ambulances	SOH pays a fixed amount per ambulance per month for unlimited transportation services to the uninsured. Ambulances are physically put at the CRU and administered by the CRU.	ARS pays for transportation services provided to its enrolees, whenever they relate to services included in the POS-S.	Same as uninsured	Very unusual, but when it occurs, ESE bills the EPS for the transportation services.	

- Light shaded areas show the services that are paid by the SOH. Dark shaded areas show the services that are paid by ARSs.

4.7.1. Acosta et al's (2000) study on irregular behaviour of hospitals

This study analysed if hospital autonomy influenced the level of irregular behaviours by analysing changes in the prices of six tracer drugs purchased at the 32 hospitals as compared to the prices of the largest distributor for the public network of hospitals. The authors hypothesised that higher levels of autonomy lead to lower levels of price

variation, because the procurement process is more transparent. To measure the effect of autonomy, the authors included several variables to reflect the extent of autonomy of each ESE: share of workers hired outside the civil service legal framework, share of revenues represented by billed services, and the existence of managerial tools. The findings are not conclusive, first, because it is difficult to conclude that autonomy is well described with these variables, and second, the pooled effect of the autonomy variables was not significant.

The study also showed that the influence of the Ministry of Health on hospitals had been strongly reduced, whereas the influence of the SOH had increased. It also found that although in theory all ESE managers had the same nature and extent of autonomy, they had taken advantage of this decision space with different levels of intensity. This was evident from the wide variation in the proportion of workers hired outside the civil service regulations, and from the share of revenues coming from services billed to third party payers other than the SOH. The soft budget constraint problem is also highlighted when they comment that no matter that the amount of billings was inferior to the operating costs of the ESE, the SOH filled the gap with block grants without any accountability. Regarding human resources management, the study found that procedures to fire civil service workers lack efficacy, which created barriers to reduce this costly component of labor.

4.7.2. McPake et al's (2003) longitudinal study of ESEs.

One of the few studies in the literature that compares hospital data before and after transition into autonomy is that of McPake et al (2003). This study analyses eight years of data in five hospitals in Bogota, including input, output, productivity and quality indicators. They challenge Preker and Harding's "market forces model" which proposes that as hospitals increase market exposure, their quality and efficiency improve. McPake et al's main argument is that transaction costs are very high due to difficulties in defining products and observing/verifying products and outcomes. These information problems are at the root of contract incompleteness and they open room for opportunistic behaviour on the hospital side. Thus, it is possible that quality decreases as opportunistic behaviour concentrates effort on observable and verifiable features of products, but not on the really important issues of technical quality. In addition, autonomous hospitals will choose to increase their production of services that show

higher price-cost ratios and reduce those with the lower ratios. This leads to changes in hospital case mix as a response to market incentives. But given that price-cost ratios are not observable, Preker and Harding's market model cannot predict these changes.

This study selected a purposive sample of five ESEs to test its hypotheses, collecting data for the period 1990 to 1998. However, as data collection went further back to 1990, increasing missing data was an important limitation to detection of clear trends in indicators. Even with these restrictions, the study finds that many trends in indicators were seen before 1997, and it was difficult to attribute the observed changes to the transition into autonomous hospitals. Although quality indicators were analysed for 1997 and 1998 only, a very short time to detect variations, the qualitative component of the study reveals concern about excess emphasis on costs and efficiency, and less emphasis on technical quality. The evidence of the study does not contradict the market forces model, but does not support other models that predict decreased quality.

4.7.3. Castano et al's (2005) study on boards of directors of ESEs

This study analysed five ESEs in Bogota, selected purposively to detect differences by level of complexity and institutional development. The study aimed at analysing the role of the boards of directors as a corporate governance body. The study found that board representatives did not clearly identify the community as the ultimate principal and did not have a clear and shared idea of the objective function of the ESE. The three principals (the government, the community and the hospital professional workers, as stated by the law) showed clear conflicting agendas and sometimes they responded to personal agendas or those of the specific group they represent, but not the agenda of the ESE. There were no devices in place to hold board representatives accountable to the groups they represent.

It was also found that the largest time of the board meetings were dedicated to issues of budget execution. This reduced the effectiveness of the board in addressing strategic issues and keeping the manager accountable, and restricted the participation of representatives that are less skilled to deal with budget issues. This lack of skills was particularly relevant among community representatives. Involvement of community representatives in corporate governance was more direct in the level I ESEs than in the higher complexity ones. The role of the board in holding the manager accountable for

hospital performance was found very weak, as no formal and routine mechanisms to evaluate the manager were found. In addition, the manager tended to show a more conciliatory behaviour when approaching reelection, as compared to the start of an office term when a more aggressive behaviour was reported.

4.8. Concluding remarks

Hospital autonomy in the Colombian context exhibits some aspects that make it different from the experiences reviewed in chapter 3. One important difference is the delegation of the purchasing function from the local health authority to third party payers. Although it did not advance to full delegation because of the non-comprehensiveness of the benefit package and the lack of universal enrolment, the role of these third parties is key to enhance the transformation of subsidies from the supply side to the demand side. Another important difference is the competitive approach to the purchasing function, which also plays a key role in increasing hospitals' exposure to market forces. A third characteristic is the extent of autonomy that has been granted to hospitals, particularly in Bogota. This experience goes far beyond the limited extent of autonomy that is seen in Sub Saharian Africa countries, but the SOH still keeps the control knobs that make it less *laissez-faire* than the Chinese experiences. It could be argued that ESEs show more market exposure than any other experiences (except for China), which gives room to explore the effects of market exposure on hospital behaviour, in the presence of some degree of control by the local health authority.

An overview of the relevant theoretical underpinnings of hospital autonomy through the prism of the experiences reviewed in chapter 3 and the case of Colombia and the Bogota ESEs, makes clear the knowledge gaps that call for further research in the topic. It is clear that hospital autonomy rests on a TCE assumption, i.e., a contract-based relationship is expected to solve the maladaptation problems of a vertically integrated structure. However, except for the cases of the United Kingdom and New Zealand, TCE has received little or no attention in both policy making and research. In addition, the evolution of the relationships between public hospitals and third party payers is expected to be influenced by transaction costs. This role of transaction costs has also been ignored in the research, and it is the main objective of this research, as shown in the next chapter.