Achieving geographic equity in the Portuguese hospital financing system

Mónica Duarte Correia de Oliveira
Department of Operational Research,
London School of Economics and Political Science

Thesis submitted to the University of London for the Degree of Doctor of Philosophy

April, 2003

To my parents and sister

ABSTRACT

The Portuguese health care system is based on a national health service structure. The Portuguese government has with various statements over time shown that it is seeking some kind of geographical equity but this has never been clearly defined. There are wide inequalities in the distribution of hospital resources in Portugal with marked concentration in urban coastal areas and little information. The objective of the research described in this thesis is to develop methods to inform the allocation of resources to Portuguese hospitals so that this can be made more equitable in both current and capital spending. The methods used are a combination of methods already used in other countries and new methods to address two questions. First, to measure inequities in hospital care in terms of capital, finance and utilisation using capitation formulas. These formulas are constructed using: a multiplicative model to measure need for hospital care; a multilevel model to estimate unavoidable costs and to disentangle allocative inefficiencies of hospital care; and a flow demand model to predict hospital geographical utilisation and to compute cross-boundary flows. Second, to indicate how redistribution of hospital supply will best improve equity of utilisation and access, using location-allocation models that were designed to consider alternative policy objectives and account for patients' choice of hospitals.

ACKNOWLEDGMENTS

I am deeply indebted to Professor Gwyn Bevan for being an extremely attentive and committed supervisor. Without all his effort and patience, all the progress and learning process throughout the last three and a half years would not have been both as challenging and pleasant as it has been. His intellectual contribution and guidance on how to carry out research have greatly marked my academic attitude.

This research has been funded by the scholarship of *Fundação para a Ciência e Tecnologia*, *Ministério da Ciência e Ensino Superior*, Portugal (schoolarship co-funded by the European Social Fund, Structural Fund III; grant BD 19972/99).

I would also like to express my gratitude to Professor Pedro Pita Barros, Mr. José Fernandez, Dr. Elias Mossialos, Professor João António Pereira and Professor Carlos Gouveia Pinto for their availability, help and insightful comments.

I am thankful to others who have always been helpful at different stages of the research: Professor Walter Holland, Ms. Céu Mateus, Mr. David McDaid, Professor Alistair McGuire, Ms. Brenda Mowlam, Dr. Adam Oliver, Ms. Filomena Parra, Dr. Susan Powell and Ms. Céu Valente.

I am grateful to the LSE Health and Social Care and to the Operational Research Department at the LSE for providing me with excellent working conditions and a very intellectual and stimulating environment in which I could develop my research. I acknowledge the financial support of the UK Health Economics Study Group, which allowed me to attend its meetings. I thank the Portuguese National Institute of Statistics and the Financial Institute for Informatics and Management of the Portuguese Ministry of Health who provided data.

I am grateful to the participants of the seminars and conferences where various studies of this thesis were presented*. I specially thank Dr. Zaid Chalabi and Dr. Andrew Street for their detailed comments as discussants of the papers presented in the meetings of the UK Health Economic Study Group.

Parts of this thesis have been published elsewhere: working paper Oliveira, M. and G. Bevan (2001). *Measuring geographic inequalities in the Portuguese health care system: An estimation of hospital care need.* Lisboa, Associação Portuguesa de Economia da Saúde: 35; Oliveira, M. (2002); and discussion paper *A flow demand model to predict hospital utilisation.* London, LSE Health and Social Care: 57. An adapted version of Chapter 5 of this thesis was accepted for publication in the Health Policy Journal (March 2003).

I thank Dina Davaki for accompanying and supporting me throughout most of this journey. I especially thank Leandro Rothmuller for giving me strength during the last stage of the thesis. I especially thank Sofia Silva for all her support in the difficult times. I thank my friends Dimitris Boucas, Ana Carvalho, Carmen Constança, Martin Cumpa, Filipe Garcia, Carsten Holbraad, Balazc Kotnyek, Heather Lannin, Peter Lutz, Yannis Mourtos, Lídia Oliveira, Pedro Oliveira, John Tan, Andrea Volfova and the *Berloquenhos* for their positive energy and encouragement.

* Parts of this thesis were presented in the following conferences and seminars: 7th National Meeting of Health Economics, organised by the Portuguese Association of Health Economics, Lisboa, November 2000; LSE Health and Social Care seminar series, May 2001; LSE Health and Social Care informal seminar series, December 2001; 60th Meeting of the UK Health Economics Study Group, Norwich, January 2002; London and South East Operational Research Society, London, March 2002; 22nd Meeting of the Spanish Health Economics Association, organised by the Spanish Association of Health Economics, Pamplona, May 2002; 4th European Conference on Health Economics, organised by the *College des Economistes de la Sante*, Paris, July 2002; LSE Health and Social Care informal seminar, November 2002; 62nd Meeting of the UK Health Economics Study Group, Leeds, January 2003.

Cada qual com seu igual. Each one with his equal.
A Portuguese expression

La première égalité, c'est l'équité. The first equality is equity.
Victor Hugo, *Les Misérables*

Las majestueuse égalité des lois, qui interdit au riche comme au pauvre de coucher sous les ponts, de mendier dans les rues et de voler du pain. The majestic equality of the law forbids the rich as well as the poor to sleep under bridges, to beg in the streets and to steal bread.
Anatole France, *Le Lys rouge*

ABBREVIATIONS

AIC: Aikeke Information Criteria

ASMR: Age Specific Mortality Rates

CBF: Cross-Boundary Flows

DBM: Distance Based Model

DEA: Data Envelopment Analysis

DHA: District Health Service Authority

DRG: Diagnostic Related Groups

EM: Entropy Model

EU: European Union

FCE: Finished Consultant Episode

FDM: Flow Demand Model

GDP: Gross Domestic Product

GLM: Generalised Linear Model

GLS: Generalised Least Squares

GP: General Practitioner

HFEM: Hierarchical Fixed Effects Model

IGIF: Institute for Financial and Informational Management

LOS: Length of Stay

MLM: Multilevel Model with random intercepts and slopes

MoH: Ministry of Health, Portugal

MP: Mathematical Programming (model)

n/a: not available

NF: Net Flow method

NHS: National Health Service

NUTS: Statistical Nomenclature of Territorial Units

OECD: Organisation for Economic Cooperation and Development

OLS: Ordinary Least Squares

PC: Primary Care

PIDDAC: Program of Investments and Expenditure for Development of the Central Administration, Portugal

PF: Proportional Flow method

PPI: Purchasing Power Index

PPP: Purchasing Power Parities

PYLL: Potential Years of Life Lost

RAWP: Resource Allocation Working Party

RHA: Regional Health Authority

RMI: Relative Mortality Index

SFM: Stochastic Frontier Model

SIM: Spatial Interaction Model

SMR: Standardised Mortality Ratio

SSMs: Sample Selection Models

TI: Treatment Intensity method

TPM: Two-Part Model

UBM: Utilisation Based Model

UC: Unavoidable Costs

UFBM: Utilisation Flows Based Model

UK: United Kingdom

USD: United States Dollar

US: United States

VHI: Voluntary Health Insurance

NOTATION

Notation	*Interpretation*
r	r is a geographic district unit (district; for Portugal, r =1,2,..18).
cap_index_r	Relative capitation index for district r, accounting for all the selected adjustments of the capitation formula.
P_r, P	Resident population in district r and total resident population.
I_{2r}	Age and additional need index for district r.
I_{3r}	CBFs index for district r.
I_{4r}	UC index for district r.
$District\_share\_1_r$	Share of need for hospital care for district r.
$District\_share\_2_r$	Share of need for hospital care, adjusted by CBFs for district r.
$District\_share\_3_r$	Share of need for hospital care, adjusted by CBFs and UCs for district r.
a	Age group a.
X_{1a}	Age (and sex) cost for age (and sex) group a.
de_{ar}	Number of deaths in area r from the age group a.
P_{ar}	Resident population of the age group a in area r.
r_{ar}	Death rate in area r from the age group a, which corresponds to the definition of age specific mortality rates for area r and for age group a.
r_a	National death rate for age group a.
$cutoff$	Age reference used in the computation of the potential years of life lost index. It is related to life expectancy.
I_a	Mid-age point of age group a (required to compute the potential years of life lost index).
SMR_r	Standardised mortality ratio index for district r.
$ASMR_{ar}$	Age specific mortality ratio index for age group a and for district r.
$PYLL_r$	Potential years of life lost index for district r.
RMI_r	Relative mortality index for district r.
h, h'	Hospital identifier ($h \neq h'$).
c	Types of hospital in the administrative (and hierarchical) classification (for Portugal: c = general central, specialised central, district, level I).
k	Geographical place of location.
l	Type of hospital in the costs' statistics classification (for Portugal: l = central,

	district, level I).
$COutput_h$	Total cost standardised by an index of hospital production. This indicator is referred to as standardised cost.
$TotCost_h$	Total cost.
$OutputIndex_h$	Equivalent patients index.
$Disch_{hl}$	Number of hospital inpatient discharges of hospital h that belongs to hospital group l.
$Outpat_{hl}$	Number of outpatient attendances of hospital h that belongs to hospital group l.
$Emerg_{hl}$	Number of emergency and accident admissions of hospital h that belongs to hospital group l.
a_l, b_l, c_l	Total unit costs from hospitals of type l, for inpatient discharges, outpatient attendances and emergency and accident admissions, respectively.
do_h	Numbers of doctors.
nu_h	Number of nurses.
be_h	Number of beds.
C, C'	Function linking the standardised cost with the covariates; and linear function linking the natural logarithm of standardised cost with the covariates.
α, β, θ	Parameters from the general hierarchical model.
$x_h^{'}$, $x_h^{''}$, x_h	Explanatory variables vector for standardised costs (x_h). $x_h^{'}$ is the sub-set of variables that have a log-linear function relationship with the dependent variable ($x_h^{'} \subset x_h$); and $x_h^{''}$ is the sub-set of variables with a semi-log function relationship with the dependent variable ($x_h^{''} \subset x_h$).
e_h	Random error for the general hierarchical model.
α_0, α_1	Coefficients of the fixed part of the HFEM (excluding the geographical and hospital group related coefficients).
g_{hk}	Dummy variables for the geographical location of hospital h in place k (HFEM and MLM).
α_{2k}	Fixed coefficients for dummies of the geographical area k (geographical related coefficients) (HFEM).
t_{hc}	Dummy variables for the hospital h in the administrative hierarchy c (HFEM).
α_{3c}	Fixed coefficients for dummies of the administrative group c (HFEM).
e_{hck}^{HFEM}	Random error for the HFEM.
$\beta_0, \beta_1, \beta_2, \beta_3$	Coefficients of the fixed part of the cost model (excluding geographical-related and hospital group related coefficients) (MLM).
β_{4k}	Fixed coefficients for dummies of the geographical area k (geographical-

	related coefficients) (MLM).
β_{0c}	Random coefficient of the random intercept of the MLM, defined at the hospital administrative group c.
β_{1c}, β_{2c}	Random coefficients of the random slopes of the MLM, defined at the hospital administrative group c; β_{1c} and β_{2c} are the random coefficients of the nurses to doctors and beds to doctors ratios, respectively.
μ_{0c}	Random component of the random coefficient of the MLM, defined at the hospital administrative group c.
μ_{1c}, μ_{2c}	Random component of the random slopes of the MLM, defined at the hospital administrative group c.
e_{hck}^{MLM}	Random error at the hospital level (MLM).
$\sigma_{\mu0}^{2}$, $\sigma_{\mu1}^{2}$, $\sigma_{\mu2}^{2}$	Variances of the random components of the model at the group level. $\sigma_{\mu0}^{2}$ is the variance of the random component of the intercept, while $\sigma_{\mu1}^{2}$ and $\sigma_{\mu1}^{2}$ is the variance of the random component of the slopes (MLM).
σ_{e0}^{2}	Variances of the error term at the hospital level (MLM).
$\sigma_{\mu0\mu1}$, $\sigma_{\mu0\mu2}$, $\sigma_{\mu1\mu2}$	Set of covariance between the random components, defined at the group level (MLM).
i, i', v and q	Population points representing small area population units. Each i, i', v and q belongs to one district r ($i,i',r,q \in r$) ($i \neq i' \neq r \neq q$).
n	n is the number of population points.
j, w and z	Hospital points representing hospital site geographic units. Each j, w and z belongs to one district r ($j,w,z \in r$) ($j \neq w \neq z$).
m	m is the number of hospital points, which is a sub-set of the total number of population points n ($m \subset n$).
U_{ij}	Utilisation flow between population point i and hospital site j.
D_j	Size of hospital site j.
$\widetilde{D}_{ij}$	Index for alternative supply to hospital site j available for population i.
$other_{ij}^{'}$	A set of other variables related with population and hospital characteristics that explains flows.
$other_{i}^{''}$	A set of population-related variables that explains flows.
$other_{j}^{'''}$	A set of hospital-related variables that explains flows.
P_i	Resident population in i.
Dem_i	Demographic characteristics of the population (age and sex) that imply higher need for hospital care for population i.
N_i	Need for hospital care for population i

X_i	Socio-economic level of population i.
G_{ij}	Accessibility costs for population i to access hospital services in j.
d_{ij}, $d_{ii'}$	Distance between population point i and hospital site j, and between population points i and i' (Euclidean distances as defined in Chapter 3).
A_i	Perceived availability of hospital care to population i.
I_{ij}	Set of institutional characteristics of the hospital system (such as hospitals hierarchy, sites with hospital teaching functions, spatial hospital subsystems, etc), to be specified below. Some of these characteristics relate to population points.
O_{ij}	Set of variables that characterise access to other sectors of health care and non-health care systems (such as welfare system and private supply) and other variables that are expected to influence demand for hospital care –such as spatial variables along the territory.
PC_i	Accessibility to primary care for population located in i.
c_j	Role of hospital j in the hospital hierarchy (for example, dummy variables for central and district hospitals).
$i1_{ij}$	Indicator of whether hospital j is the first hospital of use by population i (dummy variable).
$i2_{ij}$	Indicator of whether hospital j is the second hospital used by population i (dummy variable).
$i3_{ij}$	Indicator of whether hospital j is the central hospital used by population i (dummy variable).
$i4_j$	Vector of hospital variables that characterise hospital j outputs other than inpatient care (such as external consultations and emergencies).
$i5_j$	Vector of variables representing the hospital input mix of hospital j (labour vs. equipment vs. beds).
y	Utilisation variable as a dependent variable.
x	Set of the covariates that are hypothesised as affecting utilisation.
x' and x''	Two sub-sets of covariates of the set x ($x' \subset x$ and $x'' \subset x$).
d'_{ij}, $d''_{ii'}$	Dummy on whether hospital j is within 25 km from population point i, and dummy on whether population point i' is within 25 km from population point i.
β''	Set of coefficients of the econometric model.
e^{ε}	Residuals in the natural scale of the second part of the two-part model.
$\hat{p}_{qw}$	Predicted probability of population point q making use of hospital site w.
$\hat{U}_{qw}$	Predicted level of utilisation flows of population point q to hospital site w, [illegible]

	given that the probability of that flow being positive is positive.
$UCOutput_h$	UC index for hospital h.
I_{1r}	Age adjustment index for district r.
$Catchment_r$	Catchment population of district r.
D_r, D	Discharges from hospitals of district r; total discharges in the system.
O_r	Discharges from the resident population of district r.
W_r, W	Population need for hospital care in district r (resident population weighted by age); total population need.
$W_r^{'}$	Population need for hospital care in district r, scaled so that total need sums up total discharges in the system.
$Flow\_1_{ij}$	Dummy **variable** for expressing whether population i is served by hospital j, as a first hospital (DBM) (0 or 1 values).
$Flow\_2_{ij}$	Dummy **variable** for showing whether population i is served by hospital j as a second hospital (DBM) (0 or 1 values).
$Flow\_c_{ij}$	Dummy **variable** for denoting whether population i is served by hospital j as the closest central hospital (DBM) (0 or 1 values).
d_1_i	Distance travelled between population point i and the first hospital of use (non-negative **variable** depending on $Flow\_1_{ij}$) (DBM).
d_2_i	Distance travelled between population point i and the second hospital of use (non-negative **variable** depending on $Flow\_2_{ij}$) (DBM).
d_c_i	Distance travelled between population point i and the closest central hospital of use (non-negative **variable** depending on $Flow\_c_{ij}$) (DBM).
W_i	Needs-weighted population at population point i (DBM). This is derived from weighting resident population per age group by the age weighting index estimated in Chapter 5.
$share\_1_i$	Share (%) of population i that is assumed to go to the first hospital (DBM).
$share\_2_i$	Share (%) of population i that is assumed to go to the second hospital (DBM).
$share\_3_i$	Share (%) of population i that is assumed to go to the closest central hospital (DBM).
U_1_i	Utilisation flow by population i to the closest hospital (DBM).
U_2_i	Utilisation flow by population i to the second closest hospital (DBM).
U_3_i	Utilisation flow by population to the closest central hospital (DBM).
$\overline{U}^N$	National utilisation rate ($\overline{U}^N = U^o / W$) (DBM/UBM).
$\widetilde{p}_{ij}$	Probability of population i using hospital j, as produced by the gravity

	model, with $\sum_j \tilde{p}_{ij} = 1, \forall i$ (UBM).
U_i^N	Normative utilisation for population area i depending on total national utilisation rate (non-negative **variable**) (UBM).
D_j^0, D^0, D	Current level of supply of hospital j; total current level of supply; total level of supply, computed within the model (UBM)
U_{ij}^o, U^0	(Past) flows and (past) total level of utilisation (DBM/UBM)
a_i	Auxiliary **variable** used to obtain an absolute value of difference between utilisation and expected utilisation, per population area i (UBM).
$f(\beta_j^{'''}, d_{ij})$	Decay function that relates the effect of distance (accessibility costs) from population i to hospital j (definition in Appendix E). The decay function might differ for hospital type and the decay parameter β_j will depend on the level of attraction between hospital j and patients located at different distances from that hospital (UBM).
$\beta_j^{'''}$	Parameter that defines the elasticity of utilisation in relation to distance, for hospital j (UBM).
$f\_\min$	Proportion of current level of supply of hospital j to be kept, as a minimum (UBM/UFBM)
$f\_\max$	Proportion of current level of supply of hospital j to be increased, as a maximum (UBM/UFBM)
$\min\_D_j$	Minimum level of supply of hospital j to be maintained (UBM/UFBM).
$\max\_D_j$	Maximum level of supply to be allowed for hospital j (UBM/UFBM).
$\log U_{ij}^r$	Distribution of the natural logarithm of utilisation flows that operates as the target. This target is a distribution formulated in accordance to some type of equity principle (in this case, patients making use of the closest hospital) (UFBM).
b_{ij}	Auxiliary **variable** for defining the difference between variations in the logarithm of utilisation flows (UFBM).
$\log \hat{p}_{ij}$	Logarithm of the probability of use, generated in the first part of the estimated two-part FDM, developed in Chapter 7 (UFBM).
$\log U_{ij}^{'}$	Natural logarithm of the utilisation **variable** between hospital i and hospital j, as defined in the second part of the two-part FDM, developed in Chapter 7 (UFBM).
$DumFirst_{ij}$, $DumSecond_{ij}$ and $DumCentral_{ij}$	Dummy for whether hospital j is the closest hospital to a population i; dummy for whether hospital j is the second closest hospital to a population i; and dummy for whether j is the closest central hospital to a population i (UFBM).

$DumLisboa_{ij}$, $DumPorto_{ij}$ and $DumCoimbra_{ij}$	Dummy for the central hospital site in Lisboa and for populations from the South; dummy for the central hospital site in Coimbra and for populations from the Centre; and Dummy for the central hospital site in Porto and for populations from the North (UFBM).
$oth\hat{e}rs_{ij}$	Parameter capturing the influence on flows of all the factors from the FDM, with the exception of the variables that relate to hospital supply (UFBM).
$\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\alpha}_5$, $\hat{\alpha}_6$, $\hat{\alpha}_7$	Parameters that relate utilisation flows and hospital supply, taken from the estimated flows demand model (estimated in Chapter 7) (UFBM).

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1 CHAPTER 1 - Introducing geographic equity

1.1 Introduction

Under Constitutional law, Portugal has a democratic state that is committed to building a society based on freedom, fairness and solidarity, and promoting welfare, quality of life of citizens and equality (Assembleia da República 1992). After the ending of almost 50 years of dictatorship, the democratic government created a National Health Service (NHS) with universal coverage in 1979, but the NHS is still far from achieving its intended objectives.

The NHS has universal coverage and is financed by general taxation, in which the State assumes a critical role in provision, financing and regulation. The State is responsible for ensuring a minimal level of access to health care for all citizens and seeks to achieve some kind of equity (discussed in Chapter 2). Citizens are entitled to health protection, under a universal NHS, nearly free at the point of use, where contributions should depend upon ability to pay. Whilst it is likely that there were equity gains after the introduction of the NHS (with universal coverage and increases in health care provision and utilisation), the current system suffers from multiple inequities, for example: in health, associated with socio-economic characteristics; in finance of health care, which is regressive; in the distribution of human and material resources, with concentration of material and human resources in urban areas; and in access, with segments of the population enjoying multiple coverage.

This thesis aims to generate information for the development of policies pursuing geographical equity in the hospital financing system, in light of the absence of significant relevant information. This chapter reviews the current understanding of inequity in the theoretical literature and empirical studies of inequity in the Portuguese context, defining the research areas to which this thesis seeks to contribute and the methods used, and outlining the structure of the thesis.

1.2 Geographical equity in health policy

This section explains the interconnections between equity, geographical equity and other policy objectives, analyses the conceptual significance of geographical equity, and problematises the issues involved in measuring geographical inequities. This discussion is focused on resource allocation.

1.2.1 Defining the equity concept(s)

There is a vast body of literature on equity, equity in health, and equity in health care in particular[1]. The following analysis draws on this literature to examine the concept of equity in health policy, the importance of analysing geographical equity, as well as a range of different measures of geographical equity[2].

Equity might be defined in several ways, "depending upon the values of the person using it at the time" (Le Grand 1987), but all definitions share "some view of fairness of the distribution of something or other" (Mooney 1983). In many health care systems, it is precisely because of equity of access that health care finance does not depend on willingness to pay. But making services free at the point of delivery is necessary, but not sufficient condition for attaining equity of access and other equity objectives, as patients incur other costs too. The pioneering work of the Resource Allocation Working Party (RAWP) defined equity as "equal access for those in equal risk" (Department of Health and Social Security (United Kingdom) 1976), which was the accepted aim for resource allocation in the United Kingdom (UK) until recently[3]; in Italy, the policy of promoting equity is "... to overcome territorial inequalities in social and health

[1] The equity discussion in the health context has debated the problem of placing alternative conceptions of equity in a wider philosophical framework (Le Grand 1987), something that has not been achieved up to the moment. Only fragmented and multidisciplinary perspectives exist and these have created a sense of failure (Pereira 1993).

[2] It is important to note that equity is examined here as a macro level objective, at the top of the political system. This implies that judgements are broad and not directly related to medical practice (Culyer and Wagstaff 1992), or to other micro health care unit issues.

[3] This is one possible definition that has been very useful in the resource allocation context and is consistent with a NHS structure that implies a commitment to equity in the availability and use of health

conditions" (Rice and Smith 1999). The concept of equity of access has been systematically referred to in policy statements of most Western European countries (including Portugal) (Pereira 1993)[4]. Chapter 2 shows that equity appears to be one of the fundamental objectives of Portuguese health policy, and in its various formulations is compatible with the concept of "equal access for those in equal need".

Any definition of equity implicitly involves a comparison of different individuals across the same or different circumstances –the former case is involved in the pursuit of horizontal, while the latter in the pursuit of vertical equity (McGuire, Henderson, and Mooney 1988)[5]. For publicly financed health care systems, the two main areas of equity research have been equity in access to health care across socio-economic groups/classes and geographically (McGuire, Henderson, and Mooney 1988). This thesis focuses on the second of these research areas. Geographical equity is a useful concept for both needs assessment and planning purposes (Department of Health and Social Security (United Kingdom) 1976) and provides the basis for the allocation of resources across areas.

Various definitions of geographical equity can be used, in terms of public expenditure, final income, use, cost and income (Le Grand 1982); expenditure per capita, inputs per capita, inputs for equal need, access for equal need, utilisation for equal need, marginal met need and health (Mooney 1983). This thesis makes use of some of these definitions that are related to equity of access. As none of these equity definitions has been proved to be superior to others, the choice of which definition to use is a matter of judgment and dependent on the specific context, while the objective chosen may be in conflict with other seemingly similar objectives (Culyer and Wagstaff 1992). For example, seeking equality of utilisation may be influenced by supplier induced demand, but, in comparison with the equality of inputs, equality of utilisation has the advantage of allowing for differences in tastes and preferences (Pereira 1990).

---

care services (Black et al. 1982) (Whitehead 1995). Recently, a new and additional definition has been used in England and is analysed in the sub-section below.

[4] Equity of access (for those in equal need) has been a key definition of equity for most health systems. Mooney has defined it as equal costs to patients, and has linked it with opportunities open to individuals (Mooney 1983). Equality of access points to the factors that might distinguish different populations in the process of accessing to health care services when they perceive the need for treatment ("individuals making choices under equal constraints" (Le Grand 1987)), thus being a concept mainly concerned with the supply side.

1.2.2 Geographical equity

A number of issues ought to be considered when addressing geographical equity in health policy. First, geographical equity in health care competes with other concepts of equity, such as socio-economic equity and equity in health outcomes. This brings up the old question of whether inequity is influenced more by individual or by contextual factors (such as place) –although these are often correlated (Duncan, Jones, and Moon 1998). Geographical equity in access to health care might also be seen either as an ultimate or as an intermediate objective to achieving equity in health[6]. For example, although England has equalised health expenditure across areas (while accounting for area characteristics), inequalities in health have persisted. It has been argued that focus on spatial inequity can obscure other kinds of inequalities, such as those of race and class, and risks losing sight of the structural basis of inequality (Johnston, Gregory, and Smith 1994).

Second, one should define which geographical inequalities are inequities. As Mooney et al. observed (Mooney and McGuire 1987): "it is possible to have equitable inequalities and inequitable equalities". Inequities can be seen as 'unfair' inequalities and the relationship between inequalities and inequities is complex. In general, geographical location is expected to impact on costs to access health care in three ways (Rice and Smith 1999): variations in need, variations in health care supply and policy, and variations in the extent to which need is expressed in utilisation. For example, this thesis shows that Lisboa and Porto have a higher share of resources than their fair shares based on need for hospital care; but their shares of resources in utilisation and finance (when accounting for the impact of variations in supply) are below their fair shares. Consequently, any formulation of policies to correct inequities of access will demand a definition of which inequalities matter most. The methods used in this thesis partly test the impact of pursuing alternative definitions of equity, and show the difficulties associated with the design of policies for equity.

---

[5] In practice, the pursuit of vertical and horizontal equity objectives is often conflicting (example: measurement issues).

[6] The complex relationship between health care and health is further developed in the next sub-section.

Third, geographical equity relates to the concept of territorial justice. If there was local choice over expenditure on the NHS, then this could result in inequities in health resources across areas as an outcome of local democratic choices[7]. Hence, analysis of the geographical distribution of resources should account not only for need, but also for need and right (Powell and Boyne 2001), which would take into account the distribution of health care resources to regions as a result of devolution. In Portugal, NHS expenditure is determined centrally and equalisation of resources between geographical areas is consistent with the principle of horizontal equity of access of individuals in different jurisdictions of residence[8]. This thesis focuses on analysis at the central level to improve equity of access across geographical areas. It disregards considerations of the implications of local choice for justifying variations in the use of hospital resources. This is understandable in the context of central planning and means focusing on the creation of similar opportunity sets for health outcomes across areas (in terms of provision of hospital services).

1.2.3 Problems in formulating and using a definition of equity

Even if the objective of equity is defined as "equal opportunity of access to health care for those at equal need" (in the geographical context), difficulties still remain. Due to problems of measuring personal access costs[9] (Le Grand 1987), this objective has been redefined as equal inputs for equal need (Mooney and McGuire 1987). But there are problems in measuring need in the light of continuing debate over which proxies are best and over supply issues (in terms of capital stock, staff and variations in labour costs)[10],[11].

[7] Local decisions allow for choice and local preferences, while central decisions look at central funding, regulation and the relationship between scale efficiency and equity.

[8] The exception for this are Azores and Madeira for which there is political and financial devolution: governments of the islands receive blocks of expenditure from central government and decide how much to spend in different social areas. As explained in Chapter 2, the islands are excluded from the analysis in this thesis.

[9] Personal costs are here defined as indirect costs related with specific circumstances of a population, excluding geographic accessibility issues –e.g. the impact that living alone or having children (mainly for women) might have on health care access.

[10] There are many difficulties in comparing between people (population characteristics) and among services, on a wide range of health care variables (physical and human resources, expenditure, capital, etc). The choice of these variables is subjected to judgment.

[11] In addition, as described above, there are other policy objectives with which the principle equal opportunity of access for those at equal need might conflict, such as with economic efficiency (for example, economies of scale) (Musgrove 1999).

Additionally, and also related to the point made above on the instrumental role of equity, a focus on improving health care does not necessarily lead to better health for the population in question. While health is related to the physical and emotional well being of an individual or a defined population, health care is related to the goods, services, time, knowledge and other variables that can be seen as inputs to produce health (Folland, Goodman, and Stano 1997). Even if the ultimate objective is to improve health on the whole, it is not clear that equity of access to health care should be the main health policy objective (Pereira 1993). During the 1980s and 1990s, England accepted that tackling health care inequalities would contribute to equity improvements and the focus on health care instead of health has been based on the main argument that health care can be redistributed by health policy, while health itself cannot (Le Grand 1987). However, empirical studies have showed that despite health care policy interventions with equity objectives, the gap on health status has been widening in the UK (Macintyre 1997) and in Sweden and the Netherlands (Whitehead 1992)[12]. Even with a reduction of inequalities in access to health care, health inequalities might widen as "the health production function is complex and variables other than health care can often have higher health benefits at the margin" (Maynard 1999) and it is unknown which policies better target inequalities in health. The recent debate in the UK on equity has resulted in a shift of emphasis from equal opportunity of access to health care to contributing to the reduction of avoidable inequalities in health, thus changing the focus from resources to outcomes.

Depending on the objectives of a study, one might focus on equal (geographical) access to inputs, outputs or outcomes; on the other hand, most countries have an instrumental interest in some sort of geographical equity of access. However, for any country, a clear definition of equity is required, if policies aiming at equity are to be implemented.

[12] In the UK, even if geographic allocations have converged to the targets set in the 1980s (Holland 1986), health policy has been ineffective in decreasing standardised mortality ratio differentials in the same areas (Macintyre 1997). The evidence on convergence of geographic levels of expenditure of health care resources throughout the 1990s is much weaker (Le Grand and Vizard 1998).

1.3 Why geographical equity in the Portuguese hospital sector?

This section justifies the focus upon geographical equity in the context of the hospital sector, and describes what is and is not known about health inequalities and geographical equity in Portugal. The role of equity as a political value is discussed in Chapter 2.

1.3.1 Focus on the public hospital system

In common with other countries, Portugal has an objective of "adequacy and equity in access to some minimum of health care for all citizens" (OECD 1994) but it has lacked satisfactory means of implementing that objective (Pereira 1995)[13].

This thesis focuses on the public hospital acute care sector because it accounts for a high proportion of the public health budget[14]. The thesis includes to some extent analysis of private hospital supply, for which only limited information is available. By contrast, this thesis does not consider the psychiatric hospital sector, for which there is also limited information and which is planned differently. The next sub-section presents evidence of inequalities in the distribution of health and health care resources and of the lack of information on the hospital sector.

[13] In this respect Portugal is differentiated from: a) countries that have made the tackling of inequalities in health an explicit priority in political statements and an objective in the design of health strategies and targets (such as Australia, Canada, Finland, the Netherlands and Wales) (Benzeval, Judge, and Whitehead 1995); b) countries which explicitly pronounce the pursuit of well-defined definitions of geographic equity –that is the case of Italy ("to overcome territorial inequalities in social and health conditions") (Rice and Smith 1999). Nevertheless, it is not always clear that policy makers know what equal access for equal need means (Culyer and Wagstaff 1993).

[14] It is acknowledged that there are pros and cons in using a disaggregated component of the health care budget or targeting one health care sector for analysis. This strategy has more potential for improving equity of access to hospital care (Benzeval, Judge, and Whitehead 1995) and allows for more specific analysis, but it might exacerbate the degree of inequities in access to hospital care (Judge and Mays 1994) and might imply a loss of information of inter-relations between the hospital sector and other sectors.

1.3.2 Documenting health inequalities in Portugal

1.3.2.1 Inequality of health outcomes, finance and economic accessibility

As shown in Chapter 2, over the last two decades, the health of the Portuguese as measured in terms of life expectancy and mortality has been improving and converging with European values at a fast rate. However, as Lucas, Pereira and Giraldes have found, there exist inequalities in health by socio-economic groups. Lucas (Lucas 1986) observed (in the Lisboa area) that manual workers were three times more likely to report illness than professionals, employers and managers. Pereira found strong evidence that the distribution of ill-health is generally unfavourable in poorer income groups (Pereira 1995). Giraldes (Giraldes 1998) found that the higher the socio-economic level (measured by years of schooling, income and occupation), the lowest the level of morbidity.

Studies using different data and methods of analysis have examined inequities in the finance of health care in Portugal (Table 1.1). Most studies reported decreases in progressivity during the 1980s and found that at the beginning of the 1990s, the system of health care finance was "mildly regressive". The study by Wagstaff et al. (Wagstaff et al. 1999) of 13 developed countries found Portugal to be the only one with a regressive health care financing structure. Sensitivity analysis by Pereira (Pereira 1998) of earlier studies showed their findings on regressivity to be robust to changes in methodological assumptions.

Table 1.1: Studies on equity in health care finance

Authors	*Data under analysis*	*Conclusions*
Pereira and Pinto (Pereira and Pinto 1990)	Portuguese Family Income and Expenditure Survey -1980/81	"Mildly regressive" financing system
Pereira and Pinto (Pereira and Pinto 1993)	Family Income and Expenditure Survey -1980/1	Health care finance "slightly progressive"
Pereira (Pereira 1995)	Family Budget Surveys -1980/81 and 1989/90	Health care finance has evolved from "overall progressive to overall regressive"
Pereira (Pereira 1996)	National Statistic Institute health household budget surveys -1980/1 and 1989/90	Change from a "mildly progressive" health care financing system in 1980 towards a "moderately regressive" structure at the end of the 1980s

The main reasons why health care finance was regressive at the beginning of the 1990s was the high levels of expenditure on pharmaceuticals (with heavy copayments), the system of taxation that became less progressive in the 1980s, the structure of tax deductions on health care expenditure, as well as the levels of reimbursement of health expenditure for populations under double coverage (public and private). Chapter 2 gives detailed evidence of these changes.

Inequities in access are evident, as 25% of the population benefits from double or multiple coverage. These beneficiaries are allowed to choose their providers of care, have their expenditures reimbursed on a fee per item basis, and are highly subsidised by the state (via tax deductions). By contrast, the population with NHS coverage only has limited or no choice; this is further discussed in Chapter 2.

The next section summarises what is and is not known about geographical inequities in Portugal.

1.3.2.2 What is known and not known about geographical equity in Portugal?

Many authors have dealt with the inequitable spread of human and material resources in Portugal:

- **Staff.** Doctors are inequitably distributed across regions (Pereira et al. 1987) (Giraldes 1995b) (Pereira et al. 1999). There is a lack of doctors in remote areas, with doctors being highly concentrated in Lisboa, Porto and Coimbra (Giraldes 1995a). These three urban centres accounted for 79.5% of doctors in 1981 (compared to 75.2% ten years earlier). Campos (Campos 1984) detected shortages in certain medical specialties and in paramedical staff, and concentration of doctors in three urban centres. An uneven distribution and shortages of human resources have been problems since the creation of the NHS.
- **Equipment.** Heavy equipment is mainly located outside NHS hospitals, in private facilities, while technology is concentrated in the coastal and urban areas (Pereira et al. 1999).
- **Private provision.** Private provision is heavily concentrated in those regions where NHS supply is more extensive (Pereira et al. 1999).

Other characteristics of the health care system reinforce the uneven distribution of hospital resources, in particular low state provision of community services and little continuity of health care.

The inequitable distribution of health care resources throughout Portugal has resulted in people having to travel for certain treatments or tests (Pereira et al. 1999) and differences in accessibility have influenced utilisation (Santana 1999), which seems to be lower in areas with poorer economic conditions (Santana 1993). Pereira et al. (Pereira et al. 1987) found high variations in utilisation indicators at the district level for hospital inpatients, primary care (PC) consultations and prescribed medicines[15], and evidence on the 'inverse care law'. Moreover, coastal districts were shown to have concentrations of supply of acute care services, younger populations, the highest

[15] Comparing health care in poor industrial areas with affluent salubrious areas (in the UK), Tudor found that "the availability of good medical care tends to vary inversely with the need of the population served" (Tudor 1971). Tudor called this as the 'inverse care law' and although several authors have criticised this study, the 'inverse care law' concept has been proved powerful and has been widely used in literature that links geographic equity with resource allocation.

population growth, the highest socio-economic indicators and greater opportunities for economic growth (Oliveira and Bevan 2001).

Studies have found that the geographical distribution of PC resources is not related to need. In two studies, Giraldes (Giraldes 1988) (Giraldes 1990) has suggested that in order to achieve equity, there should be a high geographical redistribution of PC expenditure, mainly from southern districts to northern districts, which had traditionally the worst health situation. Giraldes (Giraldes 1990) has indicated that positive discrimination would favour the north of the country (Bragança, Vila Real and Viseu) and the main losers would be Évora, Santarém, Portalegre and Lisboa[16]. In another study, Giraldes (Giraldes 1989) has shown that the components of PC expenditure varied widely across districts in 1983.

The first capitation study to be used for allocating health care funds by the Ministry of Health (MoH) was applied to the 1998 PC budget, and allocated 8% of the PC budget to the five Regional Health Authorities (RHAs) (with the remaining 92% decided by incremental budgeting) (IGIF 1998). That capitation formula included an adjustment for the age structure of the beneficiary NHS population (based on the frequency of consultations in primary care centres) and has been incrementally changed in recent years by including additional adjustments to capture the burden of illness. This capitation formula also confirmed wide inequalities in the provision of PC at the health region level. A more detailed description and assessment of these formulas is given in Chapter 4. However, there has been no comprehensive study of inequities in the distribution of hospital resources.

1.4 Objectives of this thesis

This section describes the research questions to be addressed in this thesis, the methods used and the structure of the thesis.

[16] In this study, Giraldes has further decomposed primary care expenditure by sub-budget areas and Lorenz curves revealed that transfers to private hospitals were the primary care expenditure component with highest geographic inequality.

1.4.1 Research questions

As described above, the evidence of inequities in the distribution of hospital resources suggests the presence of the 'inverse care law', as supply tends to be concentrated in the most urbanised and developed areas and is unlikely to correspond to need. Briefly, this thesis attempts to contribute to two broad questions that are sub-divided into more specific research questions:

1. How to measure geographical inequities in the Portuguese hospital system:

✓ How to measure need for hospital care in the Portuguese hospital system?

✓ How to measure unavoidable costs of hospital care in the Portuguese hospital system?

✓ How to estimate cross-boundary flows in the Portuguese hospital system?

✓ How to measure geographical inequities in the Portuguese system when alternative equity concepts are used (inequities in terms of capital, utilisation and finance)?

2. How to begin redistributing supply to promote equity?

1.4.2 Methods used

The aim of this thesis is to produce quantitative information for formulating policies to correct inequities. In particular, it uses a multidisciplinary approach in modelling that draws on different disciplines, such as health policy, health economics, operational research and geography.

This work seeks to transfer methods developed in England to Portugal, when these are available and suitable for the Portuguese context. England has the longest tradition in the development of quantitative methods in resource allocation and the most sophisticated capitation formulas of NHS countries that follow an index approach (Rice and Smith 1999); this is discussed further in Chapter 4. This thesis has also developed new methods, for example, unavoidable costs are estimated using a model that deploys recent econometric techniques of multilevel modelling.

As underlined above, this thesis uses various formulations of equity. The use of different equity concepts aims at informing specific policies and at illustrating conflicts between different equity objectives (this can be seen also as a form of sensitivity analysis).

1.4.3 Structure of the thesis

The chapters of the thesis are structured in three sections addressing the problem of achieving geographical equity in the Portuguese hospital financing system and the research questions defined above.

The first section consists of Chapters 2 and 3 (Table 1.2) and deals with the following questions: How has the objective of equity been defined in the Portuguese health care policy and how should an objective of geographical equity be defined? What are the causes of inequalities and inequities in the Portuguese hospital system? How are these inequities operating in the system?

Table 1.2: Structure of the chapters of the thesis (Section I)

Chapter number	*Content*	*Area unit of analysis*	*Some of the geographical equity concepts in use*
2	The Portuguese health care system: Setting the context	Not applicable	Not applicable
3	Geographical analysis of inequalities of the hospital acute care sector	District, *concelho* and health region	Multiple concepts

The second section consists of Chapters 4, 5, 6, 7 and 8 (Table 1.3) and addresses the following questions: How to measure geographical inequities, in particular in capital, utilisation and finance? To what extent are there geographical inequities in the Portuguese hospital system? How to measure need for hospital care in Portugal? To what extent can discrepancies between the distribution of hospital resources and estimated needs be explained by legitimate components of costs (these components being unavoidable costs of hospital provision and cross-boundary flows between areas)?

Table 1.3: Structure of the chapters of the thesis (Section II)

Chapter number	*Content*	*Area unit of analysis*	*Some of the geographical equity concepts in use*
4	Setting a capitation formula to measure geographical inequities	District	Equity in capital, utilisation and finance
5	Measuring geographical need for hospital care	District	Equal opportunity of access for those in equal need
6	Modelling unavoidable costs of hospital care using a multilevel model	Hospital unit	Hospitals operating under similar budget constraints
7	Modelling geographical hospital utilisation flows	Utilisation flows between small area points (*concelhos*) and hospital sites	Current patterns of supply influence movements of populations to access care
8	Computation and analysis of geographical inequities in Portugal	District	Multiple concepts

The third section consists of Chapter 9 only (Table 1.4) and addresses the question: How do we achieve a more equitable distribution of hospital resources in terms of access and utilisation, by marginally redistributing hospital supply?

Table 1.4: Structure of the chapters of the thesis (Section III)

Chapter number	*Content*	*Area unit of analysis*	*Some of the geographical equity concepts in use*
9	(Location-Allocation) Models to improve equity by redistributing hospital supply	Hospital site for supply side and small area (*concelho*) for demand side	Equity of utilisation and access

Chapter 10 concludes with the main findings of this research.

The main results of this thesis support initial evidence that there are high inequities in the distribution of hospital resources in Portugal. They suggest that if Portugal is to improve equity in its system of hospital finance, it will have to develop new policies to correct significant inequities in the current distribution of hospital resources, which is not compatible with need for hospital care. Analysis of various geographical levels and different measures of equity were designed to illuminate different questions, but a common conclusion of excessive concentration of resources in certain areas has emerged. Pursuing different equity objectives requires different policy directions so as to correct current inequities; on the other hand, pursuing a single equity objective was shown to have negative impacts on other policy objectives. Any attempt to correct inequities should also look into policies other than the redistribution of hospital supply, such as the distribution of primary care (which was proved to interact with access to hospital care) and the extension of the hospital supply network.

A complete list of notation is presented at the beginning of the thesis, while complete notation for each chapter is presented in Appendix A. Each chapter reports new notation when it is cited for the first time.

SECTION I

2 CHAPTER 2 – The Portuguese health care system: Setting the context

2.1 Introduction

Chapter 1 has indicated the extent of geographical inequalities in the Portuguese health care system. The objective of this chapter is to provide information that addresses the question: how important is equity in Portuguese health policy, and why are there geographical inequities in the Portuguese hospital sector? The chapter describes how equity objectives have been portrayed in political statements and pursued in policies; it also provides an account of how the current characteristics of the health and hospital systems and of the resource allocation process have developed, and how they relate to the objective of equity; finally, it sets the necessary political, administrative and geographical context. The chapter is structured into five sections that:

- Outline relevant characteristics of Portugal's political system, public and health funding systems, demography, as well as variations in socio-economic, health status and needs within Portugal;
- Analyse political values and policy objectives of the health care system, and the ways in which health care was organised and financed prior to the creation of the NHS, as well as describing historical antecedents to the current health system;
- Analyse the main features of the current hospital system, including organisation and resource allocation;
- Summarise the causes of inequities in the system;
- Present concluding observations.

This chapter makes use of country-based literature obtained by searches in the main health-related electronic databases[17] and directly from the main academic and institutional libraries in Portugal; it also builds on interviews with officials from the Portuguese Ministry of Health and Portuguese researchers in health economics and health policy. Fuller descriptions of the Portuguese health system are given by (Pereira et al. 1999), (EOHCS 1999) and (Pinto and Oliveira 2001). The account given here compares the Portuguese system to other countries National Health Services with similar policy objectives, and in particular those of the UK and Spain, because:

- As described in Chapter 1, the four countries that make up the UK have had for many years policies designed to promote geographical equity[18] and some chapters in this thesis seek to transfer methods developed in these countries (in particular England) to the Portuguese context;
- Spain, although more developed, shares cultural, economic, socio-demographic and geographical characteristics with Portugal[19].

2.2 The Portuguese context

This section describes how Portugal has evolved from a dictatorship to integration into the Euro zone; however, it should be noted that the Portuguese level of economic development is still behind that of other European Union (EU) countries.

2.2.1 Political system

Portugal's territory as defined in the Constitution includes 3 territorial areas (Assembleia da República 1992): mainland Portugal, Madeira and Azores. The peaceful revolution of 1974 ended a dictatorship of 45 years and has led to the establishment of a

---

[17] Some of these electronic databases have been: Medline, International Bibliography of the Social Sciences, electronic journals database of the London School of Economics, British Medical Journal database and the electronic system of the Library of the National School of Public Health (Portugal).

[18] Since 1948, England has had the first Beveridge system, based on a national health service with universal coverage, free at the point of use and funded by general taxation. These characteristics apply to the Portuguese system.

[19] Table 2.1 (below) provides information on some socio-economic indicators for Portugal, the UK and Spain. It is acknowledged that recent developments in the Spanish health care system have made it diverge from the Portuguese, mainly because of financial and political devolution to autonomous communities.

republic with a democratic Constitution. The main institutions of the State are the President of the Republic, the Parliament (both elected by direct universal suffrage), the government and the courts. A unitary state has been maintained –the country is relatively small and homogeneous (Colomer 1996). Portugal has a multi-party system, with power shifting between the centre parties –Socialist Party (centre left) and Social Democratic Party (centre right), and has been politically stable since 1987[20]. Both the Socialists and the Social Democrats have adopted 'orthodox' economic policies aimed at stabilisation and liberalisation. Over the last fifteen years, there have been ambitious privatisation programmes intended to achieve growth through increased product market competition and improved productivity (OECD 1999). The Madeira and the Azores are two autonomous archipelagos: both have legislative and executive powers, and their own political-administrative statutes and government (Assembleia da República 1992). The Madeira and the Azores are excluded from analysis in this thesis, the reason being that their health system differs greatly from that of the mainland; and the islands' legislation and health system is decided in their own parliaments[21].

Portugal is, along with Greece, the most centralised country of the EU (Oliveira, Magone, and Pereira 2003). Excessive control from the centre and the way policies have been formulated has resulted in a two-tiered Portugal –urban coast vs. a marginalised rural interior– that can also be seen in terms of health care (Oliveira, Magone, and Pereira 2003)[22].

Following entry into the European Community in 1986 and into the Euro zone in 1999, the priority of macroeconomic policy has been to respect the Maastricht convergence criteria and make progress towards the objectives instituted in the Stability and Growth Pact[23]. In order to converge to the EU level of development, Portugal has been receiving

[20] The Social Democrat party governed between 1987 and 1995, while the Socialist party governed between 1995 and April 2002. Since April 2002, Portugal has a centre-right coalition government, where Social Democrats rule with the Popular Party (right wing party).

[21] There is a high level of devolved responsibilities to the islands' governments: a global budget is given by central government to islands' governments, which subsequently decide upon the amount to allocate to the health sector.

[22] For example, there have been systematic problems in defining and implementing decentralisation policies, as well as an over-representation of the country's most populated urban areas in political institutions at the expense of rural districts (Bruneau et al. 2001).

[23] The Maastricht criteria intended to establish stable economic conditions in the EU economies, and to promote convergence. The four criteria were: public budget deficit under 3% GDP; public debt under 60% GDP; price stability; and long-term interest rates convergence. The Stability Pact was adopted by EU members in 1997, and is an agreement with two main aspects: a preventive system to identify and

extra-funding from the EU (including funds to the health care sector). During the last decade, the State has been under external pressure to cut public spending and to implement structural reforms (Wise 2002), particularly in the fiscal system and in social welfare areas (OECD 1999).

2.2.2 Public and health care funding system

In the health sector, the State assumes the role of dominant single payer within a single mandatory insurance pool, whereby general tax contributions depend on ability to pay. The State budget prepared by the Ministry of Finance (*Ministério das Finanças*) is structured in sector budgets, contains Ministry's planning activity and is voted by Parliament. There are two health-related public sector budgets: the health current expenditure budget and the health capital budget set within the Program of Investments and Expenditure for Development of the Central Administration (*Programa de Investimentos e Despesas de Desenvolvimento da Administração Central*) (PIDDAC). The MoH is responsible for managing current expenditure on health care and submits the budget to the Ministry of Finance. PIDDAC is the government program for capital investment, which includes a health component; it is currently decided by the Ministry of Finance and managed by the Ministry of Planning, Equipment and Administration of the Territory (*Ministério do Equipamento, Planeamento e Administração do Território*).

2.2.3 Demographic and socio-economic characteristics, and need for health care

Tables 2.1 and 2.2 provide evidence on the lower levels of economic development and of health status in Portugal in comparison to Spain and the UK. Portugal is a small country, both in area terms and in terms of population. In comparison with the EU, the Portuguese population is younger, unemployment rates are lower, and females tend to have a relatively higher participation in the labour market (Eurostat, INE, and European Commission 1998).

correct "extra-size" deficits before reaching the 3% imposed by the Maastricht Treaty; and a system discouraging excessive deficits by imposing penalties for high and unjustified deficits. A slowdown of European economies in the last years has led to a more flexible application of the rules of the Stability Pact.

Measured by GDP per capita and by other socio-economic indicators, Portugal is one of the poorest countries of the EU (Table 2.1). Its industrial structure still depends on labour-intensive industries, while it has a significant percentage of its population employed in the agriculture sector (OECD 1998). The Portuguese welfare state is still underdeveloped, as indicated by per capita spending on social security and welfare, which is still well below the values for Spain and the UK (Table 2.1). Within the country, there is (Santana 2000): a north/south divide, with the north being more populated and having a younger population, and a coastal/interior dichotomy, with the coast having higher population density and being more industrialised and developed.

Table 2.1: Indicators for selected countries

1997	*Portugal*	*Spain*	*UK*
Total population (thousands)	9,950	39,323	59,009
Population above 65 over total population (%)	15.2	15.7	15.8
Population under 20 over total population (%)	24.2	23.3	25
Birth rates (crude rate per 1,000 population)	11.4	n/a	12.3
Life expectancy at birth (years)	75	78	77*
GDP per capita (USD, PPP)	15,056	16,376	20,959
Expenditure on social security and welfare per capita (USD, PPP)	1,702	2,218	3,013

Source: (OECD 2000)

*- 1996 value

On average, the health status of the Portuguese population is relatively low by EU standards (it lags behind Spain and the UK, Table 2.2). Portugal has high mortality rates, in particular for the youngest, although these rates have decreased sharply after the creation of the NHS and especially during the last 10 years (OECD 2000). These results were due mainly to improvements in living conditions and increases in coverage and utilisation of health care supply, especially for mothers and children. Production levels in the Portuguese health sector are still low (Table 2.2) and as described in Chapter 1, there is evidence of wide inequities in health outcomes by socio-economic group (Lucas 1986; Pereira 1995; Giraldes and Ribeiro 1995).

Table 2.2: Population health status and health care utilisation for selected countries

1997	*Portugal*	*Spain*	*UK*
Life expectancy females at birth (years)	78.7	82	79.7
Life expectancy males at birth (years)	71.4	74.6	74.6
All causes female mortality rate (per 100,000 population)	666.5	497.9*	603.1
All causes male mortality rate (per 100,000 population)	1,134.3	896.8*	918.5
Infant mortality rate (per 1,000 live births)	6.4	5	5.9
Perinatal mortality rate (per 1,000 total births)	7.2	6.3	8.3
Doctors consultation in ambulatory, per capita	3.4	n/a	6.1**
All causes discharge rates (per 100,000 population)	9,482.1	11,246	n/a

Source: (OECD 2000)

*-1995 value; **- 1996 value

2.3 Health system

This section clarifies the role of equity in the political values of the health care system and describes how equity objectives have been defined. Further, it summarises some historical developments that are significant for understanding current characteristics of the health care system.

2.3.1 Political goals

The creation of the NHS in 1979 constituted a crucial step towards the pursuit of equity, with the provision of universal health care, free at the point of delivery (Assembleia da República 1992). Equity objectives were the main driving force in the shaping of the Portuguese system, and have continued to feature in subsequent statements of policy (see Table 2.3).

Some of the equity objectives underlying policy directions have been: equity in health (2.3i)[24]; equity in health care (2.3ii and 2.3vi); equity in access for those in equal need, and rejection of a dependence on ability to pay (2.3iii and 2.3iv); geographical equity in

[24] This notation corresponds to Table 2.3 and bullet i.

the distribution of resources across regions (2.3iii and 2.3v). Recent statements from Socialist and Social Democrat governments stressed the need to maintain the State's responsibility in the coverage of health risks. The 1999 Socialist Party program underlined the NHS principles of universal and equitable access to health care, and the State's responsibility in "guaranteeing the access of all to health care in equitable conditions" (Presidência do Conselho de Ministros 1999). The government currently in power –composed by the Social Democrat Party and the Popular Party– has reiterated health as a key social policy required for a fairer society (Presidência do Conselho de Ministros 2002).

Statements considering equity as an objective of health care policy in Portugal have embraced quite different concepts of equity (Tables 2.3 and 2.4)[25], which are inconsistent and may be incompatible (Culyer and Wagstaff 1992) (Pereira 1995). Pereira (Pereira 1990) pointed to these inconsistencies, which are still present in the Portuguese political statements. Hence, there is a lack of a clear framework for health care policies pursuing equity objectives.

Table 2.3: Some political and policy statements concerned with equity

Pronouncements	
i.	"rights of all to the protection of their health as well as their duty to safeguard and promote it" (Assembleia da República 1992)
ii.	"access to the NHS is guaranteed to all citizens, independently of their economic and social status" (Assembleia da República 1990)
iii.	"guarantee the equity in the distribution of resources and in the utilisation of services" (Assembleia da República 1990) (since 1979 NHS law)
iv.	"(health care is to be) nearly free.... taking into account the economic and social conditions of citizens" (Assembleia da República 1992)
v.	"distribution of financial resources must follow closely a capitation basis for guarantying equity between the different regions" (IGIF 1998)
vi.	"(the NHS is charged with) guaranteeing equity of access of users, with the objective of attenuating the effects of economic, geographical and any other inequalities in the access to health care" (Assembleia da República 1990)

[25] Some time trends emerge from political statements in Tables 2.3 and 2.4. These trends are described in the historical section of this chapter (section 2.3.2).

Other inconsistencies in the normative framework for health policy objectives apply. The pronouncements of Table 2.4 show that efficiency has also a special importance as a goal of the system, particularly for the hospital sector, and it is not clear how efficiency relates with equity. For example, there might be a trade-off between the development of the private sector (2.4i and 2.4ii) and the achievement of equity in access (2.3ii). Expression 2.4i. and 2.4ii. have the implicit objective of equalising access to public and private care (promoting complementarity and competition between the public and the private), but it is not clear how this relates to equal access to health care independent of economic conditions. Expressions 2.4iii. and 2.4iv. propose the use of activity/case-mix based indicators for resource allocation that might conflict with the objective of an equitable distribution of resources across regions (2.3iii). In addition, during the 1980s and 1990s, most of the research on resource allocation in the hospital sector carried out by the MoH focused on efficiency (this research is briefly described in section 2.4.3.1).

Table 2.4: Other political and policy statements

Pronouncements	
i.	"Law 48/90 establishes a mixed health system model, instituting the complementarity and the competitive character between the private sector and the social economy of health care delivery; and the integration in the NHS of private entities and of "free" professionals that contract with the NHS all or some activities of promotion, prevention and treatment" (Ministério da Saúde 1998c)
ii.	"Support is established for the development of the private health sector, specially the initiatives of the private institutions of social solidarity, in competition with the public sector" (Assembleia da República 1990)
iii.	(on hospitals) "health care payments must be prospective, and must relate directly the activity levels with explicit prices, quantities and types of services to be provided" (IGIF 1998)
iv.	"To apply financing models of capitational basis and/or related with product and quality to all the activities of assistance" (Ministério da Saúde 1999e)

Other political objectives such as quality, accountability and devolution of power (Assembleia da República 1992) (DGS and Ministério da Sáude 1998d) are also broadly mentioned in political and policy statements.

Lack of clarity over policy objectives makes it difficult to relate research to policy. Such lack of clarity is common (Van Doorslaer, Wagstaff, and Rutten 1993) (Pereira 1993). Nevertheless, as acknowledged in Chapter 1, Portugal can be classified as sharing the objectives of "adequacy and equity in access to some minimum of health care for all citizens" (OECD 1994).

2.3.2 NHS creation and history

Equity gains in Portugal were famously achieved with the universal coverage and increases in health care provision and utilisation during the 1980s and 1990s. Nevertheless, inequities persisted in the distribution of resources within the country and in the high proportion of private finance, as shown in Chapter 1. Portugal resembles the UK in the early 1970s where, although a NHS had been created on grounds of equity since 1948, it had been taken 30 years to initiate policies to achieve equity of access to hospital services (following the RAWP report (Department of Health and Social Security (United Kingdom) 1976)).

This sub-section explains how legacies of the past help to understand the current health care system. It shows how the recent establishment of the Portuguese democracy and the concerns over equity resulted in the creation of the NHS in 1979. An incomplete transition from the Bismarckian to the NHS model explains the mixed coverage in the current system. High levels of private funding are mainly the result of policies on tax deductions and copayments/cost-sharing taken by Social Democrats between 1985 and 1995. Private provision, on the other hand, is partly explained by the historical involvement of Catholic institutions in health care provision.

The following account outlines the history of the development of the Portuguese health system in five periods:

a. Before the 1974 revolution;
b. The 1974-1979 period, which captures the context of the milestone creation of the NHS in 1979;
c. The 1979-1985 period that was marked by the (incomplete) implementation of the NHS model;
d. The 1985-1995 period, during which stable Social Democratic governments have legislated shifts in funding and provision towards the private sector;
e. The 1995-2002 period, under Socialist governments that reorganised the system.

2.3.2.1 Before the 1974 revolution

The first elements of the provision of health care in Portugal date back to the Middle Ages and are "rooted in a Christian culture" (Campos 1984). Some hospitals emerged but those initiatives were disperse, uncoordinated, and based on a concept of solidarity. Later, the management of all hospitals in the country was entrusted to the Holy Houses of Mercy to the Poor (*Santas Casas da Misericórdia*), which had the objective of protecting the poor (Lima 1998). During the eighteenth century, the State established a limited number of teaching hospitals to supplement charitable provision (OECD 1998). The nineteenth century was marked by the proliferation of new hospital establishments (as charity initiatives), in order to meet the needs of a growing population, and by the creation of the first Public Health structures as PC developments (Campos 1984).

During the twentieth century, the system evolved towards a 'Bismarckian' model[26] with significant gaps in coverage, and the role of the State was limited to that of being one insurer. Health care institutions were not available to most of the population until the end of World War II, a period during which health care consisted of a fragmented insurance system. After World War II, a State Social Insurance Service initiated the delivery of curative ambulatory medical care, using both private providers and public hospitals (Campos 1984). Insurance funds were organised on a professional basis and insurance care was provided to workers enrolled in the Social Insurance Funds (including the ones managed by the State) and to their families. Other post-war government initiatives included a 1946 act that "created a basic planning and constructional scheme with regional interrelations for the complete network of hospitals in the country", while vertically structured Public Health Institutes were created to meet the major sanitary needs (Campos 1984). These changes were designed with government having a limited role in the provision of health care; they were intended to fill the gaps left by private initiatives (but failed to do so) and to give priority to preventive services (Reis and Carvalho 1994).

[26] The Bismarckian model corresponds to a health care system predominantly financed by social insurance, whereas financing and delivery are institutionally separated, and where there are some type of contractual agreements between social insurance organisations and providers (Savas et al. 1998).

During the 1950s and 1960s, the system remained fragmented and gaps in coverage prevailed. The State carried out minor interventions in the health system, particularly in coordination and regulation. The first department to co-ordinate hospitals was instituted in 1961 –the General Directorate of Hospitals (Lima 1998), and in 1968, for the first time, the internal functioning of hospitals was regulated with the publication of the Hospital Statutes and the General Hospital Directive (Ministério da Saúde 1968). At the beginning of the 1970s, health care standards were low, as shown by high mortality rates as well as the causes of death; the health status of the population was worse than in other European countries (Reis and Carvalho 1994) and the system still consisted of independent health insurance funds.

The first crucial step towards improvements in equity through access to health services for the entire population was taken in 1971, with a legal decree on PC (Ministério da Saúde 1971). This was the first document to institute the State commitment to universal health care as a right. That decree made explicit the desire to evolve towards a national health system structure, marked the move towards greater public provision of health care, and instituted the integration between curative and preventive medical care. It also granted priority to government powers over the private sector and sought to integrate the health system into the wider context of social policy (something which did not happen). These targets were, however, accompanied by inadequate steps towards implementation. Eventually, under the system of medical care "providence" (nominated as Socio-Medical Services) 25% of the population remained without insurance coverage (Campos 1984).

2.3.2.2 *1974-1979: revolution and NHS creation*

The formal creation of the NHS followed the democratic revolution of 1974 (that ended 45 years of the dictatorship), and the new Constitutional law of 1976. Charitable hospitals (mainly *Misericórdias*) were transferred to public control in 1975-1976. This nationalisation of hospitals accompanied nationalisation of other sectors of the economy. The Constitution of 1976 explicitly recognised the collective right to health protection, by stating "the right of all to the protection of their health, as well as their duty to safeguard and promote it" and "the socialisation of medicine" (Assembleia da República 1976). Universal coverage was legally recognised in 1979 with the creation of a general system of health care open to all Portuguese, i.e., a national health service (Assembleia da República 1979).

In 1979, the NHS law was approved by the Parliament[27], formally replacing the predominantly social insurance-based system by one close to the NHS classical model. This involved universal coverage of the population and general (comprehensive) benefits, national tax financing and state ownership or control of production (with decentralised management), as well as an integrated provision of health care. Achieving equity was at the heart of the creation of the NHS, as coverage was extended and financial contributions followed the ability to pay: the Government assumed responsibility for the provision and financing of health care with the explicit commitment that "access to the NHS (should be) guaranteed to all citizens independently of their economic and social status" (Assembleia da República 1979). However, it was not clarified, nor debated, whether the available budget would be able to finance universal and comprehensive coverage, and whether it could be sufficient in the light of 'demand explosion' of an increasing population (including population returning from the ex-colonies).

2.3.2.3 *1979-1985: transition to the NHS model*

In the beginning of the 1980s[28], implementation (albeit incomplete) of the NHS model started together with the reorganisation of the system (namely the centralisation of health care provision), while there was also a 'natural' development of the private sector[29]. In 1980, health expenditures began being funded by the State budget (by general taxation), which replaced the previous financing system based on a Social Insurance Fund (Lima 1998).

[27] Interestingly, the NHS creation was legislated by means of the votes of the Socialist party (centre-left party) and of the Social and Democratic Centre party (right wing party -currently denominated Popular Party), shortly after a government made up of these two parties had ruled for a short period.
[28] The 1980s were marked by a stabilisation of the economy and by a consolidation of the Portuguese democracy, which have highly influenced the evolution of the health care sector, although the activity of the private was left unregulated.
[29] Since them, public health care services (mainly primary care) have increasingly made use of outsourcing services delivered by the private sector.

The transition to the NHS model was based on universal public coverage for all medical care but this was not completed as the NHS was unable to absorb all the existing employment insurance plans, with inevitable overlaps between health insurance plans for an individual (and his family) and NHS coverage for all (Urbano, Bentes, and Vertrees 1993). This resulted in the current patterns of multiple coverage and inequities of access described in Chapter 1. As mentioned in the previous sub-section, State social insurance dominated the provision of care before 1979 but coexisted with occupational schemes (widely known as subsystems –*subsistemas*). These subsystems covered a significant part of the population, and the largest were those for civil servants, the military and banking employees. Despite the 1975 movement towards the nationalisation of the most important economic sectors (namely banking), as a result of which almost all these subsystems became dependent on public financing, the State did not integrate all these schemes into the NHS (Urbano, Bentes, and Vertrees 1993). Since this missed opportunity, no Government has been able to establish a clear relationship between the NHS and these subsystems, and this has become a crucial obstacle to reducing inequities in access. Their beneficiaries are still allowed to choose their providers of care (while NHS enrolees had access only to a NHS family doctor), and to have their expenditures reimbursed on a fee-per-item basis, while they have been highly subsidised by the state (via tax deductions).

2.3.2.4 1985-1995: shifting towards the private

The Social Democrats that came to power in 1985 and have governed with a parliamentary majority after 1987, progressively recognised the role of the private sector (Pereira 1990) and shifted the system towards a public/private mix in provision, funding and finance. Their policies gradually introduced modifications to the classical public integrated NHS model, and shifted the financial burden of health care to the patients. These changes were influenced by the ideology of markets and disregarded implications of inequities of access. The main changes were:

- In 1989, the revision of the Constitution sought to reduce State intervention; it substituted "nearly free" for "free" NHS services at the point of use[30], as well as

[30] There is no word in the British vocabulary that captures the exact meaning of the Portuguese word "*tendencialmente*". "Nearly free" and "tend to be free" are approximate translations. Substantively, the change from 'free' to 'nearly free' allowed for user fees to be charged in public health services.

"socialisation of medicine" for "socialisation of health care costs" (Reis and Carvalho 1994);

- The new NHS law in 1990 (Assembleia da República 1990) recognised alternative forms of health care delivery (the private sector complementary to the public), accepted the principle of cost-sharing of care provided by the NHS, and allowed for private management of NHS health care units[31]. The expression National Health Service was replaced by that of a National Health System emphasising that the NHS was but one of the 'subsystems' providing health care (albeit the most important one, since it provided the only coverage to around 75% of the population) (Pereira and Pinto 1993).
- The 1993 NHS statute law (Ministério da Saúde 1993a) defined policies to implement the principles of the 1990 NHS law, such as co-payments.

The current high levels of private expenditure in funding, and the mix between the public and private sectors in provision, coverage and funding (in 2002) are outcomes of the following policies:

- *Mix in public-private provision:* acceptance of private or mixed status providers in 1990 and change in the medical law in 1993, allowing for full-time salaried doctors to engage in private practice (hence legitimising current practice) (Ministério da Saúde 1993a).
- *Mix in public-private funding:* full deduction of health expenditures from taxable income (applied since the 1989 fiscal reform) (Pereira et al. 1999); introduction of user charges by the 1990 NHS law[32]; reimbursement for drugs prescribed by doctors in their private practice, provided that they worked also for the NHS (Pereira et al. 1999) (previously only drugs prescribed in the public services were reimbursed[33]).
- *Mix in public-private coverage:* the opting-out policy established in 1993 (Ministério da Saúde 1993a) offered incentives to move from public coverage to private insurance, by payment of a premium to health insurers (although, this was

[31] This law also defined the NHS as the ordered and hierarchical set of public institutions and official providers of health care, including all private providers having agreements with the NHS, under the superintendence of the MoH. These agreements had the status of conventions (Ministério da Saúde 1998c), and steered the expansion of the private sector, mainly financed by the public sector outsourcing activity.

[32] Nonetheless, it included exemptions for the economically disadvantaged and higher risk population groups (Andrade, Branco, and Sepúlveda 1996).

[33] This has promoted private activity, although the main (explicit) objective was to promote equity of access.

not attractive to insurers and the first opting-out subsidies were only approved in 1997).

2.3.2.5 1995-2002: re-structuring the NHS model

The Socialist Party came to power in 1995 and governed until April 2001. Most changes under socialist governments were concerned with re-organisation, and the emphasis was placed on the complementarity between the public and private sectors (rather than on competition) (Presidência do Conselho de Ministros 1999). Socialists introduced legislation for a number of changes, although many of these policies have not been implemented[34]:

- The internal market model of 1997 (Ministério da Saúde 1997) is based on a public contract model that followed similar reforms in the UK, Belgium, France, Germany, the Netherlands, Canada and Japan (Hurst 1996). The internal market established a more general separation between purchasers and providers, in comparison to the UK (Mossialos and Le Grand 1999), and aimed at improving accountability in the system (with well defined contracts and responsibilities) and at decentralising functions to RHAs. Nonetheless, until 2002 the internal market was only marginally implemented. After 2002, with the new centre-right government, it is still not clear to which extent the internal market will be put to use.
- The first opting-out agreements were reached between the government and the Portugal Telecom and CTT post-office subsystems in 1997 and 1999 (Ministério da Saúde 1998a). Under a fixed monetary payment of €145 per year and per beneficiary, both subsystems are responsible for providing and/or paying all health care services to the ones enrolled. However, these decisions do not seem to be part of a strategic move towards opting-out[35].
- Other policies were incremental, mostly attempting re-organisation, clarification of the public-private mix and cost containment. Some of these policies were:

[34] Some of the reasons behind implementation problems have been: maintenance of centralised systems with low level of devolution of power, lack of influence of the MoH on the behaviour of the agents, and bureaucratic barriers (Artells 1996).

[35] These sub-systems are very small in terms of number of users. The government was forced to strike an agreement in order to make shares of Portugal Telecom attractive to investors during the privatisation process, which, in turn, forced the Ministry to make a similar agreement with CTT.

1. Development of data card technologies since 1995, such as the user card, in order to identify subsystems beneficiaries and oblige these schemes to pay for care delivered in the NHS; implementation is still far from complete.
2. Creation in 1999 of 'local health units' (*unidades locais de saúde*) and 'local health systems' (*sistemas locais de saúde*) (Ministério da Saúde 1999b) with an eye on gains in co-ordination (and thus efficiency); again, there has been no implementation.
3. Special programme to reduce waiting lists for surgery by contracting with the public and private sectors (Ministério da Saúde 1999d); implementation has been limited (Oliveira 2001a).
4. Cap on deductions of health expenditures from taxable income in 1999.
5. Changes in the regulation of the pharmaceutical market, including the recognition of the possibility of economic evaluation studies of drugs being required for decisions on reimbursement.

None of these changes have tackled the structural problems in the system (e.g. clarifying the public-private mix). In sum, the Portuguese NHS in 2002 differs substantially from the classical NHS model in three main respects: a large presence of the private sector in financing, with high contributions from families in co-payments; multiple coverage of a significant percentage of the population; and a high presence of private providers, mainly in the non-hospital sectors.

In comparison to other health systems, the Portuguese system seems to be operating with high inequities, inefficiencies and a low level of accountability of health care providers, and these problems have systematically led to severe difficulties in containing costs (OECD 1994) (Pereira et al. 1999) (Oliveira, Magone, and Pereira 2003). Some features of the Portuguese political and social system seem to explain the lack of policies to tackle major problems in the health system, for example, low participation of citizens and high mobilisation of interest groups against potential reforms (Oliveira, Magone, and Pereira 2003).

After April 2002, new policies were introduced under the new centre-right government, such as a new hospital management law (Assembleia da República 2002) and changes in the levels of cost sharing and in the rules of prescription. It is, nonetheless, too early to evaluate the impact of these reforms.

2.4 The current hospital system in Portugal

This section describes the hospital system as part of the wider health care system. Specifically, it describes the organisational structure of the NHS, the administrative hierarchy of hospitals, resource allocation, and hospital policies.

2.4.1 NHS organisational structure

The current NHS structure was designed to distribute responsibility. The following description explains how the main elements/actors of the system interact.

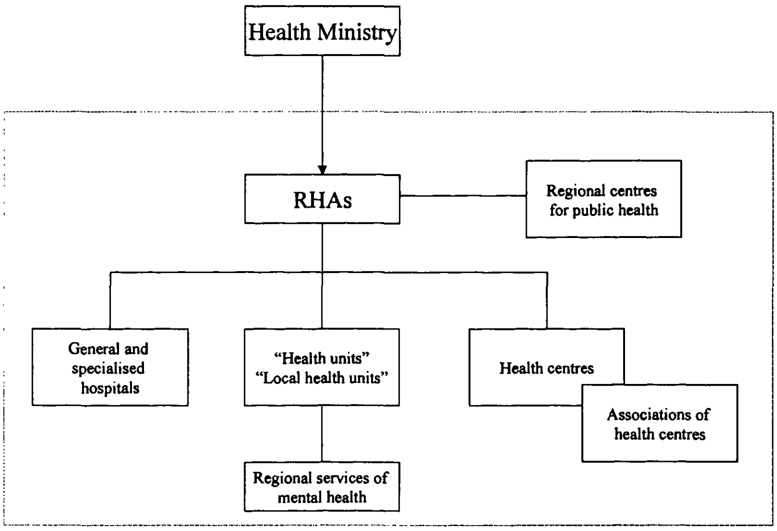
The Minister for Health has the main responsibility for the national heath strategy, namely the regulation, organisation and management of the health care system. The Minister should also oversee and assess health policy and co-ordinate health related activities with other Ministries, and in particular, the Ministry of Labour and Social Security (*Ministério da Segurança Social e do Trabalho*) and the Ministry of Science and University Studies (*Ministério da Ciência e Ensino Superior*). The Ministry of Labour and Social Security is responsible for social security benefits and for financing social care through capitation payments to not-for-profit institutions that deliver services to the young, the old and the handicapped (Teixeira, Coutinho, and Morgado 1999)[36]. The Ministry of Science and University Studies is responsible for medical education and training of health professionals. Some special training programmes are the joint responsibility of the Medical Association (the professional association where all medical doctors have to be registered to exercise their activity) and the MoH.

[36] These capitation payments depend on the type of services provided and on the characteristics of each beneficiary. On the other hand, social care is provided by entities with a special compulsory status, and those entities must be registered in a Union (*União das Mutualidades, Misericórdias e Instituições Particulares de Solidariedade Social*). In practice, charity institutions form the majority of those organisations that are closely related to the Catholic Church. In 1997, 292,418 people benefited from social care coverage, corresponding to a budget of €355,000,000 from the Ministry of Labour and Social Security (Teixeira, Coutinho, and Morgado 1999).

The organisational structure of the MoH covers central and specialised services (Ministério da Saúde 1993b). The Institute for Financial and Informational Management of Health (*Instituto para a Gestão Informática e Financeira da Saúde*) (IGIF) is a specialised service responsible for the study, orientation, evaluation and execution of the informational and management system of NHS financial resources.

The Portuguese national health system includes services managed by the NHS and supervised by the MoH. The NHS incorporates various types of health services (see Figure 2.1)[37].

Figure 2.1: NHS structure based on the NHS organic law



Source: (Ministério da Saúde 1993b)

A description of the system has to acknowledge that much of the legislation has not been implemented: the MoH retains central control of the hospital system and the hospital sector continues to dominate. Thus, legislation exists for 'local health units' and 'local health systems' as organisation units to provide continuity of health care but these have not been implemented; there has only been limited contracting following the internal market law of 1997[38], while regional centres of public health have never been created.

[37] NHS entities are defined as providing health care, being dependent on the MoH, and having a specific statute (Assembleia da República 1990). The NHS integrates both public and private entities that provide health care.

[38] Contracting agencies since 1997 overlapped with the geographic boundaries of RHAs, have had an independent status regarding RHAs (although they should cooperate with them), have assumed an

RHAs have key responsibilities for planning, allocation of funds, management of human resources, provision of technical and administrative support to health units, and assessment of health care units' performance (Ministério da Saúde 1999a). In practice, PC centres have been accountable to RHAs but other services are still managed centrally, particularly, hospital services. Patients are expected to choose their NHS general practitioner (GP) (and specialist doctors after the GP consultation), or doctors from a list published by the MoH (Mossialos and Le Grand 1999). However, the gate-keeping system operates imperfectly, since:

- There is a shortage of GPs working full-time in the NHS, especially in rural areas;
- There is a lack of co-ordination between GPs and specialists;
- The excess of demand has resulted in using emergencies to gain access to secondary care[39];
- Some population groups (mainly from some subsystems) have benefited from direct access to public hospitals (Mossialos and Le Grand 1999);
- There is little control over private activity in public hospitals and over the transfer of patients between the public and the private hospital sectors.

The recent government has introduced changes (under the new hospital management law, described below) that signal the abolition of the gatekeeping system, but it is not clear how the new system will operate.

Decisions about the State budget determine public health expenditure. There are two health budgets: the current expenditure budget and the one on capital expenses (included in PIDDAC), as described in detail in section 2.2.1.2. These budgets are decided independently within the State budget and there is no framework to link decisions taken on the two types of expenditure.

---

advisory role on resource allocation, and have contracted with health care units and independent groups of doctors.

[39] This has implied misuse of resources by treating a non-urgent demand and has reflected poor continuity between primary and secondary care (Dixon and Mossialos 2000). Recently, in central hospitals, there has been an effort to put into practice a process of skimming patients that go to the emergencies without needing emergency care (the so-called 'false emergencies'). Nonetheless, emergencies have helped to solve the system's inadequacies, namely long waiting lists in health care centres and hospital ambulatory services.

2.4.2 The administrative hierarchy of hospitals

Hospitals are classified in an administrative hierarchy (from central to level I hospitals) according to technological complexity and size of catchment areas of hospital provision (DGS and Ministério da Saúde 1998c):

- *Central and general hospitals* provide highly specialised services with advanced technology and specialised human resources.
- *Central and specialised hospitals* focus on a range of specialist services. Both general and specialised hospitals tend to be located in the main urban centres.
- *District hospitals* provide a range of specialist services, and are located in the district capital. In general, there is at least one district hospital in each geographical district.
- *District level I hospitals* are at the bottom of the hierarchy and provide internal medicine, surgery and one or two other basic specialties only. They tend to be located in small towns.

Until September 2002, NHS hospitals were public providers under public administrative law and had enjoyed administrative and financial autonomy (since the 1988 hospital management law (Ministério da Saúde 1988)). They were centrally managed and organised in a hierarchical structure. Since September 2002, following a new hospital management law approved by Parliament (Assembleia da República 2002), structural changes for hospitals are expected, and in particular the creation of new types of hospitals with financial and statutory autonomy. The new law introduces freedom of choice of providers by patients and changes the concept of the 'NHS hospital' to that of 'Network of health care providers' (*Rede de prestação de cuidados de saúde*) that includes four types of hospitals (Oliveira 2002):

1. Public providers with financial and administrative autonomy, but under public management (under public sector administrative law);
2. Public providers with administrative, financial and asset management autonomy, under (contracted) private management (also under public sector administrative law).
3. Providers under corporate law, with equity shares, having the State as the exclusive shareholder (State ownership of 100% of equity). These hospitals are informally

called 'hospital-enterprise' (*hospitais-empresa*). The State owns their capital via numerous public agencies that act as statutory shareholders.

4. Private providers with contracts with the State (under corporate law).

As part of these changes, the Minister announced that 34 hospitals (out of 114) are to be converted into 'hospital-enterprises'. For these hospitals, the levels of debts were capped at a maximum of 10% of equity, and 30% if approved by State shareholders (*O Público*, 12.10.2002) –the MoH and the Ministry of Finance. 10 hospitals are to be kept as public providers and public management (*Diário de Notícias*, 25.07.2002). This law is a path-breaking reform which introduces radical changes to the current system, but seeks to be over-ambitious because some necessary conditions for its policy directions to be successful are not satisfied (Oliveira 2002) –for example, the ability of the state to monitor and oversee hospital activity. It is too early to try to analyse the implications of the new law, as there is still no information on the underlying model for the health care system (i.e., the model based on competition or co-ordination), while there are also crucial gaps in regulation.

2.4.3 Resource allocation

Methods used to allocate resources have been poorly developed in terms of: stimulation of relative prices in the health system, clear incentives, and necessary information. In practice, financing of the system is open-ended and health care units are financed retrospectively. There have been no penalties for health units' managers that overrun budgets (OECD 1998) (there is a general failure to cap budgets). Payment systems for staff have been based on salaries and have failed to provide incentives for efficiency and cost containment (Mossialos and Le Grand 1999). Retrospective reimbursement has also perpetuated current inequities and inefficiencies.

This section describes the hospital financing system during the 1980s and 1990s in two sub-sections that summarise the history of resource allocation methods for hospitals and for other health care sectors, while key financial statistics are also analysed.

2.4.3.1 Resource allocation criteria

Diagnostic related groups (DRGs) have assumed a crucial role as a management and information tool for hospitals, and have been partly used to set budgets. Since the 1980s, following the trend of changes in hospital payment systems in EU countries, Portugal has attempted to move from incremental and historical funding of hospital activities to the establishment of budgets based on hospital activities/functions. Before the introduction of the DRG based financing system, NHS hospitals were reimbursed on the basis of actual costs (until 1980). In 1981, a financing system was introduced partly linking funding to production levels (output-based formulae) (Paiva 1993)[40]. The development of a DRG information system started in 1984 and aimed at defining a system of classifying patients and calculating a set of prices (Tranquada 1998). The DRG system has been used to charge services to non-NHS users since the beginning of the 1990s (in a similar form as in other countries, such as Finland (Hakkinen 1999) and Italy (Fattore 1999a)). Only at the beginning of the 1990s was it possible to use the DRG system for budgeting (Bentes et al. 1996).

DRGs were firstly used to set budgets between 1990 and 1992 (Bentes 1995)[41], and were deployed again for resource allocation in 1997 and 1998 (IGIF 1998). In 1998, 20% of the hospital budget was based on DRG production indicators, and the remaining on incremental budgeting (with adjustments by salary growth and inflation) (IGIF 1998). Between 1999 and 2001, the DRG system for setting budgets was gradually abandoned (Pereira et al. 1999)[42] as DRGs had had little impact[43]. This was because of budget overruns and allocations still made on the basis of incremental budgeting.

[40] This system was highly criticised for being based on crude output measures and by creating undesirable incentives concerning over-utilisation and lack of substitution of inputs.

[41] The initial purpose of DRGs as a budget-setting tool was to allocate resources to each RHA, based on the health status of its population (Bentes, Urbano et al. 1993). RHAs would then redistribute funds to providers within the region based on their production levels, adjusted for case-mix (Urbano, Bentes, and Vertrees 1993). The neutrality of the budget would be ensured by the use of a pricing factor, recursively determined, which multiplied by each hospital allocation would respect the budget constraint (Bentes, Urbano et al. 1993).

[42] In the context of the internal market reform, DRGs were supposed to keep their role as internal management tools and as providers of information for establishing hospital contracts. With the advertised changes in the new hospital management law, the role of DRGs is not clear: under the new law, hospital finance seems to be done by another unit of service and not by cases classified by DRG.

[43] A preliminary evaluation of the use of the DRG system using 1992-1993 data has indicated that hospital administrators seem to have reacted to the use of the DRG system as profit maximisers, as read in the decrease of length of stay (LOS) and in changes of LOS distributions for the five most frequent DRGs (Dismuke 1996). Nonetheless, other factors might have played a part in LOS decreases (for example, technological advances).

Moreover, neither payment by DRG nor incremental budgeting offer ways of allocating resources in accordance with population needs and cost containment (Mossialos and Le Grand 1999).

Attempts to reform the hospital financing system have thus been dominated by the objective of efficiency rather than that of equity. In addition, most of the research in the 1990s inside IGIF was focused on efficiency: the development of the DRG system, or problems of hospital organisation, auditing and information (Bentes 1992; Bentes, Gonçalves et al. 1993; Bentes et al. 1994; Bentes et al. 1996; Bentes et al. 1997; Valente et al. 1998). The development of the DRG system in Portugal is still incomplete. For instance, DRG prices still use Maryland service weights across specialties and hospital cost data, as Portuguese cost analytical data does not provide information on patient costing[44]. The use of Maryland weights is based on the assumption that Portuguese hospitals have the same relative use of resources as hospitals in the US (Bentes 1992).

Until recently, financing of non-hospital services has been mainly incremental, reflecting inflation and the annual public sector budget. In practice, an open-ended system has been operating, and the lack of use of caps for components of expenditure has contributed to the failure to keep spending within budgets. As described in Chapter 1, for the first time in 1998, a capitation formula (based on population numbers and age) was used in PC to allocate 8% of the budget, the remaining 92% being historically determined (IGIF 1998). This formula was later changed to include a correction for morbidity as measured by the burden of illness (Tranquada, Martins, and Sousa 2000). Nursing home organisations receive a fixed budget from central or local governments, based on the number of inhabitants or elderly people in their catchment area, or staff numbers (Mossialos and Le Grand 1999)[45]. A policy of capping pharmaceutical expenditure was introduced in 1997, and its impact was to decrease the rate of growth of pharmaceutical expenditure in 1997 and 1998, in comparison with previous years (INFARMED 1999). Teaching and other services are still being financed on an incremental basis (Tranquada 1998). The 1999 budget has also earmarked centralised funds for specific programs (as in the reduction of waiting lists, under the Program of Access Promotion –*Programa de Promoção de Acesso*), to be allocated to RHAs under

[44] The use of Maryland weights had been defended as a normative approach up to the point that research would produce additional information (Bentes 1992).

a capitation formula (IGIF 1998). Detailed information on the waiting lists program is presented in section 2.4.4.3.

2.4.3.2 Statistics in the hospital sector

This sub-section describes the setting in which hospitals operate, and provides some statistics. In particular, it describes monetary flows on production and financing sources in the health care system, the links of the hospital sector with other health care sectors, the sources of hospital financing and the composition of spending in the hospital sector.

Provision

In Table 2.5, agents in the health system are grouped according to public/private and financing/provider status. The table shows the importance of private supply and financing of care.

Table 2.5: Providers characterised along the public/private spectrum

	Public finance	*Private finance*
Public provision	Public hospitals (under public sector administrative law), health centres, doctors and GPs Subsystems covering public employees with own provision of services	Subsystems that do not cover public employees and private health insurers that make use of public hospitals
Private provision	"Conventioned" (complementary) care including private hospitals, specialist services, some GPs practicing in rural areas, pharmacies, laboratory tests and X-ray services, physiotherapy, renal dialysis Hospitals that belong to the State but are under corporate law	Private hospitals, private clinics and religious charities providing health care services (except for specific contracts with the NHS)

Source: Adapted from (OECD 1994) and including changes implied by new hospital management law

Supply of beds in Portugal has almost reached the EU average (EOHCS 1999), although the provision of long-term beds is well below EU values. This means that any attempt to

[45] As described before, this component is funded by the Ministry of Labour and Social Security.

make the geographical distribution of acute care beds more equitable is more likely to be based on redistribution, rather than on additional investment.

In 1996, the private hospital sector was responsible for the provision of 23% of beds and owned 45% of hospitals (Table 2.6). The number of private beds remained relatively stable during the 1985-1995 period (OECD 2000). The percentage of private hospitals beds is high in Portugal compared to other NHS countries (OECD 2000) and the role of the private sector is more important than these statistics suggest because there is no information on the scale of the private sector inside public hospitals. Psychiatric services are over-represented in the private hospital sector (Departamento de Estudos e Planeamento da Saúde 1997a). The private sector is more prominent in non-hospital health services, and it has focused on the most profitable health care areas (Campos 1984), such as ambulatory visits (25%) and diagnostic tests (66%) (1990) (Campos 1991) while disregarding others, such as births (12.4%) (Pinto, Ramos, and Pereira 2000). The private sector is mainly located in urban areas, which reinforces the already existing geographical inequalities in the distribution of public resources (evidence for that is given in Chapter 3). In 1995, 23% of the private beds were for-profit.

Table 2.6: Public/private shares in some physical resources and utilisation indicators for hospital care

1995	*Number of hospitals*	*Hospital beds*	*Average hospital size*	*Discharges*	*Inpatient days*
Private hospitals	45.3%	22.9%	92.2	13.6%	21.0%
Public hospitals	54.7%	77.1%	257.3	86.4%	79.0%
Total	100.0%	100.0%	182.5	100.0%	100.0%

Source: aggregated from the hospitals database available in (DGS and Ministério da Saúde 1996)

Financing sources

The Portuguese health care system is mainly financed through the State budget. Over the past two decades, total health expenditure increased steadily and Portugal is at present, one of the highest spenders in the EU countries with a NHS in terms of GDP, but has the lowest level of per capita health expenditure in US dollars (in purchasing power parities -PPP) among EU countries (OECD 2000). Investment in medical facilities appears to be lower than in other countries (OECD 2000). The level of expenditure is probably adequate, but it is argued that there are serious problems over how this is allocated (Pereira et al. 1999), a view that seems to be shared by the ruling government (Presidência do Conselho de Ministros 2002).

Table 2.7 presents the financing structure of health expenditure. In 1990, public and private sources were responsible for 61.2% and 38.8% of the total, respectively. This is quite different from other countries where care is provided by a NHS: for example, in Spain and the UK, more than 80% of health care financing was public (in 1995 and 1993, respectively) (Casasnovas 1999) (OECD 1998). Current high levels of private expenditure are a legacy from increases in family expenditure and tax deductions at the end of the 1980s. There is little information on the structure of private expenditure (Pereira et al. 1999). However, the main drivers of the high level of out-of-pocket expenditure seem to be lack of public supply of some services (e.g., dental care, physiotherapy), long waiting lists for specialist visits and elective surgery, and low level of reimbursement of drugs. High levels of private expenditure and generous tax deductions have led to a regressive system of finance, described in Chapter 1.

Table 2.7: Financing by source in percentage of total expenditure

Financing sources	*1980*	*1990*
General taxes	66%	55%
Social insurance	5.2%	6%
Total public	71.2%	61.2%
Private insurance	0.6%	1.4%
Direct payments	28.2%	37.4%
Total private	28.8%	38.8%

Source: (Wagstaff et al. 1999)

Coverage

Subsystems and health insurance cover around 25% of the population (OECD 1998)[46], and function in addition to NHS coverage. In 1990, 80% of subsystems beneficiaries worked in the public sector, although this share has been decreasing with the privatisation of most public-owned companies. People enrolled in most subsystems and

[46] It should be noted that there is a lack of consensus as to estimates on single covered NHS population and subsystems beneficiaries (Departamento de Estudos e Planeamento da Saúde 1997b) (OECD 1998).

in private health insurance are free to purchase services wherever they choose, most of them using the private sector for ambulatory care and the NHS for non-elective surgical interventions. Voluntary health insurance (VHI) covered around 8% of the population in 1995, and has been increasing although still representing low levels (Oliveira 2001b). Most of the insured have access to VHI under employment schemes subscription (Mossialos and Le Grand 1999).

As described above, this overlapping coverage is a legacy from the past (pre-1979 Social Insurance system) and exacerbates inequities in access, as groups under subsystem protection tend to be better-off. The civil servants subsystem (ADSE) covers 12.5% of the population (Carvalho 1998) and illustrates powerfully the reluctance of successive governments to integrate subsystems into the NHS, during the 1980s and 1990s. As described in the historical section, the government's strategy on opting-out is not clear. This lack of clarity means that there is no complete normative framework on subsystem user rights, while it creates horizontal inequities (as the capitation paid to subsystems appears to have been fixed through political negotiation), as well as inequities and distortions in the system of finance (as the State has been funding the civil servants subsystem through the Ministry of Finance budget) (Pinto and Oliveira 2001). Evidence shows that private health insurers have been selecting the healthiest (and wealthiest) groups, avoiding being comprehensive in the provision of care (Dixon and Mossialos 2000). Levels of private insurance are higher for individuals of working age (Oliveira 2001b), while private insurance occurrences are higher in well-defined areas (Pereira et al. 1999).

Analysis of current expenditure and of components of the health and hospital budgets

This section describes how the hospital sector is financed. Health care is mainly financed by public expenditure (4.5% of GDP in 1998)[47] and the weight of health on public expenditure has been stable between the 1980s and 1990s (around 10-11%) (Barreto 2000). Increases in the share of total current health expenditure in GDP have been partly due to private expenditure (Pinto and Oliveira 2001). The annual national

[47] There are reasons to expect that this amount of public expenditure is underestimated given that the state is indirectly funding the private sector in multiple ways.

health budget has been exceeded, and this has required supplementary budgets (IGIF 2000)[48].

During the 1990s, expenditure on hospitals has accounted for a higher share of the public budget. In 1995, it amounted to 56%[49], against 45% for the UK and 53% for Spain (OECD 1998). Pharmaceutical expenditure is very high, around 36% of the budget (second column, Table 2.8, includes pharmaceutical expenses and procurement of clinical products); and the level of per capita expenditure in pharmaceuticals (measured in PPP) is very high in the EU context, contrasting with a low overall health care per capita expenditure (OECD 2000). Furthermore, expenditure on pharmaceuticals increased 45% between 1990 and 1995 (Pereira et al. 1999)[50]. Pharmaceutical expenditure has made a significant contribution to regressivity in finance and inequities of access. Personnel costs account for the highest share of hospital costs (53%) (Table 2.9), although this is low by international standards (in the UK this is 70%) (Fattore 1999b). Overtime payments represent for a significant amount of personnel costs (IGIF 2000). The high percentage of external services is due to the conventioned sector (for example, laboratory exams and clinical analysis), while there is under-use of equipment in public hospitals.

[48] There is evidence that the NHS debts are under-accounted (Mossialos and Le Grand 1999). Auditors in some hospitals concluded that current costs are not being accounted for the correct year (some costs have been accounted for a lag of more than two years) (Tribunal de Contas 1999, 2000). Until 1999, evidence points to an under budgeting of the health budget, as initial budgets were lower in comparison to final spending of the previous year (Campos 2002).

[49] This figure includes a component of pharmaceutical expenditure, so it cannot be compared with the country data shown and interpreted just below.

[50] Besides cultural reasons and poor access to other types of care, this is partially explained by cost deductibility of drug expenditure in income taxes. Also, it reflects the lack of a national drug list for ambulatory care, a powerful influence of the pharmaceutical industry on doctors' decisions (EOHCS 1999), and the high presence of the private sector in non-hospital activity.

Table 2.8: Budget items

Budget decomposition I	*1998*	*Budget decomposition II*	*1998*
RHAs (without pharmaceuticals)	20%	RHAs (without pharmaceuticals, external consumption and outsourcing)	5%
Hospitals (without pharmaceuticals)	39%	Hospitals (without pharmaceuticals, external consumption and outsourcing)	30%
Psychiatry	2%	Psychiatry	2%
Others	1%	Others	1%
Central services and teaching	3%	Central services and teaching	3%
Pharmaceuticals	36%	Pharmaceuticals	36%
		External consumption and outsourcing (excluding pharmaceuticals)	24%
Total	100%	Total	100%
Amount (1,000,000 Euros)	3,879	Amount (1,000,000 Euros)	3,879

Source: (IGIF 2000)

Table 2.9: Breakdown of hospital costs

Cost structure	*1998*
Consumption	24%
External services	16%
Personnel costs	53%
Other costs	6%
Total costs	100%

Source: (IGIF 2000)

Most of hospital revenue is generated by transfers from the MoH. Although the law requires subsystems to pay the NHS for services consumed by their beneficiaries at NHS published rates, this has not happened. This is because many subsystem users are not identified in the NHS, and some of them are advised by the subsystems managers not to declare in which subsystem they are enrolled when they go to public hospitals (Pinto and Oliveira 2001)[51]. There is a lack of control for services provided to other private users in public hospitals, which implies a loss of income. General co-payments

[51] Some sub-systems have refused to pay their debts to the NHS: they consider their activities complementary to the NHS and their beneficiaries entitled to using public services since they pay general taxes. The MoH has explicitly recognised the problem and in 1999 made an agreement with the subsystems in which 50% of the debt was forgiven and the payment of the remaining 50% was re-scaled, with last payments planned for March 2001. There is no information on whether this plan was implemented.

have been low and often have not been charged. These features partly explain inequities in access to public hospital care.

2.4.4 Hospital policies

This sub-section describes hospital policies.

2.4.4.1 Capital

The lack of use of methods and explicit criteria for decisions on capital location has led to poor decisions in terms of equity, lack of control on equipment distribution and use, while it has contributed to perpetuating inefficiencies.

Hospital capital is highly concentrated in three cities: Lisboa, Porto and Coimbra. It is not clear which criteria have been used to choose the sites for new hospitals; what is clear is that these decisions have been made without regard to investment[52]. Decisions on the location of heavy medical equipment have been made independently of investment in hospitals (Ministério da Saúde 1998b). There are high regional variations in the number and age of equipment and its distribution among public and private facilities (Ministério da Saúde 1998b). Private equipment has been installed in the most urbanised areas with the lowest needs, which reinforces the pattern of geographical inequalities. Inequalities in the distribution of capital and equipment have been overall perpetuated.

Decisions on investment have not been linked to budgets for current expenditure and health care units have not had to pay for the use of capital[53]. Evidence indicates an inefficient use of capital resources, for example in terms of low occupancy rates (one of the lowest in the EU (OECD 2000)), long waiting lists and high variations and levels of

[52] About the opening of new hospitals, the MoH has recognised that there is not enough need for opening some of the hospitals on those locations and that a severe lack of human resources will restrain their opening (*O Público* 21.02.2000).
[53] I.e., the hospital financing system is working in a non-neutral form.

inefficiency estimates computed in hospital cost studies (Paiva 1993; Lima 1998; Carreira 1999)[54].

During the 1990s, Portugal continued a hospital building programme. Currently, most new hospitals (planned between 2000 and 2004)[55] are intended to replace previous hospitals and not to increase supply. This implies that the only way to correct current geographical inequities is to redistribute hospital supply. Recent attempts in health care investment have focused on increasing use of day care (Barros and Sena 1999) and have reinforced role of PC, as well as the creation of alternatives to traditional hospital care. Campos (Campos 1984) observed that the quality of the public hospital infrastructure varies greatly due to the presence of several old hospitals. Thus, despite the intention during the last two decades to replace and remodel the hospital network, hospital infrastructure conditions are very irregular (Pereira et al. 1999), while substantial investment is required to keep the infrastructure in good condition and to replace old hospitals[56].

2.4.4.2 Human resources

There are critical problems in both the distribution and supply of staff.

Although the ratio of physicians to population is close to the European average, that for nurses remains well below the European average (Table 2.10), and there is a scarcity of doctors for some specialties. The creation of new posts for doctors and nurses, within the NHS requires the approval of the Government and supply controls are applied since 1977, with the use of *numerus clausus*. Numbers of doctors are heavily influenced by the Medical Association, which limits enrolment in medical schools. However, mistakes in staff planning have led to the current widespread shortage of nurses and doctors in

[54] A more complete literature revision on efficiency and on the determinants of hospital costs is presented in Chapter 6.
[55] 12 new hospitals are to be opened, 4 in the Lisboa region, and 8 in the rest of the country (*O Público* 21.02.2000).
[56] Nonetheless, there is a lack of assessment about the need for funds to remodel and replace current supply, and how these are spread along the territory. Currently there is no public funding for building and opening some of the planned hospitals and both Socialists and Social Democrats have advocated and announced the use of private finance initiatives.

some specialities[57]. In order to fill this gap, two new medical schools have been opened in the last two years.

Table 2.10: Human resources indicators

1997	*Portugal*	*Spain*	*United Kingdom*
Certified/registered nurses per 1,000 population	3.7	4.6	4.5
General practitioners per 1,000 population	0.6	n/a	0.6
Practising physicians per 1,000 population	2.1	4.1*	1.4
Hospital employment per 1,000 population	10.1	9.4	16.4

Source: (OECD 2000)

*- 1995 value

The geographical distribution of doctors is very uneven, with a severe lack of doctors in remote areas and an excessive concentration in Lisboa, Porto and Coimbra. There has been no regulation of the distribution of doctors between areas, nor have payment systems (mainly based on salary and on number of years worked) been used to create incentives for doctors to move to rural areas. Until September 2002, NHS doctors were civil servants, paid on a salaried basis, and since 1993 they have been asked to choose one of four working regimes: part-time, full-time, extended full-time and exclusively for the NHS. In the 1993 reform, the great majority of them chose the full-time or the extended full-time regimes. This requires them to spend 35 hours or 42 hours per week in a public service, but allows for private practice, if authorised by their superiors (which is generally allowed). Those who have accepted to work exclusively for the NHS, tend to be older (and aiming at higher pensions) and younger (because of the surplus of doctors on certain specialities) (OECD 1994)[58]. Payment by salaries has led to undesirable incentives for doctors to maximise income by working overtime in the public sector and work simultaneously in the private sector. More generally, there are no financial incentives for doctors for high performance in their work in the public sector: on the contrary, there are incentives to provide low standards of care in the NHS and to transfer patients from the public to the private sector (Pereira et al. 1999).

[57] This problem is more severe given that a high number of doctors are reaching retirement age (*O Público* 21.02.2000).

[58] Medical staff wages in the public sector are more than 50% below the EU average, while the services provided in the private sector have on average prices one third higher than similar services in EU countries (OECD 1998).

Under the new hospital management law (from September 2002), individual labour contracts have replaced collective contracts and hospitals are free to recruit and to use different payment systems. This is yet to be applied and is expected to increase mobility of human resources (and potentially geographical equity) and to diminish allocative inefficiencies in hospitals with doctors constraining the use of resources. Through the national Medical Association, doctors have strongly opposed these changes in the regulation of physicians[59].

2.4.4.3 Waiting lists program

Similarly to other countries, a waiting lists program –the Promoting Access Program (*Programa de Promoção do Acesso*) started in 1999 with the intention of tackling inequities in access and the use of spare capacity of public hospitals (Ministério da Saúde 1999c). Its small budget on surgery targeted a small number of specialties for patients with waiting times that are clinically unacceptable. RHAs had to contract with hospitals of the public, social and private sectors, but priority was given to public hospitals (in order to make use of spare capacity). Nonetheless, the program of access is an example of a second best approach to solving problems of the hospital sector: for example, doctors are being paid for working extra time in the public sector but still have weak incentives for working in normal hours. Hence the program is not tackling key causes of inefficiencies in the system (Oliveira 2001a). The new government has changed the focus towards a higher use of the private sector and introduced greater freedom for patients' choice of providers (rather than focusing on the use of spare capacity in the public sector).

2.4.4.4 Hospital versus other health care sectors

Information on other sectors of the health system (other than the hospital sector) is necessary to understand the approach taken in this thesis:

[59] Past experiences on doctors payment systems (other than payment by salary) have had a very limited application. A new GPs' remuneration system was legislated in 1998, in an experimental format (Ministério da Saúde 1998d); GPs continued to be paid through salaries, but adjustments were introduced for patients' characteristics and the nature and length of doctors' work. Subsidies began to be paid for home visits and for the extension of working hours. Nonetheless, this policy was discontinued and not evaluated.

1. PC is complementary to hospital care in that in order to gain entry to hospitals, one has to go through the PC sector (via gatekeeping system), and this can reduce the need for hospital care through prevention.
2. Low provision of long-term care is expected to create problems, as hospitals have to delay discharges and thus increase length of stays (DGS 2001). Low provision of home and social care has also resulted in longer hospital stays. There is very little state provision of community care services in Portugal, particularly for long term care, day centres and social services for population groups with special needs (Pereira et al. 1999). There is a reliance on the family as the first provider of care in Portugal –this is traditional and part of the culture, and the informal network of care is particularly strong in rural areas (EOHCS 1999). However, family patterns are changing with increases in female participation in the labour market and with flows of young population to urban areas. This implies that reliance on informal care is weakening, particularly in rural areas (Santana 2000). Low provision of social and home care creates a bottleneck in the discharge of patients.
3. The low degree of continuity of health care has also been causing problems for the public sector and reinforcing inequalities in locational accessibility (Urbano, Bentes, and Vertrees 1993) (Reis and Carvalho 1994). Santana (Santana 1996) has studied the access and utilisation of emergencies and outpatient services to a hospital unit in Coimbra, using a patients' survey, and has estimated the relationship between distance and utilisation. Her results show accessibility problems, namely a negative correlation between distance (from residence to hospital) and services utilisation and a low level of referrals of patients from GPs to specialist care in hospital.

2.5 Why geographical inequities in the hospital sector?

Despite the significant progress to greater equity of access to health care following the creation of the NHS, inequities still exist as a consequence of high out-of-pocket payments, multiple coverage of risks, uneven supply of health care (public and private) and gaps in the regulation of private insurers, which is likely to result in cream skimming. The three main contributors to inequity are:

Public/private supply characteristics. Public hospitals dominate health care delivery and are unevenly distributed. Private provision offers higher quality and is partly funded by State resources (by tax deductions, by the non-identification of private utilisation of public infrastructures and by funds transfers to the conventioned sector); moreover, it tends to be located in the most urbanised areas, where there are higher concentrations of public hospitals.

Doctors dual employment and incentives. Doctors' freedom to work in both the public and private sectors seems to be a key factor for explaining inequities (lack of doctors in rural areas) and inefficiencies in the public sector (low bed occupancy rates, together with waiting lists, equipment under-use and soft budgets). Doctors escape accountability and have low incentives to work in NHS hospitals. In contrast, doctors' work in the private sector seems to be associated with higher quality.

Subsystems, VHI and population choice (i.e., multiple coverage). The population with NHS coverage only has limited choice, in comparison with populations with access to double coverage. Evidence suggests that the private insurance market is cream-skimming patients (for example, covering the youngest and the most able) and its presence is higher in urban areas, operating as complementary to public coverage and mainly insuring individuals of working age.

2.6 Concluding observations

The previous section has identified causes of inequities in the Portuguese hospital sector. This section summarises the main issues for resource allocation and for hospital policies and specifies the areas to which this thesis aims at contributing.

Firstly, the rationale for focusing on the hospital sector is its high use of resources and the lack of applied research on equity in the Portuguese hospital sector. Despite the improvements that followed the creation of the NHS (in health status, in NHS coverage and in NHS levels of supply), other inequalities remained or became worse, e.g. inequity of access and geographical inequities in the hospital sector.

Secondly, the lack of use of planning tools and the gaps in regulation has resulted in critical problems such as:

1. Hospital capital: opening of hospitals in locations where there were hospitals with low occupancy rates, and in areas without sufficient hospital doctors; failure to account for the impact of the opening of new hospitals on health revenue expenditure.
2. Hospital staffing: inadequate supply of doctors in some specialties and scarcity of nurses; doctors highly concentrated in three urban districts (Lisboa, Coimbra and Porto), and scarce in rural and remote areas; and no policies developed to motivate doctors to move to those areas and to correct a perverse private/public mix of health resources.

Thirdly, resource allocation methods have been poorly developed and decisions on investment and human resources policies have been made without paying attention to inequities; this has resulted in a failure to correct inequities. There is a lack of basic information on how resources are being allocated at all levels (locally, nationally and at intermediate geographical levels). Most hospital policies have sought to increase efficiency, with little attention to inequities between population groups and areas.

This thesis aims at developing information and tools to inform resource allocation as follows:

- Chapter 3 presents a geographical analysis of inequalities in the hospital sector;
- Chapters 4 to 8 develop a capitation formula to measure the degree of inequities of capital, finance and utilisation in the Portuguese hospital sector, by transferring methods available in international literature and creating new methods in the Portuguese context;
- Chapter 9 develops information on the redistribution of hospital supply required to improve equity of utilisation and access;
- Chapter 10 presents concluding remarks.

The next chapter gives evidence on geographical inequalities in the hospital sector.

3 CHAPTER 3 – Geographical analysis of inequalities in the hospital acute care sector

3.1 Introduction

Chapters 1 and 2 have presented evidence of the existence of inequalities in health and health care in Portugal. This chapter analyses geographical variations in hospital resources (in particular, in the hospital acute care sector) making use of data from various sources to describe degrees of inequality and deploying crude measures of inequality based on population numbers[60]. Results show that there are significant inequities and that better refinements of estimates are demanded in order to analyse the gap between demand and supply.

This chapter is organised into four sections that develop the concepts and methods for geographical analysis, apply these methods, summarise the implications of the findings for policy analysis, and present concluding remarks.

3.2 Concepts and methodological issues

This section defines and considers various definitions of geographical equity, as well as and the methodological principles that have led to the proposed research design for the geographical analysis undertaken.

[60] Geographic resident populations are used as a crude proxy for need, which neglects differences in population characteristics. As evidence in Chapter 1 pointed, there are no studies on needs variations in hospital care for Portugal. Whatever the definition of equity, the relative size of the groups for which equity is being pursued is likely to be a major determinant of expenditure distribution.

3.2.1 Defining the equity concept(s) and outputs from analysis

Chapter 1 has shown that there is a vast literature on equity, equity in health and health care. Chapter 2 showed that Portuguese policy and legislation aim at some kind of equity, although there is no clear overriding objective, nor a clear policy. This chapter analyses a set of health and health care indicators that relate to equal opportunity of access, using different definitions of equity, such as equity in utilisation, coverage and current expenditure. It provides quantitative evidence to support the description of geographical inequities presented in Chapter 1, captures the extent to which the 'inverse care law' applies to the Portuguese hospital sector, and shows which hospital variables explain allocation patterns.

3.2.2 Area level for geographical analysis

The district level (on average with populations between 50,000 and 500,000) was chosen for analysis, since it has been recommended as the appropriate level for identifying inequalities and for implementing policies to correct inequalities (WHO 1994)[61]. Districts are historically rooted in the previous territorial administrative division and are often used for planning purposes by the Portuguese MoH. Figure 3.1 illustrates the district division of the Portuguese territory, as well as the aggregation of districts to health regions (corresponding to the RHAs boundaries[62]). Inhabitants as well as territorial areas are very unevenly distributed across and within RHAs –the geographical level mainly used for planning purposes–, which is why health regions are not used as the main geographical level for analysis. Despite districts having similar area sizes, district heterogeneity (in terms of population size, geographic distribution of the population, geographic accessibility and socio-economic levels) might imply

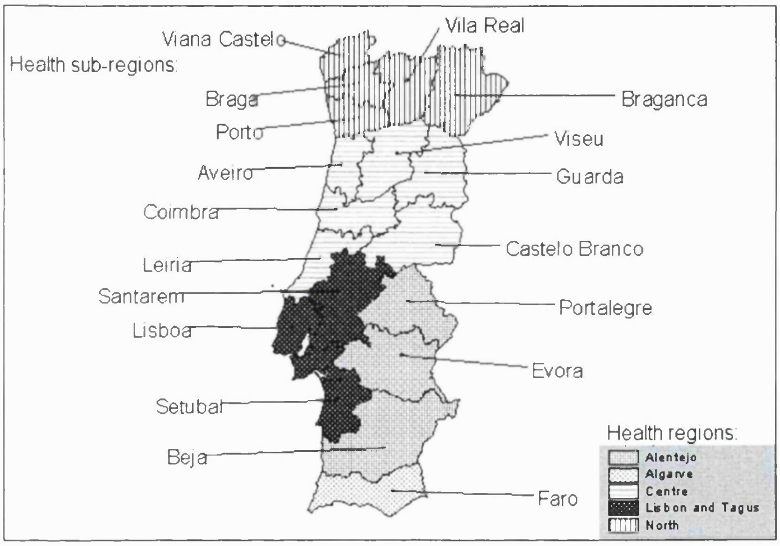
[61] Using 31.12.1997 data, Portuguese districts had an average of 500,000 inhabitants; the smallest and the largest districts had 125,680 and 2,052,330 inhabitants respectively.

[62] It should be noted that health administrative units differ from the administrative and statistical units in use; differences between health regions boundaries (RHAs) and region administrative boundaries are small.

problems with the ecological fallacy. Those characteristics should be taken into consideration in the analysis of results[63].

Analysis at the small area level and on location will make use of the *concelho* unit, which is a unit similar to the British electoral ward. This is the smallest administrative unit for which statistics are available and is a good basis for capturing variations due to geographical accessibility[64].

Figure 3.1: Geographical boundaries of 5 health regions and 18 health sub-regions (districts)[65]



3.2.3 Methods used

This sub-section describes the methods used through a description of the database, a definition of the design format and indicators in use, as well as comments on the adequacy of the measures used.

---

[63] Nevertheless, the alternative geographic units are not appropriate for the analysis of geographic inequalities: the health region level is too aggregated, whereas the *concelho* is much disaggregated (described below).
[64] Using the Portuguese administrative division in *concelhos* in 31.12.1997, there were 275 *concelhos* with an average of 34,610 inhabitants; the smallest and the largest *concelhos* had 2,137 and 609,004 inhabitants, respectively.
[65] Information on the source of the map is provided in the sub-section below that describes the database.

3.2.3.1 Building a database

The database includes only public acute care NHS hospitals owned by the MoH[66]. This exclusion implies under-estimation on the level of inequalities because most of the excluded hospitals are located in Lisboa, Porto and Coimbra, which are areas that concentrate public supply. Nonetheless, the excluded public acute beds are a negligible portion, in comparison to the number of hospital beds included in the sample. Data sources correspond to data collected and computed by different institutions, and sometimes at different times[67]: hospital capital and performance indicators refer to 1996; population figures correspond to 31.12.1996; financial flows from the MoH to hospitals are of the 1995 fiscal year; and distances from hospital acute care points of provision are computed from 1998 data. The maps were downloaded of the site from the Portuguese National Institute for Geographic Information in March 2000[68], and are the responsibility of the governmental department General Direction of the Environment (DGA 1998). The geographic information system program Arcview (Environmental Systems Research Institute 2000) was used to work out the maps.

To test whether an 'inverse care law' applies, districts were aggregated into three groups (Table 3.1): rural, semi-urban and urban. This classification was created for the purposes of this study due to the absence of available classifications and was based on information on the availability of supply, geographic accessibility to the main points of supply and socio-economic levels (as measured by illiteracy rates and the proportion of population dependent on primary sector activities, census 1991 data)[69]. Table 3.2

[66] In a few cases, hospital units correspond to hospital groups, comprising 2 or 3 units –this happens as these hospitals share the same administration, are geographically nearby, and are treated by the MoH as single units for statistical and financial purposes. Some public hospitals were excluded because of lack of data. The following groups were excluded: cancer institutes, psychiatric hospitals, military hospitals under the financial and administrative responsibility of the Ministry of Defence, and hospital institutions that are managed by the MoH in association with other entities, having as a result a special status (such as Alcoitão and Santana health units).

[67] Sources used: (IGIF 1997), (DGS 1998), (IGIF 1998), (Departamento de Estudos e Planeamento da Saúde 1997/8), (Eurostat, INE, and European Commission 1998), (Ministério da Saúde 1998b), (INE 1993a, 1993b, 1993c, 1993d, 1993e), (INE 1990/1/2/3/4/5/6/7/8).

[68] Taken from website: http://snig.cnig.pt/snig/framemg.htm.

[69] Urban districts are those with high availability of supply (or high level of accessibility to the main points of supply) and high socio-economic levels (in relation to national averages); rural districts are those with low availability of and access to supply and with low socio-economic levels; and semi-urban districts have intermediate characteristics on those indicators.

provides some information on hospital supply in these groups of districts: urban districts offer most of the hospital supply and include the largest hospitals.

Table 3.1: Districts classified in the rural/urban spectrum

Rural districts	*Semi-urban districts*	*Urban districts*
Beja, Bragança, Castelo Branco, Évora, Guarda, Portalegre, Vila Real, Viseu	Aveiro, Faro, Leiria, Santarém, Viana do Castelo	Braga, Coimbra, Lisboa, Porto, Setúbal

Table 3.2: Some indicators of hospital supply of classified hospitals (sample)

	Population %	*Official hospital beds*	*Number of hospitals*	*Average hospital size (beds)*
Rural	17.1%	4,612 (17%)	19 (22%)	242.7
Semi-urban	22.7%	4,600 (17%)	23 (26%)	200.0
Urban	60.2%	18,360 (67%)	46 (52%)	399.1

3.2.3.2 Design format and indicators in use

This sub-section defines the hospital related variables to be analysed in the chapter, and specifies the indicators for each of those variables. Separate analysis of individual variables constitutes a simplified framework as it neglects interaction between variables. By contrast, analysis of several variables attempts to include some of the multiple aspects relevant to inequality comparison (from the social judgement viewpoint) (Atkinson 1983).

The hospital related variables, to be analysed, are presented in Table 3.3 and are divided into demand, supply, and interaction of supply and demand factors.

Table 3.3: Hospital sector related variables

***Demand** side factors*	***Supply** side factors*	*Factors that result from **supply/demand** interaction*
Demographic profile Socio-economic population profile Population coverage	Hospital supply: capital (beds and equipment) and human resources (doctors and nurses) Hospital building quality Hospital administrative organisation Hospital sector related resources: private hospital care and PC	Hospital utilisation Hospital expenditure Hospital location accessibility

For each of the variable of Table 3.4, the following indicators were selected:

Table 3.4: Hospital variables and corresponding indicators for analysis

VARIABLES	*INDICATORS*
Demographic	Population density and population proportion above 65 years old
Socio-economic	Illiteracy and unemployment rates, proportion of households without electricity installation and proportion of the population dependent on the primary sector
Population coverage	Single covered (under the NHS coverage) estimated population
Hospital supply	Hospital beds, number of hospitals, size of hospitals, hospital equipment and hospital doctors and nurses
Hospital location accessibility	Hospital location and distances travelled to access hospital care
Hospital utilisation	Inpatient days, discharges, external consultations and emergencies
Hospital building quality	Hospitals' age by year of construction and by year of starting activity
Hospital expenditure	Current revenue, NHS revenue, personnel costs, revenues from the private sector and from subsystems, hospital expenditure in pharmaceuticals and clinical exams and investment on capital
Hospital administrative organisation	Supply, utilisation and expenditure analysed in the administration classification
Hospital-related resources	PC health centres and extensions, GPs, PC nurses, PC consultations, private hospital beds and private hospital location

3.2.3.3 *Tools for analysis of inequality*

Four measures of inequity are used. Their characteristics and strengths and weaknesses are:

A. *District per capita shares* are one-dimensional indicators presented in tables or in maps;

B. *Lorenz curves* incorporate fundamental principles of measuring inequality (Cowell 1995)[70]; they present information on the distribution of resources across population groups/areas in a diagrammatic form; provided that Lorenz curves do not cross, they give ranking of degrees of inequity (a curve which dominates is preferred for all social welfare functions that are non-decreasing and strictly concave (Atkinson 1989))[71];

[70] The Lorenz curves were computed with information already aggregated at the district level.

[71] In accordance with the purposes of this analysis, it was considered that Gini coefficients are not an important tool for analysis as they have implicit judgements on the weight to be put on variations in

C. *Coefficient of variation* is a single *ad hoc* measure but can be used to compare variables measured in different units;

D. *Correlation indicators* are simple statistical tools used to show relationships between different measures of supply.

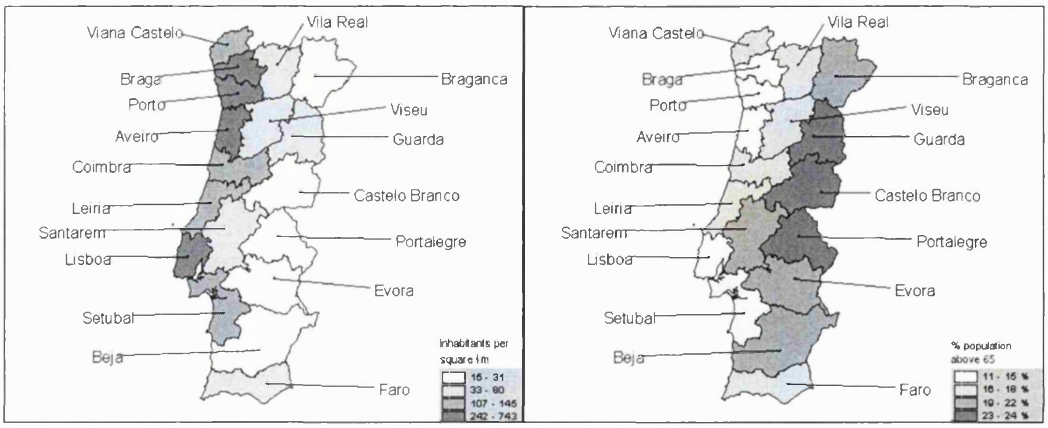
3.3 Geographical analysis

This section presents the geographical analysis applied to the Portuguese hospital sector.

3.3.1 Demographic characterisation

The Portuguese population is unevenly distributed (Figure 3.2). There are concentrations of populations in the metropolitan areas of Lisboa (21.7% of mainland population) and Porto (17.9%) and most of the rest of the population lives in Braga (8.3%), Aveiro (7.2%) and Setúbal (7.8%). Rural districts located in the interior, tend to be sparsely populated and have on average older populations (Figure 3.2). The north and the coast are inhabited by the youngest populations on average.

Figure 3.2: Population density (inhabitants in square kilometres) and population portion above 65



different parts of the variables distribution. Those weights do not seem to correspond to reality (Atkinson 1989).

3.3.2 Socio-economic information

In order to identify material deprivation and hence proxies for poor health, four census variables were selected and aggregated to the district level: unemployment rates, proportions of illiterate individuals of the 1991 resident population[72], houses without electricity[73] and adults who work in the primary sector.

Analysis in terms of material deprivation shows divisions between urban and rural areas, and between the coast and the interior. Illiteracy rates are higher in the interior districts, in relation to coastal and urbanised areas and unemployment rates are higher in the southern and northern interior districts and in two districts in the Lisboa and Tagus Valley region (Lisboa and Santarém) (Figure 3.3). The coastal/interior dichotomy is clearly marked in the proportion of houses without electricity and in the proportion of population working in the primary sector (Figure 3.4). This *ad hoc* analysis reveals that only unemployment rates do not present the north/south and coastal/interior divides[74]. These geographical patterns of the coast/interior and urban/rural differentiation have been previously documented (INE 1998) (Santana 1999). It seems that indicators from the 1991 census capture better rural deprivation rather than the concept of urban deprivation –for example, difficulties in access to education and dependence on primary sector activities are more likely in rural areas.

---

[72] Computed as the ratio between: illiterate individuals aged over 10 over resident population in 31.12.1991.

[73] Computed as the ratio between: home installations without electricity used as usual residence over resident population in 31.12.1991.

[74] Nonetheless, the determinants of unemployment are much more diverse than the determinants of the other socio-economic indicators. Unemployment depends on the economic structure of the area and on the dynamic of the industry and services sectors: while rural and interior areas rely on agriculture and have a regressive economic structure that justify high unemployment rates, urban unemployment is partly explained by sensitivity to the economic cycle, and by inadequacies between supply and demand for jobs.

Figure 3.3: Illiteracy and unemployment rates

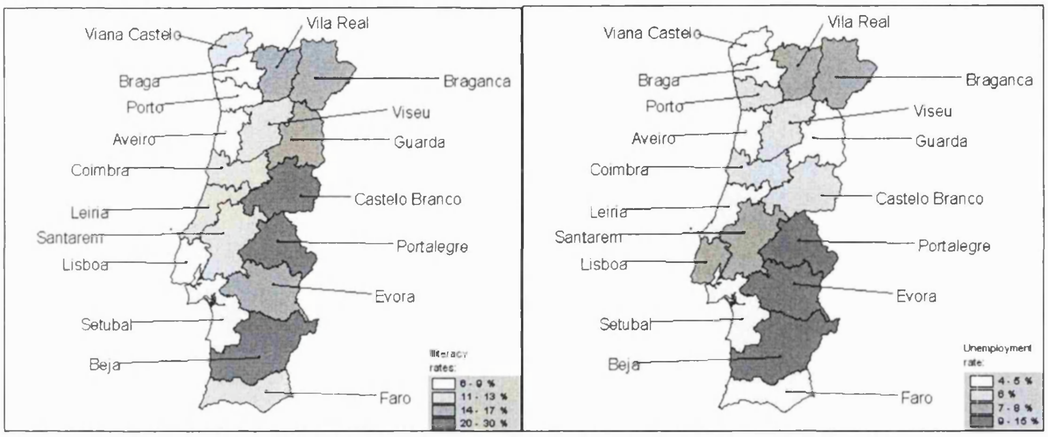
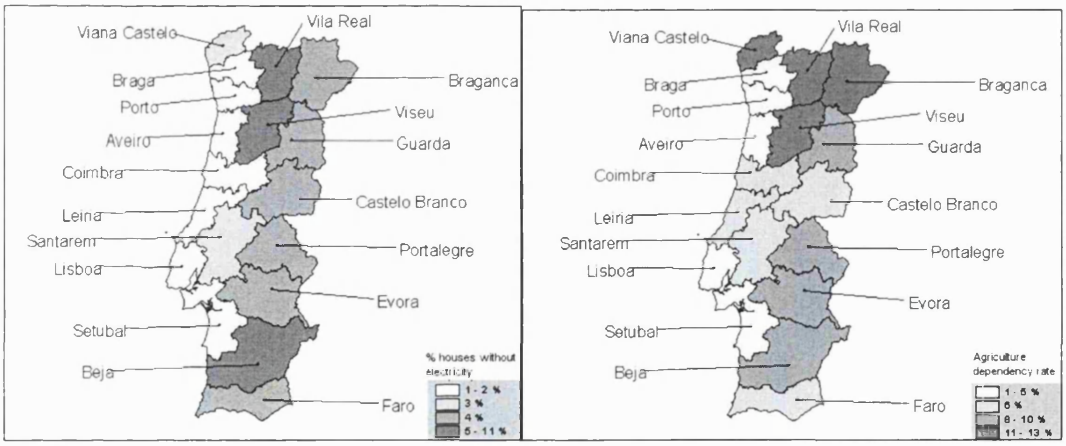


Figure 3.4: Proportion of houses without electricity installation and proportion of population dependent on primary sector activities



3.3.3 Health care coverage

Chapter 2 has shown how overlapping coverage is a key feature of the Portuguese health system. Table 3.5 shows data on 'the single-covered NHS beneficiary population' (MoH estimates), defined as population without any additional coverage to NHS coverage. Data is only available at the RHA level, but district behaviour can be inferred, as explained below[75]. The 22.2% of the population that are double/multiple-

[75] It should be acknowledged that these numbers differ from the ones produced in other sources, such as the National Health Survey (Departamento de Estudos e Planeamento da Saúde 1997b). MoH estimates were chosen because the national percentage of single covered NHS population from the first source

covered is distributed geographically as follows[76]: 47% in the Lisboa and Tagus Valley region, 25% in the North and 20% in the Centre region as shown in Table 3.5. As subsystems coverage tends to exist in tertiary sector activities, and those activities are concentrated in the main urban centres, double coverage is concentrated in the most urbanised districts. Inside the Lisboa and Tagus Valley region, tertiary sector activities are expected to be concentrated in the Lisboa and Setúbal districts (mainly Lisboa, as it concentrates most public sector activity); in the north, the Porto district is expected to be capturing the highest portion. This distribution of tertiary sector activities accounts for the geographic distribution of doubled-covered population shown in Table 3.5.

Table 3.5: Resident vs. NHS population

Regions -1996	*Resident population (share)*	*MoH estimates of NHS population (shares)*	*MoH estimates/resident population*	*Double-covered population distribution*
North	3,097,000 (33%)	2,577,103(35%)	83.2%	24.8%
Centre	2,313,460 (25%)	1,900,401(26%)	82.1%	19.7%
Lisboa and Tagus Valley	3,222,210 (34%)	2,238,286(30%)	69.5%	46.9%
Alentejo	454,670 (5%)	349,453 (5%)	76.9%	5.0%
Algarve	346,110 (4%)	271,229 (4%)	78.4%	3.6%
Total	9,433,450 (100%)	7,336,472 (100%)	77.8%	100.0%

3.3.4 Hospital supply

3.3.4.1 Capital infrastructure

Numbers of beds are commonly used to measure hospital size, but this is a crude measure because these numbers are heterogeneous and may hide other parameters and not be comparable (Berki 1972), e.g. the use of beds might be constrained by capital equipment and other hospital resources (Butler 1995).

There is an unequal distribution of hospital beds in relation to population (Figure 3.5): the number of hospital beds per 100,000 inhabitants ranges from 178 and 640. Five

better approximates data from other sources, such as the one indicated by the OECD (75%) (OECD 1998).

[76] It is unknown to which extent the population is double- or triple-covered due to cumulative benefits from several subsystems or due to combined subsystems plus insurance coverage.

urban districts have higher concentration of beds (Table 3.6), as well as a couple of rural districts –Vila Real and Castelo Branco (Figure 3.5). In contrast some districts by the coast –Aveiro, Leiria and Faro– and some rural districts –Beja and Viseu– have fewer beds. As Portugal's ratio of inpatient beds to population is slightly below the EU average, under-provided Portuguese districts are well below the average for the EU.

Figure 3.5: Hospital beds per 100,000 inhabitants

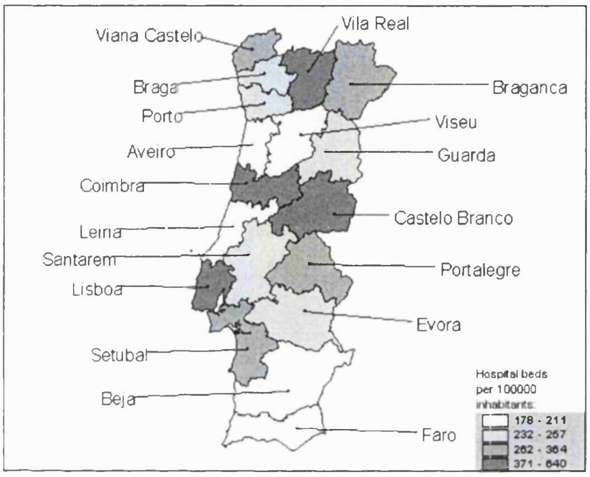


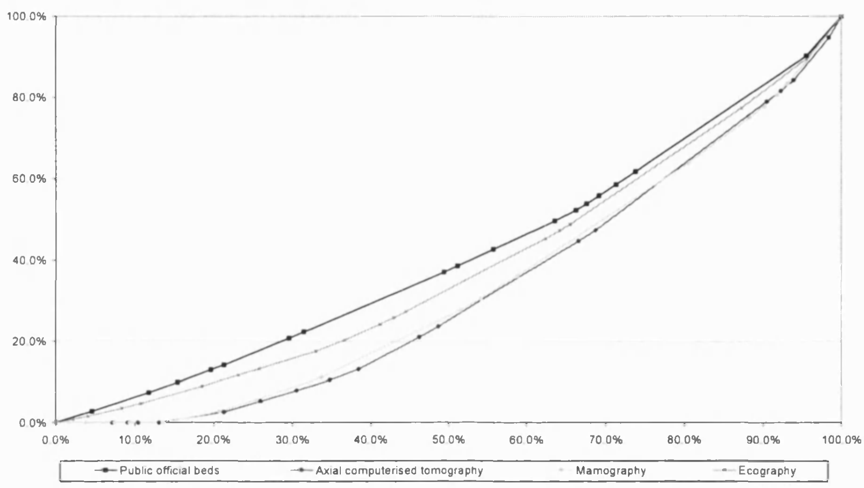
Table 3.6: Hospital beds vs. population

District	*Population %*	*Official beds*
Rural	17.1%	16.7%
Semi-urban	22.7%	16.7%
Urban	60.2%	66.6%

3.3.4.2 Equipment

Lorenz curves of three types of technology in NHS hospitals (ecography, axial computerised tomography and mammography) show this technology is distributed more unequally than beds (Figure 3.6) and is highly concentrated in Porto, Coimbra, Lisboa and Setúbal.

Figure 3.6: Lorenz curves for public hospital beds and for three types of equipment



3.3.4.3 Human resources

The distribution of doctors and nurses in NHS hospitals follows that of hospital beds. Figure 3.7 and Table 3.7 show that hospital doctors are more unequally distributed than hospital beds and that Lisboa, Porto and Coimbra account for 53% of hospital beds and 72% of hospital doctors. Figure 3.9 also shows that hospital doctors are concentrated in the main urban centres. These results mean that patients from rural areas are relatively disadvantaged with respect to access to hospital services. As Lorenz curves for hospital nurses and doctors cross, it is not possible to determine which is more inequitably distributed.

The maldistribution of nurses imposes an additional problem because they are a critically scarce resource (as shown in Chapter 2). The mix of nurses to doctors (and to beds) highly varies between districts and this raises questions of how hospital systems might be operating differently and how these variations relate to hospital allocative efficiency[77].

[77] Questions on this are reinforced by the fact that Portuguese occupancy rates are very low in the EU context, as described in Chapter 2, and highly vary within Portugal. Allocative efficiency relates to the use of the optimal mix of inputs to produce a certain level of outputs, while technical efficiency relates to the maximisation of outputs for a certain level of inputs.

Figures 3.9 and 3.5 highlight differences in the distributions of doctors and beds. This adds to descriptive evidence in that the limited availability of doctors has constrained the use of beds for hospitals located in rural areas, even if these beds are available. This means that in order to analyse the distribution of hospital supply, it is better to use the distribution of doctors, as this is a more adequate indicator of productive capacity.

Figure 3.7: Lorenz curves for public hospital resources

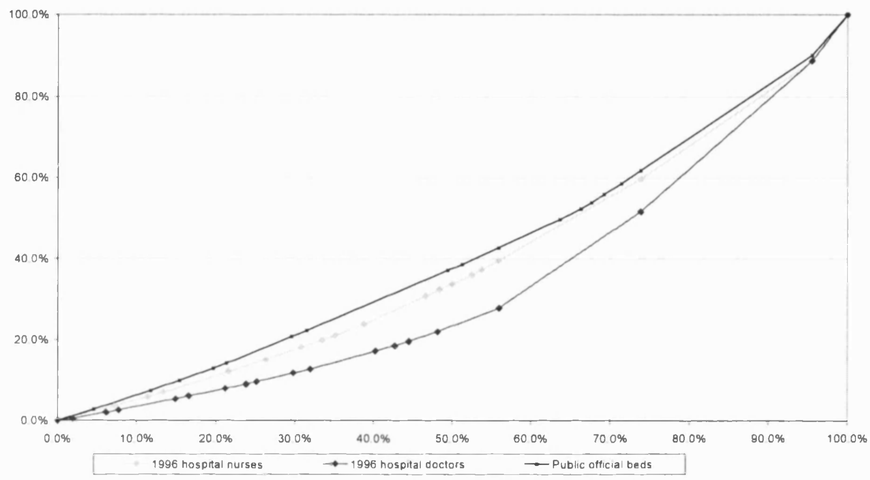


Table 3.7: Some comparative statistics

	Hospital doctors per 100,000 inhabitants	*Hospital nurses per 100,000 inhabitants*	*Hospital beds per 100,000 inhabitants*
Maximum	363.8	629.0	640.4
Minimum	41.2	123.9	177.6
Average	102.1	225.4	290.6
Coefficient of variation	0.82	0.51	0.38

Figure 3.8: Human resources per 100,000 inhabitants, normalised by the national average

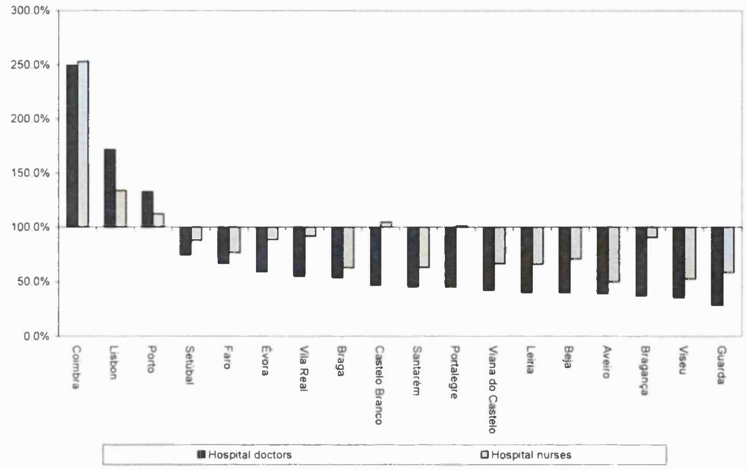
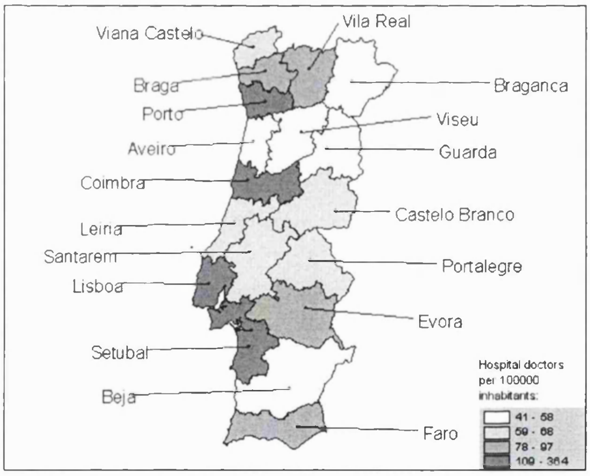


Figure 3.9: Hospital doctors per 100,000 inhabitants



3.3.5 Population accessibility

Figure 3.10, using small area *concelhos*, shows that hospital units are concentrated along the coastal areas. This means that there are large variations in average distances that patients travel to reach a hospital (Table 3.8), and hence there is inequity in accessibility. There is evidence that distance might deter utilisation –Chapter 7 provides further information on the role of physical accessibility, perceptions of availability and other geographical variables in utilisation[78].

Figure 3.10: Distribution of public hospital beds in Portugal (by *concelho*)

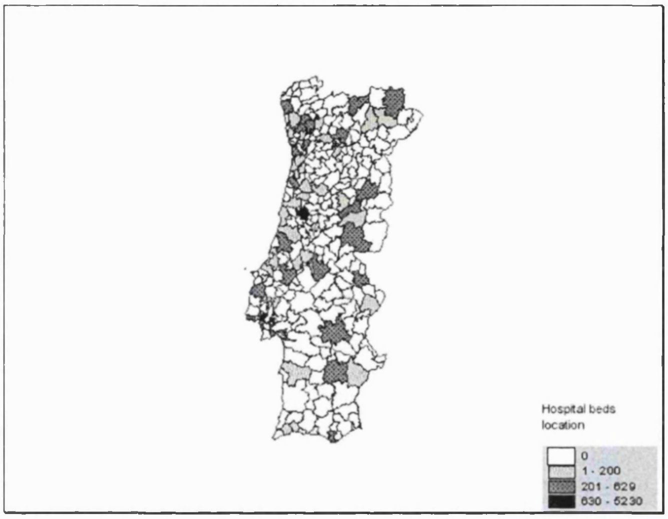


Table 3.8: Average distances (in kilometres) travelled in order to find inpatient care

1998	*Travelled distance*
Aveiro	14
Beja	51
Braga	11
Bragança	31
Castelo Branco	22
Coimbra	6
Évora	37
Faro	34
Guarda	36
Leiria	23
Lisboa	9
Portalegre	28
Porto	9
Santarém	26
Setúbal	10
Viana do Castelo	24
Vila Real	19
Viseu	22
Mainland Portugal	15

Source: DRG database 1998, provided by IGIF and Euclidean distances[79] between centroids of *concelhos* (centroids co-ordinates provided by General Direction of Environment)

[78] Nonetheless, other policy objectives (such as efficiency achieved with hospital minimal size) might

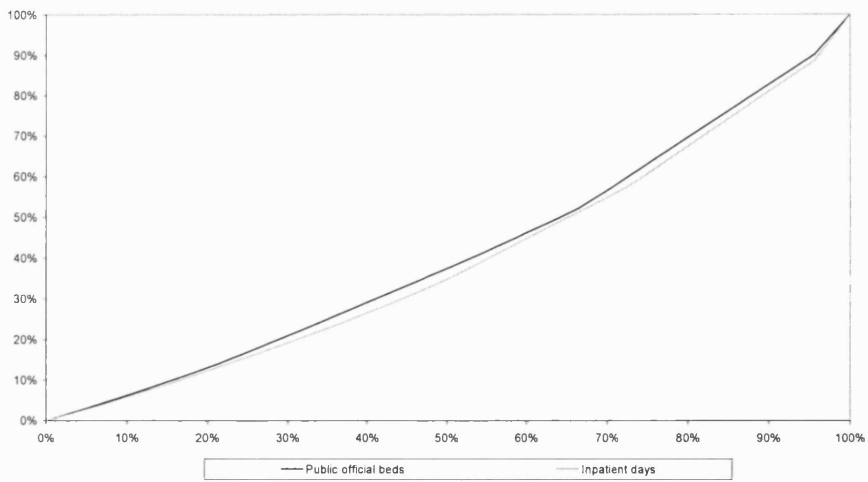
3.3.6 Utilisation of hospital care

Four hospital measures of output are analysed here: inpatient days, inpatient discharges, external consultations and emergency admissions. Measures of hospital supply (beds, doctors and nurses) and utilisation (at the district level) are strongly correlated, as expected (Table 3.9) (all correlation coefficients are greater than 95%). Inpatient days are more unevenly distributed than hospital beds (Figure 3.11).

Table 3.9: Correlations at the district level

	Hospital doctors	*Official beds*	*Inpatient days*	*Inpatient discharges*	*External consultations*	*Emergency entries*
Hospital nurses	99.7%	98.4%	99.5%	99.5%	98.8%	96.1%
Hospital doctors		98.5%	99.5%	99.1%	98.7%	95.9%
Official beds			99.0%	98.7%	95.7%	95.6%

Figure 3.11: Lorenz curves of public hospital beds and inpatient days



justify variations in accessibility.

[79] Euclidean distance is the mathematical formula used to calculate a crow-fly line between two geographic points.

Table 3.10 shows that rural districts have a lower share of utilisation than of capital (as measured by the number of beds), and their shares of beds (17%) is much higher than that for doctors (8%) –which partly explains lower occupancy rates; the 5 urban districts account for two-thirds of all external consultations because of the concentration of specialised services in them.

Table 3.10: Hospital dimension and utilisation indicators

District	*Official beds*	*Hospital doctors*	*Inpatient days*	*Discharges*	*External consultations*	*Emergencies*
Rural	16.7%	8%	15.2%	15.6%	9.7%	13.9%
Semi-urban	16.7%	10%	15.4%	16.7%	12.9%	20.8%
Urban	66.6%	82%	69.3%	67.7%	77.4%	65.3%

3.3.7 Hospitals' investment and age

There are only a few descriptive studies of the quality of hospitals infrastructure for Portugal. One crude 'proxy' for the quality of hospital infrastructure is age (Mayston 1990). Although age does not take account of structural changes since the hospital was built (or used as a hospital), it indicates the relative age of the hospital network. The year of construction and the first year of use of the main hospital building were taken as an approximate indication of the age[80]. Data linking the year of construction and beginning of activity with hospital beds in 1996 has been analysed. Data excluded from the database in use in this chapter is presented in Table 3.11.

Table 3.11: Missing data on hospital year of construction or on hospital year of beginning activity

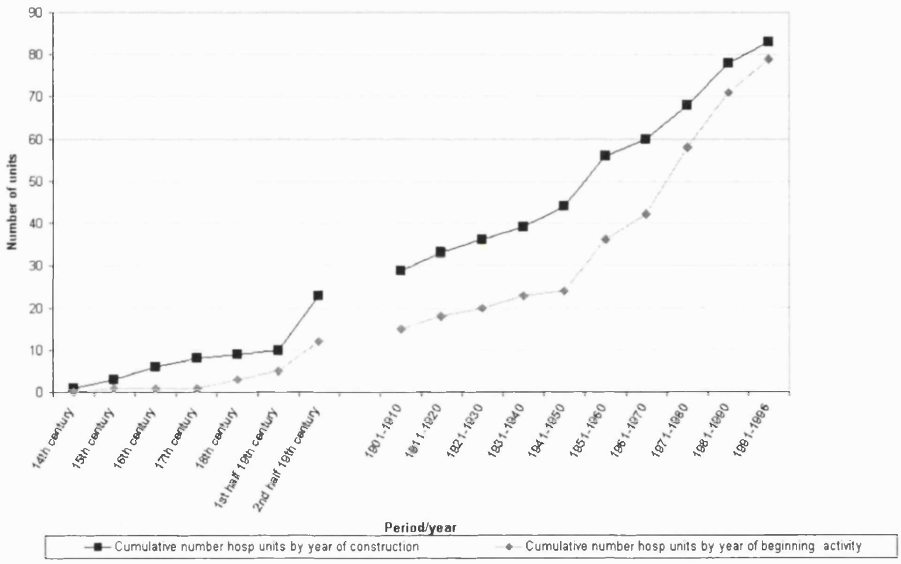
	Missing data: number of hospital beds	*Missing data: number of hospital units*	*Number beds included*	*Number of hospital units included*
Data on the year of construction[81]	1,554	5	26,018	83
Data on the year of beginning of activity	2,986	9	24,586	79

[80] When hospital units aggregate several hospitals, the age of the main infrastructure was used as the age of those hospitals. This assumption might be misleading in some cases, and is shown to influence the results for Coimbra (use of unpublished information supplied by an ex-Department of Health Planning and Statistics).

[81] The year of construction is the year when the main infrastructure building was finished. The year of beginning of activity is the year when the infrastructure was firstly used as a hospital.

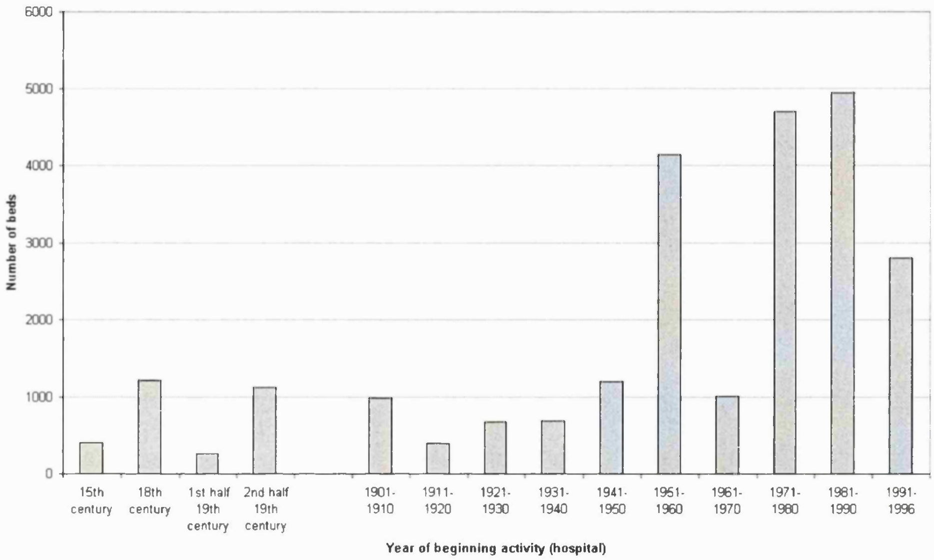
Figure 3.12 shows the cumulative number of hospitals by two criteria: year of construction of the main hospital building and year of starting activity. Figure 3.13 shows the number of hospital beds allocated to the year of starting activity. Some hospitals date back to the Middle Ages, and many hospital buildings were firstly used for other operations, being converted into hospitals later[82] (Figure 3.12). There was a second wave of new hospitals in the nineteenth century and a huge increase in hospital provision was registered in the last fifty years (Figures 3.12 and 3.13).

Figure 3.12: Cumulative number of hospital units at some points in time (by year of construction and by year of beginning activity)



[82] That is the case for buildings built by catholic institutions and initially used for religious purposes (*conventos*).

Figure 3.13: Number of beds (1996) allocated to the year of beginning activity



Most hospitals before 1900 were built mainly in coastal and urban districts, in particular, in Lisboa and Porto (Figures 3.14-3.16). The massive development of Coimbra is recent, taking place in the 1980s (although hospital supply there has partly replaced and increased previous supply). Later investments have targeted rural and semi-rural districts. During the new hospitals latest phase (1980-1996): from a total of 7,023 new hospital beds, 34% were opened in Coimbra (corresponding to the development of new hospital units with teaching facilities)[83], while 16% and 12% were opened in Setúbal and Santarém, respectively, which surround the Lisboa district (Figure 3.16). During the 1990s, Portugal continued to invest in new hospitals and in PC centres.

[83] This value is biased as part of the infrastructure had been constructed prior to 1980.

Figure 3.14: Distribution of hospital beds in new hospitals constructed <u>before 1900</u> (number of beds: 5,657)

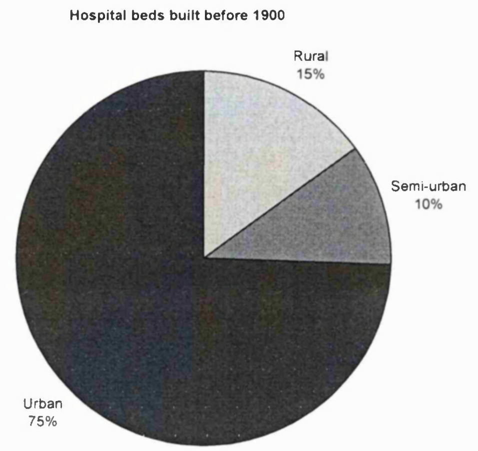


Figure 3.15: Distribution of hospital beds in new hospitals constructed <u>in the period 1900-1979</u> (number of beds: 11,278)

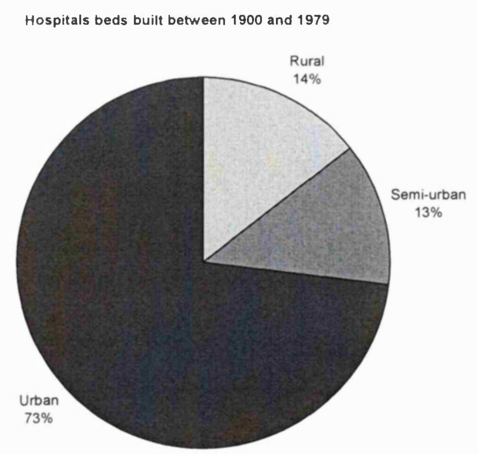
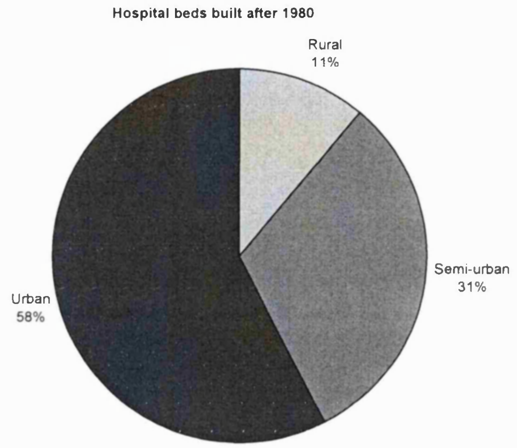


Figure 3.16: Distribution of hospital beds in new hospitals constructed in the period 1980-1996 (number of beds: 7,023)



3.3.8 Expenditure on acute care hospitals

Geographical analysis of public per capita expenditure shows high inequalities. As expected, three districts that have concentrations of supply also have the highest NHS revenue shares (Figure 3.18). In urban areas, revenue shares are higher than hospital beds and population shares; the opposite applies to rural and semi-urban districts (Table 3.12)[84]. Hospital private sector and subsystem revenues and labour costs follow closely the hospital revenue distribution.

Table 3.12: Expenditure (% distribution across districts)

Districts	*Population*	*Official public beds*	*Hospital total revenues*	*Hospital revenues from private and subsystems*	*Costs of personnel*
Rural	17.1%	16.7%	12.0%	11.4%	12.8%
Semi-urban	22.7%	16.7%	13.7%	14.4%	14.2%
Urban	60.2%	66.6%	74.3%	74.2%	73.0%

Lorenz curves (Figure 3.17) show that the current expenditure distribution is more unequal than for hospital beds, but less unequal than for hospital doctors.

[84] This analysis neglects other hospital activity variables that impact on expenditure, such as specialties differentiation, teaching status, possible economies of scale, case-mix variables, etc.

Figure 3.17: Lorenz curves –current expenditure vs. public beds vs. hospital doctors

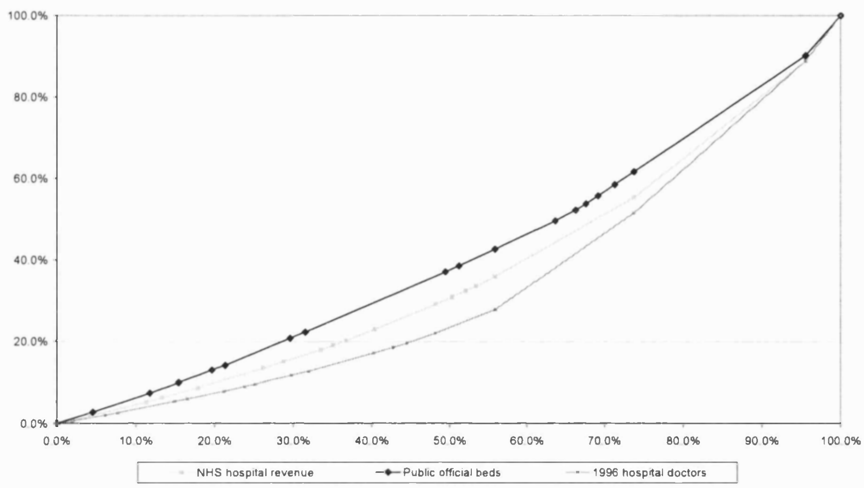
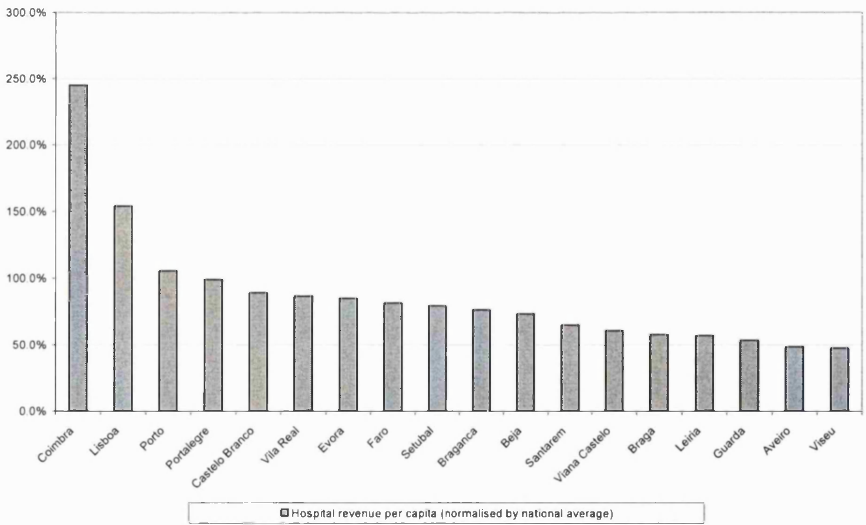


Figure 3.18 gives district rankings in terms of hospital expenditure per capita. The two extreme districts, Coimbra and Viseu, are more than 140% above and more than 50% below the national average per capita expenditure respectively.

Figure 3.18: District total revenue per capita, normalised by the national per capita average



Thus, the concentration of resources in the three centres of Lisboa, Porto and Coimbra is due to their concentration of beds, equipment and human resources. These are the districts with the main teaching units and central hospitals and with the largest hospital units. A more equitable distribution of resources would imply substantial reductions for

Lisboa and Coimbra, a relatively small reduction for Porto, and gains for all the other districts.

3.3.9 Other variables

Analysis of other variables explains the pattern of geographical inequalities in expenditure and informs the modelling in other chapters of the thesis. This analysis covers both expenditure across administrative levels of the network and inequalities in the distribution of primary and private care.

Administrative classification. The hierarchical nature of the hospital system seems to be a key explanatory factor of hospital inequalities and provides information on the hospital production and costs (relevant to Chapter 6). Hospital data based on the current administrative classification described in Chapter 2 (section 2.4.2) confirms previous findings (Tables 3.13 and 3.14):

1. Central and general hospitals are all placed in Lisboa, Porto and Coimbra, which have 9, 3 and 2 hospital units respectively; they have a high average size (668.5 beds per hospital) and contain 33% of the total number of beds; they concentrate 66% of all hospital doctors and their shares of inpatients and of external consultations are higher than their share of number of beds but lower then their share of doctors; in 1996, these 14 central hospitals received half of the NHS current expenditure and over 60% of total hospital expenditure on pharmaceuticals and expenditure in diagnostic tests.
2. Central and specialised hospitals are also located in those three districts (except for one hospital of 192 beds placed in Setúbal); this group of hospitals has an average capacity of 208 beds, and accounts for 6.8% of acute beds and 3.2% of the hospital doctors; specialised hospitals attract a high proportion of user charges, and a more than proportional share of income from subsystems.
3. District hospitals are distributed throughout the country; all the districts contain at least one hospital. They have a higher than average size –356.7 beds–, account for half of the total number of beds in the system, but have only 25% of hospital doctors. Although they account for 50% discharges and emergencies, they only represent 37% of NHS revenue. District hospitals make a high use of outsourcing.

4. Level I hospitals are located in 5 rural districts and have 93.5 beds on average. They account for 9% of the beds and 4% of hospital doctors, and for 8% of discharges and 17% of emergencies.

Table 3.13: Hospitals classified per administrative group and some descriptive indicators I

	N. hospital units	*Average size*	*% Beds*	*% Doctors*	*Inpatient days*	*Discharges*	*External consultations*	*Emergencies*
Central and general	14	668.5	33.9%	66.1%	41.2%	33.5%	45.0%	25.8%
Central and specialised	9	208.0	6.8%	3.2%	7.0%	7.5%	11.2%	3.9%
District	39	356.7	50.5%	26.7%	44.1%	50.8%	37.5%	53.8%
Level I	26	93.5	8.8%	4.0%	7.7%	8.2%	6.3%	16.5%
Total	88	313.3	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 3.14: Hospitals classified per administrative group and some descriptive indicators II

	NHS revenue	*User charges*	*Subsystems*	*Sub-contracts expenditure*	*Pharmaceuticals expenditure*	*Clinical exams expenditure*
Central and general	51.8%	31.0%	30.6%	45.0%	62.2%	61.8%
Central and specialised	6.0%	21.6%	44.7%	2.8%	3.2%	3.4%
District	36.6%	34.6%	21.6%	43.0%	31.2%	31.2%
Level I	5.6%	12.7%	3.1%	9.2%	3.4%	3.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Primary care. PC is more equitably distributed than hospital care[85]. Provision in health care centres extensions varies enormously[86] (Table 3.15). The number of doctors is used to indicate size.

Analysis shows that PC doctors follow approximately the population distribution and PC nurses are better represented in rural and semi-urban than in urban districts (Table 3.15). Lisboa and Coimbra were found to have a higher proportion of GPs than its population share (Figure 3.20) and hence this finding adds to their inequitably high

[85] As emphasised in Chapter 2, there are critical problems of continuity in the Portuguese health care system; and there is no information on whether primary care is mainly acting as complementary, or as substitutive to hospital care.

[86] Extensions are peripheral units attached to health care centres, located under the health centres areas of influence, that have the objective of providing primary care at higher proximity to patients (INE 2000).

provision of hospitals. Braga and Porto have less than their fair share of GPs. GPs are more equitably distributed than hospital doctors (Figure 3.19)[87].

Table 3.15: Primary care –distribution in terms of some parameters

	Population	*% Health Care Centres*	*% Extensions*	*% Doctors*	*% Nurses*	*% Consultations*
Rural	17.1%	34.5%	38.8%	16.9%	22.3%	17.1%
Semi-urban	22.7%	24.6%	28.2%	21.8%	24.0%	25.1%
Urban	60.2%	41.0%	33.0%	61.3%	53.7%	57.8%

Figure 3.19: Lorenz curves –GPs vs. hospital doctors

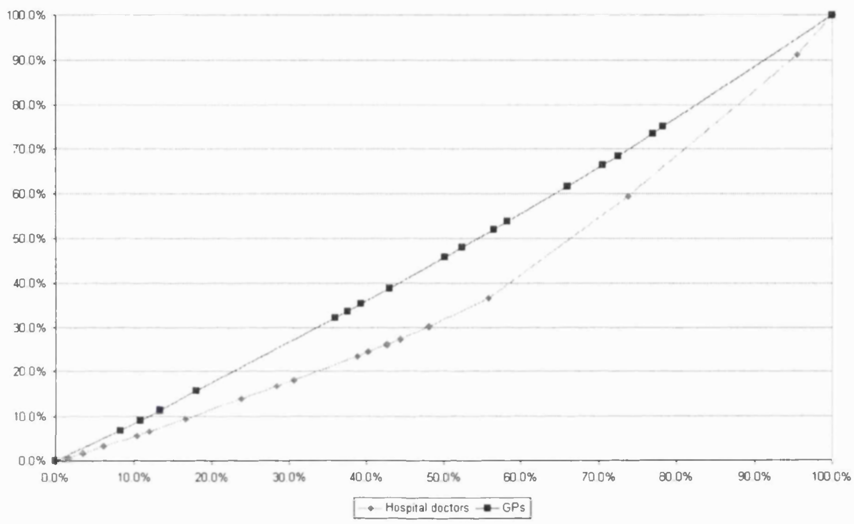
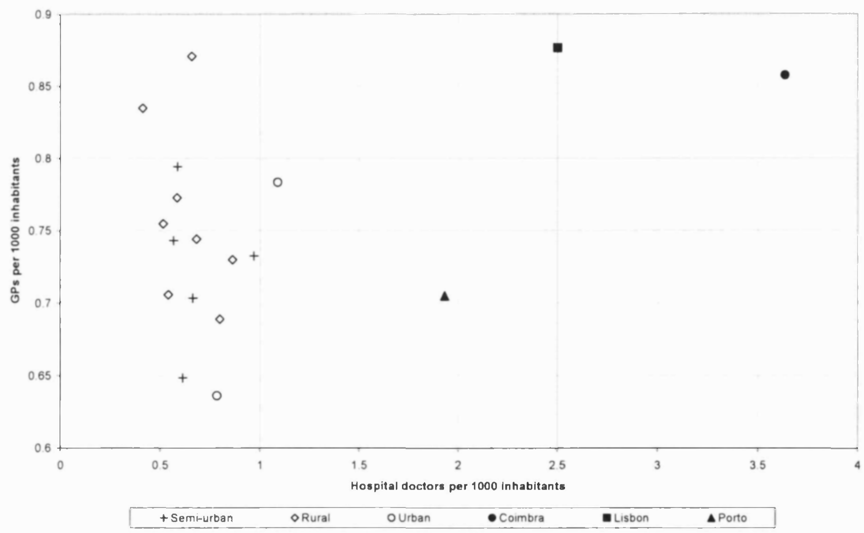


Figure 3.20: Hospital doctors vs. GPs supply per district



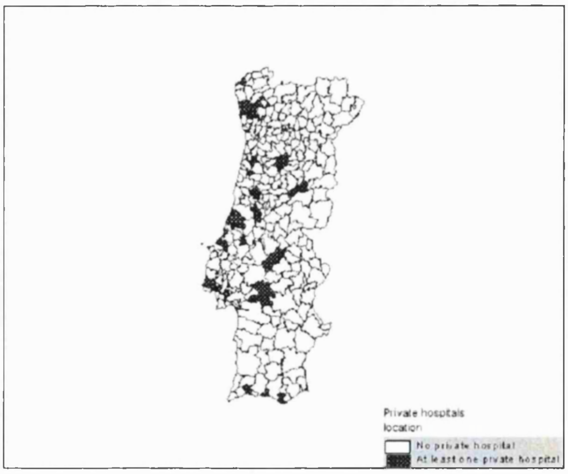
[87] Nonetheless Lorenz curves do not show how the mix hospital/primary care operates locally –as seen for Lisboa and Coimbra (Figure 3.20).

Private hospital care. The private acute care sector accounts for 13% of hospitals which are located in a few urban areas, and mainly in Lisboa and Porto. As a result, they reinforce geographical inequalities of public hospital resources. Six districts account for most private hospitals –Braga (12.3%), Porto (24.4%), Leiria (24%), Lisboa (17%), Évora (19%), and Faro (12%) (Figure 3.21). Private hospital beds are much more concentrated geographically than private hospitals: 72% of beds are concentrated in Lisboa and Porto, and 80% in these two districts and Braga (thus, the largest hospitals are concentrated in these districts). Analysis of both public and private sectors shows that private hospitals compound inequities in the distribution of NHS hospitals (Table 3.16). Private hospital provision (number of beds) is mainly not-for-profit in Braga, Porto, Leiria and Évora, while in Lisboa and Faro for-profit private supply dominates. The not-for-profit sector is the legacy of charitable organisations existing prior to the creation of the NHS. On average, private hospitals are small –63 beds.

Table 3.16: Hospital private sector indicators

	Private/total beds	*Official private beds*	*Official private beds (%)*	*Total official beds (%)*
Rural	4%	190	5%	17%
Semi-urban	10%	534	13%	17%
Urban	15%	3356	82%	67%
Total	12.9%	4080	100%	100%

Figure 3.21: Location of private hospitals



3.4 Consequences for policy

The analysis of this chapter supports the literature review in Chapter 1. Hospital provision appears to be inequitably distributed. Crude indicators show a mismatch between supply and population and if actual population needs were measured, this mismatch might be more pronounced. Evidence was found for an 'inverse care law'. Any redistribution of resources in accordance to population needs would imply a massive redistribution. Several insights for policy analysis and research development have emerged. This section summarises key findings.

First, there is a need to develop proxies for need for hospital care. Population size is a critical determinant of need but other indicators of need were shown to vary by area and justify variations in hospital resources. Urban and industrialised areas have better accessibility to a wide range and quantity of health care services, and higher choice, higher economic accessibility standards and greater choice in access, due to double-coverage. Estimates of need should account for these characteristics.

Second, the internal structure of hospital systems at the district level varies significantly and there is evidence of different levels of efficiency, as hospitals located in different areas use distinctive mixes of resources. Any measure of inequities in current expenditure should account for this. This finding also justifies the development of better measures of hospital supply, as beds proved to be a misleading indicator of productive capacity.

Third, any policy pursuing geographical equity will involve a massive redistribution of resources. The three main centres (Lisboa, Porto and Coimbra) attract a highly disproportional share of hospital resources and the majority of central hospitals. In addition, although PC resources are more equitably distributed than hospital resources, Lisboa and Coimbra still have a greater share of PC than their population shares. Private hospital care is reinforcing unevenness in the distribution of public hospital resources.

3.5 Concluding remarks

This chapter has analysed geographical variations of hospital resources, of the hospital acute care sector, in particular, while making use of readily available data and crude measures. This analysis has crucial limitations due to using crude population-based indicators and to not accounting for the impact of area variations on hospital characteristics. It has however shown a gap between demand and supply that justifies the need for further research to be developed in this thesis along the following lines:

1. In developing planning tools that consider demographic and socio-economic characteristics of the Portuguese population, in developing more accurate estimates of geographic inequities of hospital care and in making use of a capitation formula. The computation of accurate estimates of need of hospital care is critical to measure inequities. As Wright et al. (Wright and Williams 1998) have emphasised, a health care needs assessment process is appropriate when there are inequalities of access, large variability in the availability and use of health care resources, as well as when there is evidence that availability tends to be inversely related to population needs. As shown, these conditions clearly apply to the Portuguese hospital sector.
2. In creating tools to help explain the mismatch between supply and need. Understanding inequities demands the computation of unavoidable costs of hospital care (for example, information on how economies of scale and scope might be operating) and the computation of cross-boundary flows (how the concentration of supply generates flows of patients between districts).
3. In developing tools to help understand how to make a start in redistributing supply so as to improve equity.

Other relevant topics that arose from the analysis are outside the scope of this thesis. For example, hospital beds were shown to be a poor indicator of supply in the Portuguese system, and there is a need to search for better indicators of the hospital productive capacity. This is an area where there are neither satisfactory nor consensual measures in literature (Annel and Barnum 1998). Also, research on human resource policies seems to be critical to correct inequities.

In the next sections, this thesis targets two main research questions. First, developing more accurate estimates of geographical inequities in hospital care, through the development of several components of a capitation formula. Second, developing methods to identify redistribution of hospital supply that promotes improvements in the direction of geographical equity of access and utilisation.

The next chapter builds on the findings mentioned above: it sets a capitation formula to measure inequities of capital, finance and utilisation, taking into account population need, unavoidable costs of hospital care and expected cross-boundary flows between districts.

SECTION II

4 CHAPTER 4 - Setting a capitation formula to measure geographic inequities

4.1 Introduction

This chapter consists of four sections. The first section reviews literature on the usefulness of a capitation formula for countries with a NHS to analyse inequities and the methodological problems involved in modelling adjustments of a capitation formula. The second section develops a capitation formula for Portugal to measure inequities in hospital capital, finance and utilisation. The third section defines different indices of inequities. The final section summarises the content of the chapter.

4.2 Using a capitation formula to measure geographic inequities

This section describes the main historical developments and key concepts of a capitation formula, shows how capitation formula have been used to inform resource allocation, justifies their usefulness for Portugal, and summarises methodological concerns.

4.2.1 Literature review

Capitation formula have been widely used in resource allocation in developed countries (Rice and Smith 1999) and are particularly useful under certain conditions: a) for countries with a NHS using a top-down approach to resource allocation (to allocate budgets and devolving responsibility); b) when geographic equity in the distribution of resources is pursued; c) for measuring geographic health care inequities. All these conditions currently apply to Portugal.

Since the RAWP report (Department of Health and Social Security (United Kingdom)

1976) England has been using a capitation formula to distribute resources between health authorities so as to achieve greater equity. That formula has been adapted and implemented in the other UK areas and has been described as "perhaps the most sophisticated and objective system in the world for establishing the fair share of health service resources to which each administrative population is entitled" (Mays and Bevan 1987). The capitation formula for England has been changed subsequently in incremental ways, but the core principles and structure of the formula have remained. The RAWP approach has influenced other countries' approaches to resource allocation, has stimulated research, and has been adapted and developed (Persaud and Narine 1999) (Hutchison et al. 1999).

In the health sector context, capitation is "the amount of health service funds to be assigned to a person with certain characteristics for the service in question, for a specified time period and subject to any overall budget constraint" (Rice and Smith 2000). Capitation formulas have two main uses, depending on whether systems are based on a NHS or on a social insurance structure[88]. The group of countries with systems based on a NHS are the most relevant reference cases for the objectives of this study. In countries with a social health insurance structure, capitation might be simply defined as the amount of health service funds to be 'attached' to a citizen for a particular service or set of services. For instance, the Netherlands and Germany have been using capitation methods for paying sickness funds (to improve efficiency and promote competition), while the US have been using capitation for making payments from an insurance pool (to improve efficiency and contain costs) (Persaud and Narine 1999)[89]. For countries with a publicly-funded NHS, capitation formulas have been used to redistribute resources taking account of the fact that people have different needs for health care (Oliver 1999). In this group of countries, policies for allocating resources are developed in accordance with the principles of universal coverage and free access to health care at the point of use in order to achieve geographical equity of access.

[88] Capitation is typically related to processes of financial responsibility and responsibility devolution (Rice and Smith 1999). Under an alternative classification, competing insurers are capitated in social insurance systems (such as in Belgium, Israel, Germany and Russia) whereas competing providers have been capitated in the UK and US (Van Barneveld, Van Vliet, and Van de Ven 1997).

[89] Other developing/middle income countries such as Argentina, Brazil, Nicaragua and Thailand have adopted or are currently piloting capitation formulas to remunerate providers (Mills et al. 2000). For social insurance countries, formula adjusters aim at predicting health care costs per type of user (Hutchison et al. 1999) and at discouraging risk selection activity of health insurers (Persaud and Narine

Similarly, the objectives of this thesis involve the development of a capitation formula for measurement of geographical inequities in Portugal.

The thesis draws heavily on literature from UK countries (in particular on England), given the now long tradition in UK in using capitation methods to allocate resources; moreover, the formula used in England has been the most sophisticated one from the group of NHS countries that follow an index approach[90] (Rice and Smith 1999) (Hutchison et al. 1999)[91].

Capitation methods in England have changed to reflect changes in information, statistical systems, and the context of resource allocation and organisation of the health care system. They have proved to be a powerful tool for the equalisation of expenditure across areas. After the RAWP report, the English Department of Health used a formula to make allocations to 14 Regional Health Authorities (with 3 to 5 million population), then subsequently to 191 District Health Authorities (in the mid-1990s) (with populations ranging between 150,000 and 900,000 inhabitants) and is now applying the formula to 304 Primary Care Trusts (with nearer 120,000 inhabitants). The objective underlying the use of a capitation formula has been to achieve "equal opportunity of access for those in equal need" and to equalise geographic resources across areas. More recently, England has also been aiming at reducing 'avoidable inequalities' (Shaw and Smith 2001). Over the recent years, different components of the English resource allocation formula (with reference to community and hospital services) were computed under different principles[92]:

- Most resources have being allocated using a capitation formula, under the principle of 'equal opportunity of access for those in equal need';
- A budget has been earmarked for allocation directed at reducing unavoidable inequalities in health based on an index of years of life lost.

1999). In this context, the main health policy principle implicit in the use of a capitation formula is efficiency (Diderichsen, Varde, and Whitehead 1997).

[90] As explained in Chapter 5, this approach is characterised by the use of data of aggregated populations to model the adjustments of a capitation formula.

[91] Rice et al. have conducted an extensive literature review on the development of capitation formulas and their use for resource allocation in a set of developed countries; Hutchison et al. have clarified also the usefulness of capitation formula, debated some of the issues implicit in dealing with it, and described capitation formulas in use in Canadian provinces.

But England is now to return to the use of a single (capitation) formula that has been improved to account for unmet need (under the principle of equal opportunity of access for those in equal need): this formula seeks to take into consideration groups that do not receive health care services to the same level as others with similar health characteristics –for example ethnic minorities and socio-economically deprived groups.

NHS countries have mainly included the following elements in capitation formulas (examples of countries adjusting for these factors are given in the next section and in Chapters 5, 6 and 7):

1. Population numbers per geographic area. Depending on the context of use of capitation formula (for example, whether to measure need or to allocate resources) and on the available information, those population numbers might be adjusted by:
a) Population flows between geographic areas (cross-boundary flows -CBFs); this requires estimates of catchment populations. CBFs result from the geographic interaction between need and supply (as supply tends to be concentrated in some areas) and from significant flows in cities.
b) Population numbers for which responsibility for health coverage is outside the public sector (following an opting-out agreement between the public and the private or when some population groups are not treated by the NHS hospital network).
2. Relative need, due to demography and morbidity.
3. Unavoidable costs (UCs).
4. Discounting for the role of the private sector across geographic areas when there are variations in the public-private mix in provision (here the public-private mix relates to the area utilisation of private care services, which decreases the need for public hospital care, and is different from the adjustment referred in point 1.b, above).

Most of these elements are intended to capture relative needs with the adjustment for unavoidable costs being applied for the supply side of health care. Where there is no data (e.g. information for points 1a and 1b in Portugal), these adjustments might require modelling to provide estimates.

[92] This corresponds to a disintegrated approach to resource allocation whereby the overall budget is split in other sub-budgets that are allocated under different criteria.

The choice of these adjustments depends on the type of inequity to be measured and on the objectives of the study. In the Portuguese context, a capitation formula for the hospital sector can be used as a tool:

1. To measure geographic inequities –by comparing targets indicating an equitable distribution with current hospital expenditure, capital and utilisation of hospital resources.
2. To compute capitation monetary targets that can be used as a budget-setting tool in the context of a top-down budgeting system for devolution of financial and administrative responsibilities to lower levels of government.
3. To compute opting-out payments, if horizontal equity is to be pursued; as described in Chapter 2, the MoH has been paying some subsystems (health insurers, defined on an occupational basis) for the transfer of health care coverage responsibility from the NHS[93].

This thesis seeks to provide information to answer the first of these questions.

4.2.2 Methodological concerns

In the process of building capitation formulas, certain principles and problems have been widely acknowledged.

Estimating need is the most important element of capitation formulas, and ought to be based on unbiased estimates of the expected relative costs for a population unit for the chosen health care need factors. There are multiple health care need factors (Whitehead 1995) and several concepts of health care needs (Matthew 1971) (Eyles and Birch 1993) (Wright and Williams 1998) (Oliver 1999). For example, Matthew (Matthew 1971) argued that health care needs exist when an individual has an illness or disability for which there is an effective and acceptable treatment or care (where "(need) can be defined in terms of type of illness or disability causing the need or of the treatment or facilities for treatment required to meet it"). Wright (Wright and Williams 1998) has defined health care needs as those needs that can benefit from health care or from wider social and environmental changes. Alternatively, Oliver (Oliver 1999) defined that health care needs exist only when the capacity to benefit from health care treatment is positive[94]. In the context of resource allocation and under the existence of a budget constraint, the concept of relative health care needs ought to be based on the resource implications of differences in relative risk between populations.

The necessary properties of capitation formulas are (Advisory Committee on Resource Allocation (United Kingdom) 1999) (Scottish Office 1999) (Hutchison et al. 1999):

1. Technical robustness, transparency, objectivity, reliability of calculation, stability and materiality;
2. Freedom from perverse incentives, responsiveness, and the capacity for evaluation.

Building a formula for Portugal aimed at satisfying primarily the first group of properties, whilst being aware of the second group, as the main objective of the formula was to measure inequities rather than to create a formula to allocate and redistribute resources.

As recognised by the RAWP report, many of the determinants of health and needs for health care are unknown. Utilisation has been used to capture need for health care, but there are difficulties in identifying and separating legitimate and illegitimate sources of variation in health care utilisation (Rice and Smith 1999). Attempts to explain (with or without predictive purposes) health care expenditures at the individual level with individuals' characteristics (as indicators of need) has shown that traditional adjusters (such as age and sex) have a low explanatory power[95] (Newhouse et al. 1989) (Van Vliet and Van de Ven 1992). For NHS countries, formulas that have been used for devolving financial responsibility at a geographic level have been derived at aggregate levels based on a normative approach (Oliver 1999)[96]. These formulas seek to measure

[93] The capitation payment of €145 (per year and per user) for the opting-out of the Portugal Telecom and CTT subsystem members was determined by political judgement.

[94] Eyles and Birch (Eyles and Birch 1993) defined health care needs as the ability to benefit from health care as implied by reducing the risks of deterioration in health status or by improving the probabilities of health status.

[95] Functional and perceived measures of health care availability and prior utilisation measures of health services (such as prior utilisation diagnostic, disability, functional health status and chronic medical conditions, employment status and housing tenure indicators) have been shown to increase individuals' costs predictability for the US and the Netherlands (Newhouse et al. 1989) (Anderson et al. 1990) (Diderichsen, Varde, and Whitehead 1997) (Lamers 1998). However, the use of these measures in capitation formulas to allocate resources might reinforce inefficiency and give wrong incentives (Oliver 1999).

[96] Despite the indistinctive use of the terms risk and need adjustment in this thesis, a study has proposed a classification that distinguishes both and it is relevant for this context (Hutchison et al. 1999): risk adjustments try to ensure that capitation payments cover predictable future expenditures on insured health services provided to enrolees; and need adjustment is related to a funding consistent with the relative needs for services of the enrolled or geographically-defined populations.

relative needs of populations, rather than predicting costs at the individual level[97]. The development of capitation formulas has been constrained by data availability, even for developed countries. Other empirical problems arise from modelling the interaction between supply and need: these problems are reported in the literature review of the next chapters, which model the adjustments of a capitation formula for Portugal (defined below).

4.3 Capitation formula: Country application

This thesis attempts to develop methods to measure inequities, under different definitions of equity (in the Portuguese context), namely inequities on capital, utilisation and finance. The capitation formula focuses on three issues: measuring need for hospital care; modelling unavoidable costs of hospital care; and devising methods to predict hospital utilisation and to estimate CBFs.

Previous capitation studies applied in the Portuguese system (as described in Chapter 1) have been mainly limited to the primary care sector and mostly produced for academic purposes[98]. These studies have suffered from a number of weaknesses: a) they have not made clear the equity concepts used nor have they linked research with equity objectives from the Portuguese political system; b) they do not specify clearly judgements and assumptions made, and do not apply sensitivity analysis to test these assumptions. Additionally, there is less information on capitation studies for the hospital sector. As a result, although Portugal is committed to promoting equity of access, and

[97] The determinants for the population level might differ from the ones for the individual level, as discussed in Chapter 7.

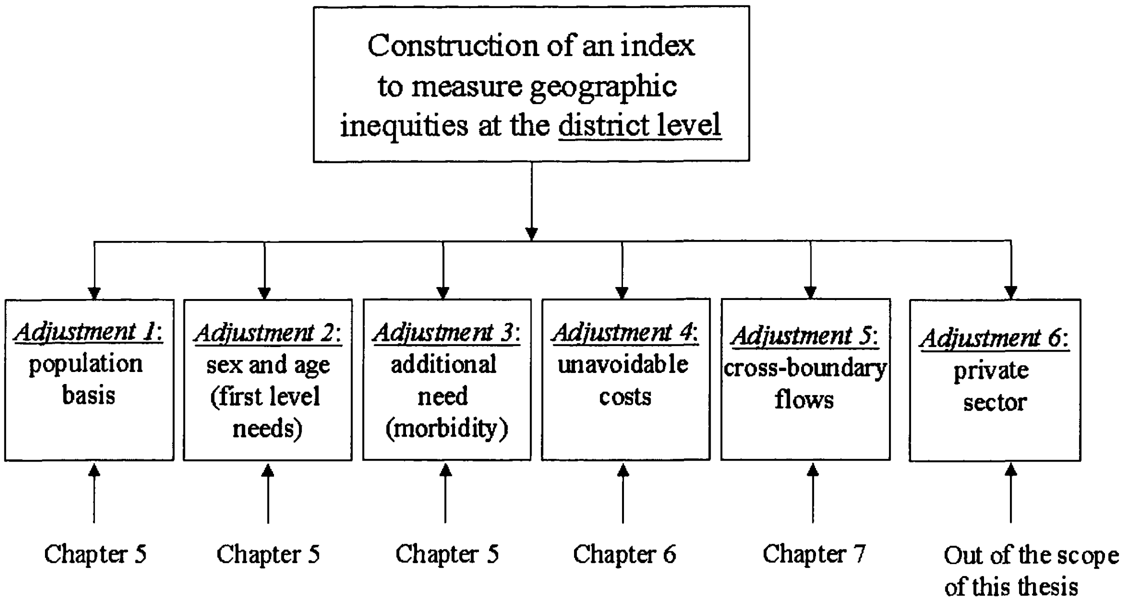
[98] Following from the information provided in Chapter 1, the 1999 resource allocation capitation formula for primary care was improved by including a correction by population age and burden of illness and was used to distribute 20% of the budget. In 2000, a new capitation formula was deployed to allocate 30% of the budget; this formula was further changed to estimate the burden of illness for four diseases (Tranquada, Martins, and Sousa 2000): hypertension, diabetes, rheumatic diseases and psychosocial stress. The 2000 formula has accounted for the prevalence of these diseases in the population, using information from the 1995/1996 National Health Survey; moreover, it has evaluated the burden of these diseases using data on related pharmaceuticals consumption and it has adjusted the pharmaceuticals consumption component to eliminate excessive prescription. For the year 2001, two additional factors were expected to be introduced –vaccination and kidney dialysis (Tranquada, Martins, and Sousa 2000). Nonetheless, it is not clear how the inclusion of additional components compares with a more global approach for needs' estimation that would involve fewer judgements within the formula and would allow for a higher level of transparency.

initial evidence suggests that it needs to do so, there are no sound estimates of the degrees of inequities.

A complete index to measure inequities in the Portuguese hospital system (using a capitation formula) should include the adjustments shown in Figure 4.1, as explained below.

The set of adjustments to be included in a capitation formula depends on the choice of the population basis (as described above), which might be: resident population, catchment populations or resident populations reduced by opted-out populations. The starting point is resident populations as used in Chapter 5 (while Chapter 7 produces information to compute catchment populations). For resource allocation in England, the RAWP report recommended using population corrected by CBFs and these were used until 1991. Subsequently, following the purchaser-provider split, resident populations were used.

Figure 4.1: Set of adjustments to be integrated in a capitation formula to measure inequities in the Portuguese hospital system



Adjustments in Figure 4.1 can be split into three main building blocks:

1. *(Resident) Population needs or risk adjusters* (adjustments 1-3). The most common adjustments have been for age and sex, and morbidity; some countries adjust for

rurality[99] (such as Scotland) and others for additional needs of disadvantaged populations (such as the New Zealand). Chapter 5 describes an approach that covers in the Portuguese context the main adjustments widely used in any capitation formula to measure need for hospital care: choice of the population basis, adjustment for age/sex and adjustment for morbidity.

2. *Unavoidable differential costs in providing health care*[100] (adjustment 4). The common practice has been to adjust for the market forces factor adjustment for staff, land values and building costs in England (Department of Health (United Kingdom) 1999), and adjustments for remoteness[101], e.g. in Northern Ireland, Finland, New South Wales, New Zealand and Scotland (Rice and Smith 1999). Estimation of UCs is addressed in Chapter 6.

3. *Other adjustments related to the interaction between supply and need and adjustments for other specific characteristics of the health care system* (adjustments 5 and 6). When health care is concentrated in some areas, some countries have made adjustments for CBFs (e.g. several Canadian states (Hutchison et al. 1999)). Australia has adjusted for substitution between the public and the private sectors. England made deductions for population groups treated outside the NHS (such as the prisoner population). Chapter 7 develops a method to estimate CBFs for Portugal. This thesis has not considered the impact that geographical variations in the provision of the private sector have on need for public hospital care. The issue of double coverage is discussed (briefly) in Chapter 5.

The rationale for the choice of each of these adjustments is explained in detail in each chapter. Adjustments are combined into an index at district level. The reasons for the use of this geographic level are the ones described in section 3.2.1.2 of Chapter 3.

The chapters that model the adjustments of a capitation formula make use of:

---

[99] In resource allocation, the rurality need component attempts to capture the higher need implied by lower locational accessibility of a population; this is different to remoteness, which is defined below.

[100] The aim is to discriminate between different health care unit costs in order to allow purchasers and providers to face an equal set of constraints when they buy or provide services.

[101] In resource allocation, remoteness corresponds to internal hospital costs related with economies of scale and scope involved in providing services to smaller populations located in remote areas.

- **Technology transfer of methods already in use in other countries** (adapted to the Portuguese context, mainly from England) (Chapter 5);
- **Development of new methods for the Portuguese context**, though these may be extended to other health care systems (Chapters 6 and 7).

The adjustments of the capitation formula are combined in a multiplicative formula as presented in Equation 4.1. Notation used in this chapter is presented in Table 4.1. The index in Equation 4.1 accounts for the impact of relative levels of need, UCs and CBFs between districts on relative costs across districts[102]. Multiplicative models have been commonly used to combine adjustments of a capitation formula (Hutchison et al. 1999); they are appropriate when relative needs and other components are to be estimated, and eventually when the capitation formula is to be used for allocating resources and when a fixed budget is to be distributed. Nevertheless, the multiplicative model might be criticised on several grounds. It is not directly derived from a theoretical formula. It results from a multiplication of a set of indices that are modelled separately, and the links/interactions between those adjustments are not taken into account. Also, as it produces an index, it does not show a clear rationale as to the adequate level of redistribution (Hutchison et al. 1999) and there is a need for judgement in determining this level. The multiplicative model for the district level might result in the problem of the ecological fallacy (Morgenstern 1982), in that values for the district level may not deal adequately with variation at a lower geographic level[103]. The alternative to the multiplicative model is an additive model, but that suffers from the same weaknesses and also requires judgments.

$$cap\_index_r = I_{2r} * I_{3r} * I_{4r} \tag{4.1}$$

[102] An index is assumed to have a central value of 100%.

[103] The ecological fallacy problem implies that the relationship between aggregate variables (at the small area level) differs from the relationship between variables at the individual level. In practice, the ecological fallacy problem means (Morgenstern 1982) that the choice of different groupings can lead to distinct results and conclusions, while there might be a reversal between cause and effect.

Table 4.1: Notation in use

Notation	*Interpretation*
r	r is a geographic district unit (district; for Portugal, $r=1,2,..18$).
cap_index_r	Relative capitation index for district r, accounting for all the selected adjustments of the capitation formula.
P_r, P	Resident population in district r and total resident population.
I_{2r}	Age and additional need index for district r.
I_{3r}	CBFs index for district r.
I_{4r}	UC index for district r.
$District\_share\_1_r$	Share of need for hospital care for district r.
$District\_share\_2_r$	Share of need for hospital care, adjusted by CBFs for district r.
$District\_share\_3_r$	Share of need for hospital care, adjusted by CBFs and UCs for district r.

4.4 Typology of geographic inequity indices

This section presents a typology of indices of different measures of inequity –such as capital, utilisation and finance– that are regarded as relevant to Portugal and also to other countries with a NHS. Estimates of these indices for Portugal are presented in Chapter 8. This section starts by describing the structure of indices and afterwards defines the formula of each index.

Any index of inequity compares the current distribution of resources (numerator) with an equitable distribution (denominator, normatively defined). The numerator can be a measure of the current distribution of resources (here, in the form of share of resources for the specific district) of current expenditure, capital, utilisation or other measures of hospital resources. The denominator is an equitable share of hospital resources for the district that combines information on the adjustments of a capitation formula. Estimates of the adjustments of the capitation formula are handled in this denominator under the multiplicative model described in the previous sub-section. These inequity indices are relative indices at the district level[104].

[104] Even if some estimates in different chapters were initially computed at a smaller geographic level than the district level, they are afterwards aggregated to the district level.

Three key types of inequity indices are discussed here, namely on capital, utilisation and finance, which correspond to four specific geographic inequity indices: see Table 4.2. The classification into inequities in capital, finance and utilisation is comparable to the structure, process and outcome framework for quality of care proposed by Donabedian (Donabedian 1988)[105]. Each index is the result of the division of the numerator (corresponding to columns) by the denominator (corresponding to rows). Indices 1 and 2 measure inequities in hospital capital, while index 3 measures inequities in utilisation and index 4 measures inequities in finance. The rationale for the use of each index follows.

Table 4.2: Geographic inequity indices at the district level

	Numerator: share of hospital supply (proxy: hospital doctors)	**Numerator**: share of hospital utilisation in hospitals of the district (proxy: hospital discharges)	**Numerator**: share of current expenditure (proxy: NHS revenue from central government)
Denominator: share based on population numbers and need	**Index 1**: captures structural geographic inequities in the **distribution of capital**, with regard to need for hospital care.	-	-
Denominator: share based on population numbers, need and CBFs	**Index 2**: captures structural geographic inequities in the **distribution of capital**, with regard to catchment populations	**Index 3**: captures short-run geographic inequities in the **distribution of utilisation**, with regard to catchment populations.	-
Denominator: share based on population numbers, need, CBFs and UCs	-	-	**Index 4**: captures short-run geographic inequities in the **distribution of finance**, with regard to an index that summarises legitimate variations in hospital costs (catchment populations corrected by need and UCs).

[105] Under Donabedian's framework, analysis of quality of care demands an analysis of the causal linkages between the structural attributes of the context in which care occurs (corresponding to inequity of capital in this chapter), the process of care (that can be related to inequity of finance), and the outcomes of care (that can be related to inequity of utilisation).

Index 1 captures geographic inequities in the distribution of hospital capital, with regard to resident populations and relative need for hospital care. The value of 100% for each district can be interpreted as the achievement of a normative target for the distribution of hospital capital. The index compares:

A) Numerator: district capital share, as measured by physical indicators (for example, the number of doctors was shown to be the best indicator of hospital supply[106]);

B) Denominator: district share of geographic population numbers corrected by demographic and additional need (as defined in Equation 4.2).

$$District\_Share\_1_r = \frac{P_r}{P} * I_{2r} \qquad (4.2)$$

Index 2 calculates geographic inequities in the distribution of hospital capital, with regard to catchment populations (population numbers corrected by need and CBFs). The use of catchment populations in the denominator acknowledges that in the short-term, an equitable share of resources should account for the implications of the unequal distribution of supply on CBFs of patients to access hospital care; and these flows are regarded as legitimate for explaining variations in the distribution of capital. The index compares:

A) Numerator: district supply share, as measured by physical indicators, for example, the number of doctors;

B) Denominator: district share of catchment populations measured by population numbers, corrected by demographic and additional need and by CBFs adjustments (as represented in Equation 4.3).

$$District\_Share\_2_r = \frac{P_r}{P} * I_{2r} * I_{3r} \qquad (4.3)$$

Index 3 calculates geographic inequities in utilisation of hospital resources, with regard to catchment populations. The index compares:

[106] The rationale for the use of doctors as a proxy for hospital productive capacity was explained in Chapter 3.

A) Numerator: district utilisation share, as measured by hospital discharges from all the hospitals of the district;
B) Denominator: district share of catchment populations, as described for Equation 4.3.

Index 4 computes inequity estimates in finance (that is, in the allocation of current expenditure), accounting for variations in population need, CBFs and UCs at the district level. The index compares:
A) Numerator: district share of current expenditure, as measured by NHS revenue transferred from central government;
B) Denominator: district share of geographic population accounting for need and for elements of the hospital system that translate on additional costs (by CBFs and UCs adjustments), as represented in Equation 4.4.

$$District\_Share\_3_r = \frac{P_r}{P} * I_{2r} * I_{3r} * I_{4r} \qquad (4.4)$$

4.5 Concluding observations

This thesis develops a capitation formula for Portugal that allows for the measurement of several types of inequities. The formula includes adjustments for need, CBFs and UCs. These adjustments are designed to reflect characteristics of the Portuguese health system and availability of data.

Different indices are proposed to measure inequities in capital, utilisation and finance, which correspond to alternative definitions of equity informing different policies.

The following chapters of the thesis develop the capitation formula described above: Chapters 5, 6 and 7 model the adjustments of a capitation formula, and Chapter 8 analyses estimates of inequity in the Portuguese context.

5 CHAPTER 5 – A capitation formula to measure need for hospital care

5.1 Introduction

Chapter 3 has given evidence of considerable inequality in the geographical distribution of hospital resources in Portugal using crude measures based on rates per capita. Chapters 1 and 2 pointed out that there is a lack of information to improve these crude estimates. Chapter 4 defined a capitation formula to measure inequities of hospital care for Portugal. The purpose of this chapter is to make a start in producing sound information required to produce more accurate estimates of the degrees of inequities; for this purpose, it models the adjustments of a capitation formula to measure geographical need for hospital care (adjustments defined in Chapter 4).

This chapter is structured into: a section that describes the role of technology transfer and other methodological options in use in this chapter; three sections that develop three components of the Portuguese capitation formula –population, demographic need and additional need– and analyse each component within the methods, and results and discussion sub-sections; and a section discussing the methods and their implications for policy and summarising the concluding remarks. The impact of each adjustment on geographical need at the district level is analysed in Chapter 8.

5.2 Methodological issues

In order to measure population need for hospital care, this chapter applies technology transfer of methods already in use in other countries, which are adapted to the Portuguese context. As described in Chapters 1 and 4, the RAWP report (Department of Health and Social Security (United Kingdom) 1976) has created the first capitation formula under the principle of achieving equal opportunity of access for those in equal need, and since then, capitation formulae have been adapted and implemented in many other countries. Capitation methods attempted to break with expenditure allocation methods that were allocating funds arbitrarily and perpetuating past inefficiencies and inequalities (Rice and Smith 1999).

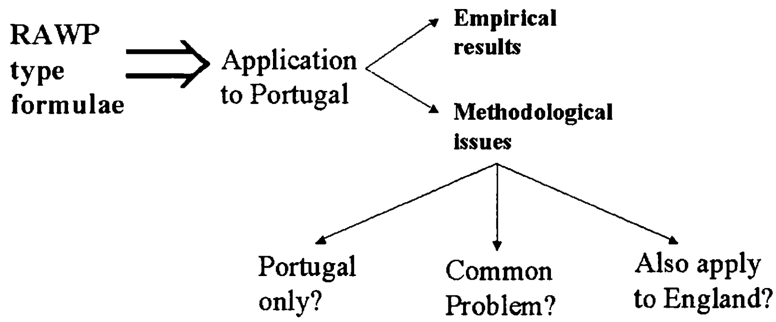
Portugal is similar to England in the 1970s: with a vague commitment to equity but no methods to achieve it. Current similarities between the Portuguese and English systems are: the NHS structure, a top-down budgeting system, the objective of equity in the territorial distribution of resources and a formal organisation model based on contracts between public purchasers and providers. Estimates of need are the most important component of a capitation formula and England has developed the most sophisticated formula in the context of NHS countries using an index approach (Rice and Smith 1999). This chapter intends to measure need for hospital care in Portugal by adapting the technology deployed in England, using an index approach. Although some countries are evolving towards a matrix approach (such as England and Sweden)[107], there is no data to develop the matrix approach in the Portuguese context. The capitation formula for Portugal follows the (index) structure of the RAWP formula.

The transfer of English capitation technology to measure need for hospital care in Portugal raises a number of questions that must be addressed (Collins, Green, and Hunter 1994), for example: What problems are involved in applying needs adjustments to Portuguese systems of health care? What difficulties arise from the lack of suitable information and/or data? Are the methodological issues posed in the Portuguese context common to other countries? In the process of applying technology transfer and of answering the questions just described, a set of outputs of the analysis is generated in this chapter (Figure 5.1), namely empirical results for Portugal and a set of

[107] The matrix approach departs from the use of contingent data available at the individual level to estimate the impact of differential risk/need on costs, and demands a well-structured and organised information system. The matrix is composed of a number of independent cells, which provide estimates of expected costs, and avoids the ecological fallacy problem, as it uses data at the individual level. Even when data is available for using of the matrix approach, its application might be difficult due to problems of confidentiality of data and due to its deterministic nature; and the matrix approach also suffers from the ecological problem. England is evolving towards the use of a matrix approach (Advisory Committee on Resource Allocation (United Kingdom) 1999); the matrix approach is dominant in social insurance-based health systems –such as in France, Israel, Germany and Switzerland.

methodological issues, which might be specific to Portugal, or applicable to England or internationally.

Figure 5.1: Process of technology transfer and outputs from analysis



The rationale behind the use of the district level is similar to the one explained in Chapter 3 (sub-section 3.2.2.1). This geographical level is appropriate for identifying inequalities and for implementing policies to correct inequalities, and has roots in the health care planning system. This chapter uses the concept of relative health care needs, which is defined as (relative) level of inputs per capita that captures the resource implications of having differences in relative risk between populations[108].

The methods used in each adjustment of the formula (for population numbers, demographic need and morbidity) have considered:

1. The traditional choice between normative (based on political judgements) and empirical approaches (based on more sophisticated regression techniques) for adjustment modelling; each approach has advantages and disadvantages (Mays 1995), and the choice highly depends on the sets of data and the information available;
2. The process of analysis of data: the modelling process generates information, and the reliability of this information might be seen as useful for validating the main outputs of modelling;
3. The properties defined in the literature review on capitation formula, such as on updateable information and transparency, and the incentives implied by the methods in use if the formula is to be used to allocate resources (in order to avoid perverse incentives). Primacy is given to the former two properties;

4. The difficulties involved in developing means to achieve equal opportunity of access for those in equal need were acknowledged in Chapter 1. This chapter uses the equity definition of equal opportunity of access for those in equal need, but because of problems in measuring personal access costs (Le Grand 1987), this objective has been commonly redefined as equal inputs for equal need (Mooney and McGuire 1987).

The next sections model the adjustments for population, demographic and additional need.

5.3 Adjustment for population

Population numbers are the main determinant of the size of the RAWP shares. There are three issues involved in measurement: the use of provider-based vs. community-based indicators, deployment of past populations vs. projections, and adjustment for opting-out populations.

5.3.1 Method

Choices related to the population numbers depend on the context of the study. This section uses resident population numbers (as community-based indicators), although it raises issues relevant with the choice between provider- and community-based indicators. The expected impact of using different indicators is discussed through analysis of available data.

There are two alternative types of population indicators: provider-based and community-based (Shaughnessy 1982). These correspond respectively to two viewpoints in resource allocation (Mays and Bevan 1987): funding of health services for what they do, and funding of health services on the basis of the geographical

[108] The discussion of the concepts of need and the definition of need used within this thesis were presented in Chapter 4 (section 4.2.2).

populations they serve[109]. Traditionally, capitation formulas have used community-based population defined geographically for the hospital sector. Some countries, such as England, in the past used resident population figures corrected by CBFs (following the RAWP report). Recently, the Advisory Committee for Resource Allocation (in England) has recommended the development of GP-based registered lists as the population base to be used in the long-term for resource allocation to the Health Authorities (Advisory Committee on Resource Allocation (United Kingdom) 1999).

For Portugal, a community-based indicator is used, but there is the question of how to adjust population estimates for those with private medical insurance and for other population groups that might be attended by health care institutions that are outside the NHS. Given the lack of data at the district level, the population numbers used do not account for double coverage[110]. Clearly, for Portugal, there are complex arguments over how to handle populations with insurance coverage for medical care provided through occupational-based subsystems[111]. On the other hand, community-based populations are suitable for measuring inequities in the distribution of hospital capital (defined in Chapter 4)[112]. Additionally, they are the preferred concept in the context of estimation of area needs for hospital care, while in the Portuguese system, RHAs are expected to receive funds for geographically defined resident populations.

In the Portuguese case, census-based resident populations are used as community-based indicators, as there are no data on registered practices at the moment[113]. Official resident

[109] These are two views of the same problem: starting from a population basis, and making adjustments for population flows, provider indicators can be obtained, and vice-versa (Wilson 1988). In practice, because of data problems in constructing these cross-boundary flows, the more uneven the distribution of supply between two geographical areas, the more obscure the relationship between the two indicators.

[110] Other populations groups could also be excluded, such as prisoners. These groups have been treated outside the NHS and their coverage is funded by other components of the public budget.

[111] As described in Chapters 2 and 3, the case for excluding subsystem populations is complex because: the State has to secure universal coverage (which is financed by general taxation); occupationally-based subsystems have been seen as a complement and supplement based on citizens' initiative, but are mainly financed by the State; subsystem users can be seen as individuals with preferential conditions in accessing health care services; the Government strategy on how to tackle the double-coverage problem is unclear; the private sector often makes use of the public network of hospitals, which implies that hospital care planning should concentrate on the public sector facilities.

[112] As discussed in Chapter 4, for equity analysis, community-based indicators are more appropriate for analysing the current distribution of capital, while provider-based indicators are more suitable for analysing inequities in utilisation and finance. In the latter case, community-based resident figures are corrected by cross-boundary flows, which are estimated in a model presented in Chapter 7. This is consistent with the use of the capitation formula to measure several types of inequities.

[113] Data on registered practices will be available in the future due to the current development of primary care patient lists and the implementation of data card technologies.

population statistics are available in two forms: estimates of past populations (from census) and projections. Although projections are more appropriate for resource allocation purposes (Mays and Bevan 1987)[114], census data is used because population projections are not available at the Portuguese district level. This implies a lag of at least two years between the data of the population estimates deployed and the year of allocation under consideration. As shown below, the use of estimates is unsatisfactory given the changing population patterns.

5.3.2 Results and discussion

Analysis of resident populations between 1990 and 1998 shows that the coastal population (mainly in northern districts) is younger and has been increasing, while population of the interior and south rural districts is older and has been decreasing (Figure 5.2 and section 3.3.1, Chapter 3). Between 1990 and 1998, the national population has increased by 1.1%, with substantial variation between districts (Figure 5.2). There were significant increases in the coastal areas, with the northern coastal districts having the highest population increases. In contrast, sparsely populated districts had decreases of 5% or more. Population growth has coincided with the urbanisation effect in the metropolitan and coastal areas over the last decades (Eurostat, INE, and European Commission 1998), and this pattern has reflected the imbalance of economic opportunities across areas[115]. Projections (Figure 5.3)[116] suggest that these trends will continue over the next two decades[117]. Thus, structural changes are expected in the long-term in the distribution of Portugal's population along its territory. The projection for the year 2000 (in comparison with estimates of past populations) for the population

[114] Projections are more appropriate for resource allocation purposes as funds are directed towards where the population is expected to be in the year for which allocations are made rather than where they were in the past. Empirical evidence for England has shown that population projections are better indicators for resident populations than population estimates (both by age group and for total population) (Advisory Committee on Resource Allocation (United Kingdom) 1999). Nonetheless, projections are liable to serious errors (Cliquet and Thienpont 1995), as they require assumptions on a number of population parameters (fertility, mortality and migration).

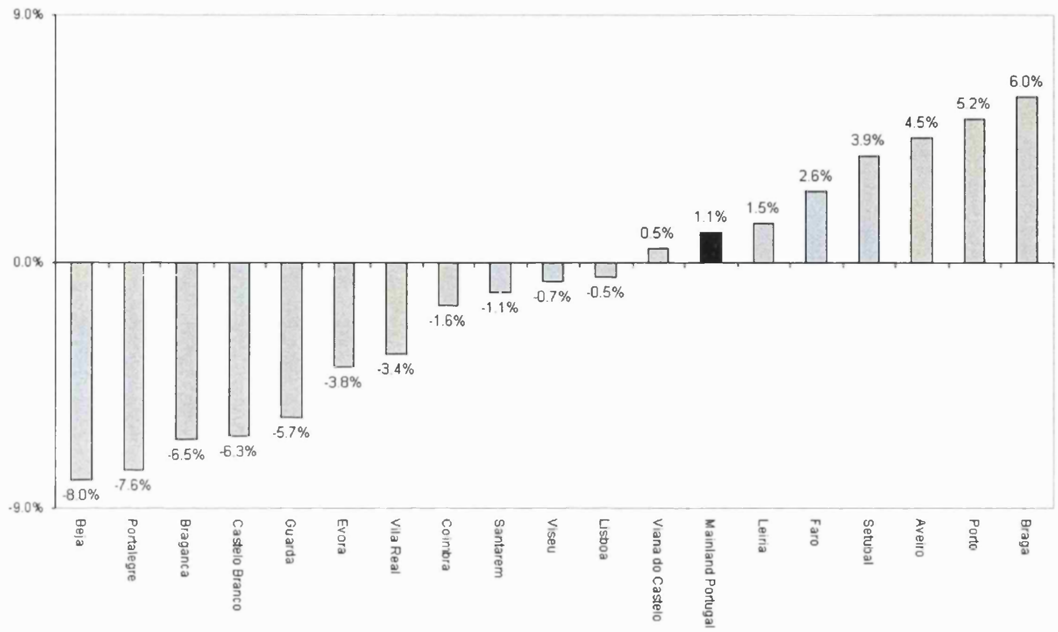
[115] District analysis hides intra-district important variations: de-population is sometimes more visible in delimited rural areas, while some of the cities of the interior continue to register a population growth (this is the case of Viseu) (Eurostat, INE, and European Commission 1998).

[116] This data is at the administrative level and not at the health region level. The difference between the two classifications is small and was described in section 3.2.2 of Chapter 3. No information is available about the assumptions made when computing these projections.

[117] The demographic structure of the population is expected to change due to migration flows, primarily towards the urban and coastal areas (INE 2002).

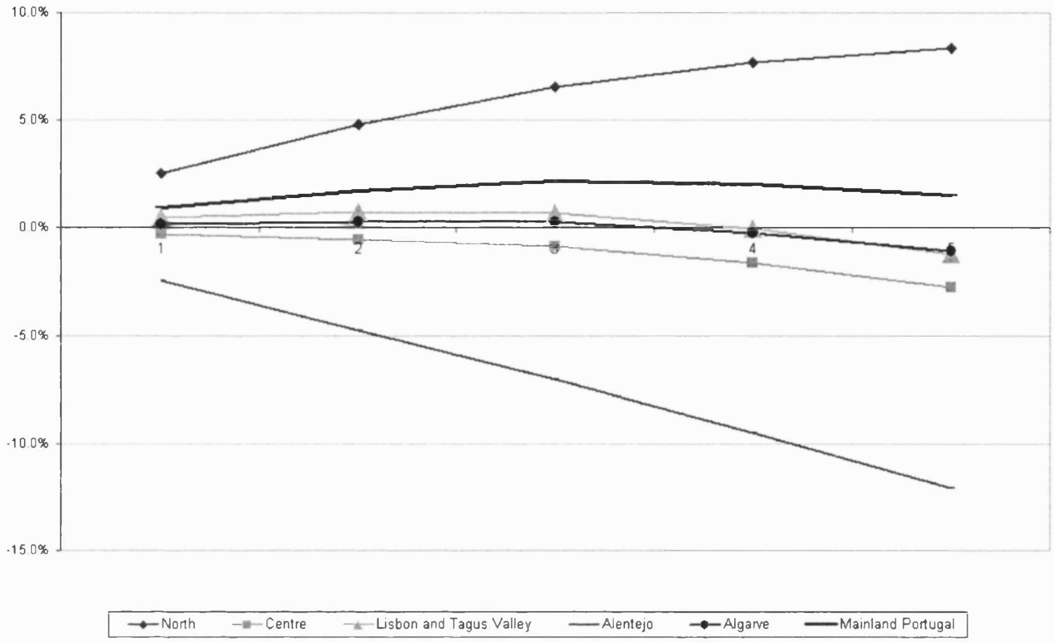
base tends to show population increases in the industrialised littoral (mainly north) and decreases in the interior and Alentejo districts (Figure 5.3). Methods used to compute population numbers seem as a result to be unable to deal with radical changes in populations.

Figure 5.2: Resident population estimates evolution 1990-98



Source: (INE 2000)

Figure 5.3: 1995 based population projections for the year 2000 at the NUTS II region level in comparison with 31.12.1995 resident population estimates



Source: (INE 2000)

Table 3.5 (Chapter 3) presented data on single-covered population proportions at the region level, i.e., populations that were uniquely under NHS coverage. Making deductions for the double-covered population would severely decrease the population share in the Lisboa and Tagus Valley region, slightly decrease the Alentejo share, and increase the population proportions of the other regions (more substantially the north and centre regions). This is expected, as employment in the tertiary sector is over-represented in the Lisboa metropolitan area and subsystems coverage is based on that sector. By contrast, the North region includes a heterogeneous mix of urban and rural areas, and has a higher share of NHS single covered population (despite the fact that Porto has also a significant share of tertiary sector activities).

To sum up, important changes are taking place in the distribution of population which necessitates the use of population projections for future resource allocation policy. Populations are increasingly moving towards urban areas with younger populations, something that current methods are not able to capture. Furthermore, the choice of indicators is constrained by information availability and more accurate data ought to be generated, in particular on the public-private mix. In order to achieve vertical equity, it is necessary to produce data on multiple coverage from subsystems, and to analyse the role of private insurance in the system, so that need estimates account for this. Given that 25% of the Portuguese population benefits from occupationally-based and private health insurance coverage and these are known to be concentrated in certain parts of Portugal, accounting for multiple coverage would obviously change population estimates for resource allocation and other health policies.

5.4 Adjustment for demography

The least controversial adjustment for health care need is the demographic component (Carr-Hill and Sheldon 1992)[118]. Most capitation formulas make an adjustment for age and many for sex, too (Rice and Smith 1999). For countries with a NHS structure and regional funding, many do not apply sex adjustments (that is the case of England) due to the similar demographic profiles between geographical areas[119].

Most countries have deployed utilisation data to compute estimates of demographic-related need. Evidence for Portugal shows that utilisation data is highly influenced by inadequacies in the provision of hospital services and problems in other health care sectors.

5.4.1 Method

There are several methods to assess the impact of age and sex on health care costs (i.e., for designing age/sex cost curves). Methods deployed have consisted of: use of national hospital utilisation rates (measured by inpatient days) by age and sex to proxy costs by age and sex (Department of Health and Social Security (United Kingdom) 1976); estimation of average costs of providing hospital care to different age groups, by means of using information on average costs for some specialties (Carr-Hill, Hardman et al.

[118] Similarly to the RAWP report and to the capitation technology used in England, modelling of age and sex, and additional need is carried separately. This has been the conventional wisdom for most countries and has the advantage that the modelling of each adjustment has a higher potential for capturing the true determinants of age and sex- and morbidity-related need. However, it has the disadvantage of neglecting interactions between age and sex and morbidity.

[119] Two other exceptional cases where adjustments for age and sex have not been used (in the past) were motivated by political (lack of consensus between health authorities in Spain) and technical reasons (similar demographic profiles between the US veterans health funding scheme users), respectively (Rice and Smith 1999).

1994); use of cost activity resources data at the national level by age and sex to build the age/sex cost curve (Scottish Office 1999)[120]; evaluation of DRG cases by age/sex groups at DRG prices to build the age/sex cost curve (in Italy) (Rice and Smith 1999)[121]; and possible use of historical expenditure data to extrapolate the impact of age/sex on health care costs, making use of regression analysis techniques (Shmueli 1999). The choice of method to model the age/sex cost curve is highly dependent on the kind of information available.

The obvious approach for Portugal is to adopt the method used in Italy, as DRGs have been used in both countries as a pricing system for hospital services (Bentes et al. 1993). The database used for estimating the age/sex curve contains 1998 data from all the public hospitals of the system, and includes all the DRG cases (nearly 1 million cases). All cases were evaluated at the same DRG price. This dataset was provided by IGIF and DRG prices were taken from the normative law *portaria* 348-B/98 (Ministério da Saúde 1998e). Nevertheless, some considerations must be taken into account when following this approach: DRG prices are set up administratively and are not updated on a regular basis, and they are computed using Maryland weights, which implies the acceptance of assumptions described in Chapter 2[122]. The estimates of the age-sex cost curve do not include care provided by the private sector and this will result in an underestimation of the costs for those of working age[123]. Also, as estimates are based on prices by DRG that do not adjust for the potential extra costs created by long-stay cases for the eldest age groups ('outliers'), costs for the elderly are bound to be underestimated.

The deployment of utilisation data may not reflect needs of individuals at different ages, as there may be implicit rationing by age; for example, the elderly may not get the health care they need (Sheldon, Smith, and Bevan 1993). There is some evidence of such implicit rationing in Portugal, related to the low hospital care accessibility of older people living in rural areas (Santana 2000). The use of utilisation data assumes that the

[120] This corresponds to the matrix approach, described in section 5.2.
[121] In Italy, DRG prices are used as a pricing system for payments to public hospitals.
[122] Although DRG prices have been applied separately to each administrative group of hospitals in the current formula of resource allocation, the adopted method neglects differentiations in prices across types of hospitals.
[123] Access to private care is expected to be higher for population with double coverage and mainly represented by employment active age groups.

current pattern of use adequately captures differentials for redistribution purposes. Little discussion has been given to the alternative approach, namely: building normative age-sex cost curves by means of other procedures departing from population or expert preferences (e.g. to deduct a normative curve). Some countries, such as Spain, have used methods accounting for the impact of the last year of life on the age/sex cost curve (Urbanos and González 2002).

The Portuguese age/sex cost curve (being the level of cost X_{1a} per age/sex group a, as defined below in the notation) was compared to the English curve, and their determinants in terms of price and quantity (per age group) were given by:

$$\text{Average cost per capita} = \text{Average cost per case} * \text{Average cases per capita} \quad (5.1)$$

These variables were compared between countries; comparison was also made with other indicators, such as the average length of stay (LOS) or expenditure shares per age group. The English data was extracted from the York report (Carr-Hill, Hardman et al. 1994) (for acute care). Note that the output hospital unit in Portugal is the DRG case, while in England it is the Finished Consultant Episode (FCE), which are distinct measures of hospital output[124].

5.4.2 Results and discussion

This sub-section first describes empirical results, and afterwards raises possible explanations for the findings. Figures 5.4 and 5.5 give the age/sex curve for Portugal and the age curves for Portugal and England. These have the expected U-shape. The deviation from the U-shape for the Portuguese female cost curve (Figure 5.4) is partly due to attributing birth costs to the relevant age group of mothers (in many countries such as Britain, birth costs are put down to the new-borns). The Portuguese age cost curve compared with that for England shows higher costs in the older groups. Table 5.1 shows that the number of cases/episodes per capita is 50% higher in England, but the average cost per case/episode is higher in Portugal; the total expenditure per capita is 10% higher in England[125].

Figure 5.4: Age/sex cost curve (expenditure per capita) (Portugal) (€'s)

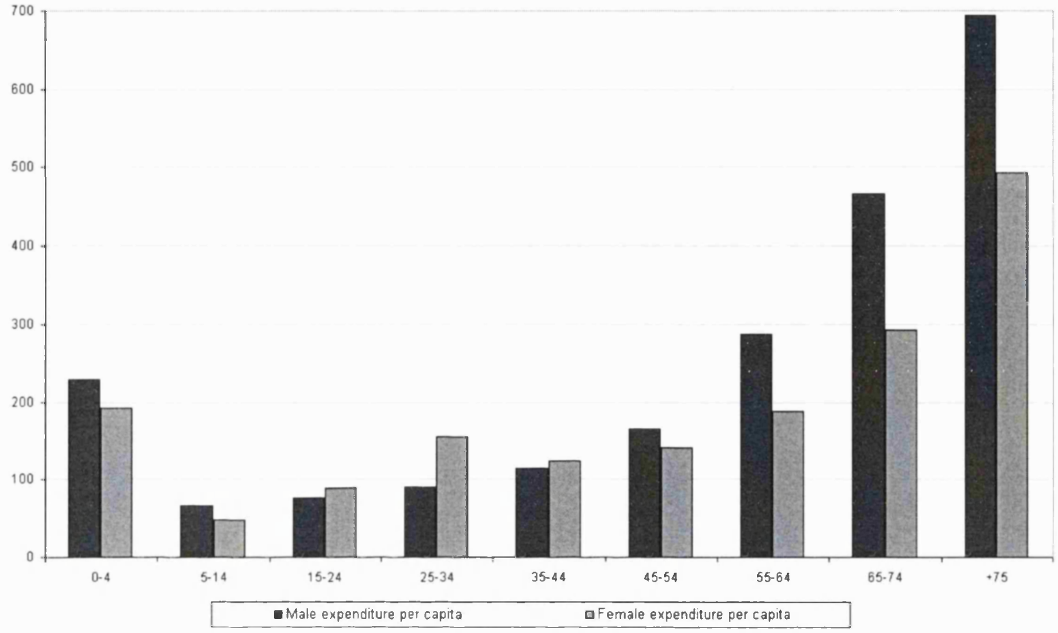
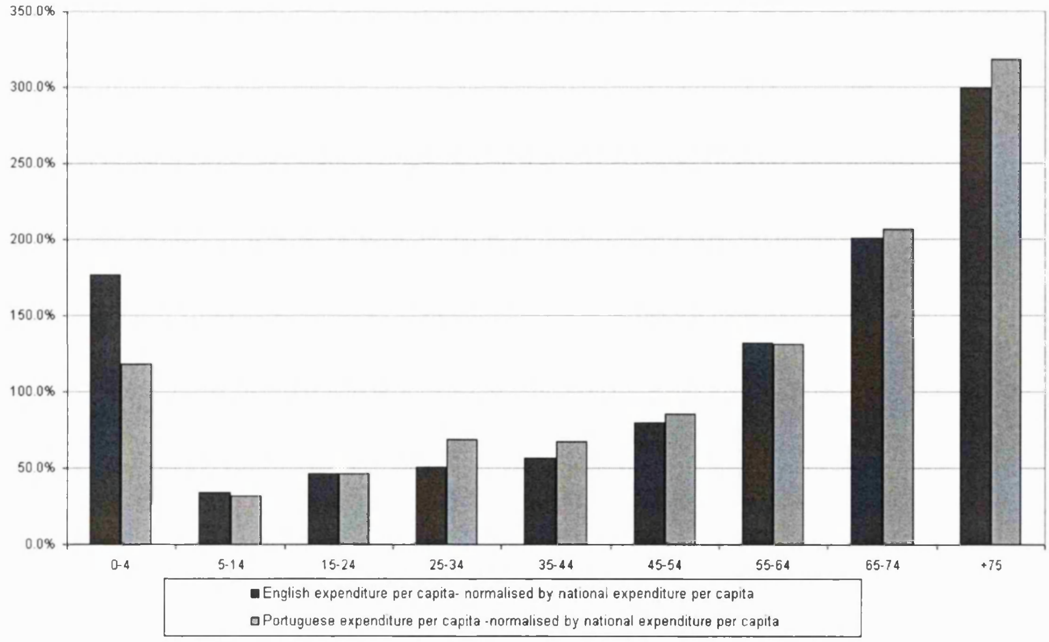


Figure 5.5: Age cost curves normalised by the national cost average (Portugal and England)



[124] Compared to DRGs, FCEs are more easily manageable and interpretable; however, FCEs do not differentiate in terms of the amount of resources expected for providing care (Fattore 1999a).
[125] Exchange rate: £ 1 = € 1.6.

Table 5.1: Expenditure and utilisation values at the national level

	Portugal	*England*
Average cost per capita	*(€'s)* 180	*(£'s)* 130
Average cases/episodes per 1000 inhabitants	98	146
Average cost per case/episode	*(€'s)* 1,805	*(£'s)* 900

Figures 5.6 and 5.7 show that for Portugal the number of cases is the most important determinant of the shape of the age cost curve, while for England it is price (the atypical behaviour for the Portuguese 25-34 age groups can be partly explained by the treatment given to births in Portugal).

Figure 5.6: Per capita expenditure decomposition by age group (Portugal) (€'s)

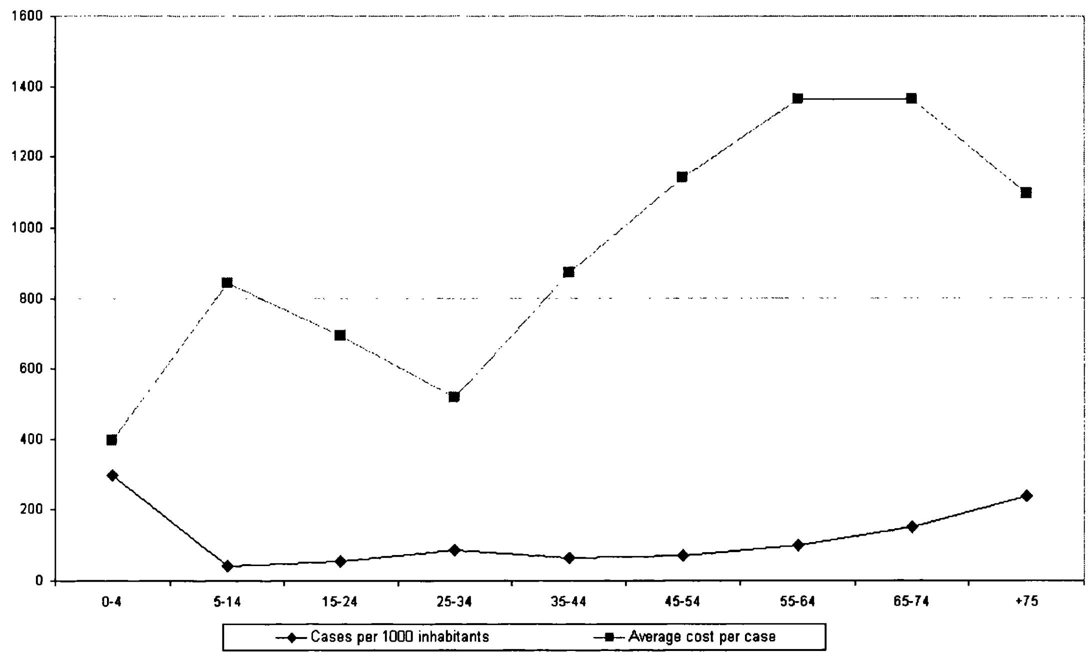
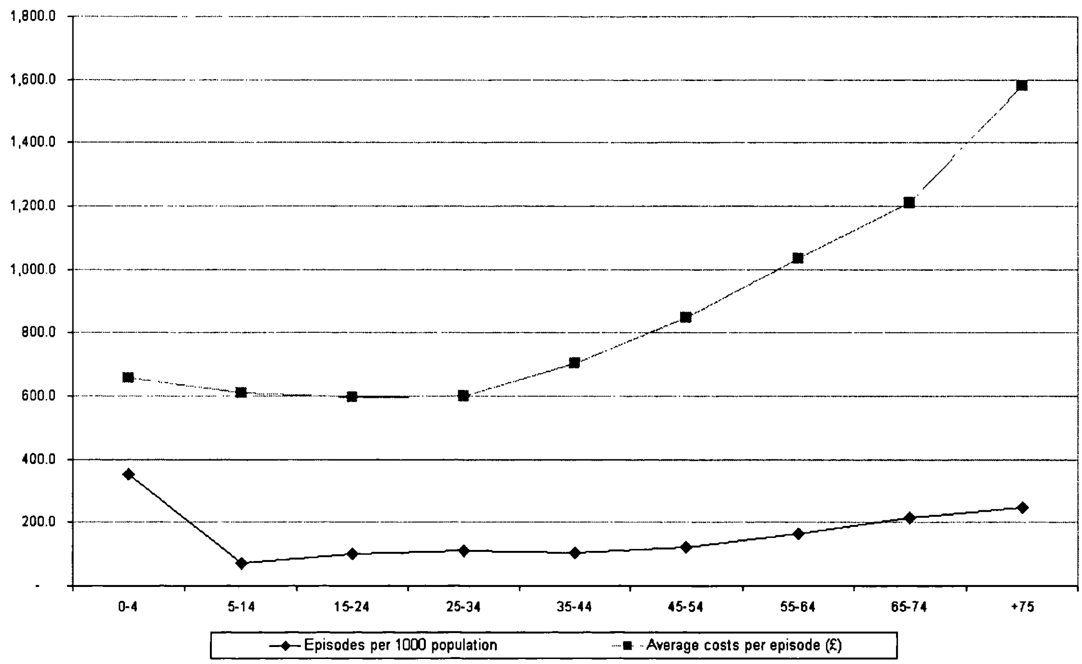


Figure 5.7: Per capita expenditure decomposition by age group (England) (£'s)



Average cost per case data (Figures 5.8 and 5.9) shows that in Portugal, male average costs per case are systematically higher than female ones; in England, males have higher average costs in the 15-54 age groups only. Moreover, the cost per case curve follows a very different pattern: in England both curves (male and female) have a convex behaviour; in Portugal, the curve for males follows a concave behaviour and for female the average cost curve has a non-monotonic behaviour (partially justified by births). In addition, there is a large gap between male and female costs per case in the 15-34 age group in Portugal. Costs for cases of the 5-24 age group for Portugal are almost 300% above those for females. For the 5-14 male group, this finding seems to require further research in the absence of any available information. For the 15-24 age group, this finding might be explained by the high levels of traffic accidents that translate both into hospital use requiring expensive inputs and into high mortality rates, described in detail in section 5.5.1 (in this age group, traffic accidents are commonly associated with motorbikes accidents). Costs for the 25-34 age group are also much higher for Portugal than for England. The same reason explained above (on traffic accidents) might apply to this finding.

Figure 5.8: Average cost by case per age/sex group (Portugal) (€'s)

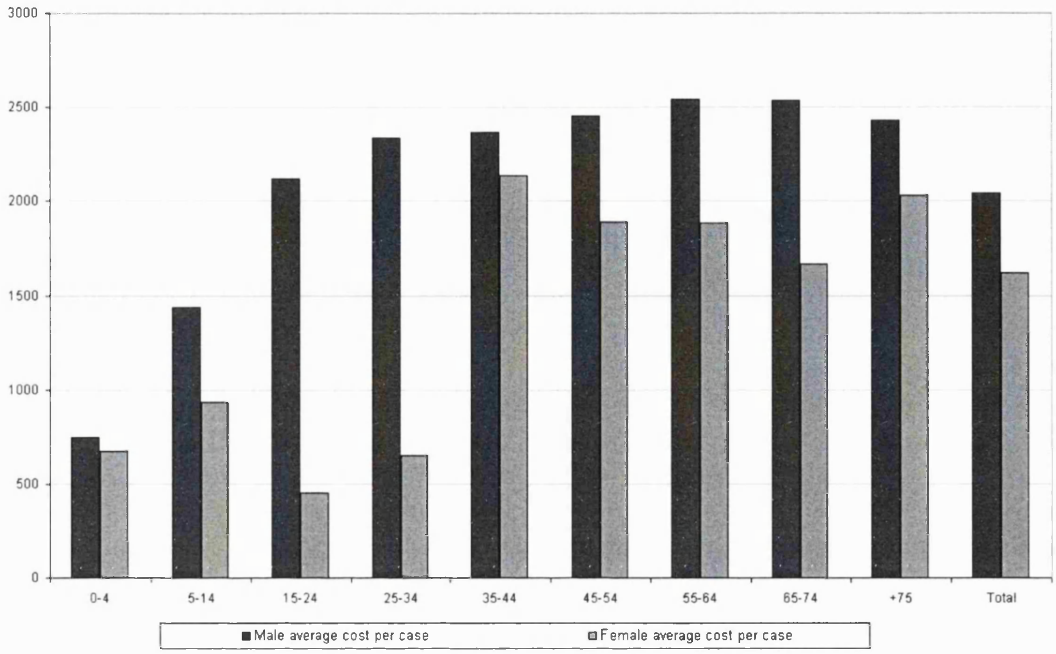
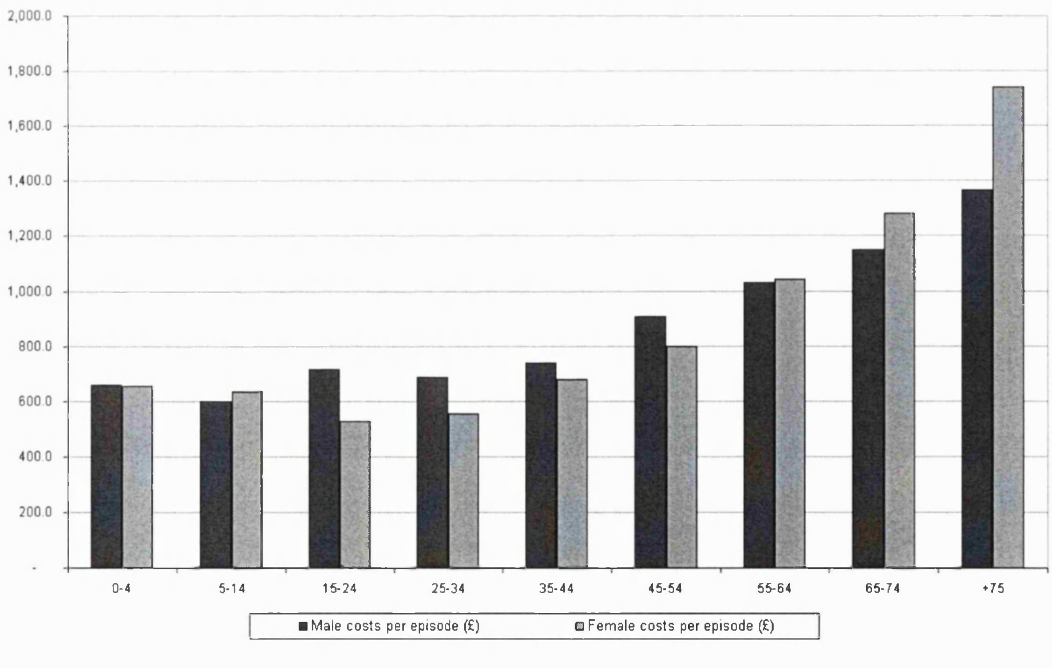


Figure 5.9: Average cost per case by age/sex group (England) (£'s)



Analysis of crude utilisation indicators (cases/episodes per capita) (Figures 5.10 and 5.11) shows a similar behaviour across age groups between the two countries: a global U-shape curve with deviations showing higher female utilisation in the 15-54 age groups.

Figure 5.10: Cases per 1,000 inhabitants by age/sex group (Portugal)

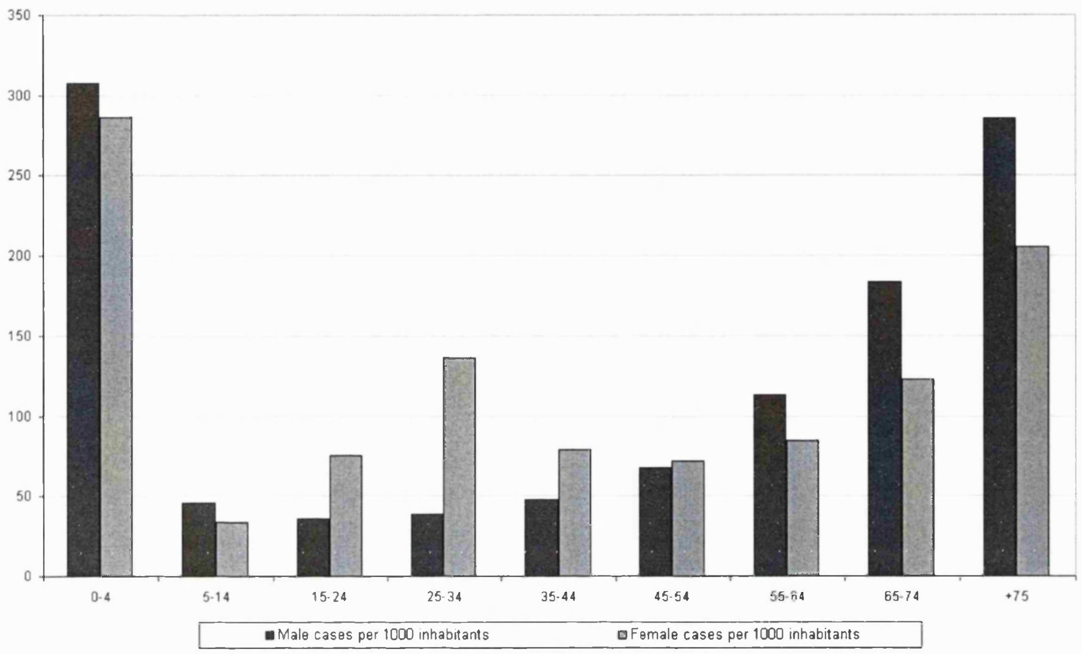
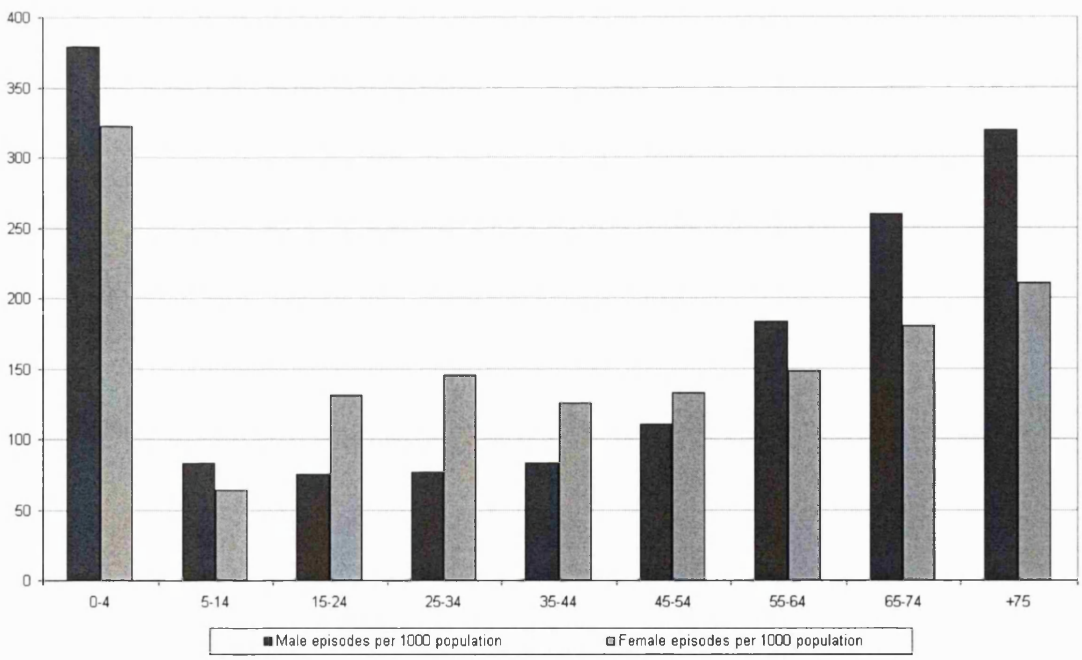


Figure 5.11: Episodes per 1,000 inhabitants by age/sex group (England)



Average length of stay data per age group differs strongly between countries (Figures 5.12 and 5.13). As expected for both countries the LOS curve follows the average cost per case curves (Figures 5.10 and 5.11), confirming the differences between the two countries. Moreover, Portuguese LOS is higher than the English LOS across all age/sex groups –the only exception is for females over 75. National expenditure shares per age group (in accordance with the DRG evaluation) were compared to the respective population shares for both countries, and have confirmed that Portuguese public hospitals are spending a comparatively higher proportion of resources on the older age groups.

Figure 5.12: Average LOS (in days) by age/sex group (Portugal)

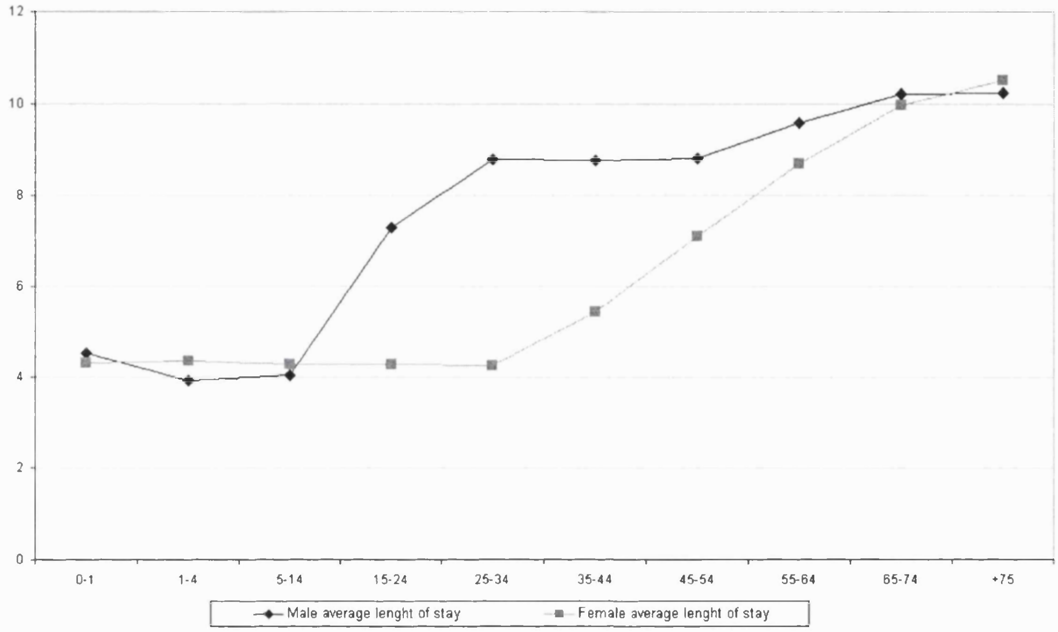
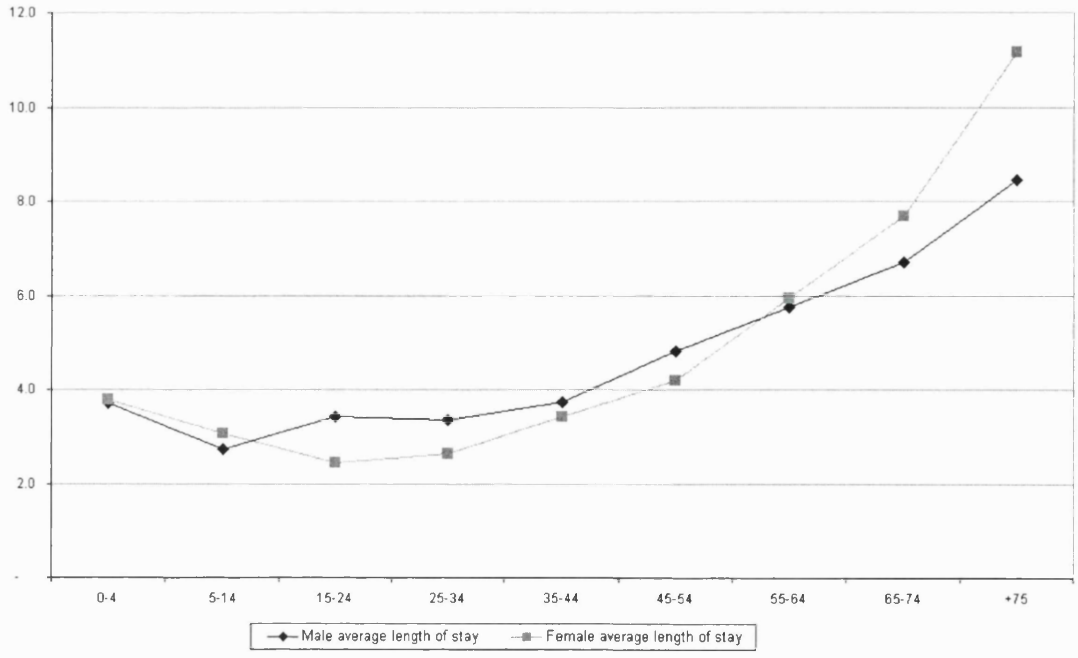


Figure 5.13: Average LOS (in days) by age/sex group (England)



Estimates of relative demographic need (age and sex) at the district level are analysed and discussed in detail in Chapter 8[126]. The district redistribution implied by this adjustment ranges between 87% and 117%. There is a negligible impact of the sex adjustment at the district population level; this is because of similar sex demographic structures across areas.

The empirical comparison of age cost curves between Portugal and England illuminates problems of the Portuguese health care system:

- The problems with the ambulatory and home care sectors, as well as inadequate social care in Portugal may explain why there is a relatively high hospital spending in the older age groups (mainly justified by quantity effects). Higher provision of other complementary sectors of health care for England (such as long term care and nursing homes) may explain why in England price effects (comparatively) are the main determinant of the age-cost behaviour of the eldest age groups. Other reasons for the relative high spending in the older age groups for Portugal might be differences in availability of supply and in doctors' behaviour.
- Higher accident rates for males of the 15-34 age group in Portugal may explain the relatively large gap between male and female costs per case in the 15-34 age groups for Portugal[127] (higher accident rates are expected to incur additional costs to hospital services).
- The relatively higher costs per case for males across all the age groups in Portugal may be attributed to cultural factors, i.e. that Portuguese men only visit hospitals when seriously ill.
- Lower level of availability of hospital care, differences in the pattern of health care delivery and lower efficiency for Portugal may explain LOS differences between Portugal and England.

Consequently, estimates of demographic need of hospital care present evidence of inadequacies in the provisions of the hospital and other health care sectors in Portugal, and seem also to reflect some cultural characteristics of the population. Current practices in the use of hospital care should be further investigated (comparing Portugal

[126] The demographic adjustment mainly favours the interior and southern districts that have older populations at the expense of the urban northern districts with younger populations, as expected.
[127] Mortality rates for this age group are also substantially higher than for other age groups (DGS and Ministério da Saúde 1994/5/6/7/8).

to England) and the question of whether to adopt a normative approach in demographic need adjustment should be addressed

5.5 Adjustment for additional need

Morbidity shows the degree of prevalence of disease in a population and thus ought to be a key component in measuring geographical need of hospital care. The problem is that there are no comprehensive morbidity data adequate for estimating health care needs while there are difficulties in specifying the factors that influence morbidity and analysing how those affect health care costs.

Modelling this adjustment shows that although standardised mortality ratios (SMRs) have been used as a proxy for morbidity in international literature (as a normative approach), they fail to meet some necessary conditions for their use in the Portuguese context. Another normative indicator, namely age specific mortality ratios (ASMRs), is preferred on the grounds of the easier epidemiological interpretation that they offer, their higher weights for deaths in the youngest age groups (in comparison to SMRs) and their robustness in comparison with other normative indicators (such as potential years of life lost).

5.5.1 Method

Two main approaches for adjusting for morbidity have been debated during the 1990s (Mays 1995): normative (based on political judgements) and empirical (based on sophisticated regression techniques). Many countries have used normative approaches, based on mortality indicators (mainly SMRs) for the additional need adjustment –e.g. New South Wales, Belgium, Wales, Northern Ireland, New Zealand and Italy (Rice and Smith 1999). Methods used in England (since 1996) and Sweden are based on results from empirical methods that deploy utilisation and supply data to estimate health care needs (Carr-Hill, Hardman et al. 1994) (Diderichsen, Varde, and Whitehead 1997). England uses regression techniques to measure the impact that need and availability of health resources have on hospital utilisation (at the small area level). The adjustment modelling uses sophisticated techniques, such as simultaneous equation regression and multilevel modelling. England has recently added an additional adjustment, dealing with unmet need (adjusting for under-utilisation in low income/ethnic minority groups). Sweden makes use of individual level data in a matrix to capture the ways in which demographic and socio-economic variables are proxies for health care need and translate into differential costs of utilisation; its modelling approach assumes that relative need for hospital services is proportional to the utilisation levels of major socio-economic groups. Recently, there has been an increasing push in favour of an empirical approach based on epidemiological modelling (Townsend 2001). The use of this approach in a resource allocation formula is nevertheless problematic, as it requires data by disease, while it is not clear how to use it in a national formula (Scottish Office 1999). Some background work on Wales has been carried out in the direction of progressing towards this approach (Townsend 2001)[128].

For Portugal, the normative approach was chosen. The use of an empirical formula based on health care utilisation data suffers from a number of problems, namely: it departs from the major assumption that some measure of use of health care can be used to predict health care needs (Carr-Hill, Sheldon et al. 1994); it calls for multiple judgements in the process of building an index; it faces technical problems caused by multicollinearity (Smith et al. 1994); it places high demands on data availability; last but not least, the relationship between health policy objectives and empirical formulas has often lacked clarity. Nonetheless, the utilisation model developed in Chapter 7 could eventually be used as an empirical approach to capture the impact of morbidity on hospital utilisation. The problems involved with this solution are addressed in the discussion section of that chapter (section 7.4.3.1).

Mortality data have a number of properties, which make them good proxies of morbidity. Mortality broadly reflects cumulative morbidity and social experience in an area, and has proven to provide more stable and comprehensive measures of morbidity compared to utilisation rates (Sheldon 1997); moreover, it is a more direct measure of needs for health care than social indicators (Mays 1987); additionally, it is a measure independent of health care supply, can be decomposed by age and sex, and is routinely

[128] Townsend proposed a new formula for Wales that makes use of updated expenditure figures by service and disease category and connects health condition to expenditure blocks. The author acknowledges that the formula makes use of survey data at the local level, but faces problems with accuracy and stability of the data and with the validity of the indicators that are assumed to capture health need.

available and periodically updateable by area of residence (Department of Health and Social Security (United Kingdom) 1976). Furthermore, mortality indicators (such as SMRs) have proven to be a statistically significant variable for explaining utilisation rates (Carr-Hill, Hardman et al. 1994). SMRs are a more direct measure of health state/status than social indicators (Mays 1987) (Newbold et al. 1998); they are moderately to strongly correlated with measures of short and long-term disability, long-term illness and self-assessed health (Hutchison et al. 1999); and their use constitutes a simple and transparent method for adjusting, that is not subject to political bargaining[129] (Holland 1998). Empirical evidence in Portugal has shown that lower socio-economic levels (which relate to morbidity) are associated with higher mortality (Lucas 1986) (Kunst et al. 1998).

Despite the extensive use of SMRs, there are several criticisms with regard to their use in capitation formula: the nature of the relationship between morbidity and mortality is (partly) unknown (Mays and Bevan 1987); there is a differentiation between the chronic conditions which are likely to generate a high use of health services and those which result in deaths (Mays and Bevan 1987), as not all illnesses are fatal (Le Grand 1982); SMRs fail to address the existence of health care resource needs associated with deprived areas which do not translate into mortality (Mays 1995); SMRs might not be responsive along time (Mays 1995); the use of SMRs does not give information about how morbidity impacts on costs, demanding a decision on the SMR weight to be given to the capitation formula (Carstairs and Morris 1991); in some cases, SMRs might not be strictly comparable between populations (Yule 1934); and the use of SMRs suffers from the 'numerator/denominator' problem, as census data and health certificates are not linked at an individual level (Macintyre 1997).

For Portugal, SMRs were computed using the method of indirect standardisation –the formula for the SMR for all ages is presented in Equation 5.2. Indirect standardisation is to be preferred when there is a small number of deaths in some geographical areas (Bowling 1997). Notation that adds to that of previous chapters is presented in Table 5.2. The SMRs use mainland Portugal as the reference population (31.12.1997); they

[129] Political bargaining has been an important issue in England: losers from the original RAWP formula have created incentives for focused research which would generate better allocations for them (Mays 1987). Most of that commissioned research has developed deprivation indexes with the aim of replacing

were computed with the 100,000 multiplier for the following age groups: 0-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, +75; and they used data taken from the Portuguese General Directorate of Health Website (http://www.dgsaude.pt/).

Table 5.2: Notation in use

Notation	*Interpretation*
a	Age group a [130].
X_{1a}	Age (and sex) cost for age (and sex) group a.
de_{ar}	Number of deaths in area r from the age group a.
P_{ar}	Resident population of the age group a in area r.
$r_{ar} = \frac{de_{ar}}{P_{ar}}$	Death rate in area r from the age group a, which corresponds to the definition of age specific mortality rates for area r and for age group a (defined below).
$r_a = \sum_r de_{ar} / \sum_r P_{ar}$	National death rate for age group a.
cutoff	Age reference used in the computation of the potential years of life lost index. It is related to life expectancy.
I_a	Mid-age point of age group a (required to compute the potential years of life lost index).
SMR_r	Standardised mortality ratio index for district r.
$ASMR_{ar}$	Age specific mortality ratio index for age group a and for district r.
$PYLL_r$	Potential years of life lost index for district r.
RMI_r	Relative mortality index for district r.
P_r	Defined in Chapter 4.

$$SMR_r = \frac{\sum_a de_{ar}}{\sum_a r_a * P_{ar}} \tag{5.2}$$

The calculation of SMRs for Portugal, raised methodological questions with regard to:

a) The number of years to be used in the computation of SMRs: the choice is crucial as SMRs instability would result from small numbers of deaths in the selected geographical areas; three-year data proved to provide stability.

SMRs in the formula (Sheldon, Smith, and Bevan 1993). The use of SMRs constitutes a normative approach based on informed judgement.

[130] The age groups in use depend on the indicator and are specified in the text.

b) The choice of the mortality causes to be included, as some causes might not justify additional need for health care resources. For instance, Portugal has had a very high number of deaths by traffic and other accidents by EU standards (OECD 1998). It was, in general, observed that deaths by external causes registered a high yearly fluctuation across district areas. Since the goal is to capture mortality related with morbidity, and that the determinants of external deaths are multiple and overall not connected with morbidity, deaths from external causes have been excluded from the SMRs[131].

c) England and Scotland have been using different thresholds for the age groups to be excluded in the SMRs. England used under-75 SMRs and Scotland under-65 ones, respectively (Scottish Office 1999) (Department of Health (United Kingdom) 1999)[132]. Such restrictions matter, as SMRs are highly influenced by the number of deaths of the elderly (Palmer et al. 1979). Surprisingly, little attention has been given to the crucial choice of threshold (Gaffey 1976). The impact of using different thresholds was investigated for Portugal.

d) There are statistical conditions necessary for the appropriate use of SMRs, such as stability of death rates per age group and stability of population structures across areas. Statistical tests and experiments previously developed by Kilpatrick and by Tsai and Wen were applied to Portuguese data (Tsai and Wen 1986) (the Kilpatrick formula was used in the version presented in the Tsai and Wen study). These tests have, however, often been neglected in the development of formulas and literature.

---

[131] There were two reasons for discounting deaths from external causes from mortality figures while the same was not done in the age cost curve. Firstly, mortality statistics provide information on deaths from external causes, while hospital DRG statistics do not provide information on the cause of entry into the hospital. Secondly, including the costs of accidents in both adjustments could possibly lead to double counting.

[132] These countries exclude different age groups in the context of capitation studies, but they combine different SMRs with distinct population bases: England applies under-75 SMRs to the whole population, while Scotland applies under-65 SMR to the population under-65 years of age. Nonetheless, Scotland is complementing the use of SMRs as a mortality indicator with other deprivation measures. Weights for the SMR and for the deprivation measures are generated by statistical analysis, which determines which deprivation indicators best explain the additional utilisation derived from the age profile (Townsend 2001).

The robustness of SMRs was tested by comparing them with a set of alternative indicators: the potential years of life lost index (PYLL, formula in Equation 5.3)[133], the relative mortality index (RMI, formula in Equation 5.5[134]) and age-specific mortality rates (ASMRs, formula in Equation 5.4[135]). ASMRs for the district are to be weighted by the age-adjusted population structures (e.g., for each district, the ASMRs presented in Equation 5.4 are weighted for the ratio of the age-adjusted population of the same age group to total age-adjusted population for that district)[136]. These mortality indicators present advantages and disadvantages compared to SMRs (in the context of resource allocation analysis), and they mainly differ from SMRs in that they give a lower weight to deaths in oldest age groups (Palmer et al. 1979). Mortality indicators were also juxtaposed with socio-economic indicators collected from the 1991 census data[137]. There was no other morbidity data available at the district level. Given that mortality rates for small age groups and for the smallest geographic areas are based on small numbers of deaths, both the ASMR and the RMI were computed using the method of indirect standardisation[138].

$$PYLL_r = \frac{\sum_a r_{ar} * p_{ar} * (cutoff - I_a)}{\sum_a r_a * p_{ar} * (cutoff - I_a)} \tag{5.3}$$

$$ASMR_{ar} = \frac{r_{ar}}{r_a} \tag{5.4}$$

$$RMI_r = \sum_a \left(\frac{r_{ar} * P_{ar}}{r_a * P_r} \right) = \sum_a \left(ASMR_r * \frac{P_{ar}}{P_r} \right) \tag{5.5}$$

---

133 In the computation of the PYLL, a cut-off of 70 years and the same definition of age groups as in SMRs were used.

134 Age groups in use: 0-14, 15-44, 45-64, 65-74 and +75.

135 Ibidem.

136 This is very similar to Equation 5.5, with the only difference that population numbers are adjusted by age.

137 Data available at the *concelho* level, and aggregated to the district level (INE 1993a, 1993b, 1993c, 1993d, 1993e).

138 It should be acknowledged that when there are small numbers of deaths, indirect standardisation has the advantage over direct standardisation in producing lower standard errors (Inskip, Beral, and Fraser 1983) (Bland 2000). Direct standardisation is useful to preserve consistency between the populations, but for Portugal, tests have shown that population structures are similar across areas (as described below).

5.5.2 Results and discussion

Analysis of the SMRs shows that:

1. Portuguese SMRs under 65 range between 88.7 and 112.5%. This is a much narrower interval than for England: English SMRs range between 80.5 and 138.1% (for 1989-1993 and for regions with similar populations to the Portuguese districts) (English data from (Department of Health (United Kingdom) 1996))[139].

2. The choice of excluding external deaths has a significant impact on Portuguese SMRs.

3. There is a weak relationship between SMR values and district rankings generated by all age, under 75 and under 65 SMRs. Table 5.3 gives Rank-Pearman correlations of SMRs at the district level (for Portugal, and for England and Wales).

Table 5.3: Rank-Pearman correlations between SMRs for Portugal, and for England and Wales

Portugal	*SMR <75*	*SMR <65*	*England and Wales*	*SMR <75*	*SMR <65*
SMR all age	84%	67%	SMR all age	93%	86%
SMR <75		93%	SMR <75		94%

Source: (Department of Health 1999) for England and Wales

Note: SMRs excluding deaths by external causes for Portugal

4. Some contiguous areas with similar socio-economic characteristics were shown to have very different SMRs. This mainly applies to contiguous districts in rural areas: Beja in comparison with Évora, and Bragança with Vila Real (Figure 5.14)[140]. Figure 5.15 shows the relationship between illiteracy rates and SMRs[141]. These

[139] The wider range of English regions might be partly explained by the larger number of regions, which leads to a higher dispersion in the distribution.

[140] Some reasons for this finding might be different trends on past mortality and on past population variations, and differences in food intakes across areas (Dias 1994). Regarding trends on past mortality and on population variations, there were some factors that might have justified 'shocks' in population and mortality dynamics: high emigration levels from rural areas in the 1960s and 1970s; the colonial war in the 1960s and the return of massive numbers of people from the ex-colonies after 1975.

[141] Illiteracy rates computed as described in Chapter 3. The same relationship was found between SMRs and other census-based socio-economic indicators, such as housing conditions and percentage of inhabitants dependent on primary sector activities.

indicators contradict the common assumption in the literature that SMRs tend to be higher in poorer areas (Whitehead 1994).

Figure 5.14: 3-year under-65 SMRs (1995/6/7) excluding external causes

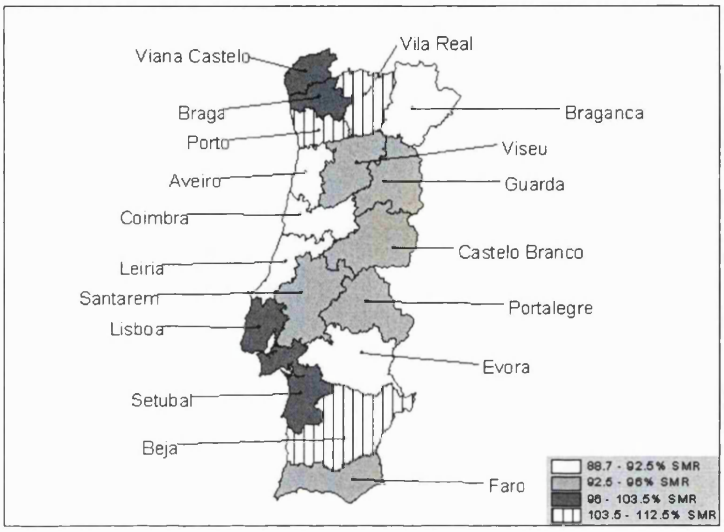
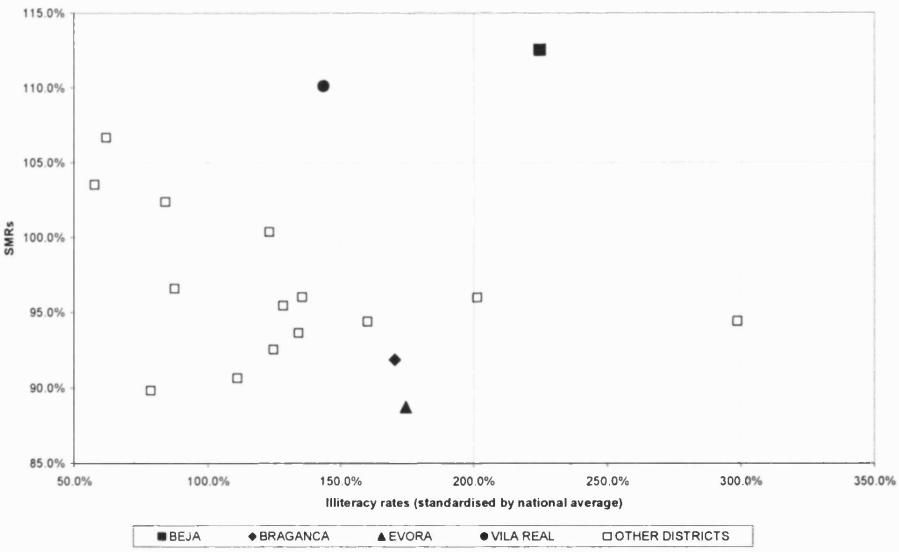


Figure 5.15: SMRs vs. illiteracy rates



5. Tests of the population age structure and mortality rates per age group indicated that age distributions of the population are similar between all districts, but mortality rates for the Lisboa district were statistically different from national mortality rates (using the Kilpatrick test). For Lisboa, the 15-44 age group present the largest death deviations from the national rate, giving evidence of a high level of premature

mortality. As the SMR is a composite indicator of the deaths of all the age groups (where its numerator and denominator are the observed and the expected number of deaths), variation in the death rate for a specific age group affects both the numerator and denominator, and it is not possible to disentangle these effects on the SMRs. Consequently, the necessary condition for the use of SMRs (stability of age-specific mortality rates across health regions) (Tsai and Wen 1986) does not apply, and Portuguese SMRs are unreliable indicators of relative need.

Examination of alternative indicators to SMRs resulted in the following findings[142]:

1. ASMRs, PYLL and RMI present very similar results (Figure 5.16 and correlations in Table 5.4). ASMRs and the RMI give very similar values, as expected: as defined above, the difference is that ASMRs are applied to the age-weighted population, while the RMI is applied to crude populations. The similarity between the rankings produced by ASMRs and the PYLL (in comparison to SMRs) constitute an indicator of robustness of these mortality indicators. Using these in resource allocation would result in a slightly higher level of redistribution than when deploying SMRs (Figure 5.16);

2. In comparison to SMRs, ASMRs and PYLL tend to favour urban areas (Figure 5.16), and they seem to capture better the concept of material/urban deprivation[143]. Urban deprivation has shown to be an important indicator of morbidity and of the need for additional resources in health care (Senior, Williams, and Higgs 2000).

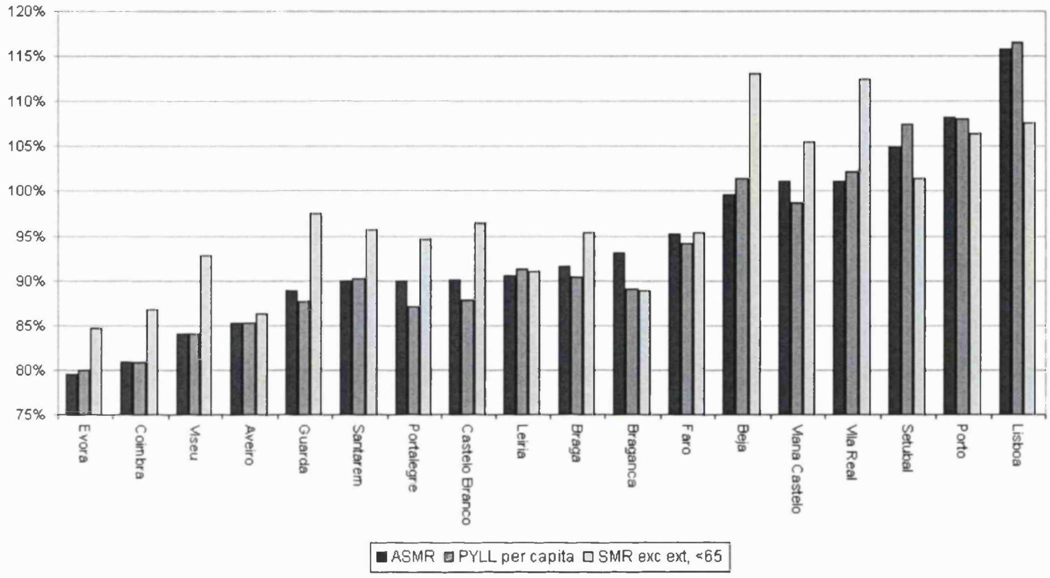
---

[142] Comparison of SMRs with other mortality indicators has shown that SMRs are very dissimilar with infant, neonatal and perinatal mortality indicators over the 1994-8 period (DGS and Ministério da Saúde 1998) (DGS and Ministério da Saúde 1994/5/6/7/8) and with 'avoidable' mortality data over the 1980-4 and 1985-9 periods (Holland and Working Group on Health Services and Avoidable Death 1991; Holland and Commission of the European Communities. Working Group on Health Services and Avoidable Death. 1997), as observed by district values and by correlation rates. Differences with infant mortality indicators are potentially explained by the restricted number of deaths in this mortality data and by the fact that the determinants of these indicators are very different. 'Avoidable' deaths are a category of total deaths from specific diseases for which mortality should be wholly or substantially reduced when appropriate medical care is sought and provided in good time (Holland and Working Group on Health Services and Avoidable Death 1991). There is not a consensus on the causes of death to be included as 'avoidable deaths' (Mackenbach, Bouvier-Colle, and Jougla 1990); and 'avoidable' mortality variations might also capture differences on coding quality, levels of incidence and prevalence of illness, and population socio-economic conditions (Treurniet et al. 1999). Differences between SMRs and 'avoidable' mortality data can be explained by: the sharp decrease of mortality rates over the period, as well as the typology of death causes; improvements on codification of deaths and on the system of data collection are also expected to have had an important impact on mortality evolution.

Table 5.4: Correlations between SMRs and alternative health outcome indicators (excluding external causes)

	PYLL per capita	*RMI/ASMRs*
SMR, <65	83%	82%
PYLL per capita		99%

Figure 5.16: PYLL vs. ASMRs vs. SMRs



ASMRs are chosen as a proxy for additional need given the empirical findings just described. ASMRs have an easy epidemiological interpretation as an indicator to be used in resource allocation: deviations on mortality rates per age group from national mortality rates per age group are taken as a measure of additional need for health care resources for that age group. ASMRs place higher weights on deaths in the youngest age groups (in comparison to SMRs). ASMRs allow for analysing the impact of a high variation in death rates for Lisboa with respect to the national rate. Lisboa, Porto and Setúbal are the districts that win with the use of ASMRs, while Évora, Coimbra and Viseu are the main losers.

ASMRs present more robust results than SMRs when there are variations in mortality rates. Nonetheless, ASMRs placed (comparatively) higher weight on the deaths of the

143 As described in Chapter 3, 1991 census-based indicators appear to express the concept of rural deprivation, but seem weaker in capturing the concept of urban deprivation.

youngest where data are less reliable (Grundy 1996)[144]. The use of ASMRs in a capitation formula to allocate resources would favour districts where there is already a concentration of supply, such as Lisboa and Porto (as shown in Chapter 3), which implies a non-correction of current inequities in the distribution of capital. It would be useful to have some evidence relating ASMRs to other socio-economic indicators for Portugal. There has been a lack of guidance in the literature on the weight to be attached to SMRs (and to other mortality statistics, such as ASMRs) in capitation formula. Different weights imply different levels of redistribution and different assumptions on how morbidity impacts on costs. In this study, in the absence of additional information, ASMRs are used with a weight of one in the capitation formula computed in Chapter 8, which means accepting a redistributive range between 80 and 116%[145]. This is a critical assumption as there is no information on how morbidity impacts on costs.

5.6 Implications for policy and concluding remarks

The use of the adjustments described in a multiplicative model to measure need for hospital care is analysed in Chapter 8. That chapter analyses in more detail the implications of the adjustments for redistribution at the district level.

Concluding observations follow based on the structure of Figure 5.1.

If Portugal is to allocate resources in accordance with geographical need for hospital care, it has to develop systems providing better information on resource allocation. Lack of data constrained the development of methods reported in this chapter (for example, the lack of population projections at the district level).

Analysis of Portuguese utilisation data supported the existence of recognised inadequacies in the hospital and health systems. The age adjustment followed a U-

[144] Other alternative indicators on mortality (available in (Inskip, Beral, and Fraser 1983)) could also be explored.

[145] As described above, SMRs were found to be a statistically significant variable for explaining utilisation rates, having attached a coefficient of 0.75 (Carr-Hill, Hardman et al. 1994). Nonetheless, the range of variation of SMRs in England and Wales is wider than in Portugal (ranging between 80 and 138%, as cited above, although this is partly expected given larger numbers of districts for England and Wales).

shape. ASMRs were chosen as the morbidity indicator and seem to relate to material deprivation.

The conventional technology in modelling the needs adjustment has been the empirical estimation of the age/sex cost curve and the normative use of SMRs. This chapter has identified problems in applying this technology and the advantages/disadvantages from a normative or empirical approach.

Firstly, the lack of information on double covered population, a common problem in many countries, conceals the role of the private sector on needs estimates and creates problems in measuring need and in monitoring equity.

Secondly, SMRs have been widely used in international literature, but have not always been tested for their suitability.

Thirdly, the deployment of utilisation data in the estimation of the age/sex cost curve has shown that it reflects cost inefficiencies and structural problems. This has implications in capitation literature as many countries use methods of resource allocation that do not link measurements of need for hospital care with measurements of need for other sectors, such as primary care and social care. This results in inadequacies in the distribution of hospital resources when using a capitation formula.

Fourthly, traditional capitation methods used to estimate current need (as captured by resident population numbers) might be outdated, as they fail to take account of sharp demographic changes.

This chapter computed indices of relative need for hospital care at the district level based on the objective of equal opportunity of access for those in equal need. The next chapter estimates the unavoidable costs adjustment of a capitation formula.

6 CHAPTER 6 - A multilevel model to estimate unavoidable costs and to disentangle causes of inefficiencies in hospital care

6.1 Objectives

This chapter presents research on modelling unavoidable costs (UCs) of hospital care. This follows from the arguments in Chapter 4 that adjustments for UCs are necessary given that different purchasers/providers ought to operate under the same set of choices, in particular when they are given a fixed budget (Wilson et al. 1996). UCs of hospital care are defined as those costs that lie outside the control of hospital management.

The ultimate objective of this chapter is to build a measure/index of the relative levels of UCs for the Portuguese district level. Estimates are first produced at hospital level and are then aggregated to district level. This study develops a different approach to that of previous studies in estimating UCs. The approach taken here recognises that variations in hospital costs can be explained by characteristics of individual hospitals, by their place in an administrative hierarchy, and by geographical location[146]. The model is also used to identify causes of allocative inefficiency. The model was developed for Portugal but can be adapted to other countries with similar characteristics, namely central control, planning and central management of key resources. The main objective is to create an index for UCs (as an adjustment of the capitation formula to measure inequities), while also taking into account hospital behaviour when this index is used for allocating resources.

[146] In the multilevel literature, compositional effects explain variations in the individual unit of analysis, that is, in individual hospital characteristics that impact on costs; and contextual effects explain variations that operate for groups of units, in our case for groups of hospitals (defined by some type of criteria), that affect costs.

This chapter consists of five further sections which: structure the problem; build hierarchical models; describe the application of the models to Portugal; discuss results and further research; and make concluding observations.

6.2 Problem structuring

This section explains why information on UCs is needed for a policy of equitable resource allocation. It describes the Portuguese context and outlines the methodological approach to be developed. The following sub-sections review previous studies, describe problems of relevant literature, such as on economies of scale and scope, review the relevant literature on the Portuguese hospital system, and summarise the proposed approach.

6.2.1 Unavoidable cost adjustments

A review of the approaches used in estimating UCs in resource allocation in different countries shows that: a) several approaches have been used; b) modelling is complex; c) the approach taken is highly influenced by the country and by systems of hospital finance. There are no clear rules for what constitutes UCs, and any classification will depend on the policy perspective (e.g. on assumptions about the short- or long-run and about the degree of managers' freedom). Most of adjustments for UCs begin by defining 'legitimate' differentials in provider costs: components of costs that hospital managers cannot reduce. 'Legitimate' differential costs are mostly explained by variations in the external environment of hospitals, and in some cases, variations in internal factors (Hutchison et al. 1999). Examples include:

1. *Costs implied by economies of scale and scope of hospital care.* Scotland is one of the few countries that has adjusted for the impact of economies of scale and scope on hospital costs, using estimates produced by a behavioural cost function model (Scottish Office 1999). The problems involved in modelling these economies are

discussed in the next sub-section and partly justify why most countries have not used this type of adjustment[147].

2. *Costs implied by variations in input prices.* Such adjustments are made because of external market forces or costs of providing services in high cost areas (Townsend 2001). The Netherlands applies regional factors, depending on levels of urbanisation (favouring urban areas) (Rice and Smith 1999); England on the other hand adjusts for differential staff costs and costs of capital, which again favours urban areas (London in particular) (Resource Allocation and Funding Team 2000). For England, the major component is captured by the labour costs adjustment, which assumes that, despite national bargaining and negotiation of salaries, urban and rural labour markets differ (Wilson et al. 1996)[148];

3. *Costs implied by different mixes on health care provision*, such as the public/private and the primary/secondary care mixes in provision. Australia for instance, has treated private provision as a substitute for public expenditure and deducted the estimated costs of this from the public budget (at DRG prices) (Rice and Smith 1999). This kind of adjustment ought to depend on the characteristics of the health care system, for example, whether there is opting-out from public coverage, or whether the private sector is operating as complementary to or as substitute for public provision.

4. *Costs of delivering health care services in rural areas.* Northern Ireland, Finland, New South Wales, New Zealand, Scotland and Wales (Rice and Smith 1999) and England (for emergency ambulance service) (Townsend 2001) have adjusted for these costs. In practice, adjustments for rurality (and also remoteness) have been applied to specific and small components of the health budget[149].

---

[147] This adjustment relates to adjustments for remoteness, defined in Chapter 4. Remoteness costs correspond to internal hospital costs related to economies of scale and scope incurred by providing services to smaller populations located in remote areas.

[148] For example, staff in urban areas has higher opportunity costs for working in the public sector as they can easily work also in the private, and incur higher living costs; and in rural areas, administrative costs relate to difficulties in recruiting labour. This adjustment on labour costs (Wilson et al. 1996) has used regression analysis to isolate the independent effect of location on earnings outside the NHS, and used the assumption that external wage variations are proportional to unavoidable provider costs within the NHS.

[149] Rurality was defined in Chapter 4. For example, Scotland is applying adjustments to selected expenditure components (e.g. community nursing), adjusting for sparcity (scores according to the proximity to a GP) and taking into consideration physical space barriers, such as footpaths and water

5. *Costs of delivering care to specific disadvantaged groups.* Australia weights costs of delivering to Aboriginal islanders by 2.5[150] and England adjusts for ethnicity (which involves the costs derived from the use of interpreter, advocacy and translation services for ethnical minority patients with English language difficulties) (Townsend 2001).

For the hospital sector, the most important components of the UCs are the first, second and third of these adjustments. Although these elements might interact (for example, analysis of economies of scale cannot disregard input prices), most of the adjustments used have focused on some components and ignored others. This is because of the complexity of modelling any adjustment given the multiple determinants of hospital costs, which makes it difficult to disentangle the effect of different elements.

This study describes an integrated approach for estimating UCs that considers the literature on cost functions and includes questions of efficiency, economies of scale and scope, and input prices. The aim is to account for the effects of these elements simultaneously. The next sub-section briefly reviews literature on economies of scale and scope, efficiency and input costs and describes the implications of that literature for the estimation of UCs.

6.2.2 Relevant literature

The presence of economies of scale and scope in the hospital sector is due to greater opportunity for the division of labour and specialisation, and to the reserves of labour and materials that are available to larger institutions (Aletras, Jones, and Sheldon 1997). In addition, there may be economies for particular services. On the other hand, diseconomies of scale might operate as managerial costs increase for large

barriers, to which it assigns double and triple weighting respectively (Scottish Office 1999); in 1999, these adjustments affected 30% of all the budget components identified as being influenced by sparcity. In Wales, since 1992, ambulatory, community services and cash limited General Medical Services are also adjusted by a sparcity factor that reflects the costs of providing services in rural areas (Townsend 2001).

[150] This adjustment in Australia departs from estimates that relate under-use of the aboriginal population group with mortality levels, and is interpreted as an adjustment to the supply side. This adjustment can be

organisations. Nevertheless, as described below there are many problems in estimating economies of scale and scope, because of the complexity of adequately measuring hospital output, input prices, and of estimating allocative and technical inefficiency[151]. In the context of a NHS system, there might be specific inefficiencies that engender higher costs and are caused by the inflexibility of hospitals in deciding upon prices and quantities (that is, these decisions are then without regard to market pressures) and by the specific incentives generated by the hospital financing system. These elements have been shown to be critical for the Portuguese case (vs. conclusions from Chapter 2), and often apply to other countries[152].

There is a diverse literature on hospital cost functions. Many studies recognise the impact of hospital characteristics (including size and scope) on hospital costs, after controlling for variations on location, external factors, etc. However, there is no consensus as to the existence and degree of importance of economies of scale and scope[153]: in 1972, Berki (Berki 1972) postulated that economies of scale 'ought to exist'; Vitaliano (Vitaliano 1987) has shown that there is a lack of agreement on the existence of an optimal size; the review by Aletras et al. (Aletras, Jones, and Sheldon 1997) have reported that the extent of existence of economies of scale is unknown; and McGuire and Hughes (McGuire and Hughes 2002) have summarised the conflicting conclusions on the existence of economies of scale and how those conclusions relate with the techniques of estimation. This lack of consensus is related to many problems, such as:

- The complexity of adequately measuring hospital output;
- A multiplicity of variables and aspects that influence the behaviour of hospital agents and thus their costs (Vitaliano 1987);
- The use of different methods, functional forms and methodological choices (Folland and Hofler 2001)[154];

seen either as an extra-needs adjustment factor above demographic factors, or as an extra-cost to deliver care by hospital units.

[151] The definitions of allocative and technical inefficiency in use are the ones given in Chapter 2.

[152] For example, for Portugal, labour price inputs (i.e., salaries) are decided at the central level and hospitals are not charged for the use of capital.

[153] Economic theory and older studies (such as (Feldstein 1988)) have postulated and given evidence on a U-shape relationship between average cost and hospital capacity. This result has not been sustained by subsequent literature.

[154] For example, it has not always been clear whether it is more appropriate to use short or long run estimations (Aletras, Jones, and Sheldon 1997) and which design options to choose (e.g. pooled vs. partitioned data). Methods of estimating input prices have been inadequate, accounting for the cost of

- The difficulty of disentangling different sources of variation in costs, such as variations in efficiency (Newhouse 1994)[155];
- Most of the studies have made the often unjustifiable assumption that hospitals behave as cost minimisers (Cremieux and Ouellette 2001)[156];
- In modelling cost functions, many studies have controlled for geographical variations that have been statistically significant (Lave and Lave 1970) (Grannemann and Brown 1986) (Vitaliano 1987) (Zuckerman, Hadley, and Iezonni 1994). However, there has not been a common framework for treating the influence of geographical and other external variables such as prices and environment, and it is not known whether these variables are important *per se*, or whether they capture the effects of other confounding variables, which have not been considered.

The two main approaches for estimating hospital cost functions taking account of inefficiency have been the two frontier methods of data envelopment analysis (DEA) and stochastic frontier methods (SFM). DEA computes the frontier practice isoquant (Folland, Goodman, and Stano 1997) and calculates distances between the hospital cost and/or output and the frontier, as a measure of technical inefficiency. The DEA method is, however, sensitive to the influence of outlier observations on the production function (Folland, Goodman, and Stano 1997) and makes critical assumptions about returns to scale. SFMs overcome DEA's weaknesses of not considering random variation, estimate the stochastic frontier using econometric modelling, and are based on the theory of the firm in order to link hospital inputs and outputs to costs. Given the objective of estimating UCs without making assumptions on returns to scale, the SFM approach is to be preferred to DEA.

---

capital is particularly problematic (Folland and Hofler 2001) and the relationship between hospital inputs is not well understood (McGuire and Hughes 2002). Estimates of cost functions and of returns to scale are very sensitive to the omission of variables (such as on hospital technology) and to the existence of incomplete data (Cremieux and Ouellette 2001).

[155] There are doubts about the feasibility of estimating efficiency parameters (Newhouse 1994). Some of the problems of making comparisons of efficiency computations using frontier analysis are similar to those problems described above (Newhouse 1994): heterogeneity of hospital outputs, potential for misspecification due to structural differences between the cost functions of groups of hospitals, the choice of cost functions, and the choice of the measures of input prices and input variables.

[156] Alternative assumptions on hospital (administrators) behaviour objectives are available in literature, though they have not been much applied in practice. Some of these alternative assumptions are: quantity maximisation, utility maximisation (such as maximisation of quality, quantity and quality, and the alternative hypothesis of the managerial expense preference model), the physician control model and the supply induced demand theory (Santerre and Neun 1996).

There are two main types of SFMs: *ad hoc* models and flexible cost functions (such as translog models). Translog models suffer from various disadvantages[157]. *Ad hoc* models seem to perform better when dealing with technical and allocative efficiency, when producing estimates for hospitals in the whole range of the hospital network and when forecasting costs, although they face other problems in imposing constraints on the technological function they assume (i.e., the link between inputs, outputs and prices)[158].

Recent studies on SFMs have indicated a preference for using fixed and random effects, which allow for adjusting the intercepts so that the cost frontier shifts to the appropriate level between groups of hospitals (Linna, Hakkinen, and Linnakko 1998). Nevertheless, there has been little theoretical guidance on the distributional assumptions used in random effects models (Linna, Hakkinen, and Linnakko 1998). Some authors (Newhouse 1994) point out that some of the techniques require strong assumptions that cannot be tested (such as in the technological production function used in some studies). SFMs also have been criticised for neglecting systematic inefficiency (Zuckerman, Hadley, and Iezonni 1994)[159].

Consequently, there are many difficulties in modelling hospital cost functions. Some additional difficulties arise in the use of this type of literature for resource allocation:

- The deployment of utilisation data without adequate control for factors such as quality and inefficiency might create perverse incentives in using the resulting estimates in a funding formula.
- It is not clear whether some of the adjustments being carried out are significant (such as the staff market forces factor for Scotland (Townsend 2001)), or meaningful, as their precise purpose is not always clear (Rice and Smith 1999).
- Determinants of costs reflect the system of incentives of key hospital actors and the characteristics of previous financing systems, and it might be difficult or impossible

[157] Some of the disadvantages of translog models are: they do not allow for distinguishing between allocative and technical efficiency (Folland and Hofler 2001) as their estimation produces small residuals and the estimates of costs are near deterministic; they provide estimates of coefficients that should be interpreted only for average values of the sample (Vita 1990) (Linna, Hakkinen, and Linnakko 1998); and they are more useful when the focus of research is on the hospital production function (Li and Rosenman 2001). Although translog models put few restrictions on the underlying technological structure, they make strong assumptions with regards to separability (McGuire and Hughes 2002).

[158] Although translog models have been more used than *ad hoc* models in recent literature, some empirical applications have shown that a Cobb-Douglas production function outperforms the production function of a restricted translog (Gerdtham et al. 1999).

[159] These criticisms apply to the model developed in this study.

to control for these effects. These elements have been shown to be critical for the Portuguese case, and often apply to other countries.

6.2.3 The Portuguese context

In order to design cost models, it is vital to take account of the characteristics of the Portuguese hospital system. This sub-section describes some aspects of the country information setting that should inform modelling of UCs.

6.2.3.1 Review of relevant literature

There have been several studies of costs and production functions of Portuguese hospitals: (Paiva 1993) (Lima 1998) (IGIF 1999) (Barros and Sena 1999) (Carreira 1999) (Lima 2000). Key characteristics of these studies are available in Appendix B. The main criteria for analysing (and assessing) those studies were taken from Aletras et al. (Aletras, Jones, and Sheldon 1997). Analysis of these studies shows:

1. Variation in the objectives and sophistication of techniques used. Most studies aimed at understanding the nature and structure of hospital costs (making use of translog models). Not all studies seem to have adequately controlled for confounding variables, such as for the use of several inputs.
2. Most studies support the idea that hospitals with an average size between 200-300 beds (mostly district hospitals) seem to be operating with economies of scale, while hospitals with average size between 600-800 beds (mostly central hospitals) seem to be operating with diseconomies of scale.
3. Some studies assume Portuguese hospitals aim at minimising costs, which is unjustified given the system within which Portuguese hospitals operate. Several studies have pointed out the great variations in terms of measures of efficiency[160].
4. Most studies have described the high variability in terms of inputs and output ratios and costs of Portuguese hospitals, even among hospitals of the same administrative group and with similar characteristics. This variability creates problems in

[160] As pointed out in Chapter 2, and as described in the next sub-section some characteristics of the system are not compatible with cost minimisation assumptions, such as the use of historical budgets, the lack of accountability of hospital managers and agency problems in doctors' incentives that imply perverse incentives for doctors in the public sector.

estimating costs, inefficiencies and economies of scale and scope (as it requires multiple controls). None of the studies hitherto encountered have accounted for quality.

6.2.3.2 Country information-setting

In designing cost models, it is vital to take into consideration the characteristics of the Portuguese hospital system. This section focuses on three areas in order to characterise the hospital sector (these were explained in detail in Chapter 2): the administrative hierarchy of the hospital system, payment systems and incentives, and efficiency-related patterns.

Administrative hierarchy. Portuguese hospitals are classified in an administrative hierarchy (from central to level I hospitals, described in Chapter 3, section 3.2.9) that shows a decreasing order of technological complexity on the treatment of illnesses, and a decreasing size of catchment areas of hospital provision:

- *Central general hospitals* provide highly specialised services with advanced technology and specialist human resources.
- *Central specialised hospitals* focus on a range of specialised services. Both general and specialised hospitals tend to be located in the main urban centres.
- *District hospitals* provide a range of specialist services and are located in the district capital. In general, there is at least one district hospital in each geographic district.
- *(District) Level I hospitals* are at the bottom of the hierarchy and provide internal medicine, surgery and one or two other basic specialties only. They tend to be located in small towns.

Hospitals at the bottom of the hierarchy send patients to hospitals at the top of the hierarchy, as they do not provide all specialties treatment. There is a referral system between GPs and hospitals, and between hospitals, but there are in practice admissions outside the referral system.

Payment systems and incentives. Until recently Portuguese hospitals are public, not for profit and expected to pursue social objectives. Hospital managers have weak incentives to operate within the hospital budget constraint: hospital budgets are determined mainly by historical reimbursement and only partly by production levels. There are no penalties for systematic budget overruns. Hospital administrators operate within a highly centralised system of planning and have little autonomy in decisions on investment and human resources. There is no charging system for capital costs. Hospital administrators have little control over hospital doctors. Doctors are paid by salary and have a dual employment status, which gives them little incentive to be productive in public hospitals, as they generate income in the private sphere and by working overtime in the public. There is limited accountability and hence room for inefficiencies and problems of cost containment. Although hospitals are expected to charge private insurers or subsystems for their services, in practice these amounts are often not charged.

Efficiency-related patterns. Evidence suggests that for smaller hospitals, a lack of doctors has constrained the use of beds (the opposite probably applies in large hospitals). The ratio of nurses to doctors is low and is expected to have negative impact on productivity levels and costs. Doctors located in urban areas tend to have a lower productivity, as they also work in the private sector. There is also evidence of various factors that impact on both allocative and technical efficiency:

- High variations in the mix of inputs of doctors/nurses/beds provide evidence of variations in allocative efficiency across hospitals (for example, see the evidence given in sub-section 3.2.4.3);
- High levels of outsourcing of services with high levels of technology are observed, and outsourcing is higher for urban hospitals;
- There is no central policy with regard to the purchase of pharmaceuticals and of goods and services, which might explain variations in these costs.

There is a lack of information on the impact of quality on costs, the levels of private activity in public hospitals, and the effect of deficiencies in long-term and home care on hospital costs. Accessibility of populations to hospitals varies highly within Portugal.

The foregoing analysis suggests that in the Portuguese context, a method for adjusting for UCs should:

1. Avoid assumptions of cost minimisation;
2. Aim to capture how payment systems and hospital organisation influence the hospital cost structure;
3. Focus on how different input mixes result in allocative inefficiencies at the hospital or at the hospital group level;

4. Deal with structural differences between:

- Hospitals at different levels of the administrative hierarchy (differences in terms of size and scope);
- Input prices;
- Geographic variations.

The next sub-section presents the rationale for the choice of the particular method for estimating UCs and summarises the key characteristics of the method developed below.

6.2.4 Methodological approach

Analysis of hospital costs shows that in the case of Portugal it is essential to use an integrated approach, rather than attempting to estimate separately the various causes of UCs (such as the market forces factor in the case of England). This is for three reasons. Firstly, whilst in England it is clear that urban and rural markets for hospital human resources are different and imply variation in UCs between hospitals, in Portugal, the complex distribution of human and other resources in urban and rural areas (a lack of nurses throughout Portugal, a lack of doctors in rural hospitals, and relative excess of beds in rural hospitals) makes it difficult to assess how these impact on hospital costs. Secondly, there is a lack of disaggregated data at the district and local level, e.g. on variations of salaries and activity within the public and private sectors, both in health and in other areas. Thirdly, it seems that in order to identify UCs in the case of Portugal, it is essential to analyse allocative inefficiency, rather than focusing on input price differentials, given the central control of key resources and inadequate variations in inputs.

The methods developed in this chapter are normative[161] and follow an integrated approach. The choice for an integrated approach to model UCs involves a simultaneous treatment of input prices, inefficiencies, economies of scale and scope and other factors in the model. The objective is to build a measure of hospital UCs at the hospital level, which accounts for individual hospital characteristics, for structural differences between

[161] E.g. an explicit framework to justify the choice of methodological options is used. This is required in the context of high variability of model results due to different methods and techniques. This approach is similar to the one used by Soderland and Jacobs (Soderlund and Jacobs 2001).

hospital types, and for variations in geographical location, while also dealing with well-specified sources of allocative inefficiency. The proposed stochastic model is based on hierarchical and multilevel techniques. The proposed model:

1. Uses **the total cost per unit of output as the dependent variable**, so as to create a standardised indicator that is compatible across areas and across hospitals.
2. Focuses on **structural differences between hospitals and between hospital groups** at different levels of the administrative hierarchy.
3. Presents **two different models: the hierarchical fixed effects model and the multilevel model**. The hierarchical fixed effects model is a simpler model that uses dummies to control for the administrative classification of the hospital; is used as a benchmark for comparing with the MLM model. The MLM uses random intercepts and slopes across different levels of the network, and the purpose of these random intercepts and slopes at the hospital group is to identify the different sources of allocative inefficiency.
4. **Controls for a wide range of variables** that impact on costs: geographical area variations, hospital size, input prices, input mixes and indicators of the hospital cost structure.
5. Makes use of an ***ad-hoc*** **approach to disentangle allocative inefficiency** effects and to estimate the level of costs. By **controlling for the influence of past hospital decisions and the historical level of funding** on hospital costs, an *ad hoc* approach is more compatible with the choice of not imposing cost minimisation assumptions.

6.3 Stochastic hierarchical models

This section presents a summary of the methodological approach, the cost and hierarchical model, and then describes the development of that model into two models, referred as: HFEM for the Hierarchical Fixed Effects Model and MLM for the multilevel model with random intercepts and slopes.

Table 6.1 contains the notation in use. Given that any hospital h belongs to a hospital group c and to a geographical area k, the indices c and k are omitted in some of the variables (that is, c and k depend on h: $c(h)$ and $k(h)$). The index h is taken as the key identifier. The index l represents an alternative hospital group classification for which information on unit costs is available.

Table 6.1: Notation in use

Notation	*Interpretation*
h , h'	Hospital identifier ($h \neq h'$).
c	Types of hospital in the administrative (and hierarchical) classification (for Portugal: c = general central, specialised central, district, level I).
k	Geographical place of location.
l	Type of hospital in the costs' statistics classification (for Portugal: l = central, district, level I)
$COutput_h$	Total cost standardised by an index of hospital production. This indicator is referred to as standardised cost.
$TotCost_h$	Total cost.
$OutputIndex_h$	Equivalent patients index.
$Disch_{hl}$	Number of hospital inpatient discharges of hospital h that belongs to hospital group l .
$Outpat_{hl}$	Number of outpatient attendances of hospital h that belongs to hospital group l .
$Emerg_{hl}$	Number of emergency and accident admissions of hospital h that belongs to hospital group l .
a_l , b_l , c_l	Total unit costs from hospitals of type l , for inpatient discharges, outpatient attendances and emergency and accident admissions, respectively.
do_h	Numbers of doctors.
nu_h	Number of nurses.
be_h	Number of beds.
C , C'	Function linking the standardised cost with the covariates; and linear function linking the natural logarithm of standardised cost with the covariates.
α , β , θ	Parameters from the general hierarchical model.
$x_h^{'}$, $x_h^{''}$, x_h	Explanatory variables vector for standardised costs (x_h). $x_h^{'}$ is the sub-set of variables that have a log-linear function relationship with the dependent variable ($x_h^{'} \subset x_h$); and $x_h^{''}$ is the sub-set of variables with a semi-log function relationship with the dependent variable ($x_h^{''} \subset x_h$).
e_h	Random error for the general hierarchical model.
α_0 , α_1	Coefficients of the fixed part of the HFEM (excluding the geographical and hospital group related coefficients).
g_{hk}	Dummy variables for the geographical location of hospital h in place k (HFEM and MLM).
α_{2k}	Fixed coefficients for dummies of the geographical area k (geographical related coefficients) (HFEM).
t_{hc}	Dummy variables for the hospital h in the administrative hierarchy c (HFEM).
α_{3c}	Fixed coefficients for dummies of the administrative group c (HFEM).

e_{hck}^{HFEM}	Random error for the HFEM.
$\beta_0, \beta_1, \beta_2, \beta_3$	Coefficients of the fixed part of the cost model (excluding geographical-related and hospital group related coefficients) (MLM).
β_{4k}	Fixed coefficients for dummies of the geographical area k (geographical-related coefficients) (MLM).
β_{0c}	Random coefficient of the random intercept of the MLM, defined at the hospital administrative group c.
β_{1c}, β_{2c}	Random coefficients of the random slopes of the MLM, defined at the hospital administrative group c; β_{1c} and β_{2c} are the random coefficients of the nurses to doctors and beds to doctors ratios, respectively.
μ_{0c}	Random component of the random coefficient of the MLM, defined at the hospital administrative group c.
μ_{1c}, μ_{2c}	Random component of the random slopes of the MLM, defined at the hospital administrative group c.
e_{hck}^{MLM}	Random error at the hospital level (MLM).
$\sigma_{\mu 0}^2$, $\sigma_{\mu 1}^2$, $\sigma_{\mu 2}^2$	Variances of the random components of the model at the group level. $\sigma_{\mu 0}^2$ is the variance of the random component of the intercept, while $\sigma_{\mu 1}^2$ and $\sigma_{\mu 1}^2$ is the variance of the random component of the slopes (MLM).
σ_{e0}^2	Variances of the error term at the hospital level (MLM).
$\sigma_{\mu 0 \mu 1}$, $\sigma_{\mu 0 \mu 2}$, $\sigma_{\mu 1 \mu 2}$	Set of covariance between the random components, defined at the group level (MLM).

6.3.1 Hierarchical cost model

This subsection describes the underlying cost model and its decomposition into the hierarchical cost model.

6.3.1.1 Cost model

Given that the objective is to compare UCs between hospitals, the dependent variable is the total cost per level of measurable output (afterwards referred to as standardised cost). Measurable output is defined as inpatient discharges, outpatient attendances and emergency and accident admissions. As hospital output is multidimensional by nature, these outputs are aggregated in an output index as presented in Equation 6.1. This index weights inpatient discharges, outpatient attendances and accident and emergency admissions by the coefficients of total unit costs per hospital type (l) for each of these outputs[162]. The standardised cost is computed by the ratio presented in Equation 6.2.

$$OutputIndex_h = \sum_l \left[\frac{Disch_{hl} * a_l + Outpat_{hl} * b_l + Emerg_{hl} * c_l}{a_l + b_l + c_l} \right] \tag{6.1}$$

$$COutput_h = \frac{TotCost_h}{OuputIndex_h} \tag{6.2}$$

The variables that affect the standardised cost are[163]: price of inputs and intermediate inputs; relative mix of raw inputs and intermediate inputs; hospital size, type of hospital (administrative group) and potential economies of scale and scope; complexity of the output; quality; hospital cost structure; hospital location; inefficiency and previous levels of funding. It is not assumed that hospitals are cost minimisers: some covariates capture inefficiencies or other avoidable costs components. This *ad hoc* model does not derive from a specific assumption of hospital behaviour.

The model developed in this study differs from the SFM approach. Mainstream SFMs have assumed a positive distribution of the error term structure (error at the hospital level) that imply: covariates capture the frontier/envelope of costs; the errors represent positive deviations from that absolute frontier/envelope and are interpreted as indicators of technical inefficiency. The methods used in this study do not assume a positive distribution of the error, but a normal distribution[164]. This is due to constraints imposed by the software available to estimate multilevel models. Given the lack of data on quality, technical efficiency and other variables, the error term should not be interpreted as a full component of technical inefficiency. The model imposes a log-linear or a semi-

[162] a_l, b_l and c_l are expected to be endogenous with the levels of $Disch_{hl}$, $Outpat_{hl}$ and $Emerg_{hl}$ for each group of hospitals. Nonetheless, this procedure is found acceptable in the context, given that it is used just to standardise total costs.

[163] This list accounts indirectly for the multidimensional nature of output and for the unobserved price of hospital output and the fact that market mechanisms are very weak and there are no explicit prices.

[164] The multilevel model differs structurally from the SFM in the approach to model inefficiencies. Nonetheless, both the multilevel and SFM models share the feature of attempting to model inefficiency (while making different assumptions on the stochastic elements) and in using random effects. This justifies the option of using the SFM as a 'benchmark' for comparing the multilevel model. However, it should be acknowledged that multilevel models are straightforward cost models as they model average cost functions (average cost function as defined in (Forsund, Lovell and Schmidt 1980)).

log relationship between the standardised cost and the covariates (functional relationships as defined in (Gujarati 1995)) as shown in Equation 6.3.

$$COutput_h = \alpha * x_h^{''\beta} * e^{\theta * x_h^{'}} * e^{e_h} \tag{6.3}$$

This cost model is developed to integrate the hierarchical structures of geographical location and of administrative types of hospitals in the following sub-sections.

6.3.1.2 Hierarchical model

The hierarchical model differs from the cost model presented in Equation 6.3 as it makes the hierarchical structures explicit. Snijders describes multilevel structures (Snijders and Bosker 1999)[165]. The model takes into account the composition and the context of each hospital in the network, and uses a multilevel structure classification, in which:

1. Hospitals are level-1 units (represented by the index h). The covariates at this level capture how the individual characteristics of hospitals translate into higher hospital costs;
2. Hospitals belong to one group of the administrative hierarchy, which corresponds to a level-2 unit. The hospital groups in the administrative hierarchy are represented by index c. The covariates at the administrative group level (c) aim at capturing how structural differences between administrative groups of hospitals impact on costs;
3. Hospitals belong to one geographical area that corresponds to an alternative (and secondary) level-2 unit. The geographical areas for hospital location are represented by index k. The covariates at this level (k) capture the influence of location on hospital costs.

Equations 6.4 and 6.5 give the generic hierarchical structure and logarithmic structure of the model. The logarithmic structure follows from the assumptions made and the

[165] At this level of analysis, the terms multilevel and hierarchical models are used interchangeably. Nevertheless, in the next sections, these two terms will be differentiated.

'expected' skewed distribution of standardised costs[166]. The logarithmic structure gives a normal distribution of costs:

$$COutput_{hck} = C_{hck}(x_{hck}) * e^{e_{hck}} \tag{6.4}$$

$$\ln(COutput_{hck}) = C'_{hck}(x_{hck}) + e_{hck} \tag{6.5}$$

The model assumes that there are systematic variations between hospitals from different groups (both in terms of hospital hierarchy and location) and that hospitals within the same group are hypothesised to share a set of characteristics. Hierarchical models can be estimated using two types of models that make use of different assumptions both on the structure of the error (e_{hck}) and on the association between hospital characteristics and standardised costs[167]. These two models are developed in detail in the next sub-sections:

- **Hierarchical fixed effects model (HFEM).** The HFEM captures variations between areas and across the hierarchy of hospitals under the use of a set of fixed effects for hospital type and for geographical area. This model assumes that the residuals (e_{hck}) behave as in the assumptions of the classic model, and estimation can be done using the traditional ordinary least squares (OLS) estimation technique[168].

- **Multilevel model with random intercepts and random slopes (MLM).** In comparison with previous models that have used random effects (i.e., random intercepts) to capture allocative inefficiency, the proposed MLM aims at capturing and identifying different types of allocative inefficiency. The model uses random intercepts and random slopes to identify sources of allocative efficiency; and it controls for spatial variations through the use of dummies for the geographical

[166] The assumptions of the model were explained above: the hypothesised function that links standardised costs and the covariates follows a mixed log-linear or a semi-log function (these are common assumptions in previous literature).
[167] These two models are not exhaustive of all the meaningful models. The HEFM is built only to compare performance with the MLM.
[168] As explained below, GLM estimation with an identity function is used in the context of this study.

area[169]. The MLM makes use of assumptions on the error term (e_{hck}) that differ from the classical assumptions used in OLS regression, as they account for intra- and inter-group correlation in the error term structure.

6.3.2 Hierarchical fixed effects model (HFEM)

Besides the set of covariates that explains variations in costs (defined above and captured by the vector of covariates x_{hc}), the HFEM deploys dummies to control for hospital administrative hierarchy (captured by the terms t_{hc}'s) and for geographical variations (captured by terms g_{hk}'s). The structure of the HFEM is presented in Equation 6.6. This corresponds to a conventional model to be estimated by OLS, with controls for geographical and hierarchical variations made by the use of fixed effects. For the Portuguese context, hospital hierarchy variations are captured in the four level classification described in section 6.2.3.2: central general, central specialised, district and level I hospitals. The spatial classification used for Portugal is presented in Table C.1 in Appendix C.

$$\ln(COutput_{hck}) = \alpha_0 + \alpha_1 * x_{hc} + \sum_k (\alpha_{2k} * g_{hk}) + \sum_c (\alpha_{3c} * t_{hc}) + e_{hck}^{HFEM} \qquad (6.6)$$

6.3.3 Multilevel random intercepts and slopes model (MLM)

Multilevel models, random coefficient models and hierarchical linear models have been used interchangeably and stand for types of statistical models that handle simultaneously (within the same model) the micro-scale of observation units and the macro-scale of contexts (Duncan, Jones, and Moon 1998). The multilevel framework has been used to analyse data that fall naturally into hierarchical structures, have been used in several health and health care areas (Rice and Jones 1997), in particular to address geographical variations (Subramanian, Kawachi, and Kennedy 2001) (Malmstrom, Johansson, and Sundquist 2001) and to analyse health care provider costs

[169] An alternative for these would be the treatment of geographical variations with random effects. However, given the state of development of MLM techniques that would be incompatible with the use of random slopes for administrative groups.

and efficiency (Carey 2000). Several studies have shown the advantages of the multilevel approach over OLS estimation (Rice and Jones 1997). SFMs have used multilevel techniques to decompose the error term into two components of allocative and technical inefficiency[170]. The multilevel approach is an alternative approach to the SFM to deal with inefficiency in modelling (econometric) cost functions; and it is useful for analysing hospital systems when the following three conditions apply[171]. First, hospitals are organised into administrative hospital groups (as in an organisational hierarchy). Second, hospital costs are affected by organisation and structure (known as 'compositional' effects) and by internal factors such as local area characteristics. Third, contextual effects influence hospital activity and costs. As hospitals within the same hospital administrative group, or with the same geographical location have similar characteristics, the covariance structure of hospital costs should consider such similarities of hospitals.

The objectives of this chapter and the characteristics just described suggest the use of a multilevel model with random intercepts and random slopes (afterwards referred to MLM). The MLM deals specifically with the relationship between costs and the mix of inputs that are expected to generate allocative inefficiencies. These allocative inefficiencies are defined in the Portuguese context: the input mix of doctors, nurses and beds was identified as a cause for allocative inefficiency. The use of random slopes is expected to be particularly useful for dealing with the high variations in input mixes across hospitals within the same administrative group (wide variations for Portuguese hospitals were reported in section 6.2.3.1). The proposed model is defined in Equations 6.7 and 6.8. Equation 6.7 gives the groups of determinants of the MLM. Equation 6.8 gives the same model, making the split between deterministic and random components explicit.

Equation 6.7: the impact of some of the covariates on standardised costs depends not only on the hospital values but also on the characteristics of the administrative group to

[170] The weaknesses of this approach were described in section 6.2.3.1; in particular: arbitrary assumptions regarding the distribution of the error and omitted variable bias and strong assumptions on the independence between the two components of allocative and technical inefficiency.

[171] Multilevel models have been regarded as an important complement to econometric techniques in analysing health care provider costs and efficiency (Carey 2000). They allow for distinguishing between sources of variations (Getzen 2000), are useful for analysing separately the variances operating at each level of the multilevel hierarchy and emphasise the implications of their differences (Carey 2000).

which the hospital belongs. This model uses random slopes for two covariates –the ratios of nurses/doctors (β_{1c}) and beds/doctors (β_{2c})-, and a random intercept (β_{0c}) as components of allocative inefficiencies. The remaining covariates (x_{hck} and g_{hk}) represent the same variations as in the HFEM model.

Equation 6.8: besides the random effects at the hospital level (e_{hck}^{MLM}), there are three types of random effects that operate at the hospital group level: one random intercept and two random slopes. Equations 6.8a-c decompose the random coefficients into a deterministic and a random component. The random intercept (β_{0c}) captures systematic variations in costs between different hospital types. Previous studies have used this component to capture allocative efficiency variations. The use of random slopes in the MLM allows for decomposing the random elements at the group level that relate to the identified allocative inefficiencies in input mix.

$$\ln(COutput_{hck}) = \beta_{0c} + \beta_{1c} * \left(\frac{nu}{do}\right)_{hc} + \beta_{2c} * \left(\frac{be}{do}\right)_{hc} + \beta_3 * x_{hck} + \sum_k (\beta_{4k} * g_{hk}) + e_{hck}^{MLM} \quad (6.7)$$

$$\ln(COutput_{hck}) = \left[\beta_0 + \beta_1 * \left(\frac{nu}{do}\right)_{hc} + \beta_2 * \left(\frac{be}{do}\right)_{hc} + \beta_3 * x_{hck} + \sum_k (\beta_{4k} * g_{hk})\right] + \left[\mu_{0c} + \mu_{1c} * \left(\frac{nu}{do}\right)_{hc} + \mu_{2c} * \left(\frac{be}{do}\right)_{hc} + e_{hck}^{MLM}\right] \quad (6.8)$$

With:

$$\beta_{0c} = \beta_0 + \mu_{0c} \quad (6.8a)$$

$$\beta_{1c} = \beta_1 + \mu_{1c} \quad (6.8b)$$

$$\beta_{2c} = \beta_2 + \mu_{2c} \quad (6.8c)$$

The proposed model uses a set of assumptions: first that the distributions of the random elements follow normal distributions (Equations 6.9a-d).

$$\mu_{0c} \approx N(0, \sigma_{\mu 0}^2) \quad (6.9a)$$

$$\mu_{1c} \approx N(0, \sigma_{\mu 1}^2) \quad (6.9b)$$

$$\mu_{2c} \approx N(0, \sigma_{\mu 2}^2) \qquad (6.9c)$$

$$e_{hck} \approx N(0, \sigma_e^2) \qquad (6.9d)$$

Second, assumptions about the covariance structure (Equations 6.10a-f): covariances between the level 2 random components and the level 1 error are null (10a); covariances between random components and covariates without random slopes are also null (10e-f). Covariances between level 2 random components are estimated within the model (10b-d).

$$\text{cov}(\mu_{0c}, e_{hck}^{MLM}) = \text{cov}(\mu_{1c}, e_{hck}^{MLM}) = \text{cov}(\mu_{2c}, e_{hck}^{MLM}) = 0 \qquad (6.10a)$$

$$\text{cov}(\mu_{0c}, \mu_{1c}) = \sigma_{\mu 0 \mu 1} \qquad (6.10b)$$

$$\text{cov}(\mu_{0c}, \mu_{2c}) = \sigma_{\mu 0 \mu 2} \qquad (6.10c)$$

$$\text{cov}(\mu_{1c}, \mu_{2c}) = \sigma_{\mu 1 \mu 2} \qquad (6.10d)$$

$$\text{cov}(e_{hck}^{MLM}, x_{hck}) = \text{cov}(\mu_{ox}, x_{hck}) = \text{cov}(\mu_{1c}, x_{hck}) = \text{cov}(\mu_{2c}, x_{hck}) = 0 \qquad (6.10e)$$

$$\text{cov}(e_{hck}^{MLM}, g_{hk}) = \text{cov}(\mu_{oc}, g_{hk}) = \text{cov}(\mu_{1c}, g_{hk}) = \text{cov}(\mu_{2c}, g_{hk}) = 0 \qquad (6.10f)$$

The use of this model implies that the estimated standardised cost for any hospital is not solely based upon its own data, but also influenced by the value for other hospitals within the same group[172]. This feature is captured by the structure of variance and covariance of the model. The derived structure of variance and covariance (between two hospitals in the same administrative group) of the MLM is shown in Equations 6.11 and 6.12. These equations show how the variance and the covariance depend both on the individual and on the group values (as noted above, this deviates from the classic assumptions of an econometric model).

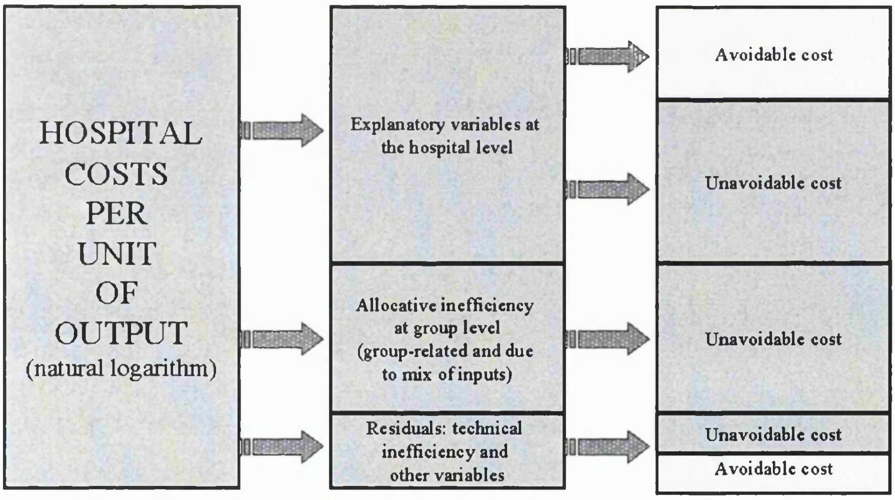
$$\text{var}[\ln(COutput_{hck})] = \sigma_{\mu 0}^2 + \left(\frac{nu}{do}\right)_{hc}^2 * \sigma_{\mu 1}^2 + \left(\frac{be}{do}\right)_{hc}^2 * \sigma_{\mu 2}^2 + 2 * \left(\frac{nu}{do}\right)_{hc} * \sigma_{\mu 0 \mu 1} + + 2 * \left(\frac{be}{do}\right)_{hc} * \sigma_{\mu 0 \mu 2} + 2 * \left(\frac{nu}{do}\right)_{hc} * \left(\frac{be}{do}\right)_{hc} * \sigma_{\mu 1 \mu 2} + \sigma_{e0}^2 \qquad (6.11)$$

[172] This is a desirable feature of the model if it is to be used in a resource allocation formula: it minimises the scope for providers' reaction in order to influence variables that impact on the estimation of UCs.

$$\operatorname{cov}ar\left[\ln(COutput_{hck}), \ln(COutput_{h'jk})\right] = \sigma^2_{\mu 0} + \left(\frac{nu}{do}\right)_{hc}\left(\frac{nu}{do}\right)_{h'c}\sigma^2_{\mu 1} + \\ + \left(\frac{be}{do}\right)_{hc} * \left(\frac{be}{do}\right)_{h'c} * \sigma^2_{\mu 2} + \left[\left(\frac{nu}{do}\right)_{hc} + \left(\frac{nu}{do}\right)_{h'c}\right] * \sigma_{\mu 1 \mu 0} + \\ + \left[\left(\frac{be}{do}\right)_{hc} + \left(\frac{be}{do}\right)_{h'c}\right] * \sigma_{\mu 2 \mu 0} + \left[\left(\frac{nu}{do}\right)_{hc} * \left(\frac{be}{do}\right)_{h'c} + \left(\frac{be}{do}\right)_{hc} * \left(\frac{nu}{do}\right)_{h'c}\right] * \sigma_{\mu 1 \mu 2} \tag{6.12}$$

The levels of standardised costs are to be adjusted in a set of avoidable and unavoidable costs, as represented in Figure 6.1. UCs are explained by variables that relate to the characteristics of hospital activity that impact on costs and that are outside the scope of management; definition of UCs depends on the empirical results of the model and is presented in the empirical section (section 6.4.3).

Figure 6.1: Decomposition between avoidable and unavoidable costs for each group of hospitals



6.4 Empirical models and results

This section describes the data, variables and sample characteristics, displays the results from the estimation of the HFEM and MLM models and presents the estimates of UCs per hospital group. It concludes with estimates of a relative index of UCs at the district level.

6.4.1 Data, variables, sample characteristics and estimation techniques

The database consists of 1998 data on: cost, expenditure and production (IGIF 2000); and an index of purchasing power at the small area level (INE, Direcção Geral do Centro, and Gabinete de Estudos Regionais 2000)[173]. The database covers 88 hospitals that until recently belonged to the NHS and were under public management status. This sample is representative of public hospitals. It only excludes psychiatric hospitals and a few number of small hospitals that are under the management of Ministries other than the Ministry of Health (for example, hospitals under management of the Ministry of Defence). Table 6.2 gives the set of independent variables at the hospital level that were included in the right hand side of the estimated models. A brief indication is given about the concept that each variable attempts to capture.

[173] Cross-sectional research was necessary due to the (limited) availability of cost data when the study was initiated. The use of panel data could highly improve the robustness of results, and is advisable for further research (data for this is now available).

Table 6.2: Variables at the hospital level

Variable	*Interpretation*
Case-mix index	Heterogeneity/complexity output and effective demand parameters.
Length of Stay	Complexity output and demand parameters.
Occupancy rate	Managerial use of beds, incentives and constraints imposed by mix of resources.
Number of doctors	Hospital size and input.
Ratio nurses to doctor	Input mix.
Ratio beds to doctor	Input mix.
Ratio other employees to doctor	Input mix.
Consumption costs as a percentage of total costs and/or consumption costs per unit of production	Intermediate input mix and intermediate input price.
Outsourcing costs as a percentage of total costs and/or outsourcing costs per unit of production	Intermediate input mix and intermediate input price.
Personnel costs as a percentage of total hospital cost and/or personnel costs per unit of production	Input mix and input price.
Other costs (apart from consumption, outsourcing or personnel costs) as a percentage of total hospital cost and/or other costs per unit of production	Input mix and input price.
Purchasing Power Index of the area where the hospital is located	Input prices.
Non-NHS revenue as a percentage of hospital revenue	Proxy for other output (work for the private sector).
Number of specialties available	Complexity of output and other hospital outputs.
Dummy for teaching activity	Other hospital outputs.
Growth in hospital expenditure in the last two years	Reflects payment systems for hospital managers, given hospital finance mostly by historical reimbursement.
Overtime payments to doctors/nurses/others, divided by the number of doctors	Reflects system of incentives for management and doctors and nurses, as well as constraints imposed by the current level and mix of resources.

The model was designed under the following assumption and using the following information:

1. Hospital output (including case complexity) is adequately captured by the output index (defined in Equation 6.1), together with the case-mix index and length of stay (covariates in the right hand side of Equations 6.6 and 6.7).
2. Classification of hospital by administrative group is as described previously: central general hospitals, central specialised hospitals (includes cancer centres), district hospitals and level I hospitals.

3. As no data was available to use as proxies for quality, technology and cost of capital, the model excludes variables that could act as proxies for these factors. There is no reliable information on quality and technology at the hospital level for Portugal; available data used to compute the cost of capital in previous studies is unreliable (Folland and Hofler 2001). Neglecting quality gives rise to the question on what is the actual output of the health care sector. Accepting the definition of medical/health care as "a process in which certain inputs or factors of production (e.g., physician services, medical instrument and equipment services, and pharmaceuticals) are combined in varying quantities, usually under a doctors' supervision, to yield an output" (Jacobs 1996), ignoring for quality of care means a failure to account for a vital dimension of care (as the output should account both for quantity and quality of care). Moreover, ignoring quality implies that estimates of the estimated models are adversely affected by a missing explanatory variable, which has consequences for econometric models described in Gujarati (Gujarati 1995). This implies an additional caution in the interpretation of the results of the model, and in particular of the residuals, as these will partly capture quality variations.
4. The number of doctors was used as a proxy for hospital size, as doctors constrain the use of other resources and are closely associated with productive capacity (Oliveira and Bevan 2001).

A summary of statistics is presented in Table C.2 in Appendix C[174].

The HFEM was estimated by GLM, using an identity link function with the natural logarithm of the standardised cost as the dependent variable. This model produces similar coefficient estimates to OLS estimation but generates statistics that are directly comparable with the results of the MLM[175]. The HFEM was estimated in the STATA

---

[174] These statistics show systematic differences between hospital administrative groups: the administrative classification clearly relates to levels of capacity, as indicated by the number of doctors; standardised costs vary across a wide range; they are higher for central and specialised hospitals and lower for level I hospitals; general and specialised hospitals have more complex case-mix and general hospitals have higher levels of LOS; occupancy rates are higher for central and general hospitals; larger hospitals are associated with a higher proportion of consumption costs in total hospital costs, and with higher levels of outsourcing per unit of output; case-mix and length of stay are correlated with cost per unit of output, and with high levels of outsourcing and consumption; the proportion of personnel costs to total costs is inversely related to hospital size; large and central hospitals are located in areas with higher purchasing power; the ratios of nurses to doctors, beds to doctors and other employees to doctors are higher for smaller hospitals.

[175] The alternative would be the use of GLM with a log link and the use of the standardised cost as the dependent variable, but results would not be comparable with outputs estimated by the MLM model.

statistical software (Stata Corporation 2001) and conventional tests for GLM models were applied[176]. The MLM model was estimated using MLWin software (Rasbach et al. 2000). The method of estimation used the restricted (or residual) maximum likelihood estimation[177]. The corresponding algorithm is the restrictive iterative generalized least squares. Hypothesis testing on single parameters and on specification was carried using the tests suggested by the literature (Snijders and Bosker 1999)[178]. The statistical comparison with regard to goodness of fit and specification between the HFEM and the MLM has made use of an adapted version of the Akaike Information Criteria, in the version suggested in the MLWin software guide (Rasbach et al. 2000)[179].

6.4.2 Results and analysis

Table 6.3 shows the estimates generated by the two models. The results need to be interpreted with caution because: a) residual variations at the hospital level might reflect the lack of controls for some variables, such as quality and non-measured outputs for the private sector carried out by public hospitals; b) control for the complexity of hospital output may not have been measured adequately[180]; and c) random intercepts and slopes might also capture systematic variations in technical inefficiency.

Estimation by GLM enables estimates to be produced of values in the original scale and of the loglikelihood of the model. The software package in use for the MLM model does not offer the possibility of carrying out GLM estimation with a log-link function. This would be the ideal estimation technique for the estimation of the MLM.

[176] The following econometric tests were applied: specification, goodness of fit, properties of residuals (including deviance) and linktest. The choice between alternative models was based on three criteria: predictive power, parsimony and expected sign of coefficients. Robust estimates of the variance of the estimators have been used (Huber-White estimates of the variance-covariance matrix).

[177] This method only differs from that of maximum likelihood estimation in the computation of variance and covariance parameters, and produces estimates with less bias (Snijders and Bosker 1999). Nevertheless, the likelihood-based tests deploy the likelihood ratio as generated by maximum likelihood estimation, as this is slightly preferable for carrying out tests based on the deviance (Snijders and Bosker 1999).

[178] These included: Wald test for hypothesis involving fixed parameters and likelihood ratio test for hypothesis involving random-effect parameters (for nested models). Residuals were checked for homoscedasticity and specification: analysis of standardised residuals (with variance equal to one); analysis of plots of standardised residuals for individual hospitals against fitted values or level 1 variables allowed for checking model specification and homoscedasticity; analysis of plots of level two residuals against fitted values or level 2 allowed for control of level two variance; comparison of residuals at level one and level two was carried out; and the model was checked for the impact of outliers.

[179] Under this version, the model with the smallest AIC should be chosen, and the AIC is equal to the sum of the loglikelihood statistic with double the number of parameters estimated in the model.

Table 6.3: Coefficient estimates of the HFEM and MLM

	A	*B*	*C*	*D*	*HFEM*
	No covariates	Control for case-mix	All covariates except geographical	All covariates	All covariates
Ln (casemix)		0.719 (0.09392)***	0.300 (0.06777)***	0.340 (0.065)***	0.264 (0.089)***
Ln (occupancy rate)			-.520 (0.08664)***	-0.500 (0.082)***	-0.380 (0.102)***
Ln (personnel costs per doctor)			0.611 (0.08446)***	0.619 (0.078)***	0.436 (0.107)***
Consumption over total costs			0.012 (0.00231)***	0.012 (0.002)***	0.013 (0.002)***
Ln (outsourcing per unit output)			0.466 (0.05907)***	0.406 (0.060)***	0.526 (0.048)***
Ln (doctors)			0.093 (0.02351)***	0.096 (0.022)***	0.070 (0.025)***
Ln (nurses per doctor)			0.251 (0.06091)***	0.233 (0.058)***	
Dummy Algarve					0.125 (0.064)*
Dummy Alentejo				0.128 (0.062)**	
Dummy interior north				0.136 (0.050)***	0.010 (0.046)**
Constant	6.067 (0.13900)***	6.020 (0.10201)***	0.112 (0.880) (*)	0.219 (0.819)(*)	1.051 (1.389)(*)
Dummy District hospital					-0.209 (0.055)***
Dummy Level I hospital					-0.239 (0.082)***
Ln (beds per doctor)			-0.124 (0.06875)**	-0.135 (0.069)**	
$\sigma^2_{\mu 0}$	0.203 (0.09394)	0.107 (0.05024)	0.0013 (0.0016)	0.0016 (0.0014)	
$\sigma^2_{\mu 2}$			0.0181 (0.0103)	0.0212 (0.0112)	
$\sigma_{\mu 0 \mu 2}$			0.0058 (0.0033)	0.0071 (0.0035)	
σ^2_{e0}	0.064 (0.01035)	0.039 (0.00634)	0.0125 (0.002)	0.0111 (0.0018)	
-2*ln(likelihood)	36.99	-8.73	-121.88	-133.13	-114.8

***- Statistically significant at 1% level.; ** -Statistically significant at 5% level; *- Statistically significant at 10% level; (*)- Not statistically significant. SEs reported in brackets.

[180] *Ad hoc* models have been criticised for the way they treat the impact of more complex cases on costs, and it has been pointed that an inadequate control for this type of biases implies an underestimation of economies of scale (McGuire and Hughes 2002).

Results from the HFEM (last column of Table 6.3) show that:

- Case-mix, outsourcing per level of output, the relative weight of consumption in the costs structure and the level of personnel costs per doctor have a positive impact on standardised costs. Occupancy rates reduce standardised costs. The coefficient for case-mix is lower than one unit, but this is not unexpected as the case-mix is correlated with outsourcing per output and consumption costs over total costs. The coefficient for occupancy rates has the expected sign: the higher the turnover, the lower the standardised cost[181].
- The number of doctors has a positive impact on standardised costs. Nevertheless, this result should be analysed together with the dummy coefficients for the district and level I hospitals, as these variables also capture hospital capacity. Analysing these dummy variables, using the central and general hospitals located in the South coast region as baseline, shows that: specialised hospitals have a similar level of costs; district hospitals have lower standardised costs in comparison to central hospitals; level I hospitals have lower standardised costs in comparison to district hospitals; hospitals located in Algarve and in the Northern interior (where there is poor access to hospital care) have comparative higher levels of costs.
- In this model, the coefficients of the variables denoting ratios of nurses to doctors and beds to doctors are found not to be statistically significant. As described below, this finding differs from the results of the MLM. The HFEM itself is statistically highly significant[182].

Results show that factors explaining standardised costs tend to reflect previous systems of finance based on historical reimbursement, for example, higher levels of consumption and outsourcing. Results indicate diseconomies of scale with hospital size[183]. This result is consistent with previous findings: there is overall agreement in studies using *ad hoc* specifications that small hospitals experience economies of scale, and large hospitals experience diseconomies of scale (McGuire and Hughes 2002).

Results from the estimation of the MLM are presented in columns A-D in Table 6.3. Columns A-C show:

[181] Nonetheless, as the occupancy rate might capture variations in quality or efficiency or economies of scale, the interpretation of this result should be treated with caution. These issues will be analysed below.

[182] The corresponding R^2 statistic from the OLS estimation is 91.5%.

[183] This finding might possibly be explained by some of the features described above, such as insufficient control for case-mix and/or no control for quality.

- Model A: when there is no control for any covariate, 76% of the random variation is explained by group variation, while the rest is explained by variation at the hospital level. This corresponds to a random intercepts multilevel model where the intra class correlation is computed as the ratio $\rho = \sigma_{u0}^2 / (\sigma_{u0}^2 + \sigma_{e0}^2) = 76\%$ (Snijders and Bosker 1999) and can be interpreted as the proportion of total variation of the dependent variable that is explained by the area level (e.g. by the administrative classification)[184].
- Model B: with an additional control for case-mix, the new level of random variations explained by group level variation is 73.3% ($\rho = 73.3\%$);
- Model C: inclusion of other covariates that are associated with hospital complexity (such as consumption over costs and outsourcing level per unit of output) implies that the relevance of the case-mix index decreases. Analysis of the coefficients of this model is similar to analysis of the HFEM (the coefficients have the same signs and similar statistical significance); the ratios bed to doctor and nurses to doctor start being statistically significant.

From model C to D, the difference is the adding of the geographical variables; the MLM shows that after controlling for internal differences between models and for variations across hospital types, two regions with low accessibility to hospital care have higher levels of costs –the Alentejo and the interior North regions. Model D is thus the most complete and final model. The results of this model are as follows:

a) When compared with the HFEM, the use of random intercepts and random slopes shows that the ratio of nurses to doctors and the ratio of beds to doctors start being statistically significant; this implies that accounting for inter and intra-group variation changes the results of the estimation;
b) A higher ratio of beds to doctors implies reduced standardised costs, while a higher ratio of nurses to doctors implies increased standardised costs; the lower costs implied by higher ratios of beds to doctors might be understood as associated with a technology less labour intensive and more intensive with regard to other inputs, which overall is characterised by lower levels of costs. The lower use of beds in rural hospitals seems to have been relevant to the availability of doctors in rural hospitals and this has implied a comparative lower level of costs for rural hospitals.

[184] This analysis can only be carried for models with random intercepts. The use of random slopes implies that the coefficient will depend on specific sample values and on the groups to be compared.

The positive coefficient for the ratio nurses to doctors might be interpreted as a result from the substitutability between nurses and doctors[185];

c) The random intercept for hospital administrative group and the random slope for the beds to doctor ratio were found to be statistically significant, while the random slope for the nurses to doctors variable was not significant. Consequently, the impact of the mix of beds to doctors on standardised costs will depend on the group of hospitals. A more detailed analysis of the results of the random coefficients is presented below.

Comparison between the HFEM and MLM on the AIC statistic has shown that the MLM outperforms the HFEM[186].

To interpret the values of the random estimates (of intercepts, slopes, and residuals at the hospital level) requires strong assumptions to be made. Random intercepts and slopes might be interpreted as capturing variations in allocative efficiency. It is more difficult to interpret residuals at the hospital level as they reflect a set of effects which the models cannot take into account. Lower and negative values of the random parameters might be interpreted as indicating higher efficiency, as they represent a negative influence on standardised costs. Figure 6.2 and Table 6.4 (and Figures C.1 and C.2, Appendix C) suggest the following:

- *Hospital level random component.* It is difficult to interpret hospital level residuals, as they include variations the model aims at controlling for: for example, variations in quality or, technical inefficiency. If this component is interpreted as technical inefficiency, then these findings suggest that the three biggest hospitals are performing well (Figure C.1, Appendix C).
- G*roup level random component.* There are two types –random intercepts and random slopes-, which capture allocative inefficiency:

---

[185] Higher standardised costs implied by higher ratios of nurses to doctors might be partly explained by the fact that nurses might partly substitute/replace doctors, but their levels of productivity when executing those functions might be lower and imply higher standardised costs (this is despite their lower wages); the lack of nurses (described in Chapter 2) means that most of them work high levels of overtime hours, for which hour payment rates are very high, which might also translate into higher standardised costs.

[186] It is worth making two additional observations on the results of the MLM model (model D). First, it has a high level of covariance between the random coefficients, which is higher than the product of the variances. This result seems striking but is expected as explained in (Snijders and Bosker 1999). Second, the deviance, computed as -2*loglikelihood, has a negative value. The negative value is explained as follows: the likelihood is a function of the probabilities and for some type of distribution the probability

a) *The random slopes component* (Figure 6.2) might be interpreted as a component of allocative inefficiency related to the ratio of beds to doctor. Table 6.3 shows that the ratio of beds to doctor has a negative impact on standardised costs; and level I and district hospitals have higher beds to doctor ratio. Nevertheless, the negative impact of the beds to doctors ratio on standardised costs is lower (more negative) for district and level I hospitals and higher (although negative) for central hospitals.

b) *The group level random intercepts component* (Table 6.4) captures systematic variations between hospital administrative groups and can be interpreted as unaccounted variations across groups of hospitals, i.e. the remaining allocative inefficiencies after the model has controlled for group variations on costs implied by the beds to doctor ratio. General hospitals are the most inefficient and are followed by specialised hospitals, while level I hospitals are the most efficient (in terms of allocative efficiency).

- Average estimates of allocative inefficiency of the group level and for the hospital level residuals are shown in Table 6.4 (though these are very approximate). General and specialised hospitals are the most inefficient; district hospitals have the lowest residuals at the hospital level while specialised hospitals have the highest positive residuals. The highest (and undesirable) impact of the beds to doctor ratio on costs is found in specialised hospitals, while the highest random intercepts are observed in general hospitals.

density function may be greater than one (when the dependent variable is continuous), and thus the loglikelihood can be positive.

Figure 6.2: Allocative inefficiency –random slope component of the random coefficients for the ratio beds to doctor

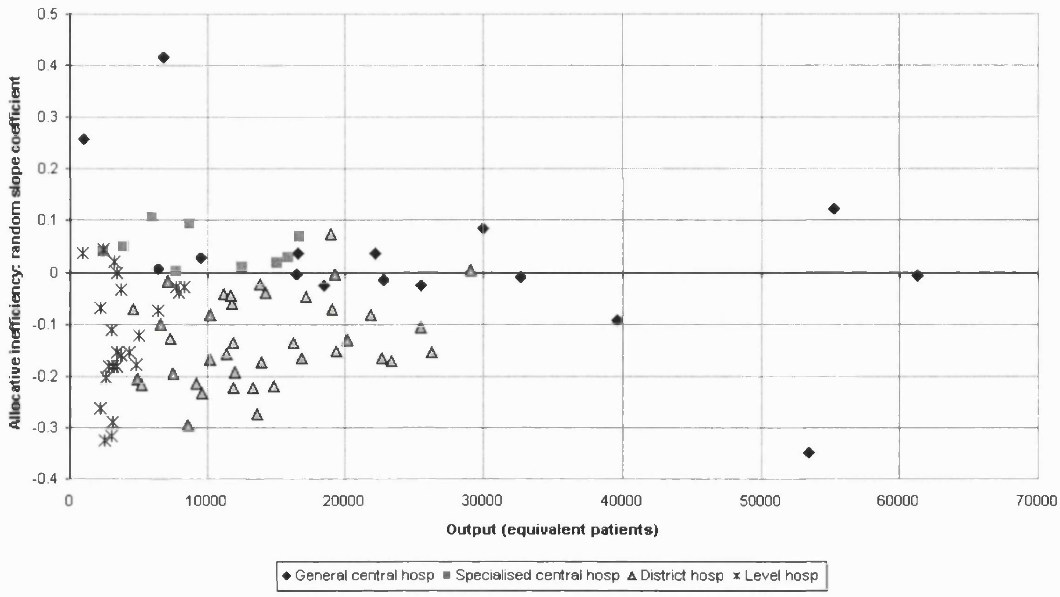


Table 6.4: Average of allocative inefficiency estimates and hospital level residuals variations at the group level[187]

	Average hospital level residual	*Average random intercept*	*Average random slope*
General hospitals	0.024	0.037	0.029
Specialised hospitals	0.032	0.019	0.046
District hospitals	-0.012	-0.025	-0.129
Level I hospitals	0.008	-0.042	-0.125

6.4.3 Estimates of unavoidable costs and geographical redistribution

This section defines UCs and presents empirical results on the levels of UCs per hospital. The impact of adjusting for UCs at the district level is presented in Chapter 8.

[187] Table 6.4 contains the simple average of inefficiency coefficients of the MLM model (random slopes, random intercepts and hospital level residuals) for each hospital administrative group. This is a crude summary of indicators. For a better interpretation, these indicators should be analysed together with Figure C.1 of Appendix C.

6.4.3.1 Definition

Literature has not provided clear rules as to which variables should be seen as avoidable or unavoidable determinants of hospital costs. The criteria for defining the components of UCs of Portuguese hospitals in this study are[188]:

- Costs that lie outside the control of hospital management or that represent short run constraints. This applies to the additional costs arising from the mix of nurses and beds in relation to doctors, and the other component of allocative inefficiency, which corresponds to the random intercepts;
- Costs that would not generate perverse incentives if hospitals were to be reimbursed for them. Analysis of results of the empirical model has shown that some of the variables that explain hospital costs relate to the current structure of incentives and to the current financing system. For example, levels of purchase (of goods and services) and outsourcing are very significant in explaining costs. If all the costs of consumption, outsourcing and personnel costs were classified as UC and if the formula was used to allocate resources, this would create perverse incentives for hospitals to increase those components of expenditure.

Consequently, model D needs to be adjusted to estimate UCs. The components of the model that are classified as UCs are:

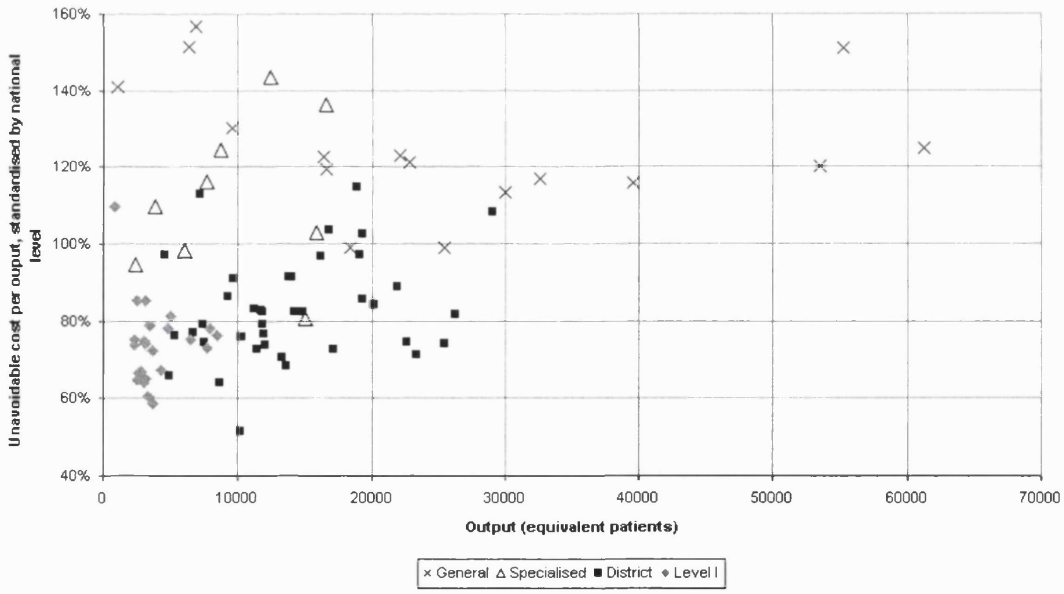
1. Allocative inefficiencies across hospital groups (that is the impact of both random intercepts and random slopes on costs);
2. Geographical variations in costs;
3. Lowest average at the group level (central, specialised, district and level I) for the variables consumption costs over total costs, outsourcing levels per unit of output, and personnel costs per doctor;
4. For all the other deterministic covariates, 100% of their value is considered UC.

It is assumed that the random component of the hospital level fully represents hospital technical inefficiency, and thus an avoidable cost. Nonetheless, this is a strong assumption, as part of the random component reflects factors for which the model has not directly controlled (such as quality).

6.4.3.2 Empirical results on unavoidable costs at the hospital level

The application of these rules to UC suggests that 78% of the national costs are unavoidable (in 1998). This means that 22% of the national cost per unit of output (or national total cost) might be explained by inefficiencies in the system. Figure 6.3 contains the UC per output at the hospital level, standardised by the national level of UC per output. It shows that general and specialised hospitals would have increased shares due to the redistribution of UCs in comparison to the national average, while level I hospitals would have decreased shares. This is clarified in Table 6.5: on average, general and central hospitals have standardised costs 23% above the national average, while level I hospitals are 26% below the national average, which implies significant redistribution between these types of hospital.

Figure 6.3: Individual hospital 'winners' and 'losers'



[188] This classification of unavoidable costs assumes that managers have little freedom to change the mix of inputs and a lot of freedom to choose levels of outsourcing and levels of overtime payments to doctors.

Table 6.5: UC per output, as a percentage of national UC per output

	Unavoidable costs per output
General hospitals	123%
Specialised hospitals	114%
District hospitals	85%
Level I hospitals	74%

An UC index for the district level weights the hospital unavoidable standardised costs for each hospital of the district by the hospital size. This index is analysed in Chapter 8.

6.5 Discussion and further research

This study has proposed a new method to estimate UCs, making use of an integrated approach and attempting to disentangle sources of allocative inefficiency. The study has produced evidence of the causes of inefficiency and the determinants of hospital expenditure, and could be used to inform policies that pursue equity and efficiency. The multilevel model with random intercepts and slopes has advantages over other methods. UCs were computed so as to avoid perverse incentives.

Results have suggested that additional costs (78% of national costs are accounted as unavoidable) are generated by the lack of flexibility of hospital managers and current incentives (such as financing based on retrospective reimbursement) and are not promoting equity in the system.

It is necessary to change the distribution of hospital resources, such as beds and doctors, to correct geographical inequities in the system and to improve allocative inefficiency. It was shown that after controlling for all the other factors, higher costs were found for smaller hospitals.

Further work could:

- Use a multilevel model with the assumption of a positive distribution of hospital level residuals, to compute the efficient frontier/envelope and develop a model using the microeconomic theory of the firm;

- Focus on the relationship between occupancy rates, staffed beds, and reservation quality and demand (reservation quality and demand relate to the need to maintain a component of spare capacity in order to answer to shocks in demand of hospital care). For example, it is expected that occupancy rates are inversely related to inefficiency (Zuckerman, Hadley, and Iezonni 1994), occupancy rates might be seen as elements of economies of scale but this is difficult to capture (Scott and Parkin 1995) and might also be related to unpredictable demand, as large hospitals benefit from lower reserve margin requirements (Aletras, Jones, and Sheldon 1997).
- Measure the cost of capital, so as building an indicator appropriate for the Portuguese context.

6.6 Concluding remarks

The approach based on a multilevel model for the estimation of UCs was developed as part of a larger study of policies to correct inequities. The multilevel model was designed to deal with systematic variations in costs across administrative groups of hospitals, geographical variations of hospital costs, and the decomposition of allocative inefficiencies (distinguishing between effects of human and capital resources).

The results of this exploratory study provide further evidence in the controversy over the existence of economies of scale and scope. Our findings indicate that policies that seek improvements in equity cannot ignore variations in UCs and incentives, and that any estimate of UCs should look into allocative inefficiencies. The planning system and methods for resource allocation should account for these findings. The estimates of UCs at the district level show that districts with concentration of the biggest hospitals and the highest levels of supply (Coimbra, Lisbon and Porto) will have a higher level of relative UCs. These results need to be taken into account in the calculation of inequities in finance. These estimates of UCs will be analysed in Chapter 8, together with the other estimated components of the capitation formula generated in Chapter 4.

7 CHAPTER 7 – A flow demand model to estimate and predict hospital utilisation

7.1 Introduction

Policies to improve geographic equity for hospital care require information on the impact of the distribution of hospital supply on utilisation. Predictive models to provide such information must take account of behaviour of patients (Folland 1983) and of the characteristics of the current system (Rushton 1987). Previous models that have tried to predict the impact of changes in supply on hospital utilisation have been inadequate in modelling the process of demand for hospital care, the impact of the hospital and health care systems on hospital utilisation, and the interaction between utilisation of different hospitals.

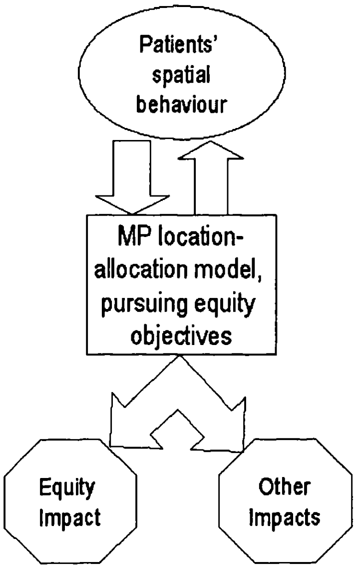
This chapter proposes a new model of demand for hospital care that predicts hospital utilisation at the small area level. The chapter has two objectives, to:

1. Generate information to compute the cross-boundary flows adjustment (I_{4r} for district r) defined in Chapter 4; and
2. Build a model to predict the impact of changes in supply on utilisation to be used in Chapter 9.

The adjustment for CBFs is justified because as Chapter 3 has shown, the concentration of hospital services in some areas means that hospitals' catchment populations are quite different from district resident populations. The computation of inequities in finance and utilisation requires estimates of CBFs (as shown in Chapter 4). CBFs also provide information on differential accessibility of populations located in different districts, as well as useful information for contracting agencies.

In Chapter 9, the predictive model is used as an input to a mathematical programming model, in the form of a constraint that captures patients' spatial behaviour in the use of hospital care. The mathematical programming model is an optimising location-allocation model intended to improve geographic equity in the system by redistributing hospital supply and to be evaluated on a set of impacts, as shown in Figure 7.1. A detailed explanation of the structure and rationale of this mathematical programming model is presented in Chapter 9.

Figure 7.1: Patients' behaviour information, as an input to a location-allocation model



The model of demand for hospital care uses the small area level, as this is considered the most appropriate geographic level for analysis (Alexander et al. 1999) and allows detailed predictions.

Modelling hospital utilisation has to take account of the characteristics of the health care system. This model is developed in the context of the Portuguese system, which is similar to those of many other countries. The key features of the Portuguese system influencing the methodological design are (characteristics described in Chapter 2): a NHS with central planning of facilities; the dominant role of public hospital provision; the fact that access to NHS hospital care is nearly free at the point of use. Other important factors that influence hospital utilisation are patients' and doctors' preferences, gatekeeping by GPs and the hierarchical structure of the hospital system. These factors are briefly outlined below.

Primary care centres are intended to provide the principal access point for patients to the health care system and thus act as gatekeepers. Patients are then free to choose a hospital doctor (after GP consultation) from a published list, although this choice may be limited. Some occupational groups have the privilege of avoiding gatekeeping, and can directly enter public hospitals, as a result of previous access to the private sector. Moreover, evidence shows that the Portuguese have been using accident and emergencies services, in order to gain access to hospital care rather than obtaining a referral via the primary care sector.

Public hospitals can be grouped into four different levels: central hospitals, specialised hospitals, general district hospitals and district level I hospitals. As observed above, central and specialised hospitals provide highly specialised services with advanced technology and specialist staff, and are mainly located in three urban centres, Lisboa, Porto and Coimbra. District hospitals provide a range of specialist services, and are located in the administrative capitals of each district. The most basic institutions, district level I hospitals provide internal medicine, surgery and one or two other basic specialties. This administrative division (and the distribution of supply) means that the hospital system is divided into three hospital subsystems, in the north, centre and south of the country, with Porto, Coimbra and Lisboa as their central supply points (being the only providers of a full set of hospital services). This hierarchical structure also relates to hospital size, as shown in Chapter 6.

Geographic utilisation in Portugal is expected to reflect the institutional structure of hospital supply (the availability and characteristics of hospitals), patients' preferences and needs, GPs' preferences, primary care provision, and other costs involved in accessing the system. The methods developed in this chapter aimed at modelling access to hospital care, so as to account for these influences on utilisation.

The study has four sections, which: a) Review models of hospital utilisation and summarise the methodology adopted; b) Develop a model of hospital utilisation and its econometric formulation; c) Apply the model to the Portuguese hospital system; d) Summarise the main conclusions.

7.2 Modelling hospital utilisation

This section reviews the diverse literature on utilisation (and CBFs), summarising some of the weaknesses and empirical findings of previous studies, and provides an outline of the approach adopted in modelling hospital utilisation.

7.2.1 Literature review

Most of the literature on CBFs at the region/district level is English and dates from the 1980s. With the development of the internal market in England and the purchaser-provider split, the financial flows allocated to purchasers started being based on administrative populations and hence there was no need for information on CBFs (Hutchison et al. 1999). England has used data on real flows of patients. CBFs have been studied in very few countries (Hutchison et al. 1999). Gravity models have been the available tool to estimate CBFs and these models are described in section 7.2.1.2. Nonetheless, the extensive literature on utilisation informs the modelling of CBFs.

Studies focusing on the determinants of utilisation are diverse but few aim at developing predictive models. Methods for interpreting hospital utilisation include: a measure of *observed demand* (Berki 1972) (Bond et al. 2000); a proxy for *need* of hospital care (Kirkup and Forster 1990) (Carr-Hill, Hardman et al. 1994); a measure of *accessibility* and an accessibility cost implied by hospital *concentration* (NHS Centre for Reviews and Dissemination 1997); and an indicator of the *effectiveness* of primary care in small areas (related to hospital utilisation for ambulatory care) (Ricketts et al. 2001). This variety of definitions illustrates that utilisation, demand and accessibility are complex inter-related concepts, and therefore should not be analysed separately. One way of linking these concepts is to argue that utilisation represents demand filtered by accessibility, and reflecting population characteristics and medical practices (need and supply, respectively) (NHS Centre for Reviews and Dissemination 1997). The following brief literature review intends to demonstrate how these concepts are related. It discusses theoretical models, empirical findings and techniques in use, as well as providing a summary of weaknesses arising from the various studies.

7.2.1.1 Theoretical models

Several conceptual frameworks have been used to model access to health care and utilisation. Most of these models have a behavioural basis[189].

The first theoretical behavioural model appeared in the late 1960s. Its objectives were to understand why families use health services, to define and measure equitable access to health care, and to assist in developing policies to promote equitable access (Andersen 1995). Two principle conceptual frameworks of the determinants of utilisation have been put forward, by Andersen and Newman (Andersen and Newman 1973) and almost twenty years later, by Evans and Stoddart (Evans and Stoddart 1990). Andersen and Newman developed the initial analytical framework, widely referred to by later work. They modelled utilisation as a behavioural process influenced by three sets of determinants: individual characteristics, enabling resources in the environment (including characteristics of the health service delivery system), and need (both perceived and professionally evaluated)[190]. They focused on how these three sets of elements interact, as utilisation can be seen to result from the interaction of individual characteristics and the societal environment. Societal determinants affect individual determinants both directly and also through the health system.

Evans and Stoddart (Evans and Stoddart 1990) modified the behavioural model to capture the dynamic and recursive nature of health services. They introduced health status outcomes (including satisfaction), as well as feedback loops showing that outcomes affect subsequent individual responses in terms of predisposing factors, perceived need for services and health behaviour. They also emphasised the powerful influence of forces outside the health care system on demand and utilisation.

In another branch of literature, a set of theoretical models was developed focusing on how individuals take decisions on consuming health care to produce health. In this approach, health care competes for resources with other activities valued by the

---

[189] Variations in these models are the result of a focus on individual vs. society factors, one-time vs. temporal factors, health status vs. health care factors, demand vs. supply factors, etc.

[190] Individual characteristics include predisposition and the level of perceived and evaluated illness; enabling resources in the environment, include the level and distribution of resources and access to and structure of the health care system; perceived and professionally evaluated need includes use of medical technologies and the social norms relating to the definition and treatment of illness.

individual (Grossman 1972) (Wagstaff 1993). The objective of this study was to focus on utilisation at the small area level[191], where organisational and institutional features of providers are expected to have a greater influence on aggregate utilisation than decisions by individuals (Alexander et al. 1999).

The models described above are from health economics and health policy literature. Literature from geography and from operational research, on the other hand, is based on the deployment of spatial interaction models (SIMs) (most commonly in the form of gravity models) to describe hospital utilisation flows between population points and hospital sites[192]. Population points refer to well-defined small geographic areas used as population catchments[193]. In their standard format gravity models use the following set of data to describe hospital flows: population numbers per area, hospital supply per hospital site, distance between hospitals and population areas, and a decay function (capturing the relationship between distance and utilisation, which might depend on the hospital). Gravity models do not account for the process of demand for hospital care. They are successful in describing the system (Cho 1998), but inadequate for predicting user flows (Porell and Adams 1995). Despite this, they have been used for that purpose, too (Hallefjord and Jornsten 1984) (Mayhew, Gibberd, and Hall 1986) (Taket 1989) (Brown 2001). SIMs model interaction unsatisfactorily, assuming that when there are changes in one hospital, all other hospitals gain in proportion to their shares of utilisation prior to that change[194]. Empirical evidence in the context of predicting utilisation behaviour shows that this assumption is false and that there is "an inherent instability in the spatial choice rule represented in the model" (McLafferty 1988). In addition, gravity models operate at the aggregate level, they do not take local health systems variations into account, nor the hierarchical and organisational structure of hospitals (e.g. considering whether there are tertiary referral hospitals within a population catchment area).

---

[191] Small area studies assess the relationships at an aggregate level but are based on individual circumstances and decisions, for example patients and physicians. Nevertheless, population-based evidence suggests that utilisation differences at the small area level have relatively less to do with patients and/or physicians, but rather with hospital market characteristics and the availability of alternative delivery (Alexander et al. 1999). In addition, the ecological fallacy applies (Folland and Stano 1990).

[192] In this stream of literature, utilisation flows are treated as "trips".

[193] Spatial models commonly assume that the population of a small area unit is located in the centroid point of the geographic area.

[194] Nonetheless, there is a scope for improvement of the SIM models in the health care context. For example, unconstrained gravity models have been used and developed in other areas of literature and could possibly overcome some of the weaknesses of previous studies.

The choice of theoretical framework (and the design of applied research) crucially depends on the unit of analysis of utilisation chosen (Andersen and Newman 1973). The most common unit of analysis has been a measure of hospital utilisation aggregated to the hospital level or to the population area level. Few studies have investigated the probability of a population making use of a specific hospital site. Few studies on SIMs in health care have modelled patient trips to hospital points.

7.2.1.2 Empirical evidence and techniques in use

This sub-section sets out the principal methodological options that have been applied in utilisation studies and provides some empirical findings. A variety of utilisation variables and measures have been used as dependent variables, such as utilisation per population area (Carr-Hill, Sheldon et al. 1994) or hospital utilisation per "hospital market area" (Alexander et al. 1999). Few studies using SIMs deploy utilisation flows between population points and hospitals as the dependent variable. Some studies have modelled the probabilities of use of different hospitals (Anas 1983)[195], or hospital market shares as measured by utilisation (Folland 1983). The most common utilisation measures have been: admissions and length of stay (Folland 1983) (Kirkup and Forster 1990) (Alexander et al. 1999), hospital days per person per year (Long 1981), crude admission rates (Kirkup and Forster 1990) (Folland 1983) and standardised admission rates (Black, Langham, and Petticrew 1995). Rohrer has shown that the model structures (including explanatory variables) and the results might differ according to whether the unit of analysis is utilisation per population area or utilisation per hospital catchment population (these units of analysis have been used as the dependent variables in econometric applications) (Rohrer 1990).

As hypothesised, empirical evidence has consistently shown that differences in utilisation may result from variations in demand (e.g. morbidity or expectations), variations in supply (e.g. availability of facilities and physician judgement) and also statistical elements (e.g. data errors or random variation) (Black, Langham, and Petticrew 1995). However, a key problem is that many studies explaining utilisation

[195] Studies focusing on probabilities have been developed mainly in the US literature.

have not adequately controlled for such confounders (Folland and Stano 1990) (NHS Centre for Reviews and Dissemination 1997). For example, supply factors have been often neglected (such as in (Long 1981), (Hanlon et al. 1998) and (Bond et al. 2000)). Neglecting key confounders in an econometric model means that econometric estimates will be biased and unreliable (Gujarati 1995). A summary of the key findings from previous utilisation studies using the small area level of analysis follows. This summary highlights problems in controlling for confounders, and also describes weaknesses characterising these studies.

Population numbers, demographic characteristics and socio-economic differences between areas are key variables that influence demand for health and hospital care for a population or hospital point. Demographic and health status characteristics are associated with place of discharge (Bond et al. 2000). Utilisation increases with population numbers (Porell and Adams 1995). Some key personal factors increase the probability of hospital admission (Hanlon et al. 1998): smoking, weight and high blood pressure. Neighbourhoods considered pockets of poverty were found to have a positive impact on hospital utilisation, both in admissions and re-admissions (Glazier et al. 2000)[196]. Material deprivation also affects utilisation (Kirkup and Forster 1990) (Carr-Hill, Hardman et al. 1994); and the case-mix seems to influence hospital market share (Folland 1983) (Porell and Adams 1995). Alexander and colleagues (Alexander et al. 1999) found that the socio-economic context is the chief determinant of small area variations in utilisation.

Service characteristics (such as hospital characteristics or size and physician availability) are important determinants of utilisation rates, place of discharge (Bond et al. 2000)[197] and market share (Folland 1983). Increased supply has been shown to increase utilisation (McLafferty 1988) (Kirkup and Forster 1990) (Carr-Hill, Hardman et al. 1994) (Black, Langham, and Petticrew 1995)[198]. Furthermore high availability of physicians also increases hospital share (Folland 1983) and utilisation rates (Kirkup and Forster 1990) (Black, Langham, and Petticrew 1995) (Bond et al. 2000), while physician affiliations and service mix also affect utilisation (McLafferty 1988). The

[196] There was inadequate control for confounders in this study.
[197] Findings for acute stroke and for hip fracture.
[198] This is in accordance with Roemer's Law, which postulates that areas with more hospital beds per capita have more hospital utilisation per capita (Rohrer 1990).

importance for hospital utilisation of the provision of alternatives to inpatient care (e.g. outpatient care) on hospital utilisation depends on the existing nature of the illness (Alexander et al. 1999).

There is mixed evidence on the effect of distance to hospitals on utilisation (NHS Centre for Reviews and Dissemination 1997). This is dependent on the health care sector, tending to be smaller for curative care in comparison with prevention programmes, for specialist compared to generalist care, hospital versus primary care, and severe versus mild illness (NHS Centre for Reviews and Dissemination 1997). There is conflicting evidence on the trade-off between hospital concentration (related to increased distance to access hospital care) and utilisation for inpatient services. Most studies, however, present some evidence of a distance-decay effect, with evidence stronger in the UK than in the US (Porell and Adams 1995) (NHS Centre for Reviews and Dissemination 1997).

Although other factors seem to influence hospital utilisation, there is only limited empirical evidence. Use is affected by the quality and reputation of health care units, and the perceived threat of illness can operate as "propulsion" for utilisation (Lynch, Edington, and Johnson 1996) (NHS Centre for Reviews and Dissemination 1997). Hospital roles in the hierarchy are key variables in explaining hospital utilisation (Porell and Adams 1995). Social support networks are associated with discharge location (Bond et al. 2000). Some evidence also indicates that hospital use might often be inappropriate (Folland and Stano 1990). Determinants of utilisation chosen in the literature depend on the local health care system: e.g. while income, insurance and the concept of trade areas have been used in the US (Folland 1983), there has been greater attention given to need, supply and organisational issues in the four UK countries.

Some studies have drawn attention to multicollinearity between demand, need, or supply variables (Long 1981) (Folland and Stano 1990) (Bond et al. 2000). Many of these variables are interconnected and there is evidence that local supply and demand conditions interact and influence utilisation[199]. Although this implies that some explanatory variables might interact, this has not been given much attention in the literature. Some examples of expected interactions could include private supply being

[199] This is supported by evidence from SIMs, which include interactions between supply and demand and are appropriate for explaining flows.

associated with more densely populated areas and higher socio-economic classes, while hospital size tends to be strongly correlated with the number of specialties available (McLafferty 1988)[200].

This review has shown that any utilisation model ought to explore how the determinants need, supply, health care system and the interactions between these determinants explain and predict hospital utilisation. The relative importance of each type of factor remains unclear.

The techniques used in utilisation studies have been dominated by econometric and statistical modelling. These techniques are adequate for decomposing effects, testing hypotheses (parametric tools), describing relationships between variables, making predictions, and are common tools in health economics and health policy literature. The choice of technique should depend on the objectives of the study, data available and unit of analysis[201]. An alternative approach is to use simulation and system dynamic techniques. These techniques have not been used in hospitals literature at the national level, but they have been used to model flows within and between small groups of health care units (Ridge et al. 1998) (Cote 1999) (El-Darzi et al. 2000). Simulation and system dynamic techniques have great potential for predicting the effect of change in the system and offer flexibility in modelling, since they require neither the restrictive assumptions nor the simplifying generalisations commonly found in more analytic approaches (Cote 1999). Unfortunately, most information systems do not produce the data required to build these models[202].

7.2.1.3 Weaknesses of previous studies

Earlier studies attempting to make predictions had a number of weaknesses.

---

[200] Another example of interaction effects: distance in terms of elapsed time before receiving health care might be expected in some cases to influence outcome and need, and the effect of two accessibility barriers –perceived importance of need and perceived ability of a service to produce results- is greatest for the most disadvantaged (NHS Centre for Reviews and Dissemination 1997).

[201] Some of the statistical and econometric techniques have been: multifactor ANOVA with survey statistics (Folland 1983) and with district level data (Black, Langham, and Petticrew 1995), logistic regression to predict place of discharge with cohort data (Bond et al. 2000), multiattribute conditional logit model using survey data (Folland 1983), Poisson regression models with DRG-based data (Alexander et al. 1999).

[202] These modelling techniques have a high demand of data for estimating parameters.

Firstly, these models of utilisation have not accounted properly for the characteristics of hospitals and health care systems (NHS Centre for Reviews and Dissemination 1997); for example:

- Most models have not captured the spatial interactions between hospital capacity and the utilisation levels of other hospitals[203]. SIMs/gravity models are an exception, as they have managed to model interaction but in a way inadequate for predicting radical changes[204];
- Few studies have considered hospital systems in the wider context of the health care system (e.g. accounting for the interface between hospital and primary care, between public and private hospital supply);
- Most studies have not put enough emphasis on the role of organisational and institutional factors (Andersen 1995).

Secondly, econometric studies that have modelled utilisation using population area as the unit of analysis have often failed to control for simultaneity between supply and demand (Carr-Hill, Hardman et al. 1994) and between supply variables (Folland and Stano 1990)[205]. Failing to control for simultaneity leads to biased estimates (Carr-Hill, Hardman et al. 1994).

Thirdly, utilisation data typically has a positively skewed distribution (Blough, Madden, and Hornbrook 1999)[206]. Most studies, however, have not taken this problem into consideration, and hence have not controlled for this effect[207]. Estimates that are

[203] Traditional gravity models respect the property of independence of irrelevant axioms, i.e. that the flows to any destination are independent of other destinations (Congdon 2001).

[204] A recent study from Congdon (Congdon 2001) has improved the interaction mechanism of gravity models in the context of modelling emergency flows. He has adapted a gravity model to make it more responsive to changes in the patterns of supply, by using Bayesian methods to re-estimate some of the parameters of the model (in order to represent new accessibility scores given supply changes), and afterwards re-running the model with the new parameters. This approach, however, requires local knowledge to specify the new parameters for changes in supply and is thus difficult to use for other than small local studies.

[205] For example, hospitals may have been planned either to match need, as in the case of the English NHS or in co-ordination with primary and tertiary care supply.

[206] Utilisation indicators (such as number of medical visits, medical expenses, etc) have a skewed and mixed distribution and present the following characteristics (Blough, Madden, and Hornbrook 1999): flows are non-negative, there is a high proportion of zero flows, positively skewed empirical distribution of the non-zero flows (corresponding to a heavy-tailed distribution), and non-constant variance.

[207] An exception to this was the study carried out by (Long 1981) that recognised the greatest utilisation frequency appearing at the extreme left, at zero, which implied the use of econometric techniques to account for this (so as to stabilise the variance, as will be analysed later).

generated without using methods that tackle the skewed distribution of utilisation lead to biased estimates (Duan et al. 1983).

Finally, there is no conclusive evidence about the relationship between concentration and utilisation. This requires a complete analysis of the hierarchical organisation of hospital systems, the concentration of supply in central hospitals, and how these characteristics impact on utilisation and accessibility.

7.2.2 Methodological approach: Overview

The following sub-section presents a brief overview of the methodological approach used to build a predictive model for hospital utilisation at the small area level. The suggested model is referred to as the *flow demand model -FDM*[208].

As observed above, the choice of a theoretical framework depends on the choice of unit of analysis because this entails choices between different representations of space. In this study, utilisation is taken as a flow variable; and flows refer to patient movements to access hospital care.

The expression 'flow demand model' focuses on flows as the unit of analysis and on the process of demand for hospital care. It is called a 'demand model', because utilisation is taken as a measure of observed demand. Demand is considered at an aggregate level, but is seen as resulting from the sum of individual demands, in which patients maximise their utility subject to accessibility costs, levels of perceived need, size/configuration of health care supply, etc. The FDM for hospital care also examines the wider context of the health care system[209] and organisational and institutional factors in the hospital system using routinely collected national level data.

The approach to modelling the patient flow to hospitals adopted in this study is different from the conventional method reported in the health economics and health policy

[208] The rationale for the use of this term is explained below.
[209] This includes the hospital sector and other variables from the health care system that influence hospital utilisation, as well as examining the role of income, socio-economic, private care, primary care and social care indicators in the hospital sector.

literature, which treats utilisation as a stock variable. The conventional method is based on a spatial representation of a system in which populations are taken to be concentrated in demand points for hospital care. The FDM represents a system that considers both population points and hospital points, and demand is regarded as a concept that relates each population point to each supply point. This approach differs from the previous approach used in the geographic literature based on SIMs, which have aimed at describing patient flows between areas without an adequate model of health care demand. In their most common formulation, gravity models depict flows using data on population numbers, hospital supply levels, distance between points and a decay function (which represents the relationship between utilisation levels and distance, for different types of hospitals). Since these models do not consider how other components of the health care system impact on hospital utilisation, they are unsatisfactory. Furthermore, as described earlier, they treat demand for alternative hospitals inadequately. The FDM uses the small area level, as it is expected to improve accuracy of estimates and predictions.

The next two sub-sections describe two key features of the FDM, namely a combination of the concepts of utilisation flow and an index of alternative hospital supply, as well as the rationale behind the choice of an econometric model as the estimation technique.

7.2.2.1 Flows and alternative hospital index

Although the conventional approach has been to model utilisation as a stock variable, the flow demand concept used in the model developed here has in fact been available for some time (in literature on network analysis). Hodgson et al. define "flow demand" as a stream between origin and destinations, inside a system that interacts as a whole, and where the flow terminates in the destination sites (Hodgson, Rosing, and Storrier 1996). This is a more adequate way of representing hospital system activity. Utilisation flows constitute the dependent variable whereas a specific independent variable represents the alternative supply to that population point, so that flows to a hospital site interact with the level of supply available from alternative hospital sites. That covariate assumes the form of an index. The combined use of an alternative supply index for an area as an independent variable and of utilisation as a flow dependent variable allows for interactivity in the flows utilisation matrix, and thus overcomes previous problems of capturing interaction, which is important in the context of prediction. For prediction and simulation purposes, when there is an increase in capacity in a hospital, utilisation flows for that hospital are expected to increase while utilisation flows for alternative hospitals in the area are expected to decrease. An alternative hospital index is defined in relation to a population point and a hospital location and it measures the ratio between the alternative hospital supply for that hospital site within the population point and the population competing for the use of that hospital site (competing population located nearby that population point and using that hospital site).

7.2.2.2 *Econometric modelling*

Econometric modelling is an appropriate method for dealing with the features of flow data and for estimating the FDM because it captures a stochastic component from small area variations[210], and deals with imperfections of past data[211]. The FDM operates at a smaller geographical scale than conventional studies, which means that it is less vulnerable to the problem of ecological fallacy and increases the number of observations for the econometric estimation. The covariates of the FDM include population- and supply-related variables. Consequently, the econometric model captures a set of fixed effects on the population and at the hospital level. Previous 'stock' demand models had to control for simultaneity in conjoint determination between supply, need and demand, which introduced complexity and lack of transparency (for policy analysis) into the methods. Using utilisation flows as the dependent variable minimises problems of simultaneity in estimation, as the scale is much smaller than if measures of demand and supply were used[212].

The econometric regression isolates the impact of hospital supply factors on utilisation flows, after controlling for variations in other factors, as shown in Equation 7.1, and can

[210] Random variations may be interpreted as local variations in patterns of health care delivery, different clinical judgements of access to primary care, hospital care and other health related services, and other variables not controlled in the econometric model. Nevertheless, they are not seen as being of critical importance for analysing aggregate flows.
[211] Applied to Portuguese hospitals, incomplete DRG forms (for some hospitals) implies that the utilisation matrix is incomplete; and for small areas, there is a high variation due to small numbers.
[212] For example, utilisation flows relate to population and to hospital points, hospital supply relates to hospital points, need for hospital care relates to population points, access to primary care relates to population points.

be used to predict utilisation when one changes hospital supply. All the notation used in this chapter that adds to previous notation is defined in Table 7.1.

$$U_{ij} = f(D_j, \widetilde{D}_{ij}, other'_{ij}, other''_i, other'''_j) \tag{7.1}$$

Table 7.1: Notation in use

Notation	*Interpretation*
i, i', v and q	Population points representing small area population units. Each i, i', v and q belongs to one district r $(i,i',r,q \in r)$ $(i \neq i' \neq r \neq q)$.
n	n is the number of population points
j, w and z	Hospital points representing hospital site geographic units. Each j, w and z belongs to one district r $(j,w,z \in r)$ $(j \neq w \neq z)$.
m	m is the number of hospital points, which is a sub-set of the total number of population points n $(m \subset n)$.
U_{ij}	Utilisation flow between population point i and hospital site j.
D_j	Size of hospital site j.
$\widetilde{D}_{ij}$	Index for alternative supply to hospital site j available for population i.
$other'_{ij}$	A set of other variables related with population and hospital characteristics that explains flows.
$other''_i$	A set of population-related variables that explains flows.
$other'''_j$	A set of hospital-related variables that explains flows.
P_i	Resident population in i.
Dem_i	Demographic characteristics of the population (age and sex) that imply higher need for hospital care for population i.
N_i	Need for hospital care for population i.
X_i	Socio-economic level of population i.
G_{ij}	Accessibility costs for population i to access hospital services in j.
d_{ij}, $d_{ii'}$	Distance between population point i and hospital site j, and between population points i and i' (Euclidean distances as defined in Chapter 3).
A_i	Perceived availability of hospital care to population i.
I_{ij}	Set of institutional characteristics of the hospital system (such as hospitals hierarchy, sites with hospital teaching functions, spatial hospital subsystems, etc), to be specified below. Some of these characteristics relate to population points.
O_{ij}	Set of variables that characterise access to other sectors of health care and non-health care systems (such as welfare system and private supply) and other variables that are expected to influence demand for hospital care –such as spatial variables along the territory.
PC_i	Accessibility to primary care for population located in i.

c_j	Role of hospital j in the hospital hierarchy (for example, dummy variables for central and district hospitals).
$i1_{ij}$	Indicator of whether hospital j is the first hospital of use by population i (dummy variable).
$i2_{ij}$	Indicator of whether hospital j is the second hospital used by population i (dummy variable).
$i3_{ij}$	Indicator of whether hospital j is the central hospital used by population i (dummy variable).
$i4_j$	Vector of hospital variables that characterise hospital j outputs other than inpatient care (such as external consultations and emergencies).
$i5_j$	Vector of variables representing the hospital input mix of hospital j (labour vs. equipment vs. beds).
y	Utilisation variable as a dependent variable.
x	Set of the covariates that are hypothesised as affecting utilisation.
x' and x''	Two sub-sets of covariates of the set x ($x' \subset x$ and $x'' \subset x$).
$d'_{ij}, d''_{ii'}$	Dummy on whether hospital j is within 25 km from population point i, and dummy on whether population point i' is within 25 km from population point i.
β''	Set of coefficients of the econometric model.
e^{ε}	Residuals in the natural scale of the second part of the two-part model.
$\hat{p}_{qw}$	Predicted probability of population point q making use of hospital site w.
$\hat{U}_{qw}$	Predicted level of utilisation flows of population point q to hospital site w, given that the probability of that flow being positive is positive.

Similarly to other utilisation data, flow data has a skewed distribution, and can be represented in a matrix that has a well-defined structure. A two-part model is shown to be consistent with that structure, allowing for an intuitive interpretation of spatial utilisation flows, and suitable for treatment of the mixed nature of distribution flows data (problem of skewed data, as described earlier).

The process of building the econometric model should attempt to maintain linear relationships between supply variables in the right side of the econometric regressions[213].

7.3 A flow demand model for hospital care

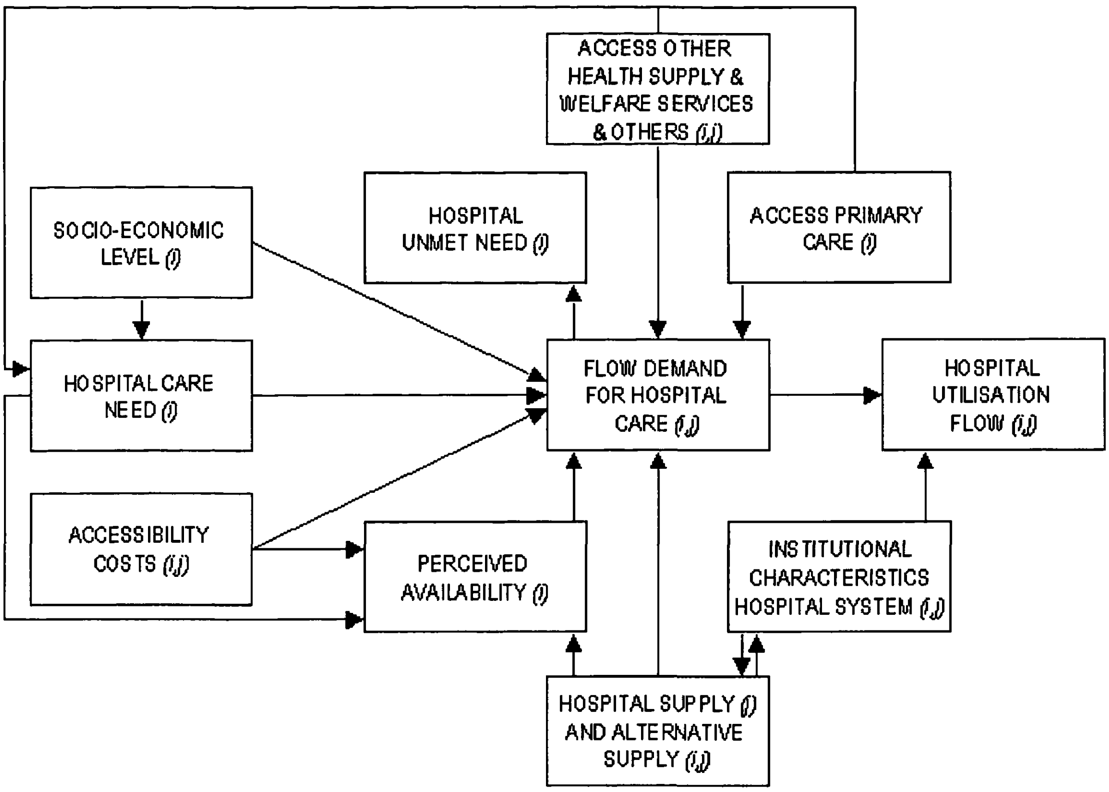
This section presents the FDM for hospital care and the corresponding econometric model. The model is set within the context of the Portuguese hospital system.

7.3.1 Conceptual model

The process of (geographic) demand for hospital care is depicted in Figure 7.2, developed from the demand for health care model previously constructed by (Carr-Hill, Hardman et al. 1994). The FDM differs from that model in four ways:

a) It uses the flow demand concept (as opposed to demand by population points);
b) It integrates the hospital system into the health care system (instead of modelling demand for health care);
c) It models institutional and organisational factors in the hospital system;
d) Finally, it deals with interaction between utilisation levels and the supply of alternative hospitals for a population area.

Figure 7.2: A flow demand model for public hospital care at the small area level



One interpretation of Figure 7.2 might consider that flow demand and utilisation flows are at the centre of the model. All the variables with a bi-dimensional index relate to

[213] This mainly applies to the second part of the model, to be used in Chapter 9, as will be shown later.

both a population point and a hospital site and incorporate, to some extent, the concept of geographic accessibility. Demand for hospital care is influenced by three main sets of variables:

- Individual factors (need for hospital care, socio-economic level and accessibility costs).
- Hospital supply related factors: perceived availability, hospital supply, and institutional characteristics of hospital systems. The latter interact with supply levels and together influence utilisation, which implies that analysis of supply levels and institutional variables cannot be separated.
- Other health care and social policy related factors (such as access to primary care, or other health related services).

The key factors that relate to utilisation and that differentiate utilisation flows from demand flows are supply, the institutional characteristics of the hospital system and unmet need[214].

A more detailed description of how all the variables in the system are linked and how they are expected to impact on utilisation flows is now set out. Hypothesised relationships are in accordance with empirical evidence and with the postulates of theoretical models described previously.

Equation 7.2 represents all the variables of Figure 7.2 that directly affect utilisation flows or demand (constituting thus a more detailed version of Equation 7.1).

$$U_{ij} = f_1(N_i, X_i, G_{ij}, A_i, D_j, I_{ij}, \widetilde{D}_{ij}, O_{ij}, PC_i) \qquad (7.2)$$

Equation 7.3 hypothesises the determinants of the need for hospital care (N_i).

$$N_i = f_2(P_i^+, Dem_i^+, X_i^{+/-}, O_{ij}^{+/-}, PC_i^-) \qquad (7.3)$$

The determinants of need (N_i) directly influence utilisation flows. Population numbers (P_i) and demographic need[215] (Dem_i) increase need and affect flows. The impact of

[214] As observed before, utilisation is taken as a measure of observed demand.

socio-economic levels (X_i) on need and demand is unknown. This is because there are two effects: higher socio-economic levels on the one hand decrease need and are associated with increasing use of private facilities, thus decreasing utilisation of public hospitals, but on the other hand they are accompanied by better access to information and knowledge on hospital care, increasing thus perceived need for care and utilisation as a result. Access to primary care (PC_i) decreases need, as GPs' activity might replace hospital care activity and prevent subsequent use of hospitals. Access to other health care supply and welfare services (variables included in O_{ij}) decrease need (for example, when there are higher levels of home care and minimal income subsidies, the need for hospital care declines). The impact of other variables on need, included in the O_{ij} such as geographic variations within any territory, is unknown.

Equation 7.4 captures the positive impact of distance (d_{ij}) on accessibility costs – distance is the most widely available proxy for accessibility costs in most health care systems. Higher accessibility costs (G_{ij}) are assumed to reduce utilisation flows although the evidence from previous studies is inconclusive.

$$G_{ij} = f_3(d_{ij}^{+}) \qquad (7.4)$$

Equation 7.5 describes the determinants of perceptions of hospital care availability (A_i); such perceptions increase utilisation.

$$A_i = f_4(N_i^{+}, G_{ij}^{-}, D_j^{+}, \tilde{D}_{ij}^{-}, I_{ij}^{+/-}) \qquad (7.5)$$

Relative perceptions of the availability of hospital care (Equation 7.5) vary positively with need (N_i) and hospital supply (D_j) and negatively with accessibility costs (G_{ij}) and with the availability of alternative hospital supply ($\tilde{D}_{ij}$). Institutional characteristics affect perceived availability, but their impact is unclear.

215 Demographic need is represented by the index capturing the impact of the population age and sex structure on need for hospital care (in the empirical application, this index is calculated in Chapter 5).

Equations 7.6 and 7.7 capture the interdependence between hospital capacities and institutional characteristics of the system.

$$(D_j, \widetilde{D}_{ij}) = f_5(I_{ij}) \tag{7.6}$$

$$I_{ij} = f_6(D_j, \widetilde{D}_{ij}, c_j, i1_{ij}, i2_{ij}, i3_{ij}, i4_j, i5_j) \tag{7.7}$$

Hospital size (D_j) and alternative supply ($\widetilde{D}_{ij}$) depend on the institutional and organisation characteristics of the system (I_{ij}) and increase and reduce flows respectively. Supply and institutional characteristics influence utilisation both directly and indirectly; utilisation flows are constrained by the level of hospital supply, 'connections' between hospitals and by the hierarchy and organisation of the hospital system. Institutional factors include characteristics of: hospital hierarchy ($c_j, i1_{ij}, i2_{ij}$), central sites with special teaching functions ($c_j, i3_{ij}$), other hospital activities and characteristics ($i4_j, i5_j$), regional hospital system structures, hospital size, distance between hospitals, etc. It is hypothesised that these variables impact on flows as follows:

- There are increased flows for the first and second hospitals of use ($i1_{ij}, i2_{ij}$) and for hospitals at the top of the hierarchy (c_j); these variables indirectly capture the characteristics of the referral system;
- As central hospitals ($c_j, i3_{ij}$) have a specific role in the system (unique providers of a range of services) they have increased flows of patients;
- The impact of other hospital outputs on flows is unclear: outpatient care (mainly accident and emergency admissions and outpatient attendances) might be seen as a substitute (e.g. day patients) for or as an entry point to inpatient care;
- More intensive labour inputs increase utilisation (such as the number of doctors per bed). For Portugal, this is particularly important, as hospitals are unable to take advantage of existing beds and equipment because of labour shortages (Pinto and Oliveira 2001).

Some other elements of Figure 7.2 also influence utilisation. Access to primary care (PC_i) is an entry point to the hospital system (within a gatekeeping system) and increases utilisation. Some variables included in O_{ij} affect demand directly:

- Access to private hospitals, which are a substitute for public hospitals, decreases demand to public hospital care;
- Access to other health and welfare services may generate information and increase demand for hospital care;
- Other factors such as variations in climate, health behaviour, spatial divisions, geographic divisions of the health system, as well as health and related policies might affect demand.

It is difficult to predict the impact of these variables on flows. The use of variables representing geographic divisions of the territory of a country (such as divisions based on physical, cultural, political and administrative characteristics) should be interpreted as an attempt to account for other system characteristics that influence flows. The impact of other variables not included in the model (such as variations in medical patterns) will be captured by residuals in the estimation.

The hypothesised impact of all the 'measurable' variables on utilisation is shown in Equation 7.8, which is a more disaggregated version of Equation 7.2.

$$U_{ij} = f_1(P_i^+, Dem_i^+, X_i^{+/-}, d_{ij}^-, A_i^+, D_j^+, \tilde{D}_{ij}^-, c_j^{+/-}, i1_{ij}^+, i2_{ij}^+, i3_{ij}^+, i4_j^{+/-}, i5_j^{+/-}, O_{ij}^{+/-}, PC_i^{+/-}) \quad (7.8)$$

The specification for the alternative hospital supply index is presented in Equation 7.9. Its numerator represents alternative supply to hospital j available within 25 km to population i[216]. Its denominator represents the whole population within 25 km from population point i, and indicates competing demand. The dummy structure of the

[216] The use of 25 kms as the geographic threshold to capture alternative supply and competing populations was based on the following rationale: 25 kms is an average distance from the centre of the main urban cities to the geographic limit of their metropolitan areas. This is a parameter specific to the health care system. In the empirical application, alternative values for this threshold were tested and it was found that for values between 20 and 30 kms, the coefficient for the index was highly significant and assumed stable values.

distance-related dummies from Equation 7.9 ($d_{ij}^{'}$ and $d_{ij}^{''}$) is defined in Equations 7.9a and 7.9b.

$$\widetilde{D}_{ij} = \frac{\sum_{j} D_j * d_{ij}^{'} - D_j * d_{ij}^{'}}{\sum_{i'} d_{ii'}^{''} * P_{i'}} \qquad (7.9)$$

$$d_{ij}^{'} = \begin{cases} 1, d_{ij} \leq 25km \\ 0, d_{ij} > 25km \end{cases} \qquad (7.9a)$$

$$d_{ii'}^{''} = \begin{cases} 1, d_{ii'} \leq 25km \\ 0, d_{ii'} > 25km \end{cases} \qquad (7.9b)$$

7.3.2 The flow-demand econometric model

The econometric application of the FDM has three advantages in comparison with previous econometric models. Firstly, it uses an increased number of observations (and hence has more degrees of freedom) as the analysis is conducted at a lower geographic level. Secondly, it captures fixed effects at the hospital, population and population/hospital levels. Thirdly, it makes unnecessary the need to control for simultaneity in conjoint determination between supply, need and demand, something that was required in previous models that sought to explain utilisation by population area. Those models have used utilisation, supply and need variables at the same geographic reference –the population area. The problem of simultaneous determination is not expected to apply to the FDM, since utilisation flows, supply and need have distinct geographic references[217]: utilisation flows relate to population and hospital; supply variables relate to hospital points; and need variables relate to population.

The following sub-sections describe the rationale for the use of a two-part model and the choice of estimation techniques; in addition, they raise some methodological questions relevant to empirical applications of the FDM.

[217] Furthermore, in the case of the Portuguese system, the determinants of the levels of hospital supply have been highly political, and not as much informed by a planning system and by assessed need.

7.3.2.1 Two-part model

Equation 7.10 presents flows in an utilisation flows matrix. This format helps to analyse the characteristics and structure of flow data, which constrain the choice of econometric model.

$$\text{Utilisation flows matrix} = \begin{bmatrix} U_{11} & \dots & U_{1w} & U_{1z} & \dots & U_{1m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ U_{v1} & \dots & U_{vw} & U_{vz} & \dots & U_{vm} \\ U_{q1} & \dots & U_{qw} & U_{qz} & \dots & U_{qm} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ U_{n1} & U_{n2} & U_{nw} & U_{nz} & \dots & U_{nm} \end{bmatrix} \tag{7.10}$$

where: $i = 1,2,\dots,v,q,\dots,n$; $j = 1,2,\dots,w,z,\dots,m$.

Each observation in the matrix represents a flow between a population i and a hospital site j; and each row represents all the flows that originate from population i. The matrix can be described as follows:

a) Each population area (i) tends to use a reduced number of hospital points (for Portugal, on average 6 hospital points)[218], which means that the number of positive observations in each row tends to be limited, while there are a high number of zeros;
b) If a hospital point w coincides with population point v, that is the highest flow from that population point;
c) If population points v and y are contiguous, they make use of the same hospital points;
d) Two hospitals located in contiguous points - w and z - are used by the same populations;
e) If w is a central hospital, most of the flows to that hospital will be positive (corresponding to observations in column w).

[218] In the Portuguese case, each population point makes use of an average of 8 hospital sites. This figure is affected by very small numbers: if one places the threshold at 5 patients for regarding a flow as positive, each population uses on average 6 hospitals. In terms of the distribution of flows, 15% of flows are positive, 8% above 5 patients, 5% above 30 patients and 3% above 100 patients.

These characteristics translate into a highly skewed distribution of flow data[219]. The most common procedure for dealing with a distribution dominated by a high number of zeros or small numbers is to stabilise the variance by using a logarithmic transformation (Manning and Mullahy 2001) but this is inadequate for flow data with such a level of skewness. Three statistical models described below have been used to deal with dependent variables with highly skewed data: sample selection models, log-linear models using positive observations and two-part models. The two-part model seems to be the most appropriate model in this context.

Sample selection models (SSMs) have been used in studies that involve data with censuring and selection bias characteristics, and make assumptions about the generation process for positive observations (Leung and Yu 1996). SSMs are inadequate in this context: as flows refer to geographic units, there is no reason to assume that positive flows are the result of a self-selection process from population areas; the interpretation of a zero flow is that one should not expect a particular population area to use a specific hospital, not that they are false zeros. SSMs are not appropriate when there is evidence of collinearity (Leung and Yu 1996).

Log-linear models, which make use of positive observations of data (and neglect the zero observations), are inadequate since they produce biased estimates and therefore are not appropriate for prediction (Manning and Mullahy 2001).

Two-part models (TPM) deal with skewed data and are appropriate when there are upper tails and/or 'high-end' outliers (Mullahy 1998). This appears to be evident for the utilisation levels of central hospitals. A TPM that decomposes flows in two independent econometric models is presented in Equations 7.11, 7.11a and 7.11b.

$$\mathrm{E}[U_{ij} / x] = \Pr(U_{ij} > 0 / x') * \mathrm{E}[U_{ij} / U_{ij} > 0, x''] \qquad (7.11)$$

---

[219] In the Portuguese case, utilisation flows present the following statistics that confirm those characteristics (information taken from the database built for this study and described below): 18,700 observations; around 85% of zero observations; sample mean of 48; standard deviation of 732; maximum value of 70,674; skewness of 57; and kurtosis of 4,881. Some descriptive statistics are presented in Appendix D.

$$Part\ A = \Pr(U_{ij} > 0 / x') \qquad (7.11a)$$

$$Part\ B = \mathrm{E}[U_{ij} / U_{ij} > 0, x''] \qquad (7.11b)$$ [220]

TPMs capture the ways in which geographic determinants impact on hospital utilisation in that they express utilisation by multiplying the probability of hospital use by the level of use in the hospital by a population point i. Utilisation flows can thus be seen as a two-stage process of decision-making (Pohlmeier and Ulrich 1994) (Mullahy 1998). TPMs also are consistent with the view that the geographic determinants of accessing a hospital might be different to the determinants influencing the level of utilisation.

Estimated flows from the TPM (Equation 7.11) result from the multiplication of both estimates of the two parts of the model (Equations 7.11a and 7.11b). *Part A* (Equation 7.11a) represents a dichotomous model used to predict the probability of a population from area i making use of hospital site j. It may be interpreted in one of two ways, either as a model to express a populations' choice of a set of hospitals that will be used, or as indicating whether a population point belongs to the catchment area of a hospital. Logit and probit models have been used to estimate *Part A* of the model (Mullahy 1998) and the probability of utilisation is explained by the set of covariates x'.

Part B (Equation 7.11b) provides a model to predict the level of flow, given that it is greater than zero and flows are explained by a set of covariates x''. The choice of econometric technique for this part of the model must specifically consider the way in which utilisation flows data are counted[221] and the long tails of their distribution, which require a procedure to stabilise the variance (which arises from the fact that there are many flows with small numbers). Two suitable methods can be used to estimate the second part of a model characterised by this type of data, either a log-linear model estimated by OLS, or a generalised linear model (GLM) with a log link (Blough,

---

[220] x' and x'' are two sub-sets of covariates from the set x of covariates that might explain flows (x is dictated by the conceptual model); and $\mathrm{E}[y/x] = \Pr(y > 0/x') * \mathrm{E}[y/y > 0, x'']$ is the generic formulation of two-part model (with $y = U_{ij}$).

[221] I.e., flows assume integer numbers as they represent numbers of patients.

Madden, and Hornbrook 1999). These two models have different strengths and weaknesses, outlined below.

The determinants of the two parts of the econometric model might differ, as distinct determinants might be involved (economic, political, geographic), while their policy implications might also differ (Pohlmeier and Ulrich 1994) (Mullahy 1998). This implies using the sub-sets of covariates from x, with a sub-set for each of the parts of the TPM: x' and x''. x is expected to integrate all the variables set out in the conceptual FDM, that influence flows.

Equation 7.12 illustrates a breakdown of the predicted utilisation flows matrix into two parts (estimates of Equation 7.10).

$$\begin{bmatrix} \hat{p}_{11} & \cdots & \hat{p}_{1w} & \hat{p}_{1z} & \cdots & \hat{p}_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{p}_{v1} & \cdots & \hat{p}_{vw} & \hat{p}_{vz} & \cdots & \hat{p}_{vm} \\ \hat{p}_{q1} & \cdots & \hat{p}_{qw} & \hat{p}_{qz} & \cdots & \hat{p}_{qm} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{p}_{n1} & \hat{p}_{n2} & \hat{p}_{nw} & \hat{p}_{nz} & \cdots & \hat{p}_{nm} \end{bmatrix} * \begin{bmatrix} \hat{U}_{11}/>0 & \cdots & \hat{U}_{1w}/>0 & \hat{U}_{1z}/>0 & \cdots & \hat{U}_{1m}/>0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{U}_{v1}/>0 & \cdots & \hat{U}_{vw}/>0 & \hat{U}_{vz}/>0 & \cdots & \hat{U}_{vm}/>0 \\ \hat{U}_{q1}/>0 & \cdots & \hat{U}_{qw}/>0 & \hat{U}_{qz}/>0 & \cdots & \hat{U}_{qm}/>0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{U}_{n1}/>0 & \hat{U}_{n2}/>0 & \hat{U}_{nw}/>0 & \hat{U}_{nz}/>0 & \cdots & \hat{U}_{nm}/>0 \end{bmatrix} \qquad (7.12)$$

7.3.2.2 *Estimation techniques*

The choice of econometric technique should satisfy three objectives. Firstly, it ought to make predictions. Secondly, it should allow for analysis of model outputs (i.e., flows) in their natural units[222]. Thirdly, the estimation technique should produce acceptable estimates across the whole range of the utilisation distribution. For the first part of the model, it is convenient to use a logit or a probit model. For the second part of the model, the choice is more complex, and one should compare the strengths and weaknesses of log-linear and generalised linear models. Preliminary analysis suggests that GLM estimation seems to be the appropriate choice.

---

[222] As shown below, some econometric techniques use mechanisms to stabilise the variance (such as the log-linear model estimated by ordinary least squares) that produce predictions which have to be adjusted when they are retransformed from the log-linear prediction to the natural scale of flows.

A log-linear model estimated by OLS is appropriate if the distribution of utilisation has a long tail[223], but there are severe problems in the estimation when the initial model is heteroscedastic (Manning 1998)[224]. To test for heteroscedasticity, errors must be checked against groups of data and against combinations of covariates. To correct for heteroscedasticity, the most used and appropriate procedures have been generalised least squares (GLS) estimators and Huber/White estimates of the variance-covariance matrix, in order to get consistent statistics for inference (Manning 1998). Whether or not there are heteroscedasticity problems, the model should be retransformed, but the specific transformation procedure depends on whether there is heteroscedasticity. Retransformation is the conversion of the regression to the natural scale of the dependent variable (since this is a log-linear model). It is mathematically difficult (Blough, Madden, and Hornbrook 1999) and, if not correctly applied, will produce biased estimates (Manning and Mullahy 2001). In the absence of heteroscedasticity, in the case of no adjustment, the following bias in the estimated response of interest will occur (using general notation):

$$E(y \mid x) = e^{(\beta'' x)} E(e^{\varepsilon}) \neq e^{(\beta'' x)} \qquad (7.13)$$

with $E(e^{\varepsilon})$ as the residuals in the natural scale and y as a utilisation dependent variable. In the case of log-scaled residuals that are heteroscedastic, a specific retransformation is required to estimate the dependent variable on the raw scale value (Ai and Norton 2000). There are several alternatives (Manning 1998):

a) If the distribution is known, then the expectation of the exponential error must be derived directly ($E(e^{\varepsilon})$);
b) If the distribution of residuals is unknown, one available nonparametric alternative is the smearing estimator, which uses the average of the exponential residuals to estimate the expectation of exponential error term (Duan et al. 1983).

This discussion shows that retransformation is complex and poses serious problems, which need to be resolved using the log-linear model to make predictions.

---

[223] Table D.1 in Appendix D gives evidence on the skewed distribution of utilisation flows for the second part of the TPM.

Generalised linear models (GLM) with a log link have been widely used for stabilising variance when it is important to retain the original scale of data and to avoid retransformation. This method is preferable when there is evidence of heteroscedasticity and only minimal assumptions. GLM models seem to be better than log-linear models for FDMs. Estimation requires the choice of two parameters: a family function for the structure of the variance and a link function to relate the dependent variable with the function that aggregates the independent variable. The link function for this type of skewed utilisation data is the natural logarithmic function. The Poisson distribution is expected to be the most appropriate family function for the structure of the variance, because of the counting process and the discrete nature of flows data (McCullagh and Nelder 1983). An extended Park test proposed by Manning and Mullahy (Manning and Mullahy 2001) must be conducted on raw-scale residuals to ensure adequacy in the use of the Poisson distribution. In one case, however, GLM estimation might be biased (but consistent) compared with a log-linear model, if the distribution of log-scaled errors has a long tail (Manning and Mullahy 2001)[225]; each model application should test for this. Notwithstanding this limitation, the GLM technique with a Poisson distribution for the structure of variance and a log link function has been selected as most suitable for the Portuguese data, and tests for adequacy should be carried out.

The move towards a hierarchical (and more disaggregated) model increases the number of observations and, more crucially, it (partly) obviates the need to control for simultaneity in determining supply and need given that these variables have different geographic references.

7.3.2.3 Methodological issues

Three further issues are worth noting. Firstly, hospital outliers are defined as urban hospitals that have a role as centres of excellence, are the only providers of some services, and are larger than average. This definition is used for the alternative hypotheses looking into the impact of hospital outliers on flows, which merits special

[224] As evidence below shows, in the application to Portuguese hospitals, the running of a preliminary log-linear model for the second part of the model has shown strong indications of the presence of heteroscedasticity between residuals and hospital size.
[225] However, in the case of the use of the FDM for prediction, the weakness of GLM should be also compared with the gains of overcoming problems with the retransformation procedure.

attention. Three such hospitals are located in Lisboa, Coimbra and Porto. Secondly, the strategy required to construct a parsimonious model needs to be clearly defined. Finally, it should be kept in mind that the FDM makes a number of assumptions.

Treatment for centres of excellence

Modelling the impact of hospitals that operate as centres of excellence is difficult due to expected multicollinearity between hospital size and variables that capture their special status, while recognising the need to form a hypothesis on the quantitative impact that these centres will have on utilisation rates. Alternative specifications for the econometric model are defined for the Portuguese setting[226]:

1. *Does hospital size have a linear relationship with utilisation flows[227]?* To ascertain this one needs to test whether hospital size might influence flows under a non-linear function, for example a piecewise linear function (with a different role for more sizeable hospitals), against an alternative hypothesis where the impact of hospital size is fully linear[228].

2. *Does hospital capacity in central hospitals affect flows in a fixed or in a variable way (i.e. depending on hospital size)[229]?* The latter is expected to be a better fit. Testing for this hypothesis requires the use of dummies for the central hospital sites (for fixed effect) and/or dummies multiplied by hospital size (for variable impact).

3. *Do all central hospitals have the same impact on flows, or does this vary across hospitals?* This requires testing whether all central hospitals have the same degree of attraction for patients in different areas (using one dummy for central hospital), or whether each central hospital has different effects (which would demand the use of individual dummies for each central hospital).

---

[226] These hypotheses are more meaningful for the second part of the TPM.
[227] Linearity is here defined in relation to the logarithm of flows, as the second part of the model uses a log link function.
[228] Ibid.
[229] Ibid, fixed or variable effect in relation to the logarithm of flows.

Strategy for building a parsimonious model

A significant level of multicollinearity between some covariates is to be expected. Textbook procedures to deal with expected problems should be used, such as the elimination of some correlated covariates, construction of composite variables (multiplicative covariates) and standardisation of variables (Gujarati 1995). The following strategy was used in choosing between covariates:

- Include in the first model the whole set of variables (shown in Equation 7.8 and as defined by x) and the interaction effects known to exist between some need-related, supply-related and distance variables;
- Exclude from the first model (iteratively) variables that are not statistically significant until attaining a parsimonious model, and construct composite variables.

Dummy variables can be freely tested in the second part of the model. But in the first part, as the model has a dichotomous dependent variable, an extensive use of dummies is not appropriate. The set of covariates that explain the probability of use (first part of the TPM) is expected to be simpler in comparison to the set of covariates that explain the second part of the model.

Model assumptions

Given the predictive objectives of the FDM and the available datasets of information, the model makes the following assumptions:

a) The population lives within a population catchment area and their utility functions are assumed to be homogeneous;
b) The impact of utilisation on outcomes and patients' satisfaction levels, and the ways in which these influence the process of need for hospital care is neglected. Modelling the influence of utilisation on health outcomes would be particularly useful if one attempted to link utilisation and resource allocation with health outcomes;
c) The impact of waiting lists and provision of tertiary care on utilisation flows are not assumed to be significant. These assumptions are required given the absence of data to test assumptions on the impact of those variables on utilisation;

d) The impact of important, but difficult to measure, variables affecting hospital admissions such as social networks, social interactions and culture is not taken into account (Andersen 1995).

The use of a model deploying logarithms of flows (log link function for the second part of the TPM) implies that results are in the form of geometric means (Manning 1998). Consequently, the impact of variation of a covariate on flows will depend on the current level of flows, and must be accounted for interpretation.

7.4 Application to the Portuguese hospital system

The FDM can be applied to the Portuguese hospital system, in order to produce a model to predict hospital utilisation and generate information on determinants of hospital utilisation. In what follows, the construction of the database is outlined, measurement issues are highlighted, and finally results are presented and discussed.

7.4.1 Dataset building and measurement issues

The *concelho* level is used as the small area unit of geographic analysis for population areas (this geographic level is described in Chapter 3, section 3.2.2). The hospital group is used to represent the hospital level because data tend to be published for hospital groups rather than for individual hospitals[230]. A hospital group consists of a set of individual hospitals under common administration. When a hospital group includes hospitals that provide services for several *concelhos*, it is the *concelho* of the largest hospital that is used as the point of location of the hospital site.

The dataset built for this study consists of 18,700 statistical units with 275 population geographic points (i's) and 68 hospital geographic sites (j's) (as of 1999). Utilisation data were taken from the hospital discharges DRG information system (nearly 1 million

[230] Hospitals in the same *concelho* are grouped together and treated as a single hospital geographic point. This was necessary, as several indicators were only available at the hospital group level, and many hospitals of the same *concelho* belong to the same group.

discharges in 1999). Additional data were obtained from many sources, specified in Table 7.2.

The covariates included in the database and their expected sign in econometric regressions (based on the hypothesised relationships of the conceptual model) are presented in Table 7.3. For some variables, squared coefficients seem to be meaningful: for example, beyond a certain distance level, the impact of distance on flows might be negligible[231]. Predicted signs for the interaction terms are excluded from the table[232]. Some descriptive statistics of the dataset and for the two parts of the TPM are presented in Appendix D (Tables D.1 and D.2).

The following methodological assumptions were made:

- The number of hospital discharges was taken as a proxy for hospital capacity, given the findings of Chapter 3, as beds are a poor indicator of supply. For some observations, a close relationship between flows and aggregate hospital supply is expected (mainly for populations and hospitals located in the same geographic point). A one-to-one relationship between those two variables was not anticipated, as each hospital served several population areas and other covariates captured other related effects, such as the dummy for closest hospital. Moreover, there were no alternative data to generate indicators of extra hospital capacity such as those based on local waiting lists (this information would highly improve the model).
- Chapter 5 has used SMRs and ASMRs as indicators for quantifying socio-economic and/or morbidity levels. These indicators were not used as covariates, as their capacity to be a proxy for need at the small area level have not been validated.
- Euclidean distance was used as a proxy for accessibility costs, given that there was no data on travel time.
- Perceived availability was measured by a score for each population area, calculated as the output of the attraction-constrained gravity model; the formula for this model is given in Appendix E.

[231] Santana has shown that the relationship between distance and utilisation of some hospital services in Portugal depended on the type of service, the severity of illness and the level of provision of certain specialties of the other hospitals in the area (Santana 1996). Nevertheless, this study did not control for other confounders.

[232] Multiplicative terms were applied to the following variables: distance, population and hospital size multiplied by some other variables. Their coefficients are easily interpretable; for instance, besides the influence of supply and population on utilisation, it might be assumed that high population levels and high supply can translate into higher utilisation flows.

- Preliminary analysis of the database has confirmed the anticipated problems of multicollinearity. There were high correlations between: population, purchasing power, private care supply and socio-economic variables, as well as between the number of specialties available and the number of discharges. To deal with these problems, some covariates were eliminated, some composite variables were used (such as population numbers multiplied by illiteracy rates), while other variables were standardised (such as primary care utilisation per capita, outpatient attendances by hospital discharge and accident and emergency admissions by hospital discharge).

Table 7.2: Database sources

Indicator	*Source and/or method of computation*
Utilisation flows: inpatient discharges per population point and hospital site. Hospital size: inpatient discharges per hospital site.	DRG information system, 1999 data, provided by the Financial Institute for Informatics and Financial Management.
Population need: resident population per population point.	Resident population estimates, 1999 data (INE 2000).
Population need: demographic need index per population point.	Needs index at the small area level that results from evaluating resident population by age groups by the age cost curve computed in Chapter 5; 1999 resident population data (INE 2000).
Socio-economic indicators: a) Ratio of illiterates with more than ten years to resident population, unemployment rates, ratio of homes used as usual residence and without electricity to homes used as usual residence, percentage of resident population dependent on the primary sector, per population point. b) Purchasing power index, per population point.	a) Census figures, 1991 data (INE 1993a, 1993b, 1993c, 1993d, 1993e). b) 2000 data (INE, Direcção Geral do Centro, and Gabinete de Estudos Regionais 2000).
Accessibility costs: distance between population points and hospital sites.	Euclidean distances computed using centroid coordinates provided by the Environment Directorate, from the Portuguese Ministry of Environment and Organisation of Territory.
Access to primary care: utilisation as measured by consultations on primary care, per population point.	Primary care consultations, 1999 data, provided by the National Institute of Statistics, under research protocol.
Availability of private hospital care: number of private hospital beds, per population point.	Addresses and number of beds of private hospitals, 1996 data (Departamento de Estudos e Planeamento da Saúde 1997a).
Other hospital outputs: outpatient attendances and accident and emergency admissions, per hospital site.	Number of outpatient attendances and accident and emergency admissions, provided by the General Directorate of Health, 1999 data.
Diversity of hospital outputs: average number of specialties available per hospital site	Number of specialties available, 1996 data (DGS and Ministério da Saúde 1998a, 1998b, 1998c)

Table 7.3: Covariates (or conceptual explanatory variables) included in the database (excluding interaction terms) and hypothesised behaviour

Explanatory variable	*Indicators*	*Predictive behaviour*
Population	Resident population *(i)*	Positive up to a threshold (hospital size is finite, and for highly populated areas, it can be expected that there is a threshold for the increasing impact of population on flows).
Demographic need	Demographic need index *(i)*	Positive.
Socio-economic level	Purchasing power index *(i)*	Positive if higher income implies better access to hospital care, which offsets the impact of high income on a lower need for hospital care and/or higher use of private hospitals; negative if the impact of higher income on better accessibility is offset by a lower need for hospital care and/or higher use of private beds.
	Illiteracy rates *(i)* and other census indicators *(i)*	Effect depends on whether better education, better housing conditions, etc mainly imply better information and increased utilisation, and whether lower education implies higher need and higher use.
Distance	Euclidean distance *(i,j)*	Negative up to a threshold (insensitive to high distances, after a certain point).
Perception of availability	Accessibility coefficients from a gravity model *(i)*	Positive up to a threshold (after a certain point, insensitivity to high perceptions on availability).
Geographic area variations	Dummy for population located in the north health region *(i)*	Not defined.
	Dummy for population located in the centre health region *(i)* (the South region is the baseline)	Not defined.
Hospital size	Discharges *(j)*	Positive up to a threshold.
	Number of hospital units in the site *(j)*	Positive.
	Average number of specialties available in the site *(j)*	Positive.
Alternative hospitals supply	Alternative hospital supply index *(i,j)*	Negative.
Private hospital supply	Number of private beds in the site *(i)* /Resident population *(i)*	Negative.

Primary care supply	Primary care utilisation *(i)* /Resident population *(i)*	Positive as an entry point; negative if substitute for inpatient care.
Closest hospitals supply	Dummy if closest hospital *(i,j)* *Discharges *(j)* and/or Dummy if closest hospital *(i,j)* Dummy if second closest hospital *(i,j)* *Discharges *(j)* and/or Dummy if second closest hospital *(i,j)*	Positive. Positive.
Central hospitals and hospital system areas of supply	Dummy if closest central hospital *(i,j)* and/or Dummy if closest central hospital *(i,j)* *Discharges *(j)* Dummy if Lisboa central hospital for population from the south region *(i,j)* *Discharges*(j)* Dummy if Porto central hospital for population from the north region *(i,j)* *Discharges*(j)* Dummy if Coimbra central hospital for population from the centre region *(i,j)* *Discharges*(j)*	Not defined (It depends on the relationship between hospital size and the role of central hospitals). Not defined. Not defined. Not defined.
Hospital supply (excluding inpatient care)	Hospital outpatient attendances *(j)* /Hospital discharges *(j)* Hospital accident and emergency admissions *(j)* /Hospital discharges *(j)*	Positive if outpatient attendances are mainly an entry point; negative if outpatient attendances operate as substitute for inpatient care. Positive if accident and emergency admissions are mainly an entry point; negative if accident and emergency admissions operate as substitute for inpatient care.

7.4.2 Econometric results

Some results from the two-part econometric model are reported in this section. STATA 7.0 statistical software (Stata Corporation 2001) was used and econometric controls were made including specification, goodness of fit, properties of residuals (deviance, in the case of the GLM model) and linktest. The choice between alternative models was based on three criteria: predictive power, parsimony (models with smaller number of variables were preferred) and expected sign of coefficients. Robust estimates of the variance of the estimators have been used (Huber-White estimates of the variance-covariance matrix).

In general, the relationships between covariates and flows (or probabilities, in the case of the first part of the TPM) were observed as expected in the conceptual FDM, including:

a) The impact of need, availability of hospital supply and perceptions of availability on flows are positive, while distance and primary care utilisation are negative.
b) The value of the negative coefficient for the alternative hospital index is highly significant, which makes the model important for prediction purposes.
c) The institutional characteristics of the hospital system have an important impact on flows: the coefficients of institutional variables are highly significant, while each central hospital has a different impact on flows.
d) There are regional area variations in utilisation.
e) Primary care acts as a substitute for hospital care.

The determinants of the probability and of the level of flow are different, as expected. Key findings, as well as the results from the two parts of the model follow below.

7.4.2.1 First part model

This model has shown that the probability of a positive utilisation flow is positively influenced by population need and perceptions of accessibility and availability of supply, while it is negatively influenced by distance, primary care utilisation, as well as by alternative availability of hospital supply (Table 7.4).

The three variables related with supply and with the institutional characteristics of the hospital system (last row of Table 7.4) should be interpreted in conjunction with the coefficient of supply availability (Table 7.4). Those variables reflect the specific attractiveness of each of the hospital sites in the system (or alternatively, the propensity in each region for the use of the central hospital[233]). The logit model is highly robust as proven by the indicators of goodness-of-fit (which are the percentage of correctly predicted classifications, the pseudo-R2 and the Wald test statistic).

Table 7.4: First part LOGIT model

Indicator	*Variable*	*Coefficient*	*Z*
Other	Constant	-2.339868***	-3.31
Distance	Distance(i,j)	-0.0191674***	-6.42
Perceptions of availability	Gravity accessibility index(i)	1.418926*	0.77
Demographic need	Population (i)*Demographic need(i)	0.0000174***	9.91
Primary care	Primary care utilisation(i)/population(i)	-0.0000014**	-1.96
Private care	Private care supply(i)	-0.0008909*	-1.69
Availability of supply	Discharges(j)	0.000131***	14.76
Alternative public hospital supply	Hospital competition "index" (i,j)	-0.5453832***	-7.48
Institutional factors	Dummy for the population in the north using the Porto hospital site(i,j)*discharges(j)	-0.0000767***	-8.47
	Dummy for the population in the centre using the Coimbra hospital site(i,j)*discharges(j)	-0.0000603***	-7.35
	Dummy for the population in the south using the Lisboa hospital site(i,j)*discharges(j)	-0.0000841***	-8.68

Model summary: 18,700 observations

Diagnosis: 94.94% correctly predicted classifications (0.5 cut-off[234]); Pseudo R2= 63.03%; Wald Chi2(10): 2,152.

***- Statistically significant at 1% level; ** Statistically significant at 5% level; *-Not statistically significant.

[233] This is higher for the Centre and North regions, in relation to the South region.

[234] The value of 0.5 was used to compute the percentage of correct predictions of the 1's and 0's of observed data by the predictions of the Logit model.

7.4.2.2 Second part model

Results of the GLM model are recorded in Table 7.5. The model that best depicts the role of the centres of excellence uses a linear effect of discharges on the logarithm of utilisation flows (although the slope and the fixed effect for central hospitals differ in relation to other hospitals). The best performing empirical model is one with a variable treatment of the central hospital dimension on flows; this was the preferred model on grounds of responsiveness of utilisation to changes in supply (see section 3.2.3).

The main findings might be summarised as follows:

- Utilisation is positively related with population numbers, demographic need and socio-economic status. In order to overcome multicollinearity problems, a composite indicator for population need was used (multiplying population by demographic need and illiteracy). A threshold effect was found to apply to the impact of population numbers on utilisation. This is explained by the fact that hospital size is finite, therefore for highly populated areas, there is a threshold on the positive impact of population on flows.
- Perceptions of availability have a positive impact on utilisation and are more influential when associated with densely populated areas. This is represented in the model by the product of the composite indicator population by the perceptions of availability index.
- Distance is a deterrent for hospital utilisation but has a limited value. Previous evidence about the effect of distance on utilisation was inconclusive; this model demonstrates that it has a crucial effect.
- The higher the level of alternative hospital supply the lower the flows to any given hospital. This is seen as a key finding on the predictive capacity of the model.
- Supply availability and institutional factors play a paramount (and positive) role on flows, and the size of the closest hospital is also positively related to flows[235].
- Primary care acts as a substitute for public hospital care, which has implications for policy.
- Geographic variations were found between the health regions. The relevant variables are expected to depict the effect of non-controlled factors, such as variations in health policies, welfare systems or geographical latitude. It is not

obvious why the centre and northern regions have a lower propensity for hospital use. These results should be analysed alongside the finding that each central hospital has a different role with regard to flows.

- A resident of the central region is less likely to use any hospital but more likely to use the central hospital in Coimbra, and the opposite applies in the southern region.

Table 7.5: Second part GLM model (Poisson distribution of errors and log link)

Indicator	*Variable*	*Coefficient*	*Z*
Other	Constant	6.468201***	30.44
Distance	Distance(i,j)	-0.0423718***	-16.94
	Distance(i,j)*distance(i,j)	0.0000776***	13.95
Perceptions	Population(i)*gravity accessibility index(i)	2.95e-07***	15.73
Demographic and socio-economic-related need	Population(i)*demographic need index(i)*Illiteracy rate(i)	0.0000354***	4.14
	Population(i)*population(i)	-2.97e-12***	-3.48
Geographic variations	Dummy population in north region(i)	-0.2022617**	2.10
	Dummy population in centre region(i)	-0.3838279***	-3.37
Primary care	Primary care utilisation(i)/population(i)	-0.1160661***	-2.10
Supply availability	Discharges(j)	0.0000352***	6.56
Alternative public hospital supply	Hospital competition "index" (i,j)	-0.1873067***	-3.57
Institutional factors	Dummy for first hospital(i,j)*discharges(j)	0.000231***	6.54
	Dummy for second hospital(i,j)*discharges(j)		
	Dummy for closest central hospital(i,j)	0.0000141***	4.33
	Dummy for population in the north using the	-4.304794***	-4.30
	Porto hospital site(i,j)*discharges(j)	0.0000158***	2.70
	Dummy for population in the centre using the Coimbra hospital site(i,j)*discharges(j)	0.0000255***	2.99
	Dummy for population in the south using the Lisboa hospital site(i,j)*discharges(i)	-0.0000136***	-5.79

Model summary: 2,217 observations; LogLikelihood=-258,759.3

*** Statistically significant at 1% level; ** Statistically significant at 5% level; * Not statistically significant. Predictive power (calculated as the ratio of predicted utilisation –obtained by multiplying the two parts of the model- and observed utilisation): total utilisation 97.6%; total utilisation of central hospitals: 99.6%.

The Poisson distribution is used as the family for the error distribution in the GLM model. The extended Park test provided a value of 1.5 for λ, on the borderline between

[235] These variables can be interpreted as proxies for the hospitals of reference to a population point.

the use of a Poisson and a Gamma distribution (adequate for values of $\lambda = 1$ and $\lambda = 2$, respectively) (Manning and Mullahy 2001). This justifies the use of the Poisson distribution as the family distribution for residuals.

7.4.3 Discussion of results

7.4.3.1 Implications of results

The reasons for the different values for central hospitals in each region are not well understood, and results may be interpreted as an indicator of inequalities in access. It seems that central hospitals play a different role for populations located in different areas, depending on the pattern of health care delivery.

The alternative hospital supply index to a population area appears to be a key variable in explaining utilisation levels (both the probability of use and conditional flows of utilisation). This is a crucial mechanism for predicting utilisation and for analysing the impact of supply change on utilisation flows.

This application of econometrics suggests the following implications for Portuguese health policy. Firstly, the role of primary care as a substitute for hospital care deserves more attention as an alternative instrument (to hospital supply) to influence hospital utilisation. Secondly, the system permits supplier-induced demand of some kind, as perceptions of availability increase hospital flows. If perceptions of availability impact on flows, this should be incorporated into the design of policies, for instance by improving information within the health care system, since this affects equity of access. Thirdly, there is evidence of inequalities related to location accessibility for populations in different areas. This information should be related to area variations in health outcomes. Finally, the model could be used as a launchpad to define alternative policies that might change utilisation levels. Governments might, for instance, change primary care provision, hospital supply or institutional characteristics in the system. The model could be used to analyse whether use of such policy tools is likely to be effective in changing utilisation.

The econometric application could be improved, if there were more data on needs and on geographic flows, such as those for primary care utilisation, private hospital utilisation and other variables within the welfare system. This would help overcome multicollinearity problems and improve estimates of substitution or complementarities between health care sectors.

In the literature, many studies have adopted utilisation as a proxy of need for hospital care, subsequently encountering difficulties in disentangling the determinants of utilisation related to need or other characteristics. Some have questioned the deployment of utilisation data to measure need as a starting point because such data are influenced by supply (Morgan, Mays, and Holland 1987). The FDM attempts to separate effects and to model interaction between demand and supply factors. Nonetheless, although the FDM was built to predict hospital utilisation, it can only be used for predicting marginal changes in hospital supply when changes in the institutional context are minimal.

The FDM could possibly be used to produce estimates for the morbidity adjustment of Chapter 5. This approach would be similar to the one currently in use in England and described in section 5.5.1. It would correspond to the use of predictors for need to estimate utilisation (while disregarding the supply related predictors) and would provide estimates of the component of flows justified by need. Nevertheless, this approach would be complex given certain problems with dealing with the two-part model structure of the FDM, difficulties in disentangling between needs and supply factors (defined in the previous paragraph), and difficulties in excluding covariates in a model using a logarithmic link. These reasons justify why the computation of those estimates (using the FDM) was not performed.

7.4.3.2 How the flow demand model could be further developed

The use of multilevel modelling techniques to differentiate the impact of covariates on utilisation per hospital, per district or per health region has not been addressed; this is beyond the scope of this study. Neglecting such multilevel effects might imply biases (Carr-Hill, Hardman et al. 1994), and both random intercepts and/or random slopes models could be used to test the impact on utilisation of variations between groups of population areas or groups of hospitals (Snijders and Bosker 1999). Further research is required, but hierarchical generalised linear models and specific software programs might be deployed to estimate multilevel effects (Snijders and Bosker 1999). Nevertheless, the FDM developed here takes account of multilevel variations through the use of fixed effects (population and hospital-based) and area dummies at the health region level. These covariates capture similarities across hospitals and health regions.

The FDM is applied at the small area level, where small area units within one geographic area can share properties: for example, neighbouring areas have some common characteristics and may depend on similar factors (Glazier et al. 2000), such as similar socio-economic characteristics or dietary intake. One can perform tests for spatial dependence between regression error terms. Spatial heterogeneity and correlation often implies heteroscedasticity, random coefficient variation and switching regressions, and unbiased but inefficient estimators (Anselin 1988). The FDM takes account of some types of spatial autocorrelation, controlling for variables like distance and regional variations (e.g. by using dummies per health region). Such variables capture accessibility and differences between geographic areas, the most critical types of autocorrelation. However, there is scope for further research in this area.

Several versions of the FDM can be used, depending on the methodological approach. For example, the dependent variable (utilisation flows) could be replaced by patient length of stay in a population point that use a hospital site, or utilisation flows could be corrected by case-mix. Although the FDM was first conceived for applications at the general hospital level, it can also be applied to the specialist sector.

Other methodological issues that merit further attention include challenging the assumption that the same pattern of substitution exists between hospitals within an area, as well as investigating the impact of high levels of multicollinearity problems on predictive ability.

7.5 Conclusions

The FDM serves as an alternative model for analysing hospital utilisation at the small area level, and the econometric model chosen in this study seems to be suitable for prediction purposes. It also tackled some weaknesses from previous utilisation models.

It can be concluded that the use of a multidisciplinary approach is useful. It linked the flows concept commonly used in operational research and geographical literature, with the process of demand for health care, an area where health economics and policy literature has concentrated.

Moreover, it has been shown that many variables influence utilisation levels in public hospitals. Thus, any policy that attempts to improve equity in hospital utilisation, e.g. in the acute sector, might make use of other health-sector factors, such as primary care, as well as non-health-related sector policies. Evidence suggests that in Portugal the current structure of the hospital sector and the distribution of supply are producing inequalities in access. Future policies aiming at promoting equity should also pay attention to institutional factors.

Methods testing the redistribution potential of hospital capacity to improve geographic equity are developed in Chapter 9.

8 CHAPTER 8 – Geographic inequity estimates in the Portuguese hospital financing system

8.1 Introduction

This chapter completes a set of chapters producing information on adjustments of a capitation formula, introduced in Chapter 4. These adjustments aimed at computing estimates of inequities of capital, utilisation and finance under the indices presented in Chapter 4. This chapter carries out an analysis of those indices of inequities in the geographic distribution of hospital resources in Portugal.

The chapter is structured in three sections, which analyse implications of redistribution as implied by each adjustment of the capitation formula, analyse inequity estimates for Portugal, and discuss results and methods and draw conclusions.

The additional notation used in this chapter is presented in Table 8.1 and the data are from sources described in previous chapters.

Table 8.1: Notation in use

Symbols	*Intuition and explanation of the choices made*
$UCOutput_h$	UC index for hospital h.
I_{1r}	Age adjustment index for district r.
$Catchment_r$	Catchment population of district r.
D_r, D	Discharges from hospitals of district r; total discharges in the system.
O_r	Discharges from the resident population of district r.
W_r, W	Population need for hospital care in district r (resident population weighted by age); total population need.
$W_r^{'}$	Population need for hospital care in district r, scaled so that total need sums up total discharges in the system.

r, I_{2r}, I_{3r}, I_{4r}, cap_index_r, P_r	Defined in Chapter 4.
X_{1a}, P_{ar}, $ASMR_{ar}$	Defined in Chapter 5.
h, do_h	Defined in Chapter 6.
U_{ij}	Defined in Chapter 7.

8.2 Results from adjustments at the district level

Chapters 5, 6 and 7 modelled different adjustments of the capitation formula defined in Chapter 4: need adjustments (Chapter 5), unavoidable costs (Chapter 6) and cross-boundary flows (Chapter 7). This section combines these adjustments to give indices of needs, UCs and CBFs by district and discusses issues that arise from these calculations.

8.2.1 Needs adjustment index

Equations 8.1 and 8.2 formulate the age index (I_{1r}) and the age and additional need multiplier index (I_{2r}) at the district level. Both indices are proxies for the expected impact of these factors on costs for hospital care at the district level[236].

$$I_{1r} = \sum_a \frac{P_{ar} * X_{1a}}{P_r} \tag{8.1}$$

$$I_{2r} = I_{1r} * \sum_a \frac{P_{ar} * ASMR_{ar}}{P_r} \tag{8.2}$$

[236] As remarked in Chapter 5, the sex adjustment was not applied to the needs index, as it was found that it had little impact on district shares. This happens as districts have similar demographic profiles on their sex distribution (the same has been found for England).

Table 8.2 shows how these indices operate for the shares of resources of Portuguese districts:

- The age adjustment increases the age-related needs for shares of resources of districts (and thus of normative shares of districts) with older populations (such as Castelo Branco, Portalegre and Beja) and decreases shares of urban and northern districts (such as Braga and Porto), with the youngest populations, as expected (second column);
- The additional need adjustment increases the shares of resources of the most urbanised areas and decreases the shares of a mix of semi-peripheral urban and rural areas, as expected (third column in comparison with second);
- The combined adjustment for age and additional need implies that a mix of urban and rural areas have higher need for hospital resources than the national value (mainly Beja and Lisboa). Lower relative need is found for smaller urban and coastal districts of the north, mainly Aveiro and Braga;
- The needs index corresponds to a (substantial) redistributive range between 81% and 117%.

Table 8.2: Need adjustment indices

	Age index (age population shares weighted by age cost curve) (I_{1r})	*Age and additional need multiplicative index (age population shares weighted by age cost curve and by ASMRs)* (I_{2r})
Aveiro	95%	81%
Beja	113%	113%
Braga	89%	81%
Bragança	110%	102%
Castelo Branco	117%	105%
Coimbra	106%	86%
Évora	112%	89%
Faro	107%	102%
Guarda	115%	102%
Leiria	102%	92%
Lisboa	101%	117%
Portalegre	118%	106%
Porto	92%	99%
Santarém	109%	98%
Setúbal	97%	102%
Viana Castelo	104%	105%
Vila Real	101%	102%
Viseu	103%	86%
Min	89%	81%
Max	118%	117%

8.2.2 Unavoidable costs index

The UCs index weights the estimated unavoidable standardised costs for each hospital of the district by the hospital size (UCs were defined and estimated in Chapter 6, section 6.4.3.2), when size is proxied by the number of doctors[237] (Equation 8.3). Estimates of the UCs index are computed using index of unavoidable standardised costs ($UCOutput_h$),

[237] The rationale for the use of doctors as a proxy of hospital productive capacity was presented in previous chapters.

as generated by the multilevel model with random intercepts and slopes (MLM) produced in Chapter 6[238].

$$I_{3r} = \frac{\dfrac{\sum_{h \in r} (UCOutput_h * do_h)}{\sum_{h \in r} do_h}}{\dfrac{\sum_{h} (UCOutput_h * do_h)}{\sum_{h} do_h}} \quad (8.3)$$

The redistribution suggested by the UC index increases relative shares for urban areas with large concentration of supply and decreases shares of resources for areas with the smallest hospitals. Values above 100% in Figure 8.1 show that the UC index for the district is above the national average, and that if the index were to be used for redistribution, these districts (with costs above 100%) would be 'net' winners from redistribution. District values reflect two elements: the UC index (which implies higher than average shares of the largest hospitals) and the characteristics of the district hospital system. Complementary information on the structure of each hospital system (that is, for each district) is presented in Table 8.3. The UC index for Coimbra, Lisboa and Porto are 28%, 13% and 11% above the national average, respectively; and for Aveiro, Leiria and Guarda, 25%, 26% and 32% below the national average. The higher value for Coimbra reflects a hospital structure dominated by large hospitals; the comparative lower value for Lisboa and Porto reflects a mix of large and small hospitals (see Table 8.3). The district index implies a significant level of redistribution: 68% to 128% (Figure 8.1).

[238] The values obtained for the UC indices generated by the HFEM at the district level were computed and compared with the ones generated by the MLM. Comparison between the two alternative indices has shown that results from the two models differ greatly. These differences mainly reflect the more adequate treatment of allocative inefficiency and differences between hospital groups in the MLM. The results of the MLM were selected for analysis.

Figure 8.1: UC indices for MLM model

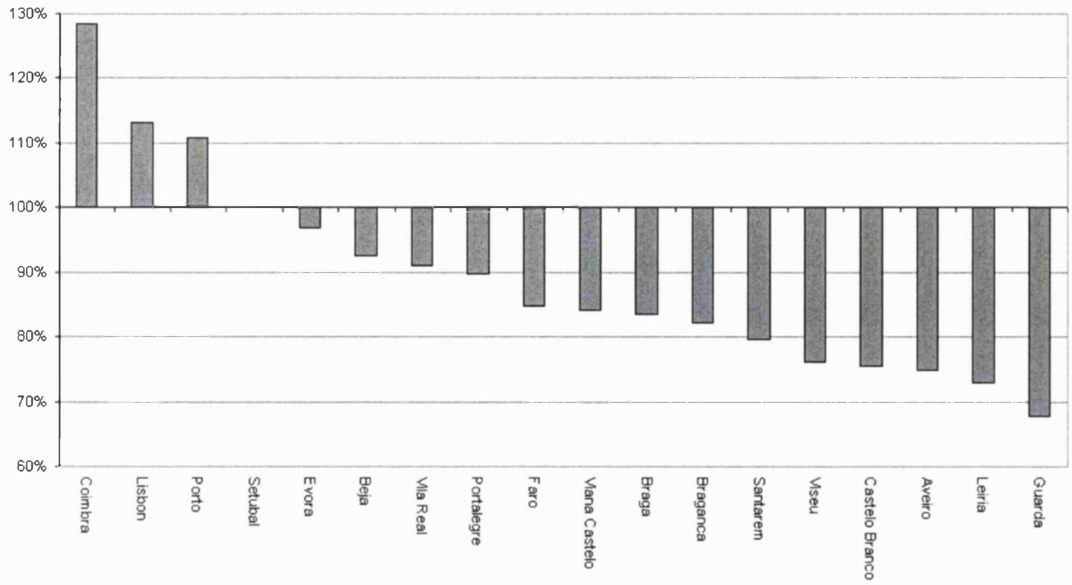


Table 8.3: Indicators of the hospital system structure

	Average number discharges per hospital	*No. doctors biggest hospital*	*No. doctors smallest hospital*	*Number hospitals*
Aveiro	7,358	170	8	9
Beja	7,433	96	8	2
Braga	15,775	283	15	5
Bragança	6,756	55	14	3
Castelo Branco	10,377	89	8	3
Coimbra	21,705	913	10	5
Évora	16,218	158	158	1
Faro	11,537	280	9	3
Guarda	8,064	68	5	2
Leiria	8,965	150	7	5
Lisboa	15,429	1180	34	17
Portalegre	6,917	66	44	2
Porto	16,579	1380	14	14
Santarém	11,196	183	31	4
Setúbal	15,164	481	13	5
Viana Castelo	11,142	177	21	2
Vila Real	10,727	119	15	3
Viseu	12,665	216	10	3

8.2.3 Cross-boundary flows index

The construction of an index of CBFs for the district level is required for the computation of catchment populations. Catchment populations of a district are defined as the amount of population need (for health care) from the health care system that is expected to use the district hospitals for treatment (Mays and Bevan 1987). There are several methods to aggregate flows in catchment populations[239]. Each method makes assumptions about admission rates and when these vary across districts different methods produce different results (Bevan and Ingram 1987). The assumptions on admission thresholds by hospital clinicians imply different incentives when they are used to allocate resources (Wilson 1988) (Bevan 1988)[240]. Some problems of incentives arise because hospitals might be able to manipulate their admission policies so as to maximise their catchment populations, and so influence their future allocations that might diverge from the desired allocations (Bevan 1988). There is no consensus as to the best method to compute catchment populations for the purpose of resource allocation and all the methods used for resource allocation can create perverse incentives (Bevan and Ingram 1987).

In the capitation formula developed in this thesis, the main objective is to analyse how variations in district supply characteristics impact on flows of patients between districts (captured in CBFs). CBFs are to be used to explain inequities in finance and utilisation, while the question of whether the methods might create perverse incentives if used in resource allocation is not central. Nonetheless, when making methodological options, some of those implications will be taken into account.

Three main methods are available to compute catchment populations. These have been presented both in mathematical format (Wilson 1988) and in a descriptive format, in the context of hospital systems (Bevan and Ingram 1987). In the Portuguese context, there

[239] Previous studies have used flows generated by the gravity model to compute catchment populations. This study uses flows produced in the FDM (i.e., information of the predicted utilisation flows matrix - U_{ij} -, defined in Chapter 7).

[240] For example, it is desirable that a resource allocation mechanism for CBFs creates incentives so that oversupplied regions will progressively decrease their levels of provision (mainly reducing inflows from other areas) and undersupplied regions will progressively increase their levels of provision, to be used by their residents. However, not all methods allow for this.

is a case for using the proportionate flow method (PF). The arguments for that follow a brief description of each of these methods:

1. **The net flow method (NF)** assumes that national average admission rates apply to cases crossing district boundaries, and the admission rate of the district to cases treated within districts (Bevan and Ingram 1987). This was the method used by the initial RAWP report (with flows valued by a costing system) (Department of Health and Social Security (United Kingdom) 1976). NF is the best method for aggregating data from different geographic levels[241].

2. **The proportionate flow method (PF)** assumes that the admission rate of the district of residence applies whenever cases are treated (Senn and Samson 1981) (i.e. independently of the hospital of treatment). This assumption is not necessarily valid, and problems arise, as districts might be not responsible for their residents' use of other districts (Bevan and Ingram 1987). The assumptions of the PF are more compatible with systems where decisions on admission rates are more influenced by GPs and less by hospital doctors.

3. **The treatment intensity method (TI)** assumes that hospital admission rates are defined by hospital of treatment. This method creates incentives for treating patients regardless of place of residence (Bevan and Ingram 1987) and it accepts implicitly that the location of the hospital is an important determinant of admission rates (Wilson 1988). Moreover, this is the method that requires the least assumptions and places the least demands on data (Senn and Samson 1981). The TI method is the best method, in that it motivates hospitals to decide on the basis of admissions and not on the basis of patients' addresses, and is more adequate for systems where decisions of hospital doctors on admissions are more important than decisions of GPs (Wilson 1988). Nevertheless, the TI formula is only advisable when CBFs are small (Wilson 1988) because it is sensitive to random fluctuations. The TI method generates the most extreme level of CBFs between areas (in comparison to other methods), which implies that patients are assumed to travel longer distances to access hospital care (Wilson 1988).

[241] Aggregation errors for the other two methods are greater, but of small amount if populations are not too far from homogeneous (Wilson 1988).

Although these methods are conceptually different, studies have found that their application creates only minor differences in practice (for example, for England see (Bevan and Ingram 1987)). In the Portuguese context there is a case for using the PF method for the following reasons: first, the TI is inadequate when CBFs are not small (which is the case in Portugal). Second, the NF is inappropriate when admission rates vary widely (which is also the case in Portugal). Third, on normative grounds, the PF is in accordance with a system in which GPs should have a higher influence on hospital admission rates than hospital doctors, which is also true for the Portuguese system. Under the PF, catchment populations are computed as defined in Equation 8.4. The NF method is selected for comparison just to show how sensitive results are to the methods chosen (NF model computed as defined in Equation 8.5).

$$Catchment_r = \sum_{j \in r} \sum_{i} U_{ij} * W_r / O_r \tag{8.4}$$

$$Catchment_r = W_r^{'} + (D_r - O_r) \tag{8.5}$$

The index of CBFs for the district level (I_{4r}) is defined in the following equation:

$$I_{4r} = \frac{Catchment_r}{W_r} \tag{8.6}$$

The CBF index was estimated with crude utilisation numbers that did not take into consideration the impact of different types of users on costs (U_{ij} s were not weighted for costs and did not account for variations in case-mix)[242], and neglected the impact of the private sector on utilisation. Analysis of estimates of the CBFs index (computed using Equation 8.6 and deploying catchment populations calculated by the PF or NF models) shows that:

- There are significant differences between indices derived using the PF and the NF methods (Figure 8.2) for districts with the most extreme values; this finding

[242] If possible, it would be desirable to value physical CBFs by a costing procedure, as currently valued in some Canadian provinces (Hutchison et al. 1999).

contradicts previous findings in which application of different methods creates minor differences in practice;

- Using estimates from the PF, only five districts operate as net suppliers (Coimbra, Lisboa, Porto, Braga and Vila Real). The unequal distribution of supply implies that a significant proportion of patients of some districts are 'treated' outside the district of residence (implying inequalities in access costs): more than 20% of the patients from Leiria, Viseu, Viana do Castelo and Évora are 'treated' in hospitals of other districts.
- The CBFs index calculated by PF and the level of supply of a district (measured by discharges from hospitals of the district) are highly correlated (Figure 8.3). This is expected, as found in previous studies (Taket and Mayhew 1981).

Figure 8.2: CBFs indices computed under PF and NF

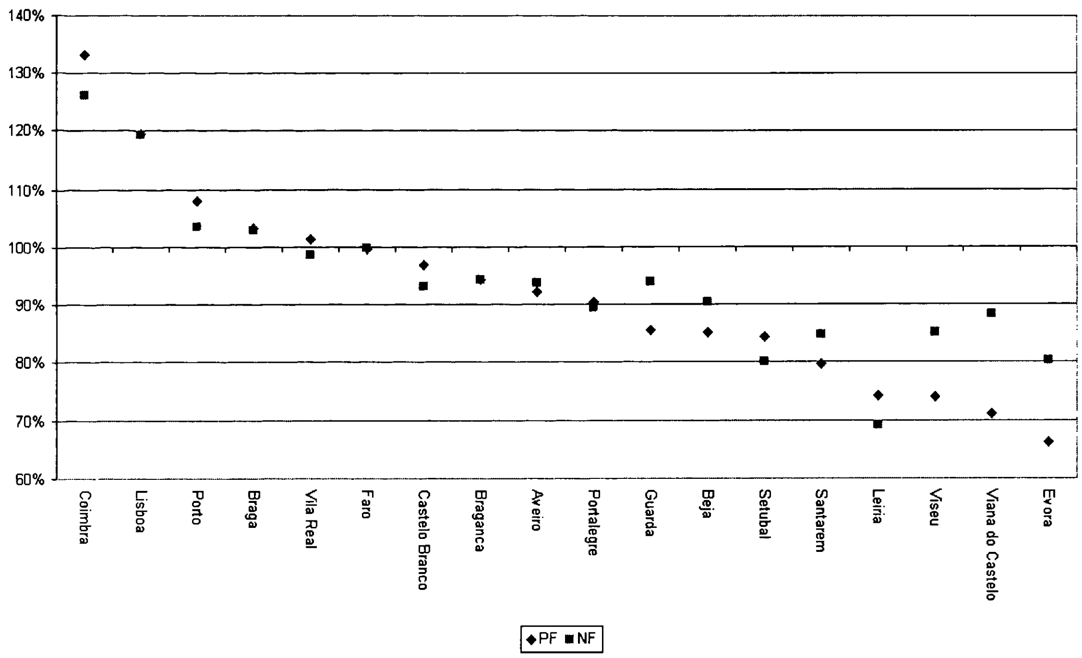
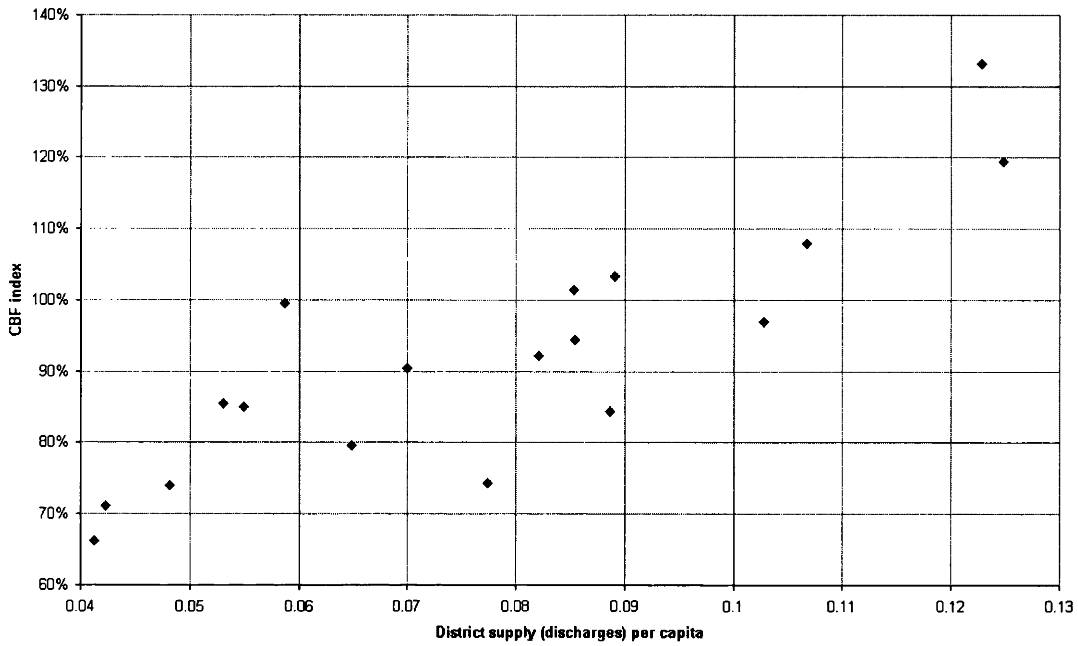


Figure 8.3: CBFs index vs. supply at the district level



8.2.4 Synthesis of all adjustments at the district level

All adjustments of the capitation formula imply legitimate (and relative) variation in hospital costs (at the district level) as defined in Equation 8.7. Implicit in this capitation formula is the multiplicative model defined in Chapter 4.

$$cap\_index_r = I_{2r} * I_{3r} * I_{4r} \tag{8.7}$$

This implies that Lisboa has a level of legitimate variation on costs 59% above the national average, while for Viseu the same value is 51% below the national average (Table 8.4). The three districts that concentrate supply are the ones with higher levels of legitimate variation.

Table 8.4: Adjustments (need, CBFs and UCs) in the multiplicative model

	Multiplied capitation
Aveiro	56%
Beja	89%
Braga	70%

Bragança	79%
Castelo Branco	77%
Coimbra	147%
Évora	57%
Faro	86%
Guarda	59%
Leiria	50%
Lisboa	159%
Portalegre	86%
Porto	118%
Santarém	62%
Setúbal	86%
Viana Castelo	63%
Vila Real	94%
Viseu	49%
Maximum	159%
Minimum	49%

8.3 Results of estimates of inequity indices

Table 8.5 presents estimates of the four indices of geographic inequities introduced in Chapter 4.

All indices compare the current distribution of resources with a distribution that would account for some or all adjustments of the capitation formula. Index 1 compares the district share of hospital doctors with a district share of resources that accounts for population numbers, demographic need and additional need for hospital care. Index 2 compares the district share of hospital doctors with a district share that accounts for population numbers, demographic need and additional need for hospital care and for CBFs. Index 3 compares the district share of hospital utilisation with a district share that accounts for population numbers, demographic need and additional need for hospital care and for CBFs. Index 4 compares the district share of hospital current expenditure with a district share that accounts for population numbers, demographic need and additional need for hospital care, CBFs and UCs of hospital care.

Table 8.5: Geographic inequity estimates

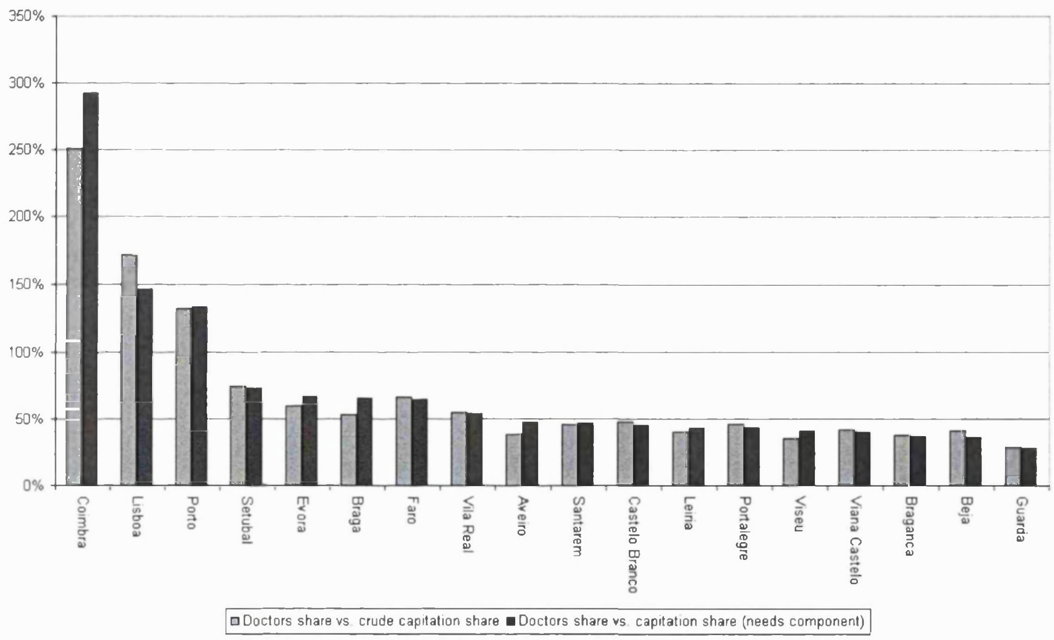
	Index 1: inequities in doctors	*Index 2: inequities in doctors (accounting for CBFs)*	*Index 3: inequities in utilisation*	*Index 4: inequities in finance*
Aveiro	48%	52%	101%	87%
Beja	36%	43%	80%	83%
Braga	66%	63%	104%	82%
Bragança	37%	39%	117%	96%
Castelo Branco	45%	46%	110%	115%
Coimbra	293%	220%	144%	167%
Évora	67%	101%	133%	148%
Faro	65%	65%	85%	95%
Guarda	28%	33%	92%	90%
Leiria	44%	59%	120%	114%
Lisboa	146%	122%	88%	97%
Portalegre	43%	48%	93%	115%
Porto	133%	123%	98%	89%
Santarém	47%	59%	104%	104%
Setúbal	73%	86%	101%	92%
Viana Castelo	40%	56%	104%	96%
Vila Real	54%	53%	107%	91%
Viseu	41%	55%	94%	97%
Maximum	293%	220%	144%	167%
Minimum	28%	33%	80%	82%

Table 8.5 shows that there are huge inequities in the distribution of capital, as measured by the number of doctors (index 1). The populations of Coimbra, Lisboa and Porto benefit from higher accessibility to hospital doctors (and to hospital care resources), with shares 193%, 46% and 33% above their fair share, respectively. Populations from Beja, Bragança and Guarda are the most disadvantaged, with an accessibility to hospital doctors 64%, 63% and 72% below their fair share. This shows the huge extent of inequities in the distribution of hospital capital in Portugal. Only three districts have resources above the national average –Lisboa, Coimbra and Porto, and these are the

urban districts where central hospitals are located and human resources are concentrated.

Figure 8.4 shows that the use of need adjustments increases estimates of inequities in the distribution of hospital doctors (in comparison to crude population numbers as analysed in Chapter 3): the differences between the two columns are due to the use of need estimates in addition to population numbers.

Figure 8.4: Hospital doctors per capita (normalised by national average) vs. inequities in capital (index 1)



After adjusting for expected levels of CBFs (index 2, Table 8.6), Coimbra, Lisboa and Porto still have more than their fair shares of productive capacity (measured by number of doctors) above their fair shares. When accounting for CBFs, Évora appears now with a share of doctors approximate to its fair share. For most districts, expected levels of CBFs do not fully explain variations in the distribution of productive capacity.

Estimates of inequity in utilisation (index 3, Table 8.6) are smaller than inequities in capital (index 2, Table 8.6); and given the expected CBFs, some rural districts have higher levels of utilisation than their fair share, despite inequities in the distribution of doctors/productive capacity (Figure 8.5). Nonetheless, even after accounting for catchment populations, utilisation in Coimbra is still above its fair share. The opposite applies to Lisboa and Porto. These results might be due to: relative difficulty of urban populations in accessing services in Lisboa and Porto, better physical accessibility of rural populations to hospitals in Lisboa and Porto (in comparison to Coimbra), and the different role of the private sector across districts.

Figure 8.5: Inequities in utilisation (index 3)

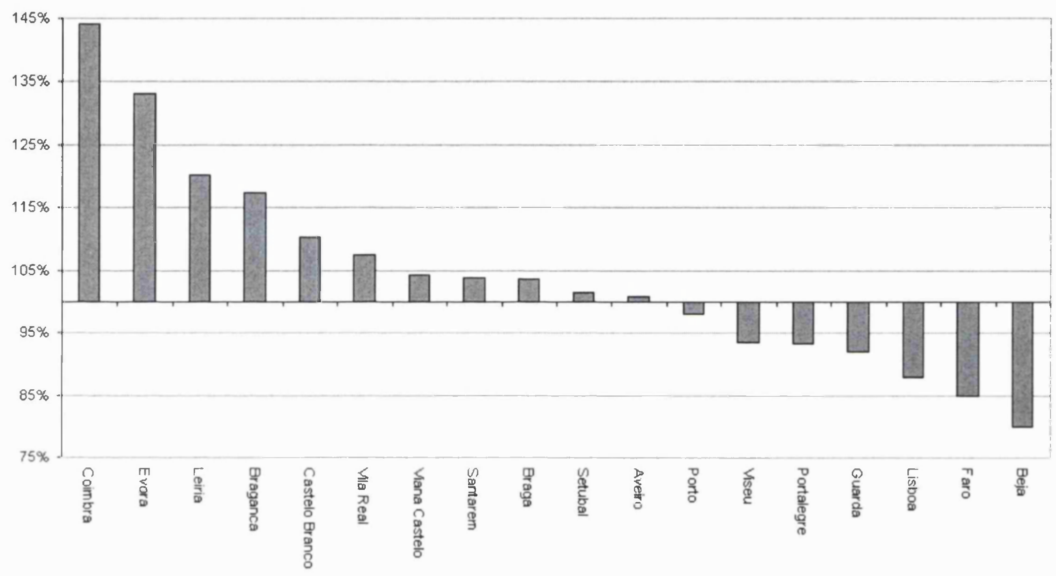
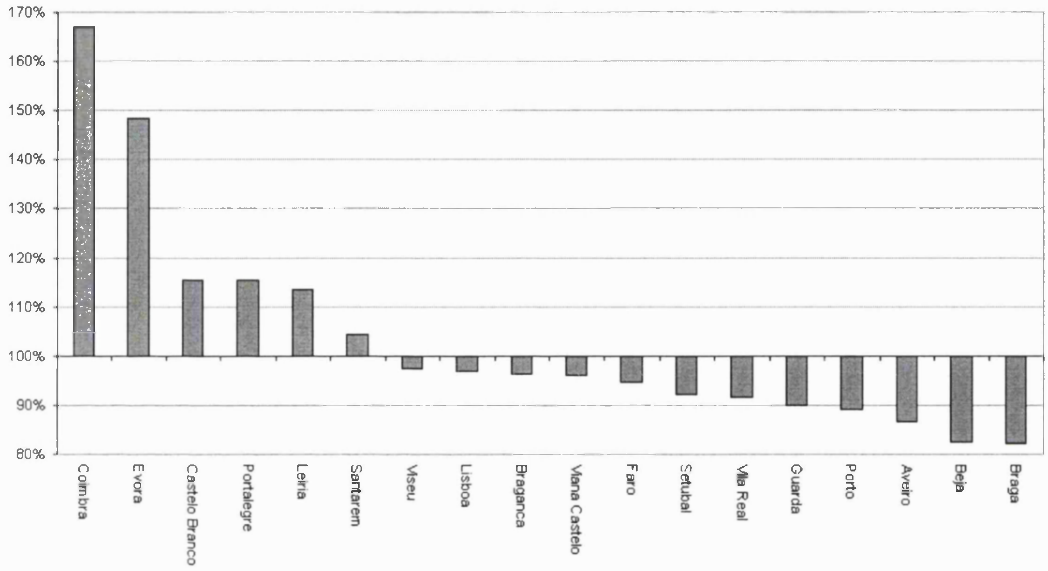


Figure 8.6 shows the extent of inequities in finance (i.e. in current expenditure, index 4, Table 8.6). They differ from inequities in capital and in utilisation as index 4 also accounts for UCs. Six districts use more than their fair share of financial resources with Coimbra being the most extreme case. The other five districts are mainly rural and located in the central and southern regions. Figure 8.6 shows that after controlling for need, UCs and CBFs, Lisboa and Porto are using less than their fair share of financial resources. But these results have not taken account of the role of the private sector, which would be expected to show that these districts in fact are using more than their fair share of resources. The situation in certain under-supplied districts (in terms of capital) also differs: while Beja is presenting low shares of doctors, utilisation and supply, Évora is under-supplied in doctors but is receiving current expenditure and is using resources above its fair share.

The consequences of these results are discussed in the next section.

Figure 8.6: Inequities in finance (index 4)



8.4 Discussion and concluding remarks

This section is organised in three sub-sections: consequences for Portuguese policies, discussion of methods and concluding remarks.

8.4.1 Consequences for Portuguese hospital policies

Current systems of finance and investment have not addressed the need to correct for geographical inequities:

a. Coimbra always appears as an outlier with substantially more than its fair share of resources.

b. The populations of Porto and Lisboa have more than their fair shares of capital in relation to their resident populations and relative need for hospital care, but less than their fair shares in terms of utilisation and finance.

c. Some districts have a share of capital below their fair shares but utilisation and current expenditure above their fair shares[243], while other districts have shares below

[243] That applies to Castelo Branco, Évora, Leiria and Santarém.

their fair shares in all indices except utilisation[244], one district has shares below their fair share in all indices except in current expenditure[245];

d. Some districts have shares below their fair shares for all the indices (Beja, Faro, Guarda and Viseu).

Depending on the importance attached to different types of inequities, different redistribution policies might be formulated. Analysis of inequities should acknowledge the weaknesses of each of the inequity indices in use: for example, although inequities in utilisation are smaller than the ones for capital, they assume that it is acceptable for populations to travel long distances in order to use hospital services.

Coimbra differs from Lisboa and Porto. Due to historical developments, all these districts concentrate hospital supply which represents inequities in access. Nevertheless, there have been substantial movements of populations to Lisboa and Porto and improvements in transportation, in particular, which have resulted in these districts having a supply share in excess of the needs of their resident populations, but utilisation and current expenditure shares below their fair shares. But Coimbra is different and has an inequitable high level of supply and utilisation.

If the capitation formula –including the needs, UCs and CBFs adjustments -were to be used for resource allocation to the district level, it would create incentives that are not compatible with progressing towards a distribution of capital in line with needs: Lisboa and Porto would benefit from the system, which would create incentives for greater concentration of supply (which is incompatible with the correction of inequities in capital), to maintain the largest (and most expensive) hospitals and to make people having to travel to access hospital care in these districts. This shows how the pursuit of different concepts of equity leads to different results and that the use of UCs and CBFs in a capitation formula to allocate resources can create incentives to perpetuate current inequities in the distribution of hospital capital.

The distribution of hospital productive capacity (measured by the number of hospital doctors) is also out of line with needs. Total hospital supply is close to the average for the EU in terms of per capita provision and there are no funds for building new

[244] That applies to Aveiro, Braga, Bragança, Setúbal, Viana do Castelo and Vila Real.
[245] That is the case for Portalegre.

hospitals (described in Chapter 2). Thus, it seems that correction of geographic inequities on capital must be done mainly by redistribution.

The resource allocation system for hospitals does not account for legitimate variations on hospital costs, as described in Chapter 2 and shown by inequities in finance. If geographic equity is to be achieved, the funding system should be redesigned. If the capitation formula is to be used to allocate resources, research is needed on developing appropriate financial incentives to progressively move towards a more equitable distribution.

The estimates of resources on capital, utilisation and current expenditure are subject to three important caveats. The hospital sample (used to compute current expenditure and number of doctors) excludes certain types of hospitals that are mainly located in Lisboa and Porto (mainly private hospitals as there were no data on these); the capitation formula has made no adjustments for alternative or complementary services (such as social care); and there was no detailed data to compute the impact of double-covered population on estimates. The problems of accounting for the influence of the private sector and for the impact of double covered populations were emphasised in Chapter 4. Their resolution requires further analysis of the role of the private sector in the Portuguese health care system. The finding that the residents of Porto and Lisboa use less than their fair shares in terms of utilisation and current expenditure may well be due in part to high levels of supply by the private sector in these two districts.

8.4.2 Discussion of methods

The above results show that the use of different measures of equity produces different inequity estimates. Chapters 1 and 3 illustrated the complexities of theoretical and empirical analysis of equity. Results from this chapter give more evidence of these complexities: different concepts give quite different estimates of inequities and hence would suggest different policies to achieve equity. Nonetheless, districts might be grouped under similar patterns of inequities (for example, districts with unfair shares in the four indices).

The use of more sophisticated methods to measure need in relation to crude population has changed the range of variation of inequities (increasing the range for Portugal) and shown that the use of crude populations is inadequate.

The use of the district level for geographic unit hides intra-district heterogeneity. District areas are not homogeneous and within districts there are variations in access to hospital care. Hence, the ecological fallacy might apply. The use of district estimates enables a start to be made on correcting inequities but there is also a need for studies of inequities within districts.

As expected, adjustments for CBFs and UCs tend to reduce the relative use of services by populations in those districts where supply is concentrated. But accounting for UCs and CBFs improves inequity estimates for districts that have concentrations of supply, and were highly sensitive to the choice of methods, in particular to the estimation of CBFs.

8.4.3 Concluding observations

The application of capitation formulas shows considerable inequities in the distribution of hospital resources. Central hospitals and hospital human resources are concentrated in three urban areas. Estimates of the fair shares of resources in accordance to need can be used to inform future policies on capital investment and deployment of staff.

The current financing system does not create incentives to tackle current inequities. Hospitals are funded on an incremental basis and investment policies have not been informed by evidence of inequities. An implementation of a capitation formula to redistribute resources on current expenditure would mean a significant redistribution.

The final point, of accounting for the private sector, is likely to become increasingly important. This concerns both use of services outside the NHS and the complex mix of insurance in Portugal. The difficulties in obtaining information mean that estimates of inequity were underestimated. This neglect in official statistics of health insurance and care outside the NHS may have had some justification when there was at least a pretence that this was a transitional position for a sector of declining importance. What now seems to be happening, however, is increased recognition of the role and importance of the private sector, in Portugal and in other countries. Under these circumstances, governments can no longer maintain a pretence of the transitory nature of the private sector and it is vital that systems are developed to provide information on its coverage and supply, if we are to maintain and develop policies to achieve equity of access to health care within pluralistic systems.

This chapter has reported estimates of inequities at the population/district level and shown which type of inequities are operating for each district. The next chapter addresses the question: how can we redistribute hospital supply to best improve equity of utilisation and access? It makes use of location-allocation models that consider alternative policy objectives and account for different assumptions on patients' choice of hospitals. The models work at the small area level, and the context for developing this type of study is stated in Chapter 9.

SECTION III

9 CHAPTER 9 – Location-allocation models to redistribute hospital supply

9.1 Objectives

This chapter aims to identify how to improve equity in access and utilisation of hospital care by redistributing hospital supply. This is a key issue in Portugal because:

1. There are inequities in the distribution of hospital resources in relation to the population and in actual and fair capitation shares of resources between districts (Chapters 3 and 8);
2. Investment decisions of the MoH on hospitals have been made without being informed by relevant evidence (Chapter 2);
3. As there is no prospect of increasing the total resources available (Chapter 2) future hospital policies are expected to focus on redistribution and on marginal changes (including replacement of current hospital capital);
4. Changes in hospital supply are difficult and slow to implement, require long-term planning, and are constrained by other resources. It is thus essential to identify changes that are feasible and marginal.

This chapter develops models that test how redistribution of hospital supply can be used to target improvements in equity of access and utilisation. This chapter uses three formulations of the concept of equity of access and utilisation, which are compatible with stated objectives in Portuguese health policy. Those three formulations correspond to three alternative models that not only relate to different concepts of equity but also test different assumptions about patients' behaviour and illustrate the impacts of different assumptions on methodological choices (as a type of sensitivity analysis of results in the use of alternative models).

Chapter 5 produced relative needs-based estimates that may be interpreted as indicative of the optimal distribution of hospital resources in proportion to relative need for hospital care. But these estimates did not take account of information crucial in changing the existing distribution of supply. To do this, it is vital to take into consideration the characteristics of the current hospital network and of the local mix of the primary and private care sectors. Furthermore, even if equity of access for those in equal need were achieved, there might be major inequities in utilisation, given the behaviour of the population when using hospitals. It is necessary to produce estimates of changes in hospital supply at the hospital level (needs-based estimates were produced for the district level).

This chapter develops estimates for redistribution of hospital supply at the hospital level that aim at improving geographic equity of access and utilisation. As the problem is one of redistribution and the current characteristics of the system should be taken into account, the proposed models for determining marginal improvements follow a second best approach[246]. The estimates for redistribution of hospital supply produced in this chapter (computed at the hospital level and aggregated to the district level) are compared with the district estimates of need for hospital care produced in Chapter 5 (and analysed in Chapter 8): it is desirable that both estimates (estimates are computed under different policy objectives) lead to the formulation of compatible policies for redistribution.

The proposed models assume a NHS institutional setting where the MoH has the power to enforce or influence changes in hospital capacities. Inequities in utilisation also relate to the hierarchical structure of Portuguese hospitals and to the concentration of resources in coastal and urban areas, and the MoH has power to change these 'policy variables'. The models attempt to account for these characteristics of the hospital system that seem to be resulting in significant variation in access and utilisation.

This chapter starts with a review of the relevant literature. This indicates that previous models for the redistribution of hospital supply have been weak in three respects: in

[246] An optimal (first best) approach would consider how to determine the optimal level of supply in accordance with principles of geographic equity, as if a new network of hospitals was to be built. A second best approach makes use of information about the current distribution of hospital care provision

modelling patients' behaviour, in modelling interaction between the use of alternative hospitals, and in accounting for the process of demand for hospital care. The following section presents the experimental design used in building redistributive models and sets an analytical framework for the analysis of outputs, so as to compare implications of alternative models. The next section then proposes a set of alternative models to analyse hospital redistribution, addressing the weaknesses of previous models. One of these models uses the information from the FDM developed in Chapter 7 so as to represent patients' behaviour in the choice of hospitals. The next section applies the models to the Portuguese hospital system, and a final section summarises the main arguments of this chapter.

9.2 Literature review

This review describes location-modelling literature and frames the choice of a modelling approach. It is structured in three sub-sections which discuss: methodological issues in the literature on location, multi-spatial objectives and available models, and specific issues in the application of models in the area of health care, as well as challenges and objectives for new models in the area.

9.2.1 Methodological issues for location literature

Two different modelling approaches have been used to inform decisions on locations and capacities between facilities located in different geographic areas in geographic and operational research literature: location and location-allocation. *Location models* optimally locate systems of facilities for a defined set of providers, but consumers' responses to location factors are made independently of provider conditions (Rushton 1987). *Location-allocation models* endeavour optimally to locate systems of facilities and to allocate simultaneous consumer demand to those facilities[247] (Hodgson, Rosing, and Storrier 1996). Location and location-allocation models differ in how they interpret consumers' decisions on choice of provider (Love, Morris, and Wesolowsky 1988), as

---

and the characteristics of the hospital network, and defines directions for change that improve geographical equity, in the context of the existing distribution.

[247] Consumers' responses depend on provider conditions.

shown in Table 9.1. Since in the health sector the behaviour of patients with regard to changes in hospital location and size is a key element for analysing variations in utilisation, the approach of location-allocation models is obviously better.

Table 9.1: Location and location-allocation models vs. variables to be calculated within the models

Models	*Individual hospital supply*	*Individual demand*	*Total demand*
Location	Calculated within model	Independent of supply	Fixed total demand
Location-allocation	Calculated within model	Dependent on supply	Calculated within model

To adequately solve a location-allocation problem a model with four elements is required (Ghosh and Rushton 1987)[248]:

- Decision criteria on the objective or objectives (where there are multiple criteria) to be attained;
- Rules for consumer behaviour with respect to the spatial choice of hospital;
- A representation of the environment, such as the choice of the geographic level of analysis, with the need to determine travel costs, times or distances;
- A choice between deterministic model and stochastic model.

Most of these analytical choices entail judgements (Mandell 1991), which are discussed below.

9.2.2 Multi-spatial objectives

Objectives for location-allocation models differ for the public and the private sectors (Erlenkotter 1983) (Current, Min, and Schilling 1990): for the private sector the objectives are efficiency and minimising cost (for profit maximisation and for meeting client demands which are assumed to be fixed); for the public sector (as for the Portuguese NHS) the objectives are improving equity and reducing access costs to potential consumers (Rahman and Smith 2000).

[248] These elements of a location-allocation problem can be converted into a generic mathematical formulation, where: the objective is to minimise a multi-criteria/single-criteria function: $\{f(x): x \in Q\}$; there is a map $f = (f1, f2, ...fm)$ that converts the decision space $X = R^n$ into the criteria space $Y = R^m$ and that captures a set of constraints of the system; a feasible set of location patterns is defined inside the

Several objectives have been used in location-allocation models for public facilities. In a review of the literature, Current et al. (Current, Min, and Schilling 1990) identified two main objectives: minimisation of accessibility costs (traditional objective, whether defined by distance, travelling costs, etc) and maximisation of demand coverage (such as to minimise variations in proportions of population covered by public services across areas). For cost minimisation objectives incorporated include: minimising the sum of distances to be travelled by users, minimising the maximum distance to be travelled by users and minimising the number of facilities. For coverage of demand objectives incorporated include: maximising the demand assigned to a facility and maximising the demand covered (Current, Min, and Schilling 1990).

These objectives pursued by location-allocation models have been applied at different geographic levels and/or for different population groups. Different indices of equity have been used as objectives implying different models of preference in the distribution of resources, depending on the concept of equity, the weight to be given to different population groups, etc (Erkut 1993) (Marsh and Schilling 1994) (Kostreva and Ogryczac 1999) (Ogryczak 2000). Marsh and Schilling (Marsh and Schilling 1994) have summarised a set of properties that ought to be satisfied by equity indices. These include: analytical tractability for problem size and computational requirements; appropriateness of interpretation; not discriminating between the (geographic) groups being evaluated; and the principle of transfers[249]. As different equity objectives might lead to different results, this chapter examines the impact of different models on the analysis of hospital supply.

Different studies have defined efficiency in terms of travel distances and time for users, or maximisation of demand coverage (Mayhew and Leonardi 1982) (Cho 1998), or total user travelling costs (Hansen, Peeters, and Thisse 1983). However, this study follows the health care literature that includes these definitions as equity definitions.

full set of possible locations: $Q \subset X$; and $x \in X$ is taken as the vector of decision variables (for example, hospital capacity).

[249] The role of the principle of transfers in the literature on equity indices is the following: the measure of inequity decreases when there are transfers from the best to the worst-off groups.

9.2.3 Spatial models

Three main methods have been used for analysing geographic distribution and redistribution of public facilities (including hospital supply): *spatial interaction, entropy* and *mathematical programming models*[250]. These models represent variations in the elements described above that: characterise a location-allocation model (for example, in the decision criteria in use); use distinct types of information on spatial behaviour; and have different potentials for describing or predicting behaviour. A brief description of the strengths and weaknesses of each group of models follows.

9.2.3.1 Spatial interaction models

Spatial interaction models (SIMs) (such as gravity models) were introduced in Chapter 7. They constitute a form of probability interaction modelling (O'Kelly 1987), and use information on: population numbers, hospital size, distance and a decay function to reproduce flows of consumers (or patients). A summarised presentation of the assumptions incurred and of the problems that arise from the use of SIMs (in particular, when used for prediction) has been given in Chapter 7. Gravity models can be seen as a variant of MP models – when the cost exponent in a gravity model tends to infinity (that is, the elasticity of utilisation to distance in the decay function is infinite), the total distance travelled tends towards to a minimum, and the trip distribution tends to a linear programming assignment with patients travelling to the closest point (O'Kelly 1987). As described in Chapter 7:

a) SIMs have been used as reliable models for replicating the current pattern of patient flows between demand and treatment zones (Cho 1998) but are inadequate in predicting user flows in response to supply changes (McLafferty 1988) (Porell and Adams 1995). Despite this, they have been used for that purpose (Hallefjord and Jornsten 1984) (Mayhew, Gibberd, and Hall 1986) (Taket 1989)[251] (Brown 2001).

[250] Evaluation and appraisal techniques such as cost benefit analysis, cost effectiveness analysis and cost utility analysis models (broadly analysed in Drummond et al. (Drummond et al. 1999)) are seen as not appropriate for comparing alternative improvements to the hospital network. This is because these techniques would imply comparison of multiple scenarios of redistribution, and would not provide a simple tool for analysing changes in a network of hospitals.

[251] For example, in this study, gravity models were used as a simulation model in order to explore different options for the future provision of inpatient hospital facilities at the English Regional Health Authority level (Taket 1989).

b) When SIMs are used for prediction, they assume that when there are changes in one hospital all the other hospitals gain in comparison with their shares of utilisation prior to that change. Empirical evidence shows this assumption to be false (McLafferty 1988).

c) As SIMs operate at the aggregate level, they consider neither local variations (from the local health system), nor the hierarchical and organisational structure of hospitals (for example, they do not consider whether there are tertiary referral hospitals in a population area).

d) SIMs require a definition of a decay function. Building the decay function requires a choice of function for different groups of hospitals (exponential, power or Tanner being the most common functions), which copes with methodological problems (McLafferty 1988). SIMs assume that populations use hospitals in accordance with the decay parameter and that both urban and rural populations are able to use both urban and rural hospitals. However, evidence for Portugal shows that mobility of urban populations to rural areas is not verified in practice[252].

However, as discussed in Chapter 7, there is a scope for improving SIM models, for example by developing unconstrained gravity models that have already been used in other areas of the literature.

9.2.3.2 *Entropy models*

Entropy models (EMs) have been used in different contexts and disciplines (such as thermodynamics, statistical dynamics, statistics and information theory (Wu 1997) (Fang, Kajasekera, and Tsao 1997) (Arndt 2001)) and are a type of mathematical programming model that makes use of the first principle of data reduction (Wu 1997): when there are incomplete data, the solution must include and be consistent with all relevant available data. Mainstream EMs replicate the macro-properties of the system with information on users, on levels of supply and on accessibility costs among users[253].

---

[252] Analysis of the database used in Chapter 7 shows that almost 100% of population of urban areas are attended in urban hospitals, and thus do not seem to be very willing to use rural hospitals.

[253] The generic formulation of the origin and destination constrained entropy model is: maximise $-\sum_i \sum_j U_{ij} * \ln U_{ij}$, subject to $\sum_i U_{ij} = D_j, \forall j$, $\sum_j U_{ij} = O_i, \forall i$ and $\sum_i \sum_j c_{ij} * U_{ij} = C$ (with c_{ij} being a measure of accessibility costs incurred in travelling from i to j, O_i the number of discharges for population living in i, C the total costs of the system, and other notation as interpreted in previous chapters). This model has a correspondence to a double constrained gravity model.

EMs (Webber 1978) emphasise short-term prediction (given fixed supply); do not take account of the determinants of individuals' decisions (i.e., consumer choice); and depend highly on the formulation of consumers' travelling costs. EMs are adequate when there is a lack of theoretical understanding of what to include in the model (Anas 1983). Also, they avoid the assumption of micro-economic models based on utility theory[254], may be derived from a theoretical set-up and examine only small components of the decisions of individuals (Webber 1978). The weaknesses of EMs lie in that they follow a holistic view, which imposes system constraints that do not consider the ways in which groups make spatial choices (Nijkamp 1978) –i.e., they do not account for the process of health care demand. Some variants of the entropy model are shown in (Erlander 1977) and (Hallefjord and Jornsten 1984).

9.2.3.3 *Mathematical programming models*

Mathematical programming (MP) models have been widely used for locating and allocating public facilities. They maximise a certain kind of equity concept (for example, an equity index based on distance travelled by users), assume some type of user behaviour (for example, patients travelling to the closest hospital), and use constraints that reflect characteristics of the health care system. Table 9.2 presents (and explains) a number of choices of MP models based on some technical variants that are relevant for the public sector: the choice between a single or a multiple objective function; optimal or non-optimal solution models; and nested or hierarchical models. Multiple objective MP models might be more realistic but entail several judgements on weighting the criteria and applying complex algorithms (Martin et al. 2000) (Gonzalez 2001). Heuristic models are required when the algorithms built are not efficient or unable to compute an optimal solution in reasonable time. Hierarchical models capture variations in the facilities network, deploy a complex formulation and often require heuristic models to arrive at a solution (heuristic models are defined in Table 9.2).

[254] EMs are an alternative to micro-economic models. Micro-economic models depart from the rational theory of consumer behaviour, make assumptions on utility functions, such as patients maximising a utility function based on health, income, time, etc, and also use a set of budget constraints.

The strengths of the MP approach are: flexibility in the choice of the objective function and the constraints on the system, and MP ability to provide a global solution[255]. Weaknesses of MP models have included the use of crude assumptions about users' behaviour, such as users travelling to the closest facility (Current, Min, and Schilling 1990); also the modelling and structure of MP has been highly restricted by the assumption of linearity, although mathematical programming is still capable of producing efficient approximation models.

Table 9.2: Variety of MP models for public and health care facilities location

MP models under several classifications	*Description*	*Examples from literature*
Single- vs. multiple objective models	*Single objective models* minimise a single objective, such as: distance from users to the closest provider point, or maximum distance; or total travelling costs.	Single-objective models (Mohan 1983)
	Multiple-objective models consider several objectives; examples of multiple objectives models are goal programming, multi-criteria utility models or bi-criteria models[256].	Multi-objective models (Rushton 1987)
Optimising vs. heuristic models	*Optimising models* provide an optimal solution.	Optimising models (Mohan 1983)
	Heuristic models produce satisfactory results that may not be an optimal solution, but a second best solution (given technical difficulties in solving these models).	Heuristic models (Bennet 1981)
Nested vs. hierarchical models	*Nested models* represent facilities at different hierarchical levels and treat all the facilities of the hierarchy as if they were at the same level (i.e. as if no hierarchy applied) assuming that all the facilities provide the same set of services.	Nested models (references in (Marianov and Serra 2001))
	Hierarchical models represent a referral system, such as top-down and bottom-up models of referral formulation; and might account for total travel costs incurred in accessing different levels of facilities.	Hierarchical models (Church and Eaton 1987) (Rahman and Smith 2000)

255 As shown below, some MP models such as quadratic models might produce local and non-global optimums (in some cases).

256 Goal programming models allow decision-makers to assign weights to the realization of each goal that is included in the objective function and carry out sensitivity analysis. Multi-criteria utility models quantify trade-offs and test alternatives under changes in the objective function (Mayhew and Leonardi 1982) (Cho 1998). Bi-criteria models use a constrained approach to multi-objective programming (Mandell 1991).

Specific versions of the MP, EM and SIM can be applied and have previously been shown to produce similar results, mainly when the models are used to reproduce behaviour[257]. Gravity models have been shown to be the result of an entropy mathematical formulation (Wilson 1970) (examples shown in (Fang, Kajasekera, and Tsao 1997)), and when the travelling costs in a system tend to infinite, then the gravity model tends to a MP model, as described above (O'Kelly 1987).

9.2.4 Specific issues in health research

Some location-allocation models have been applied to the health sector, and their objectives and context of application have influenced the structures of the models and the selection of techniques. The choice of model depends upon prior beliefs with regard to *demand for health care function, the equity objectives to be pursued and the constraints that characterise the health care system.*

A variety of studies have been applied to health care focusing on different aspects. For example, MP hierarchical models have been tested for combinations of facilities in two tiers of a health care system. They have modelled a referral system with top-down and bottom-up models of referral formulation (Church and Eaton 1987) (Rahman and Smith 2000). Some MP models have calculated the optimal spatial distribution of future hospital capacity, assuming a specific pattern and distribution of demand for hospital facilities: that is the case for users travelling to the nearest facility (Mohan 1983) and for models minimising total travel or total transportation and facility costs incurred in accessing different levels of health care facilities (Church and Eaton 1987). Some models have determined the location, number and size of centres of health care supply, using heuristic techniques, minimising the aggregated user costs as measured by distances travelled, and imposing constraints on the maximum allowable travel distances or the service capacities of the supply centres (Bennet 1981). SIMs have also been used to compare the impact of pursuing different (and alternative) objectives, such as equity of utilisation and of access (Mayhew and Leonardi 1982). Nevertheless, none

[257] In the context of description of flows, these models can also be translated into random utility models: the SIM/gravity model may be applied from random utility theory (characterised by rational choice behaviour) (Williams and Senior 1978), while a maximum entropy formulation can be a starting point for building a logit/utility theory model (Jornsten and Lundgren 1989).

of the available models in the literature is suitable to meet the objectives set in this chapter.

All earlier methods (applied to the health sector) suffer from a number of weaknesses. First, most of the models in use are location models, which have assumed simple rules of allocation of demand and have not adequately captured patients' (or doctors') behaviour. For example, Cho (Cho 1998) has pointed out that previous models for locating medical facilities have not questioned their modelling structures in terms of assumptions made about the utilisation pattern of medical facilities. Models that use crude assumptions about the criteria of patients' choice of hospital cannot be used to predict changes in utilisation (Bennet 1981) (Leonardi 1983) (Mohan 1983) (Rushton 1987) (McLafferty 1988) (Avella et al. 1998). Moreover, most models have ignored interaction between hospital supply and utilisation of alternative hospitals (Porell and Adams 1995)[258] which is vital for prediction purposes. Although gravity models deal with interaction, they do so in an unsatisfactory way (as described in Chapter 7 and in the literature review of this chapter). Furthermore, models of hospital supply have not taken account of the characteristics of the health care system in estimating demand for hospital care, nor of how the process of hospital demand is formulated. Thus, none of the previous studies meets the stated objectives of the chapter, and a different set of models that give some insight into the problem of redistribution of hospital supply is required.

9.3 Experimental design

The key objective of this chapter is to tackle the problem identified by Rushton (Rushton 1987) that location-allocation models need to identify current behavioural patterns and must define the main goals to be achieved by providers and/or consumers.

[258] Traditional gravity models respect the property of independence of irrelevant axioms, in that the flows to any destination are independent of other destinations. A recent study from Congdon (Congdon 2001) (study recalled in Chapter 7) has improved the interaction mechanism of gravity models in the context of modelling emergency flows. As described in Chapter 7, he has adapted a gravity model that is more responsive to changes in the patterns of supply, using Bayesian methods to re-estimate some of the parameters of the gravity model (in order to represent new accessibility scores given supply changes), and afterwards re-running the gravity model with the new parameters. This approach, however, requires local knowledge to specify the new parameters for changes in supply and is thus difficult to use for other than small local studies.

It is assumed that information on patients' behaviour is crucial to identifying mechanisms for inducing desirable changes.

The approach developed is a *multi-modelling one* based on *different mathematical programming models* with *single objective functions*[259]. As observed above, the MP approach is flexible in the choice of objective function and of constraints of the health care system and on limits to redistribution. A multi-modelling approach corresponds to the use of a set of models for analysing the redistribution problem (see Table 9.3). The single objective function is based on equity improvements in only one dimension, as multiple-objective functions create problems in the solution of the algorithms. The focus here is to improve current models with better information on patients' behaviour (this also implies changes in the algorithms in use).

The three models have distinct (and alternative) equity objectives, make different interpretations of the redistribution problem, and use different assumptions about the utilisation behaviour of patients with regard to different levels of hospital supply. Three models are justified below. Each model implies different conceptualisations of the health care system (e.g. the connection between hospital and primary care) and uses different constraints on redistribution (such as using a lower bound for change in individual hospital capacity). Each model also uses different concepts of equity. Any single objective function tends by its very nature to be incomplete (Rushton 1987) as policy makers pursue several objectives. Analysis of results of models that pursue different objectives indicates the degree of similarity of results using different assumptions, and may be seen as an exercise in sensitivity analysis.

Table 9.3: Three MP models

Model	*Abbreviation*
Distance-based model	DBM
Utilisation-based model	UBM
Utilisation flows-based model	UFBM

[259] The use of a SIM approach (more specifically, the gravity model) for predicting hospital utilisation was excluded because of its pitfalls in the context of prediction. EMs lack the flexibility to integrate more complex information on the health care system and on the process of demand for health care.

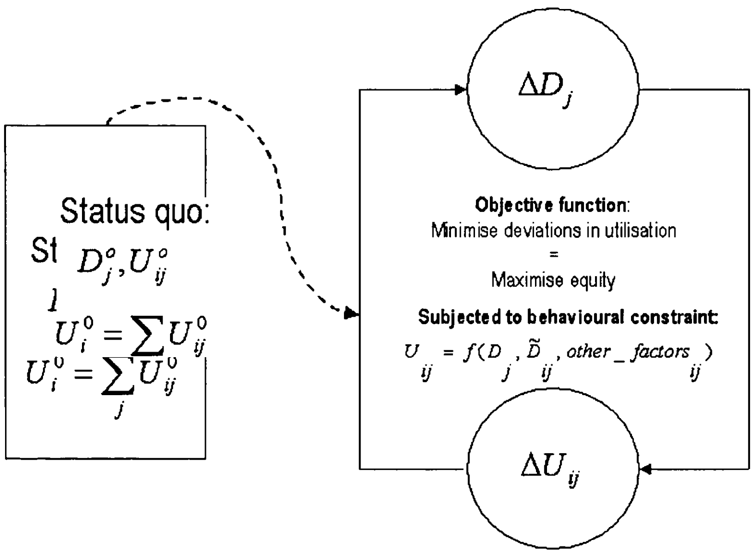
The modelling rationale is now described throughout the presentation of the generic structure and the reasoning for the use of a mixed modelling strategy, a description and comparison of the three MP models, and the definition of a framework for analysis and comparison of outputs of the models.

9.3.1 Generic structure

Figure 9.1 presents the general structure of the MP models, organised according to the following principles (the notation used is consistent with that of previous chapters):

1. The starting point is information on the *status quo* with regard to location of hospital facilities and on the current patterns of utilisation, as the models aim to redistribute hospital supply (D_j^0, U_{ij}^0 and U_i^0 represent current values of these variables - notation in use is explained below in Table 9.5);
2. The objective is to improve equity in the utilisation or access of public hospitals by the redistribution of public hospital supply (D_js). The D_js are the decision variables;
3. Changes in the D_js imply new levels of utilisation flows (U_{ij}s).
4. The distribution of new levels of U_{ij} must be analysed according to a set of desirable equity criteria. As shown in Chapters 1 and 2, different objectives of geographic equity have been promulgated in policies and these objectives correspond to different measures and indicators. These criteria are used as objective functions of alternative models, and they are synthesised in the framework for analysis of hospital outputs described below;
5. The underlying behaviour of patients is made explicit in the three models, with behaviour represented in different constraints of the MP models. Each model aims to capture the process of demand for hospital care in a different way, given current system characteristics.

Figure 9.1: Simultaneous spatial redistribution of hospital supply and variations of utilisation flows in a location-allocation model



The three models were chosen for the following reasons:

1. The distance-based and utilisation-based models (DBM and UBM) are based on previous models that have been described above. These models have been adapted to meet the specific aims of this chapter. The DBM changes the assumption that patients make use of a single hospital and hence is an improvement on the nearest centre model, which minimises distances travelled by assuming that patients use the closest hospital. The UBM makes use of information on patient travelling behaviour taken from the gravity model. The version of the gravity model used is defined in Appendix E. In comparison to previous MP models that have used SIM model information within a MP model, the UBM differs in that it uses an objective function that is often evoked by policy-makers (in that it minimises variations in utilisation rates by population area). The assumptions of the SIM model were discussed in section 9.2.3.1.

2. The utilisation flows-based model (UFBM) takes a quite different approach to existing models. It focuses on realistic assumptions about patients' utilisation behaviour in response to changes in hospital supply levels; and accounts for the process of demand for hospital care. The UFBM makes use of the FDM for predicting hospital utilisation (described in Chapter 7), and uses a distinct objective function based on a principle of location accessibility.

The development and choice of models is guided by the need to find algorithms that produce efficient solutions. This is a problem both because of the computational burden imposed by the large number of potential locations and population points (68 and 275, respectively, in Portugal) and the difficulty of solving certain types of models (for example, difficulties of solving MP models with integer variables).

A framework is developed for the analysis and synthesis of outputs from the three models. This framework is useful for comparing the proposed levels of supply and utilisation with respect to their degree of achievement in terms of different policy objectives and for assessing trade-offs in pursuing different objectives.

The models are described and compared in detail in the following sub-section, followed by the framework for analysis of outputs.

9.3.2 The three alternative models

The three models deployed share a set of features for the modelling of redistribution. They structure a multi-hospital system with hospital facilities 'competing' for limited capacity fixed at the current level[260]. They also use a predetermined and finite set of potential locations based on current hospital sites. In fact, they can be interpreted as optimisation models (optimising an equity objective) or as simulation models (testing the consequences of changes in hospital supply on the objective function). The three models can be compared across four characteristics, structured in four columns in Table 9.4:

1. *Whether they are location or allocation models*;
2. *Choice of equity index in the objective function.* There are various possible indices for summarising geographic variations that represent different preference models. As noted in the literature review section, the choice of objective function ought to take into account the principles which a measure of equity should obey (for example, all the indices respect the principle of transfers, as described above);

[260] As of 1999 in the Portuguese context.

3. *Use of normative vs. prescriptive assumptions for patient behaviour*: normative studies seek to describe ideal patient flows based on social criteria, such as the minimisation of patient travel times (Folland 1983), while prescriptive models integrate patients' behaviour with regard to supply changes;
4. *Use of constraints that impose the redistributive nature of the model and choice of a mechanism for capturing the interaction* between hospital supply and utilisation levels of alternative hospitals. For example, the total supply is fixed in all three models, while limits to levels of redistribution (per hospital) only apply to some models.

Table 9.4: Qualitative description of MP models

Model (and typology)	*Key equity concept*	*Patients' behaviour constraint*	*Supply interaction and other system constraints*
DBM -Distance-based model Location model	Equity of access: minimisation of total distances travelled by patients to (closest) hospitals.	Assumption that patients travel to closest hospitals, and patients are allocated to three types of hospitals in accordance with past quotas (normative).	No interaction between hospital supply and utilisation of alternative hospitals. Patients make use of three hospitals: one central hospital and two other hospitals of the network.
UBM -Utilisation-based model Location-allocation model	Equity of utilisation by population area: minimisation of sum of variations between predicted and normative utilisation per population area (according to need).	Assumption that patients in each geographic area use hospitals based on fixed conditional probabilities (prescriptive).	No interaction between hospital supply and utilisation of alternative hospitals. System constraints: upper and lower limits to variation in the supply of each hospital; fixed total supply. Conditional probability of use of hospital generated by a gravity model.
UFBM -Utilisation flows-based model Location-allocation model	Equity of utilisation flows between population areas and hospitals with the equity target defined as: flows if patients were treated in the closest hospital.	Assumption that patients in each geographic area use hospitals in accordance with the FDM developed in Chapter 7 (prescriptive).	Interaction constraint as captured by the FDM constraint. System constraints: upper and lower limits to variations in the supply of each hospital; fixed total supply. Fixed probabilities of a population making use of a hospital (prediction taken from the first part of the FDM)

A brief description of the reasoning behind the use of each model follows. Each of these models is described in more detail in the following section. The notation used is presented in Table 9.5; as in previous chapters, i and j stand for population point and hospital site.

Distance-based model (DBM). As noted in the review section, classic objective functions from the location literature are too restrictive and unsatisfactory for applications in the health care area, being restricted to objectives of minimisation of total or maximum distances, and making the assumption that patients use one hospital only. The DBM model follows a second best approach assuming that patients want to travel to the closest hospital and that they go to a small number of hospitals determined by past data[261]. As the behaviour of patients is not influenced by supply levels, this is a location model. In this model:

- The objective function minimises the total distance travelled by patients;
- Patients visit three hospitals: the closest, the second closest and the closest central hospital;
- The population quotas (from each population area) using the three hospitals are estimated on the basis of empirical information. Details of this procedure are described in section 9.3.3.1 below.

Utilisation-based model (UBM). The gravity model was applied to Portuguese data to estimate CBFs and generated a set of utilisation flows between populations and hospitals (the model is described in Appendix E). The UBM assumes that patients' behaviour is fixed, as described by the gravity model. Though the use of a gravity model is inadequate to predict radical changes, as the UBM is a model for only marginal redistribution, the probabilities of flows of populations to hospitals of the gravity model are used as a possible method to predict patients' 'travelling' behaviour. The behavioural constraint makes the simplistic assumption that when there are changes in the size of one hospital, all the other hospitals gain in proportion to their shares of utilisation prior to that change. The use of this constraint is compatible with the use of

[261] The characteristics of past data (flows utilisation data) are described in Chapter 7, with reference to the characteristics of the utilisation flows matrix.

an objective function promoted by policy-makers. The following can be said about the model:

- It minimises variations between predicted and expected utilisation rates per population. Expected utilisation rates are defined as the level of hospital utilisation for that population if use were at the national rate;
- Utilisation per population area results from the sum of utilisation flows to all hospitals of the area while initial utilisation flows relate to hospital supply as described in Equation 9.1a. The probabilities of a population using a specific hospital ($\widetilde{p}_{ij}$s) are fixed. Those probabilities are taken as outputs from the gravity model applied in accordance to the definitions described in Appendix E (the attraction-constrained version of the gravity model calibrated with 1999 data from the DRG database of the Portuguese system). Equation 9.1.b makes the implicit assumption of a destination-varying model, in that utilisation flows vary proportionally with the prior conditional probabilities and depend on the capacity of the destination hospital (Wilson and Gibberd 1990).

$$D^0 = \sum_i \sum_j U_{ij}^0 \quad (9.1a)$$

$$U_{ij} = \widetilde{p}_{ij} * D_j, \forall i, j \quad (9.1b)$$

- The model uses constraints on the range of variation to be allowed in the size of each hospital/hospital site.

Utilisation flows-based model (UFBM). This model addresses the question of how to sufficiently encompass complex patient behaviour following changes in hospital supply. The UFBM makes use of a behavioural constraint adapted from the econometric application of the FDM constructed in Chapter 7. The FDM captured how the population behave in relation to their location and levels of hospital supply in their area. The UFBM takes account of the process of demand for health care. Using an index for alternative supply of hospital care, it also captures the interaction between utilisation flows and supply of alternative hospitals. The construction of this MP model uses the two-part model (TPM) structure of the FDM. The second part of the flow demand model is used as a constraint that captures flows behaviour resulting from changes in hospital supply. This overcomes the problem of previous location studies, which have often ignored the impact of zero flows on the estimation techniques (Porell and Adams 1995). The probabilities of a population area using a hospital are generated by the first part of the TPM model and considered fixed. The UFBM requires the use of a new equity index to accommodate the characteristics of patients' behaviour captured by the second part of the TPM. This index summarises variations of flows against an equity target based on an equitable distribution. The choice of objective function for the UFBM was intended to be compatible with the use of the realistic behavioural constraint (derived from the FDM) in a MP model, and to represent an equity objective of the policy-maker. This model uses constraints on the range of variation in current capacities of hospital sites, and on total supply.

Each model has different strengths: the DBM has a simple structure that generates changes in supply in order to improve location accessibility; the UBM uses an objective function that is clearly consistent with objectives of health policy; and the UFBM accounts for the most realistic information on patients' use of hospitals following changes in supply. The UFBM is the preferred model because it tackles the weaknesses detected in the location-allocation models literature. Nevertheless, a better behavioural function to predict utilisation flows has the trade-off of not allowing for the free choice of objective function , as will be shown later[262].

[262] As described below, the modelling of non-linear relations between variables with MP models is very complex, or even impossible in certain cases, therefore constraining the construction of the model.

Table 9.5: Notation in use (and model where variable is used)

Notation	*Interpretation*
$Flow\_1_{ij}$	Dummy **variable** for expressing whether population i is served by hospital j, as a first hospital (DBM) (0 or 1 values).
$Flow\_2_{ij}$	Dummy **variable** for showing whether population i is served by hospital j as a second hospital (DBM) (0 or 1 values).
$Flow\_c_{ij}$	Dummy **variable** for denoting whether population i is served by hospital j as the closest central hospital (DBM) (0 or 1 values).
d_1_i	Distance travelled between population point i and the first hospital of use (non-negative **variable** depending on $Flow\_1_{ij}$) (DBM).
d_2_i	Distance travelled between population point i and the second hospital of use (non-negative **variable** depending on $Flow\_2_{ij}$) (DBM).
d_c_i	Distance travelled between population point i and the closest central hospital of use (non-negative **variable** depending on $Flow\_c_{ij}$) (DBM).
W_i	Needs-weighted population at population point i (DBM). This is derived from weighting resident population per age group by the age weighting index estimated in Chapter 5.
$share\_1_i$	Share (%) of population i that is assumed to go to the first hospital (DBM).
$share\_2_i$	Share (%) of population i that is assumed to go to the second hospital (DBM).
$share\_3_i$	Share (%) of population i that is assumed to go to the closest central hospital (DBM).
U_1_i	Utilisation flow by population i to the closest hospital (DBM).
U_2_i	Utilisation flow by population i to the second closest hospital (DBM).
U_3_i	Utilisation flow by population to the closest central hospital (DBM).
$\overline{U}^N$	National utilisation rate ($\overline{U}^N = U^o / W$) (DBM/UBM).
$\widetilde{p}_{ij}$	Probability of population i using hospital j, as produced by the gravity model, with $\sum_j \widetilde{p}_{ij} = 1, \forall i$ (UBM).
U_i^N	Normative utilisation for population area i depending on total national utilisation rate (non-negative **variable**) (UBM).
D_j^0, D^0, D	Current level of supply of hospital j; total current level of supply; total level of supply, computed within the model (UBM)
U_{ij}^o, U^0	(Past) flows and (past) total level of utilisation (DBM/UBM)
a_i	Auxiliary **variable** used to obtain an absolute value of difference between utilisation and expected utilisation, per population area i (UBM).
$f(\beta_j^{'''}, d_{ij})$	Decay function that relates the effect of distance (accessibility costs) from population i to hospital j (definition in Appendix E). The decay function might differ for hospital type

	and the decay parameter β_j will depend on the level of attraction between hospital j and patients located at different distances from that hospital (UBM).
β_j'''	Parameter that defines the elasticity of utilisation in relation to distance, for hospital j (UBM).
f_min	Proportion of current level of supply of hospital j to be kept, as a minimum (UBM/UFBM)
f_max	Proportion of current level of supply of hospital j to be increased, as a maximum (UBM/UFBM)
$\min\_D_j$	Minimum level of supply of hospital j to be maintained (UBM/UFBM).
$\max\_D_j$	Maximum level of supply to be allowed for hospital j (UBM/UFBM).
$\log U_{ij}^r$	Distribution of the natural logarithm of utilisation flows that operates as the target. This target is a distribution formulated in accordance to some type of equity principle (in this case, patients making use of the closest hospital) (UFBM).
b_{ij}	Auxiliary **variable** for defining the difference between variations in the logarithm of utilisation flows (UFBM).
$\log \hat{p}_{ij}$	Logarithm of the probability of use, generated in the first part of the estimated two-part FDM, developed in Chapter 7 (UFBM).
$\log U_{ij}'$	Natural logarithm of the utilisation **variable** between hospital i and hospital j, as defined in the second part of the two-part FDM, developed in Chapter 7 (UFBM).
$DumFirst_{ij}$, $DumSecond_{ij}$ and $DumCentral_{ij}$	Dummy for whether hospital j is the closest hospital to a population i; dummy for whether hospital j is the second closest hospital to a population i; and dummy for whether j is the closest central hospital to a population i (UFBM).
$DumLisboa_{ij}$, $DumPorto_{ij}$ and $DumCoimbra_{ij}$	Dummy for the central hospital site in Lisboa and for populations from the South; dummy for the central hospital site in Coimbra and for populations from the Centre; and Dummy for the central hospital site in Porto and for populations from the North (UFBM).
$\hat{others}_{ij}$	Parameter capturing the influence on flows of all the factors from the FDM, with the exception of the variables that relate to hospital supply (UFBM).
$\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\alpha}_5$, $\hat{\alpha}_6$, $\hat{\alpha}_7$	Parameters that relate utilisation flows and hospital supply, taken from the estimated flows demand model (estimated in Chapter 7) (UFBM).
$\log U_{ij}$, D_j, $\tilde{D}_j$, U_i, d_{ij}, m, n	Notation presented in Chapter 7.
W	Notation presented in Chapter 8.

9.3.3 Framework for analysis of outputs

The results from the three models are analysed within the framework presented in Figure 9.2 and Table 9.6 (below). The use of a framework for the analysis of outputs is crucial because decisions on redistribution involve not only equity, but also efficiency and various scales of redistribution. The pursuit of these objectives is expected to involve trade-offs. Figure 9.2 structures the set of alternative objectives to be achieved in a value tree, which is a hierarchical representation of objectives structured from general and abstract ones (at the top) to measurable and well-specified criteria (at the bottom) (Goodwin and Wright 1998). Three main objectives are pursued. These correspond to the top objectives of the value tree (the corresponding measurement indicators are described in Table 9.6):

a) *Maximisation of equity improvements.* This objective can be divided into three more specific goals:

- *Maximum equity of utilisation.* This means minimising variations of utilisation rates between population areas (an explicit equity objective for many countries).
- *Minimisation of differences between supply and need at district level.* Since equal opportunity of access for those in equal need across geographic areas is also a policy objective, it is desirable that redistribution of supply in the MP models minimises differences between supply and need at the district level. Hence, the outputs from each model are juxtaposed with the distribution based on needs-based estimates for the district level, presented in Chapter 8.
- *Minimisation of distance travelled.* Shorter travel distance to access hospital care implies lower costs for patients to access health services. Both total distance and distance travelled per population area are important in this context.

b) *Maximisation of efficiency.* Efficiency is specifically defined here as the total level of (potential or predicted) production in the system, as measured by utilisation. It has been demonstrated in Chapter 7 that: hospital utilisation by patients is negatively influenced by distance, the relation between utilisation and distance varies by hospital type, and flows depend on hospital size. Therefore, it is anticipated that redistribution of supply will impact on patient flows and on total utilisation in the system (for the UBM and UFBM). Higher total utilisation also means overall better access while reduced total utilisation in the system means poorer access on the whole.

c) *Minimisation of the scale of redistribution.* As redistribution involves increasing and decreasing supply per hospital, it is important to analyse the number and location of 'winners' and 'losers' that result from such changes in supply.

Comparison of the results of the different models in terms of different objectives enables conclusions to be drawn from differences on redistribution results, and gives insights as to how redistribution affects other objectives in the system. These comparisons can also be interpreted as an application of sensitivity analysis to the use of alternative models of redistribution.

Figure 9.2: Value tree representing equity and policy-related criteria

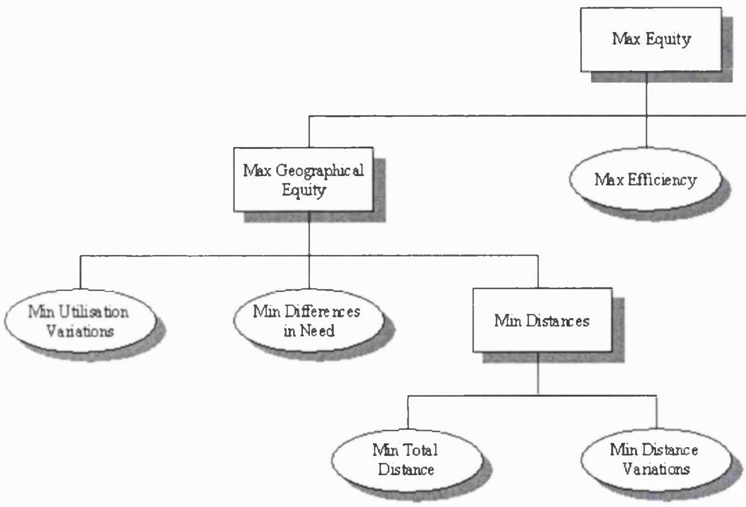


Table 9.6: Measures of equity and policy related criteria

Criteria	*Measure*
Equity- utilisation	Utilisation rates by small area
Equity – supply/need	Variations in the levels of supply at the district level in comparison to needs-based estimates
Equity -distance	Average distance to be travelled by population area Total average distance in the system
Efficiency	Total utilisation in the system
Redistribution	Numbers of 'winners' and 'losers' as a result of redistribution at the hospital level

9.4 Models for improving geographic equity

Each of the three proposed models is discussed next, beginning with an intuitive account, then a description of the structure of the MP program, and then the formulation of the model. Table 9.7 gives a summary of the key features of the quantitative structure of the three models.

Table 9.7: Quantitative differences between models (but key features)

Objective function	*Behavioural constraint*	*Other constraints*
$\min \sum_i \begin{bmatrix} U\_1_i * d\_1_i + U\_2_i * d\_2_i + \\ U\_c_i * d\_c_i \end{bmatrix}$ (equity in location accessibility) Decision variables: d_1_i, d_2_i, d_c_i, D_j Where, for example: $d\_1_i = \sum_j Flow\_1_{ij} * d_{ij}$	$Flow\_1_{ij} = \begin{cases} 1: if\_j\_first\_hospital\_use \\ 0: otherwise \end{cases}$ $Flow\_2_{ij} = \begin{cases} 1: if\_j\_second\_hospital\_use \\ 0: otherwise \end{cases}$ $Flow\_c_{ij} = \begin{cases} 1: if\_j\_central\_hospital\_use \\ 0: otherwise \end{cases}$ Fixed: $U^0 = \sum_i U_i$	Fixed utilisation per area in proportion to national utilisation rate: $U_i = \overline{U}^N * W_i$ Utilisation per hospital type (deterministic): $U\_1_i = share\_1_i * U_i$, $U\_2_i = share\_2_i * U_i$, $U\_c_i = share\_c_i * U_i$
$\min \sum_i \lvert U_i - U_i^N \rvert$ (equity of access per population area) Decision variables: D_j, U_{ij}	$U_{ij} = \tilde{p}_{ij} * D_j$ U_i and U are unconstrained.	Upper and lower limits in variation of hospital supply: $\min\_D_j \le D_j \le \max\_D_j$ Fixed total supply: $\sum_j D_j = D$. Constraint defining total utilisation: $U_i = \sum_j U_{ij}$. Normative utilisation: $U_i^N = W_i * U^o / W$
$\min \sum_i \sum_j \left(\log U_{ij} - \log U_{ij}^r\right)^2$ (equity of utilisation) Decision variables: D_j, $\log U_{ij}^{'}$	$\log U_{ij} = \log \hat{p}_{ij} + \log U_{ij}^{'}$ and $\log U_{ij}^{'} = other\hat{s}_{ij} + \hat{\alpha}_0 * D_j + \hat{\alpha}_1 * \tilde{D}_{ij} + \hat{\alpha}_2 * DumFirst_{ij} * D_j$ $+ \hat{\alpha}_3 * DumSecond_{ij} * D_j + \hat{\alpha}_4 * DumCentral_{ij}$ $+ \hat{\alpha}_5 * DumPorto_{ij} * D_j + \hat{\alpha}_6 * DumCoimbra_{ij} * D_j$ $+ \hat{\alpha}_7 * DumLisboa_{ij} * D_j$ U_i and U are unconstrained.	Upper and lower limits for variation of hospital supply: $\min\_D_j \le D_j \le \max\_D_j$ Fixed total supply: $\sum_j D_j = D$. The α's and $\log \hat{p}_{ij}$ are parameters estimated in the FDM.

9.4.1 Distance-based model -DBM

This model uses information from the utilisation matrix (described in Chapter 7), in which each population area makes use of a small number of hospital sites, and all population areas use central hospital sites (which are the only providers for highly specialised services). The DBM aims at answering the following question: which supply changes would be required for a higher proportion of demand to be met by the closest hospitals?

The DBM follows the classical structure of the 'p-median' model that minimises distance weighted populations, but it differs from this, in that it minimises the total distances (or the weighted total distances) travelled by patients to reach three hospitals – the closest, second closest and closest central hospitals- rather than just the closest hospital. The objective function minimises distances travelled by patients to access the three hospitals used, and this formulation is deployed because empirical evidence contradicts the rule that patients' behaviour is supposedly based on the nearest center. In the p-median model, total distance is minimised and patients are supposed to be treated in one hospital. These assumptions are unrealistic as they are incompatible with characteristics of hospital systems, such as of diversification of supply, different attractiveness of hospitals and inelastic demand characteristics that might justify the use of more than one hospital (O'Kelly 1987). In the DBM, users are allocated to the three hospitals, under previous fixed quotas. The quota of patients allocated to the closest hospital is increased in comparison to past quotas, which implies that this model represents improvements in current accessibility as supply from other hospitals is transferred to the closest hospitals. Details of the quotas for the three hospitals used are given in the description of the constraints in the next sub-section.

This model is simple and could be produced using spreadsheet modelling: the model allocates people to three hospitals (two closest hospitals and the closest central hospital), while total hospital supply and utilisation are constrained to past levels. The use of mathematical programming to formulate this model is useful for structuring the problem and allows for comparison with the other mathematical models presented (same language in use).

The model produces: new supply levels and new distances to be travelled by patients as outputs. Utilisation levels are fixed: it is assumed that all the populations use (and demand) hospital services at the national 1999 level.

9.4.1.1 Structure of the program

Objective function. The model minimises the distances to be travelled by patients to reach the three hospitals used, weighted by utilisation numbers[263]. Utilisation flows depend on needs-weighted population and on past utilisation quotas. The DBM is a linear mixed integer-programming model, as some decision variables are dummies, while one is continuous.

Constraints. The population uses hospital services at the national utilisation rate. Utilisation per population area is divided into fixed shares for the three types of hospitals (computed from empirical data):

1. At a national level, 10% of area discharges/utilisation on average are served by central hospitals (computed from the database described in Chapter 7);
2. The discharges allocated to the second closest hospital are those observed in 1999 (for example, 15% of patients from population area i used the second closest hospital);
3. The remaining discharges are allocated to the closest hospital for any population area (this quota is the result of subtracting the quota of patients of the second closest and of the closest central hospital from 100%). This implies a marginal redistribution of capacity towards the hospital point closest to the population.

9.4.1.2 Formulation

The model takes as *decision variables* the following:

- $Flow\_1_{ij}$: dummy variable for whether population i is served by hospital j, as a first hospital (0,1 values);

[263] This objective function respects the principle of transfers in terms of distance.

- $Flow\_2_{ij}$: dummy variable for whether population i is served by hospital j, as a second hospital (0,1 values);
- $Flow\_c_{ij}$: dummy variable for whether population i is served by hospital j, as the central hospital (0,1 values);
- d_1_i: distance travelled between population point i and the first hospital (non-negative);
- d_2_i: distance travelled between population point i and the second hospital (non-negative);
- d_c_i: distance travelled between population point i and the closest central hospital (non-negative).

And as *parameters*:

- d_{ij}: Euclidean distance between population point i and hospital site j;
- W_i: needs-weighted population for population point i;
- $share\_1_i$: share of population i that is assumed to go to the first hospital;
- $share\_2_i$: share of population i that is assumed to go to the second hospital;
- $share\_3_i$: share of population that is assumed to go to the closest central hospital.
- U_1_i: utilisation numbers for population i and its closest hospital;
- U_2_i: utilisation numbers for population i and its second closest hospital;
- U_3_i: utilisation numbers for population i and its closest central hospital;
- U^0: current (past) level of utilisation, on aggregate (*status quo*);
- $\overline{U}^N$: national utilisation rate.

The model minimises total distance-weighted utilisation, as set in Equation 9.I.1.

$$\min \sum_i \left[U\_1_i * d\_1_i + U\_2_i * d\_2_i + U\_c_i * d\_c_i \right] \quad (9.I.1)$$

The model is subjected to fifteen sets of constraints summarised in eight categories. First, total hospital utilisation by a population is assumed to be at the national utilisation rate (Equations 9.I.2 and 9.I.3).

$$\overline{U}^{N} = U^{o} / W \tag{9.I.2}$$

$$U_i^N = W_i * \overline{U}^N, \forall i \tag{9.I.3}$$

Second, utilisation/demand from a population to each of the three hospitals –closest, second closest and closest central– are fixed (Equations 9.I.4, 9.I.5 and 9.I.6):

$$U\_1_i = U_i^N * share\_1_i, \forall i \tag{9.I.4}$$

$$U\_2_i = U_i^N * share\_2_i, \forall i \tag{9.I.5}$$

$$U\_c_i = U_i^N * share\_c_i, \forall i \tag{9.I.6}$$

Third, each population point has to use three hospitals:

$$\sum_j \left[Flow\_1_{ij} + Flow\_2_{ij} + Flow\_c_{ij}\right] = 3, \forall i \tag{9.I.7}$$

Fourth, for each area, the second hospital has to be different from the first:

$$Flow\_1_{ij} + Flow\_2_{ij} \leq 1, \forall i, j \tag{9.I.8}$$

Fifth, each population point can only have access to one first, one second and one central hospital (Equations 9.I.8, 9.I.9 and 9.II.10).

$$\sum_j Flow\_1_{ij} = 1, \forall i \tag{9.I.9}$$

$$\sum_j Flow\_2_{ij} = 1, \forall i \tag{9.I.10}$$

$$\sum_j Flow\_c_{ij} = 1, \forall i \tag{9.I.11}$$

Sixth, central hospitals have limited locations:

$$Flow\_c_{ij} = 0, \forall i, \text{ if } j \text{ is not a central hospital} \quad (9.I.12)$$

Seventh, hospital supply (measured by discharges) is defined as the total utilisation by the populations allocated to it, assuming that utilisation by population area is proportional to the national utilisation rate.

$$D_j = \sum_i \left[Flow\_1_{ij} * U\_1_i + Flow\_2_{ij} * U\_2_i + Flow\_c_{ij} * U\_c_i\right], \forall j \quad (9.I.13)$$

Lastly, distances travelled by patients to access the closest, second closest and closest central hospitals are computed as expressed in Equations 9.I.14, 9.I.15 and 9.I.16.

$$d\_1_i = \sum_j Flow\_1_{ij} * d_{ij}, \forall i \quad (9.I.14)$$

$$d\_2_i = \sum_j Flow\_2_{ij} * d_{ij}, \forall i \quad (9.I.15)$$

$$d\_c_i = \sum_j Flow\_c_{ij} * d_{ij}, \forall i \quad (9.I.16)$$

9.4.2 Utilisation-based model -UBM

This model uses an objective function that corresponds to a common health policy objective, namely the minimisation of differences between predicted and normative utilisation per population area. Predicted utilisation is utilisation predicted by a behavioural equation that relates supply (D_j) to utilisation flows (U_{ij}), as expressed in Equation 9.II.1 ($\tilde{p}_{ij}$ is a conditional probability of population in i making use of hospital j). Normative utilisation is the desirable level of utilisation per population area, if population used resources at the national average utilisation rate.

$$U_{ij} = \tilde{p}_{ij} * D_j, \forall i, j \qquad (9.II.1)$$

Probabilities ($\tilde{p}_{i,j}$) are obtained from the application of the gravity model described above. The problems of using the gravity model to make predictions were outlined in Chapter 7 and in the literature review presented in this chapter[264]. However, information taken from the gravity model can be used in a different way within the MP model to describe spatial behaviour by patients, as this is a method for analysing marginal redistribution. The set of conditional probabilities - $\tilde{p}_{ij}$ - is one of the outputs of the gravity model and represents a measure of the propensity of a population to make use of one hospital (the equation is explained below). The behavioural constraint 9.II.1 is an improvement on the DBM, as the UBM assumes that utilisation depends on supply.

As this is a model for redistribution, a set of upper and lower limits for the maximum allowed variation in supply per hospital are imposed.

9.4.2.1 Structure of the problem

Objective function. It minimises differences between predicted and normative utilisation (Equations 9.II.2 and 9.II.3). Mayhew and Leonardi (Mayhew and Leonardi 1982) have used a similar model with an objective function that differs in that it uses the square of the difference between observed/predicted and expected/normative utilisation. The absolute difference between predicted and normative utilisation per population area is preferred, as a utilisation unit has the same value across geographic areas[265].

$$Minimise \sum_i \left| U_i - U_i^N \right| \qquad (9.II.2)$$

$$U_i^N = W_i * \overline{U}^N, \forall i \qquad (9.II.3)$$

[264] Given that a gravity model uses probabilities to explain geographic behaviour, the way the UBM deals with interaction between utilisation flows and supply of alternative hospitals is unsatisfactory: the utilisation by populations of one hospital is not affected by the existence of alternative/close hospitals.
[265] This objective function respects the principle of transfers with regard to utilisation numbers.

Constraints. The probabilities ($\tilde{p}_{i,j}$) are outputs of the gravity model and fixed in the behavioural constraint defined in Equation 9.II.1. This equation used the assumptions of a destination-varying model[266], and the $\tilde{p}_{ij}$ should be interpreted as the probability that a randomly chosen patient living in zone i is treated in hospital j (Wilson and Gibberd 1990). The formula generating the probabilities is defined in Equation 9.II.4, which corresponds to the version of the gravity model presented in Equation 9.II.5. Since $\tilde{p}_{ij}$s are conditional probabilities, their sum is one as in Equation 9.II.6. The gravity model used is presented in greater detail in Appendix E.

$$\tilde{p}_{ij} = W_i * f(\beta_j''', d_{ij}) * \left[\sum_i W_i f(\beta_j''', d_{ij}) \right]^{-1}, \forall i, j \qquad (9.II.4)$$

$$U_{ij} = \left[W_i * f(\beta_j''', d_{ij}) * \left[\sum_i W_i f(\beta_j''', d_{ij}) \right]^{-1} \right] * D_j, \forall i, j \qquad (9.II.5)$$

$$\sum_j \tilde{p}_{ij} = 1, \forall i \qquad (9.II.6)$$

Lower and upper limits are set for reductions and increases in hospital capacities[267]. These constraints are consistent with the behaviour of countries that will attempt to correct geographical inequities in the redistribution of hospital supply, and that already have a network of sufficient size. In the Portuguese context, total supply capacity is fixed at the 1999 level.

9.4.2.2 *Formulation*

The model takes as *decision variables* the following:

- U_i: utilisation level of population in site i (non-negative);

[266] Under the destination-varying hypothesis, utilisation flows depend on the prior probabilities from the utilisation matrix of the gravity model, and the constant of proportionality depends upon the capacity of the destination (i.e. hospital capacity) (Wilson and Gibberd 1990).

[267] As equation 9.II.1 has a limited ability to predict wider changes in supply, upper and lower limits may also be seen as a mechanism of control for avoiding inconsistencies in prediction values (Hallefjord and Jornsten 1984).

- U_{ij} : discharges of population i, attended in hospital j (non-negative);
- U : total utilisation (non-negative);
- U_i^N : normative utilisation for population site i (non-negative);
- D_j : hospital discharges from hospital site j (non-negative);
- a_i : auxiliary and definitional variable used to obtain an absolute value of differences between utilisation and expected utilisation, per population area i (non-negative).

And as parameters:

- D^o : total supply within the current system, given by $D^0 = \sum_j D_j^0$.
- W_i : needs-weighted population in site i ($W = \sum_i W_i$);
- $\tilde{p}_{ij}$: probability of a patient from site i making use of hospital j;
- $f\_\min$: minimum acceptable percentage of variation in capacity per hospital that is to be maintained;
- $f\_\max$: maximum acceptable percentage of variation in capacity per hospital, as a function of current supply;
- $\min\_D_j$: lower limit of the size of hospital j, computed under the following relationship: $\min\_D_j = f\_\min * D_j^0$;
- $\max\_D_j$: upper limit of the size of hospital j, computed under the following relationship: $\max\_D_j = f\_\max * D_j^0$.

The model minimises absolute differences between predicted utilisation and normative utilisation per population i, as expressed in Equation 9.II.7, following the definitions in Equations 9.II.7A and 9.II.7B. The responsiveness of utilisation flows to changes in hospital supply is predicted by the behavioural constraint presented in Equation 9.II.1. Normative utilisation is the utilisation level for the geographic area that is proportional to the national utilisation rate computed within the model.

$$\min \sum_i a_i \,,\ a_i \geq 0 \qquad (9.II.7)$$

$$a_i = \left|U_i - U_i^N\right|, \forall i \tag{9.II.7A}$$

$$U_i^N = W_i * \overline{U}^N, \forall i \tag{9.II.7B}$$

The model imposes seven constraints (grouped in five categories). First, a constraint capturing the behaviour of patient flows for different sets of supply distribution (Equation 9.II.1), which corresponds to the assumptions of a destination-varying model.

Second, two constraints aggregating flows of utilisation per population area and total utilisation (Equations 9.II.8a and 9.II.8b).

$$U_i = \sum_j U_{ij}, \forall i \tag{9.II.8a}$$

$$U = \sum_i U_i \tag{9.II.8b}$$

Third, two constraints for setting of upper and lower limits on changes in supply (Equations 9.II.9 and 9.II.10).

$$D_j \geq min\_D_j, \forall j \tag{9.II.9}$$

$$D_j \leq \max\_D_j, \forall j \tag{9.II.10}$$

Fourth, a constraint maintaining the current total level of supply in the system (Equation 9.II.11).

$$\sum_j D_j = D^0 \tag{9.II.11}$$

Fifth, two mathematical constraints for defining the objective function as an absolute value. These constraints are required in order to generate an absolute value in a minimisation problem (Williams 1993a).

$$U_i - U_i^N \leq a_i, \forall i \qquad (9.II.13)$$

$$-U_i + U_i^N \leq a_i, \forall i \qquad (9.II.14)$$

9.4.3 Utilisation flows-based model -UFBM

The UFBM makes use of a realistic behavioural constraint through the econometric application of the FDM, which provides a behavioural constraint to predict utilisation flows as a response to changes in hospital supply. The FDM is the most satisfactory model for capturing spatial behaviour as regards the process of demand for health care and models the interaction between hospital supply and use of alternative hospitals in a predictive model.

The structure of the UFBM is designed to accommodate the characteristics of the empirical application of the FDM, which follows a TPM structure. As described in Chapter 7, the TPM predicts utilisation flows with two components. First, a model that predicts the probability of a population area making use of a hospital. Second, a model for predicting the level of utilisation, given that the probability of use is positive. The UFBM assumes that the probability is fixed at the 1999 level[268] and does not change with changes in levels of supply, and uses the second part of the model to predict patients' behaviour.

The second part of the econometric model produces data in logarithms, something which has constrained the structure and formulation of the MP model. Since there is no index in the literature for relating the logarithm of flows to an equity objective, a new index was created for the objective function.

Similarly to the UBM, this model imposes constraints on variations in supply at the hospital level and on lower and upper limits for redistribution. There are no constraints on variation in patient flows, which changes in response to supply modifications.

[268] The impact of this assumption is discussed below.

9.4.3.1 Structure of the problem

Objective function. Indices in the literature aggregate one-dimensional information, such as utilisation per population area, into a composite index summarising information on predicted and normative utilisation. There is no bi-dimensional index based on population area and hospital site. Consequently, a new equity index had to be developed to sum up deviations between utilisation flows and some (more equitable) target distribution (or more precisely, differences between the natural logarithm of flows, given the behavioural constraint). This index ought to satisfy properties often cited in the literature on location of facilities and described in the literature review section (Marsh and Schilling 1994). These include: analytical tractability for problem size and computational requirements, appropriateness for interpretation, non-discrimination between the (geographic) groups being evaluated, and the principle of transfers. Following the variance of the logarithms index version presented in Marsh and Schilling (Marsh and Schilling 1994) (originally used by Theil (Theil 1967) as described below), Equation 9.III.1 was created[269].

$$\frac{\sum_i \sum_j \left(\log U_{ij} - \log U_{ij}^r\right)^2}{n * m} \qquad (9.III.1)$$

The log nature is required as flows were generated by the TPM, for which the second part follows a log-linear structure and is used as a constraint of the MP model. The probabilities of the first part of the TPM are assumed as fixed, which allows for the natural logarithm of the flows to be divided as analysed in Equation 9.III.2.

$$\log U_{ij} = \log\left(\hat{p}_{ij} * U_{ij}^{'}\right) = \log \hat{p}_{ij} + \log U_{ij}^{'}, \forall i, j \qquad (9.III.2)$$

Variations in utilisation are compared to a distribution of utilisation formulated under an equity principle ($\log U_{ij}^r$) (as seen in Equation 9.III.1). The proposed index is a measure of the dispersion of the logarithm of utilisation flows against an equity target –i.e., a measure of the variance of the logarithm. This agrees with Theil's proposition (Theil

[269] This objective function respects the principle of transfers, with respect to utilisation flows. *n* and *m* as defined in Chapter 7.

1967)[270] that the variance decomposition is useful when the variable is approximately lognormally distributed. Theil has applied this in the context of study of the distribution of income. The use of the logarithm of flows as a variable in a MP model implies that it is impossible to link flows with total utilisation per population area inside the MP model. This must be taken into account in the analysis of results.

As the logarithm of zero is minus infinity, areas with zero logarithmic flow took the value of one as in Equation 9.III.3.

$$\log U_{ij}^{r} = 0 \Rightarrow U_{ij}^{r} = 1, \forall i, j \qquad (9.III.3)$$

In summary, the objective function minimises expression 9.III.4: it gives the quadratic function of a linear index of the logarithm of utilisation flows. It represents the same relationship as Equation 9.III.1; Equation 9.III.4B shows how the objective function is linked to the predicted values generated by the TPM.

$$\sum_{i}\sum_{j}\frac{b_{ij}^{2}}{n*m} \qquad (9.III.4)$$

$$b_{ij} = \log U_{ij} - \log U_{ij}^{r} = \log\left(\frac{U_{ij}}{U_{ij}^{r}}\right), \forall i, j \qquad (9.III.4A)$$

$$\log\left(\frac{U_{ij}}{U_{ij}^{r}}\right) = \log\left(\frac{\hat{p}_{ij} * U_{ij}^{'}}{U_{ij}^{r}}\right) = \log \hat{p}_{ij} + \log U_{ij}^{'} - \log U_{ij}^{r}, \forall i, j \qquad (9.III.4B)$$

The component of probabilities ($\log \hat{p}_{ij}$) consists of point estimates that are fixed for the current (1999) levels of hospital supply and are not allowed to vary in the MP model. This is a restrictive assumption because when hospital capacities vary, probabilities ought to change. Nevertheless, as this is a model for redistribution, hospital capacities

270 The original formula of the variance of the logarithms suggested by Theil was: $\frac{1}{n} * \sum_{i}\left(\log U_{i} - \log \overline{U}^{N}\right)^{2}$.

might vary within a limited range, which implies that changes in probabilities would be small.

An equity distribution target was generated in accordance with a definition of location accessibility ($\log U_{ij}^{r}$). This target is a set of estimates of an equitable distribution of utilisation by populations. This distribution was constructed using populations going to the closest hospital by means of a MP program that minimises weighted distance between population points and hospital sites[271].

The use of the objective function from Equation 9.III.4 requires the use of a quadratic MP model[272]. The quadratic function needs to be convex so that the solution of the MP program is a global (instead of local) optimum (Williams 1993a). As all the decision variables of the squared utilisation variable of expression have positive coefficients (as can be seen in the coefficients of $\log U_{i,j}^{'}$ when taking the squares of Equation 9.III.1, and given that $\log \hat{p}_{ij}$ and $\log U_{ij}^{r}$ are fixed parameters), the Kuhn-Tucker conditions that guarantee global optimality are satisfied (Williams 1993b).

Constraints. Besides the constraint that patient behaviour must be simulated on the basis of the FDM, the UFBM uses a set of constraints that are similar to the constraints of the UBM, i.e. on the lower and upper limits of supply distribution, and on maintaining the total level of supply. The model only considers flows for which the first part of the FDM was predicted as positive.

9.4.3.2 *Formulation*

Decision variables:

- $\log U_{ij}^{'}$ as the logarithm of utilisation flows, given that the probability of utilisation between population i and hospital j is positive (free variable);
- D_j as the size of hospital j (non-negative);

[271] This distribution could be computed in a simple spreadsheet.

[272] A quadratic mathematical programming model is a model with a quadratic objective function and linear constraints.

- b_{ij} as a definitional variable for the objective function (free variable).

Parameters:

- $oth\hat{e}r_{ij}$ as the parameter from the behavioural constraint that captures the impact of other factors of the FDM on hospital utilisation flows (factors unrelated to the variables of the UFBM);
- $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\alpha}_5$, $\hat{\alpha}_6$, $\hat{\alpha}_7$ as parameters estimated in the FDM, which are required to relate hospital supply and utilisation flows;
- $\log \hat{p}_{ij}$ as the logarithm of the probability of a population point i making use of a hospital j;
- $\log U_{ij}^r$ as the reference utilisation flow that operates as an equity target;
- $\min\_D_j$, $\max\_D_j$ and D as defined before in the UBM;
- $DumFirst_{ij}$, $DumSecond_{ij}$ and $DumCentral_{ij}$ as a dummy parameter for the first hospital of use, second hospital of use and central hospital;
- $DumLisboa_{ij}$, $DumPorto_{ij}$ and $DumCoimbra_{ij}$ as a dummy parameter for the central hospitals of Lisboa, Porto and Coimbra.

The objective is to minimise the squared difference between predicted utilisation flows and target flows based on an equity concept, as explained below, corresponding to Equations 9.III.5 and 9.III.5A.

$$\min \sum_i \sum_j \frac{b_{ij}^2}{n*m} \qquad (9.III.5)$$

$$b_{ij} = \log \hat{p}_{ij} + \log U'_{ij} - \log U_{ij}^r, \forall i, j \qquad (9.III.5A)$$

This model uses four constraints. The first constraint is on the predicted level of flows in response to changes in supply, as defined in Equation 9.III.6. The formulation of this constraint has relied on the specific econometric application of the FDM developed in Chapter 7. Equation 9.III.6 assumes that flows depend on hospital capacity, on alternative supply, and on whether this is the closest, second closest or closest central hospital.

$$\log U'_{ij} = oth\hat{e}r_{ij} + \hat{\alpha}_0 * D_j + \hat{\alpha}_1 * \widetilde{D}_{ij} + \\ \hat{\alpha}_2 * DumFirst_{ij} * D_j + \hat{\alpha}_3 * DumSecond_{ij} * D_j + \\ \hat{\alpha}_4 * DumCentral_{ij} + \hat{\alpha}_5 * DumPorto_{ij} * D_j + \\ \hat{\alpha}_6 * DumCoimbra_{ij} * D_j + \hat{\alpha}_7 * DumLisboa_{ij} * D_j, \forall i, j \tag{9.III.6}$$

with $\hat{\alpha}_0 = 0.0000352$, $\hat{\alpha}_1 = -0.1873067$, $\hat{\alpha}_2 = 0.0000231$, $\hat{\alpha}_3 = 0.0000141$, $\hat{\alpha}_4 = -4.304794$, $\hat{\alpha}_5 = 0.0000158$, $\hat{\alpha}_6 = 0.0000255$ and $\hat{\alpha}_7 = -0.0000136$ for the Portuguese system.

Second, lower and upper limits for changes in supply are imposed:

$$D_j \geq \min\_D_j, \forall j \tag{9.III.7}$$

$$D_j \leq \max\_D_j, \forall j \tag{9.III.8}$$

Third, there is a constraint on the maximum capacity in the system:

$$\sum_j D_j = D^0 \tag{9.III.9}$$

9.5 Results and discussion

The three MP models were solved with the AIMMS software package (version 3.1) (Paragon Decision Technology 2000).

The application to Portuguese data has been based on the following methodological choices: population numbers have been weighted by demographic need estimates generated in Chapter 5, while Euclidean distances between centroids of the *concelho* geographic units were taken as proxies for travelling costs (in the same way as used in Chapters 3 and 7). The UBM and the UFBM were run with a lower limit set at 80% and an upper limit set at 120% of the current capacity. Data used in this chapter has been based on discharges data from the DRG system for 1999, resident population estimates from the Portuguese National Institute of Statistics, and estimates of the FDM from Chapter 7.

This section describes and compares the results of the three models; describes the policy implications of results; and suggests further developments of the models.

9.5.1 Comparison of the three models

The results are presented in Tables 9.8-9.12 using 1999 data as reference point.

Table 9.8: Utilisation impacts with a redistribution of 20% of supply both on lower and upper bound

Model	*Total utilisation*	*Maximum utilisation rate*	*Minimum utilisation rate*	*Standard deviation utilisation rate*
Past data (1999)	901,229	645%	29%	0.46
DBM	901,229	100%	100%	NA
UBM	858,426	136%	37%	0.22
UFBM	675,306	327%	12%	0.48

Table 9.9: Distance impacts with a redistribution of 20% of supply both on lower and upper bound

Model	*Average distance travelled*	*Maximum average distance*	*Minimum average distance*	*Standard deviation average distance*
Past data (1999)	22.4	164.9	1.1	33.5
DBM	16.8	193.9	0.0	21.3
UBM	53.5	107.6	18.4	15.1
UFBM	14.6	91.94	2.2	17.0

Note: Average distances computed for all the patients from a population area

Table 9.10: Number of winners and losers with a redistribution of 20% of supply both on lower and upper bound

Model	*Number of hospital 'winners'*	*Number of hospital 'losers'*
DBM	42	16
UBM	29	39
UFBM	59	9

Table 9.11: New levels of supply

	Initial hospital supply	*DBM*	*UBM*	*UFBM*	*Needs-based capitation shares*[273]
Aveiro	5.3%	7.1%	4.3%	6.4%	5.9%
Beja	1.2%	1.3%	3.1%	1.5%	1.8%
Braga	7.1%	7.3%	5.7%	8.1%	6.8%
Bragança	1.7%	1.6%	3.9%	2.1%	1.6%
Castelo Branco	2.4%	1.5%	3.5%	2.3%	2.2%
Coimbra	9.3%	6.8%	7.4%	7.8%	3.8%
Évora	1.4%	1.0%	1.1%	1.6%	1.6%
Faro	3.1%	3.6%	4.7%	3.7%	3.8%
Guarda	1.5%	2.0%	3.5%	1.8%	1.9%
Leiria	3.7%	5.5%	3.8%	3.9%	4.2%
Lisboa	26.2%	24.2%	21.0%	22.2%	25.5%
Portalegre	1.1%	1.3%	4.2%	1.1%	1.4%
Porto	18.5%	16.4%	14.8%	17.9%	17.8%
Santarém	3.7%	4.7%	5.9%	3.6%	4.5%
Setúbal	6.7%	7.6%	5.3%	8.0%	8.0%
Viana do Castelo	2.0%	2.4%	1.6%	2.4%	2.8%
Vila Real	2.7%	2.0%	4.1%	3.2%	2.5%
Viseu	2.5%	3.6%	2.0%	2.3%	3.6%

Table 9.12: Variation in utilisation by district (UFBM)

	New utilisation	*Previous utilisation*	*Variation*
Aveiro	56,647	65,621	-14%
Beja	11,999	14,332	-16%
Braga	61,335	70,497	-13%
Bragança	15,805	18,460	-14%
Castelo Branco	24,109	25,132	-4%
Coimbra	21,423	29,932	-28%
Évora	11,584	14,706	-21%
Faro	25,222	31,712	-20%
Guarda	12,181	16,299	-25%
Leiria	34,736	45,288	-23%
Lisboa	10,2310	212,505	-52%
Portalegre	11,708	13,960	-16%

273 Information taken from the results of index I_{2r}, Chapter 8.

Porto	116,771	152,476	-23%
Santarém	35,331	43,428	-19%
Setúbal	76,745	70,839	8%
Viana do Castelo	15,354	21,225	-28%
Vila Real	20,992	25,736	-18%
Viseu	21,055	29,081	-28%
Mainland Portugal	675,307	901,229	-25%

DBM: Results show an improvement in location accessibility, as measured by distance travelled (Table 9.9). The Lisboa, Coimbra and Porto districts are the main losers in hospital supply, and redistribution favours urban districts peripheral to Lisboa and Porto (Table 9.11). There is no impact on total utilisation or on utilisation rates (Table 9.7), as the DBM assumes fixed utilisation rates per population at the small area level. Although there is a decrease in the average distance and in the standard deviation of average distances across areas, the maximum average distance from a population area is increased in the DBM (Table 9.9). The redistribution proposed by this model is consistent with progression towards a needs-based distribution, as there is a convergence on the results for redistribution for most districts (interpreted as similar sign of variation) –there is divergence for only two districts (Braga and Évora) (Table 9.11).

UBM. Results show a substantial decrease in the variation of the utilisation rates (Table 9.8), as explicitly pursued in the objective function of this model. These gains are counterbalanced, however, by increases in distances travelled (Table 9.9) and by a reduction in total utilisation (Table 9.8). The UBM also produces substantial redistribution of hospital supply (Table 9.10), with a decrease in supply in central and other urban hospitals located in coastal areas and increases in hospitals in the interior and the south of the country. This is expected, as the probabilities of travelling to hospitals are based on current data and the model attempts to equalise utilisation rates by redistributing supply and utilisation so as to increase flows for hospitals with lower levels of accessibility. The increase in distances travelled is due to the fact that the gravity model makes the unrealistic assumption that urban populations are prepared to travel to rural hospitals[274] and focuses on equalisation of utilisation rates. In comparison

[274] Analysis of current data revealed a very low probability of this happening.

with a needs-based distribution of hospital supply, the redistribution suggested by the UBM is incompatible with the needs-based distribution of Table 9.11.

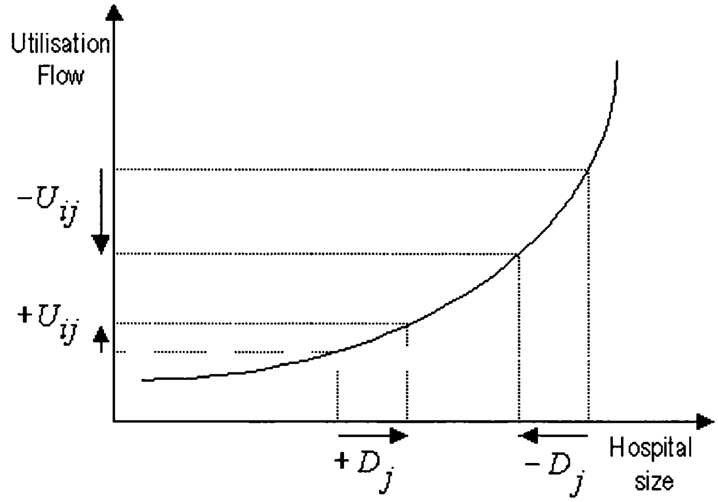
UFBM. The redistribution of supply by this model results in the highest gains in geographic accessibility (Table 9.9). This is expected, since the equity target (included in the objective function) was based on patients having access to the closest hospital. The model proposes redistribution of supply towards peripheral hospitals in urban areas and rural hospitals in the interior and the south. In general, the proposed redistribution is consistent with progressing towards equity in terms of need of hospital care (in terms of sign of variation), except for five districts: Braga, Bragança, Santarém, Vila Real and Viseu (Table 9.11). This model reduces supply in a small number of hospitals and discriminates positively in favour of a large number of hospitals (Table 9.10). There is a substantial decrease in total utilisation, which might be explained as follows:

1. The large hospitals attract patients over large distances; when the capacity of large hospitals is decreased and that of small hospitals is increased, the reduction in flows to the central hospitals is not offset by increases in flows to smaller hospitals. This result might be partly interpreted as a reduction in supplier-induced demand.
2. The logarithmic structure of the behavioural constraint implies that flows are linked to hospital capacity under an exponential function, as shown in Figure 9.3. It shows that decrease in the capacity of a big hospital implies a reduction in flows that is higher (in absolute value) than the increase of flows implied by an increase in capacity of a small hospital. This is explained by the fact that the impact of hospital capacity on flows depends on the current hospital size and flows[275].

The district that loses most in its level of utilisation by resident populations is Lisboa (Table 9.12). Moreover, redistribution of supply implies a loss in equity of utilisation per small area (as variations in utilisation rates by population area increase) (Table 9.8).

[275] This result can be confirmed by analysing the behaviour of the first derivative of $U'_{i,j}$ as a result of a differential variation in D_j, to be estimated by Equation 9.III.6.

Figure 9.3: Redistribution of supply vs. utilisation flows at the small area



The UFBM is the most realistic model in that it accounts for the process of demand for hospital care and models interaction, but not so realistic in terms of reduction in total capacity. Results were found to be robust to alternative econometric formulations of the FDM; and the use of alternative equity targets –such as the use of the outputs of the DBM (instead of patients using the closest hospitals). The UFBM can be used to rank hospitals in terms of their potential for improving geographic equity of utilisation (in the case of increase in hospital capacity), by using the following sequence: a) Run the model for small levels of redistribution of supply to identify the first hospital whose capacity is to be increased; b) Put the hospital at the top in the ranking; c) Re-run the model with a further constraint of not changing the level of supply for that specific hospital; d) The next hospital for which an increase in supply is identified occupies the next place in the ranking; e) Repeat the sequence until all the hospitals are placed in the ranking.

Another interesting result from the UFBM is that when the upper limit for redistribution is increased (above 20%, i.e. the upper limit for changes in current capacity), the model tends to concentrate supply in a reduced number of hospitals. This result may indicate that the highest improvements in accessibility are achieved through big changes in hospital supply for a very small number of hospitals. This might imply that improving equity of access and utilisation requires high changes in supply, and that, at the national level changes in hospital supply are insufficient to tackle geographic inequities in utilisation and other policy tools to influence utilisation should be deployed.

Comparing the results of the three models (generic results are summarised in Table 9.13) shows that:

- The two models that use realistic models of patients' behaviour (UBM and UFBM) result in reductions in total utilisation.
- All the models point out the need to reduce hospital supply in the three districts with central provision (Lisboa, Porto and Coimbra).
- In all the models, complete progress towards capitation-based allocations is not achieved for all districts, though the DBM and the UFBM perform comparatively well in this direction (Table 9.10).
- The pattern of redistribution of supply from the UFBM lies between the patterns from the two other models: reduction of capacities in a small number of central hospital sites in the UFBM is combined with increases in urban hospitals in peripheral areas and in the interior and southern areas, while the other models redistribute more towards rural and southern areas.
- Each model focuses on the achievement in one dimension of equity at the expense of others. For example, the UFBM implies gains in accessibility and losses in total utilisation and in equity of utilisation; the UBM involves gains in equity of utilisation and losses in accessibility and in total utilisation; the DBM implies gains in locational accessibility but neglects variation in utilisation.
- Both the UBM and the UFBM involve losses in the total level of utilisation. This result has not been reported in some of the previous MP models that have used behavioural information generated by gravity models (i.e. models with a structure similar to the UBM). This result shows that improvements in location accessibility or in equity of utilisation rates at the expense of total utilisation. The implications of this finding are discussed below.
- Past data shows that there have been wide variations in utilisation rates across small areas (Table 9.8): some areas use hospital resources at a rate five times higher than the national rate while other areas use hospital resources at one third of the national rate. The UBM improves equity of utilisation in terms of decrease in variations in utilisation rates while the UFBM does not achieve this result (it leads to increase in variations in utilisation rates).
- Implementation of redistribution seems to be easiest in the UFBM as the scale of redistribution is smaller (Table 9.10).

Table 9.13: Comparative results from the location-allocation models

	DBM	*UBM*	*UFBM*
Variation in utilisation rates	No change	+	-
Differences to needs-based distribution	+	-	+
Variations in distances travelled	+	+	+
Total distances travelled	+	-	+
Total utilisation	No change	-	-
Redistribution level	-	-	+

+ denotes positive effect; - denotes negative effect (evaluation on a comparative basis)

9.5.2 Implications for policy analysis

Each model has both strengths and weaknesses. Results show that the pursuit of different equity objectives involves trade-offs, such as between equity of geographic access and of utilisation or between equity of geographic access and efficiency (when this is measured as total utilisation in the system). The use of more realistic assumptions about patients' behaviour makes the models increasingly complex and less transparent.

These results suggest some general conclusions. First, whichever equity objective is pursued, a redistribution of supply should favour hospitals in the interior, the south, and peripheral hospitals in urban areas. As shown in other chapters, the high level of centralisation of services in central hospitals in three urban districts creates geographic inequities. Second, reductions in the capacities of central hospitals result in a decrease in total utilisation in the system. This result is partly due to the capacity of central hospitals to attract patients over long distances as compared to smaller hospitals.

The UFBM tends to redistribute towards a small number of hospitals, which suggests that this is the best way to improve equity of utilisation. As explained above, this may lead to the conclusion that hospital supply might not be the best policy tool for improvement in geographic equity of utilisation at the national level. Consequently, other policy tools should be put into use so as to assess their impact on equity of utilisation.

Results show that when one takes into account for population behaviour with regard to the use of hospital services (UBM and UFBM), there are high equity costs in terms of losses in total utilisation. Such losses are likely to be unacceptable for the policy-maker. Thus improvements in geographic equity of utilisation might be better achieved by other types of policies, such as by marginally increasing hospital supply in the system rather than by redistribution. These policies should be explored through further research.

9.5.3 Improving the models

All the models rely on a set of assumptions, some of which could be modified in future research.

Certain assumptions are common to all models. First, all hospitals provide the same set of services. The models could be adapted for the specialty level. Second, discharges were used as a good proxy for capacity, which assumes the same levels of productivity across hospitals at the 1999 level, ignores variations in prestige and quality of hospitals, as well as the heterogeneity of the current stock of hospitals (in terms of buildings, or medical equipment). Better proxies for hospital production capacity should be developed as underlined in Chapter 3. Third, the proposed models operate with a view to a second best approach and do not question the optimal number or the location of facilities. Improvements in equity might be achieved by closing or building new hospitals. Fourth, the models consider a restricted view of location accessibility as they only account for the patients' accessibility costs and do not consider other types of accessibility costs (such as journeys by visitors and staff). This could be changed, for example, by adding distances travelled by staff and visitors in the DBM, or by changing a single objective function to a multiple objective function in the UBM. Fifth, the three models do not account for the impact of the redeployment of resources on costs and the political context in which changes in capacities are carried out (that is vital to implement changes) (Mohan 1983)[276].

Examples of other one-step incremental improvements for each model can be suggested:

[276] For example, conflicting objectives between RHAs and local health authorities might create problems for implementation. In addition, local social geography and institutional behaviour may change the relationship between changes in supply and changes in utilisation (McLafferty 1988).

- The referral process is only indirectly accounted for in the UFBM[277], and all models could be improved to explicitly capture the referral system.
- Both the UBM and the UFBM assume that past patterns of flows can be used to predict future flows, but there are uncertainties and risks associated with future behaviour. As a result, the models could be improved to specifically model uncertainty in some parameters.
- The mechanism modelling interaction in the UFBM imposes a fixed parameter of substitution between all hospitals across the country (this relates to the use of a single alternative hospital supply index for the whole country in the application of the FDM in Chapter 7). This assumption could be relaxed by permitting a variation in the index by geographic area.
- The DBM could be extended with constraints on indivisibilities in changes to hospital capacity (such as assuming the following values: 0%, 10% or 20%). This might be relevant, as changes in hospital capacities imply fixed costs, which might justify the use of thresholds for variations in capacities.

An alternative extension of the models would be to introduce costs into the objective function, so as to redistribute supply while minimising hospital costs (this would possibly move away from incremental analysis and move towards a more general equilibrium solution in the analysis of alternative models).

9.6 Concluding remarks

The use of alternative models has illuminated the discussion of redistributive policies about hospital supply. There is neither a single objective nor optimal model to answer the redistribution problem and no single model can fully address the broad concept of equity. Some key conclusions follow.

First, it seems important to include in analysis the assumptions about patient's behaviour. The use of 'more pragmatic' behavioural assumptions has shown that there is a trade-off between gains in geographic accessibility and gains in equity of utilisation.

[277] It considers only the impact of utilisation in the primary care sector on hospital utilisation.

Second, it appears that policies within the Portuguese system for improving geographic equity will result in decreases in the capacity of central urban hospitals within the system and increases in both peripheral and urban hospitals along with rural hospitals. Nevertheless, given the costs of redistribution (due to decreases in utilisation), other policies aiming to improve equity in utilisation should be explored (such as attempts to correct inequalities by marginally increasing hospital supply).

Third, the UBFM was shown to be an alternative location-allocation model that tackled both the problems of unrealistic assumptions about patients' behaviour with regard to hospital utilisation and of interaction between utilisation flows and supply for alternative hospitals. It has also accounted for the process of demand for hospital care. The UFBM is the best model for meeting the objectives set in this chapter, and it has been shown that if one aims to improve accessibility for populations located at different points, the best policy tool might be to look at alternative instruments to hospital supply. The innovative formulation of the UFBM has used an interdisciplinary approach between health economics and operational research. Health economics was used to develop realistic assumptions about patients' behaviour in the demand for hospital care. Operational research was used to provide a holistic model of the whole system, using the MP models, algorithms and software.

Fourth, variations in utilisation levels and in admission rates may have implications for costs, quality and hence for health outcomes, which were not analysed.

Finally, even if the equity concept is specifically defined, different measurements have shown to generate different redistributive results. This shows again the need for policy-makers to clearly define the equity objectives to be pursued.

10 CHAPTER 10 - Concluding remarks

10.1 Overview

This thesis has sought to answer two main research questions. The first question was how to measure geographic inequities in the Portuguese hospital system (in terms of capital, utilisation and finance). This required answering to three subsidiary questions: how to measure need for hospital care in Portugal; how to estimate unavoidable costs for Portuguese hospitals; and how to estimate cross-boundary flows of patients between Portuguese districts. These subsidiary questions were addressed by modelling the adjustments of a capitation formula for Portugal, and by computing alternative indices of inequities based on the outputs of the capitation formula. The second main question was how to begin to correct inequities by making marginal redistribution in supply. This required the development of location-allocation models that considered access, utilisation, and patients' choice over the use of hospitals.

Research has shown that if Portugal were to improve equity in its system of hospital finance, it will have to develop new policies to correct wide inequities in the current distribution of hospital resources. This is because distribution does not match need for health care: resources are concentrated in urban areas, while populations in rural areas have poorer access to hospital services. Capital is excessively concentrated in Lisboa, Porto and Coimbra, and the population of Coimbra is using more than its fair share of resources. To improve equity of utilisation at the small area level, resources need to be redistributed from urban hospitals to semi-urban and rural hospitals. However, this would result in a reduction in the total utilisation of hospital services.

Analysis of various definitions of geographical areas were carried out and different measures of equity were designed to illuminate different questions, and all led to the conclusion of excessive concentration of resources in certain areas. The following paragraphs summarise the objectives, methods, implications for Portuguese policy and methodological implications of each chapter. Sections 10.2 and 10.3 summarise implications for policy analysis, make recommendations, and suggest further research.

Chapter 1 analysed the importance of studying geographic equity in the Portuguese hospital sector drawing upon published analysis. It showed that despite the lack of clarity in the formulation of equity and geographical equity objectives in Portuguese health policy, several political statements support the idea that Portuguese health care policy ought to pursue some concept of equity of access. Despite a lack of studies of the hospital sector, empirical evidence suggests wide inequalities in the distribution of hospital resources. As the literature on equity in health and health care lacks a clear framework to analyse equity and a standard definition of the concept of equity of access, various equity concepts were used in this thesis to inform different policies.

Chapter 2 described the context of the Portuguese hospital and health care systems. Analysis based on literature review and interviews shows lack of research and information on equity in the hospital sector. Given the absence of policies to promote equity in planning, regulation, resource allocation and policies to define a clear role of the private sector, there is a long way for the Portuguese health care system to go if it is to deliver its equity objectives.

Chapter 3 presented analysis of geographical inequalities in the Portuguese hospital acute care sector. Using readily available information and crude measures of inequality based on population numbers, it showed: a) that there is a mismatch between supply and demand for hospital resources; b) that this mismatch is exacerbated if need is taken into account; c) that there is evidence of an 'inverse care law'; d) that a distribution of resources in accordance to need would demand a massive redistribution; e) that there is evidence of high variations in efficiency across districts.

Chapter 4 described the structure of a capitation formula for Portugal and introduced several indices to measure geographic inequities. The following adjustments of a capitation formula for the Portuguese hospital sector were defined: population numbers, demographic need, additional need, unavoidable costs and cross-boundary flows. The choice of these adjustments was based on characteristics of the Portuguese health system and availability of data. Four indices were developed to measure inequities in hospital capital, finance and utilisation and to inform policies accordingly.

Chapter 5 developed a capitation formula to measure need for Portuguese hospitals, transferring the technology of methods previously used in England. Estimates of need for hospital care accounted for size of populations, age/sex and additional need. The population adjustment used resident populations but showed problems of accounting for double coverage and for future changes in populations. The demographic adjustment showed that the age/sex cost curve redistributes towards the youngest and eldest; and on the other hand, comparisons with England revealed inadequacies in the current utilisation of hospital care in Portugal. It was shown that SMRs are misleading indicators of additional need in the Portuguese context, as necessary conditions for their reliable use are not fulfilled. ASMRs are more robust indicators for additional need and seem to capture better the effects of urban and rural deprivation on need for hospital care.

The four main methodological findings of Chapter 5 were as follows: a) The need to take account of changes in population: forecasts have been used but there are still problems as these are subject to errors. b) There is scope for normative approaches to estimate the age/sex cost curve, so as to avoid that this curve is highly influenced by the current characteristics of the health care system. c) For many countries, there is a need to develop information systems to produce information on regional coverage of the private sector, as this is crucial for analysing regional need. d) SMRs have been widely used in international literature without proper attention to necessary conditions that need to be satisfied for their use. ASMRs and other morbidity indicators should be considered as alternatives.

Chapter 6 developed a multilevel model to estimate unavoidable costs of hospital care by disentangling causes of allocative inefficiency. It showed the following empirical findings for Portugal: larger hospitals have diseconomies of scale and are mostly affected by allocative inefficiencies, while the ratio of beds to doctors is a cause of allocative inefficiency. There is little flexibility for local management tools, and there are perverse incentives in the system. Inequities generated by an uneven distribution of doctors also create allocative inefficiencies.

The method developed in Chapter 6 to estimate unavoidable costs suggests that hospital costs are affected by the hierarchical structure, geographical location and current incentives. The empirical application of the model has shown the complexity of measuring economies of scale and scope in hospital cost functions and contributed to this literature. Typical approaches to estimate unavoidable costs have sought to estimate costs of individual characteristics, while the model developed has accounted simultaneously for factors such as economies of scale, organisation, input prices and geographic location. Moreover, this model can be applied to other health care systems with central control over planning, management and key resources.

Chapter 7 modelled geographic utilisation flows of hospital care by using a flow demand model to estimate and predict hospital utilisation. Some key findings for Portugal were: a) primary care is a substitute for hospital care; b) there are wide variations in accessibility for different populations; c) the level of supply and the type of hospital have a crucial impact on utilisation flows; d) the index capturing the interaction between utilisation of different hospitals was highly significant.

The proposed demand model estimated flows of patients between areas and hospital sites, at the small area level. This model was intended to overcome weaknesses of earlier models by taking account of the interaction between hospital supply and the utilisation of other hospitals, and of the process of demand for hospital care. Estimates were derived using a two-part model. The model can be applied to other countries, in particular to countries with hospital policies defined at the central level.

Chapter 8 reported results of empirical analysis of geographic redistribution using the various adjustments of the capitation formula and estimates of four inequity indices defined in Chapter 4. A central empirical finding is that there are huge inequities in the distribution of hospital capital, with resources concentrated in Lisboa, Coimbra and Porto. Moreover, inequities in the distribution of utilisation and finance are smaller than for capital. Additionally, when inequity estimates account for the impact of unavoidable costs and cross-boundary flows, the populations of Lisboa and Porto are using approximately their fair shares of resources; by contrast, all estimates shows inequitably high use of resources by the populations of Coimbra. The current system of hospital finance is not creating incentives to correct inequities. Inequities are underestimated given lack of data on activities of the private sector. Finally, any policy to redistribute resources would imply a massive level of redistribution.

Chapter 8 showed that pursuing different equity objectives involve different policy directions for correcting inequities; on the other hand, pursuing one single equity objective was shown to create undesirable impacts on others. Lack of data has constrained the application of methods, in particular in accounting for the role of the private sector, which is also a problem facing other countries. Given the changes in the role of the private sector that many countries (including Portugal) experience, these countries need to develop their statistical systems to produce this data.

Chapter 9 indicated changes in the distribution of hospital supply to improve equity of utilisation and access by developing location-allocation models to redistribute hospital supply. These models were developed to take account of different policy objectives and different assumptions of patients' behaviour. All the location-allocation models pointed towards reductions in the size of central hospitals. The 'best' model (utilisation flows based model -UFBM) indicated increases in the supply of interior, southern and semi-urban hospitals. The net effect was, however, a decrease in total utilisation, which would be unacceptable. The UFBM suggested redistribution towards a small number of hospitals, which shows that improvements in equity require high increases in supply in a small number of sites, and raised the question of whether an increase in supply might be preferable to redistribution for improving equity.

The methods used in Chapter 9 showed that the pursuit of different equity objectives (a different objective for each model) produced different results and there were trade-offs when different redistributive results were evaluated under different equity objectives. Redistributive results were sensitive to certain assumptions of patients' behaviour and to the role of central hospitals. The models can be adapted and applied to hospital systems in other countries.

10.2 Implications for policy analysis

It was shown that there are huge inequities in the distribution of hospital capital in Portugal, with resources concentrated in urban areas with higher socio-economic levels and better access for patients who live there. Redistributing hospital capital in accordance with need would mean reductions in supply in three districts –Lisboa, Porto and Coimbra- and increases in all the other districts using the suggested models. In order to improve equity of access and utilisation, productive capacity should be moved from urban hospitals to semi-urban and rural hospitals. This, however, would lead to a decrease in total utilisation in the hospital system and as a result other policies ought to be explored (such as using the primary sector to compensate for variations in hospital supply, and increasing hospital supply).

Accounting for UCs and CBFs decreases the range of estimates of inequities, while as mentioned before, inequities in utilisation and finance are smaller than for capital. While the populations of Lisboa and Porto districts use less than their fair share of resources in utilisation and finance, the population of Coimbra uses more than its fair share. For certain districts, inequities in capital translate into inequities in utilisation, and these cases should be given higher priority for policy.

Results might suggest that there is a strong case for reducing supply in Coimbra and redistributing it towards other districts. The case for Lisboa and Porto is more complicated: it seems that levels of supply in Lisboa and Porto should be maintained, given that they are justifying their levels of resources, in terms of utilisation and finance, and reducing supply in these districts would potentially decrease total utilisation. Movements of population are likely to result in less of a mismatch between the distribution of supply and need in these districts. On the other hand, estimates of inequities for these districts are expected to downplay the real dimensions of inequity given the neglect of the role of the private sector.

Any attempt to correct inequities should also look into the distribution of hospital doctors, as this was shown to interact with access to hospital care. Low levels of doctors in rural areas lead to inequities (as their numbers are not related to need) and allocative inefficiencies (as they imply an under-utilisation of beds and equipment in rural hospitals). Correcting geographical inequities in the distribution of doctors is expected to rectify inefficiencies in the system.

Hospital policies should also look into the role of primary care, as well as interactions between the primary and hospital sectors. Utilisation of primary care was shown to be substituting for hospital care (Chapter 7) and the high relative levels of primary care provision in the Coimbra and Lisboa districts (Chapter 3) signify inequities, given the high levels of hospital supply in these districts.

10.3 Further research

This thesis suggests the need for further research into two categories: research that directly follows from the findings of the thesis and related research areas that have been out of scope of this thesis.

Some research areas that directly follow from this thesis are:
a) Modelling an adjustment of the capitation formula to take account of the role of the private sector. In the light of the increasing role for the private sector in Portugal (as in other countries), modelling this adjustment seems to be of crucial importance;
b) Adapting the capitation formula so that it can be used in the redistribution of hospital resources in Portugal; and
c) Relating future investment policies with allocation of current expenditure in order to correct inequities in capital and finance.

Some areas that would provide complementary information for equity-pursuing policies in the distribution of hospital care could also be researched. First, policies to correct inequities in geographic distribution of hospital doctors could be designed. Second, this thesis has modelled health care but has not dealt with impact on health outcomes; future research could link some models with information on health outcomes, for example building location-allocation models that pursue equity in health objectives. Third, research could focus on the possible implications of redistribution on costs: changing, for instance, the size of hospitals is likely to modify unit costs and such effects need to be considered in resource allocation.

BIBLIOGRAPHY

Advisory Committee on Resource Allocation (United Kingdom). 1999. Report of the Advisory Committee on Resource Allocation: Advisory Committee on Resource Allocation.

Ai, C., and E. C. Norton. 2000. Standard errors for the retransformation problem with heteroscedasticity. *Journal of Health Economics* 19:697-718.

AIMMS Academic (Network) - Unrestricted 3.1.468p, CONOPT Solver 2.070G (DLL). Paragon Decision Technology, Netherlands.

Aletras, V., A. Jones, and T. Sheldon. 1997. Economies of scale and scope. In *Concentration and Choice in Healthcare*, edited by B. Ferguson, T. Sheldon and J. Posnett. London: Jonh Harper.

Alexander, J. A., S. Y. D. Lee, J. R. Griffith, S. S. Mick, X. Lin, and J. B. Holl. 1999. Do market-level hospital and physician resources affect small area variation in hospital use? *Medical Care Research Review* 56 (1):94-117.

Anas, A. 1983. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research -B* 17:13-23.

Andersen, R. M. 1995. Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior* 36 (1):1-10.

Andersen, R. M., and J. F. Newman. 1973. Societal and individual determinants of medical utilisation in the U.S. *Milbank Memorial Fund Quaterly* 51 (1):95-124.

Anderson, G. P., E. P. Steinberg, N. R. Powe, S. Antebi, J. Whittle, S. Horn, and R. Herbert. 1990. Setting payment rates for capitated systems: a comparison of various alternatives. *Inquiry* (27):225-33.

Andrade, A., C. F. Branco, and R. Sepúlveda. 1996. Taxas moderadoras: sucessos e fracassos, 4 anos depois. Research Paper, Escola Nacional de Sáude Pública, Universidade Nova de Lisboa, Lisboa.

Annel, A., and H. Barnum. 1998. The allocation of capital and health sector reform. In *Critical Challenges for Health Care in Europe*, edited by J. Figueras, R. Saltman and C. Sakellarides. Buckingham: Open University Press.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*, *Studies in Operational Regional Science*. The Netherlands: Kluwer Academic Publishers.

Arcview GIS 3.2a (Geographic Information Systems).

Arndt. 2001. *Information Measures: Information and its Description in Science and Engineering -With 64 Figures*. Berlin: Springer-Verlag.

Artells, J. 1996. La inexistente reforma del S.N.S. en España. In *As Reformas dos Serviços de Sáude*, edited by A. Vaz, C. G. Pinto, F. Ramos and J. Pereira. Lisboa: Associação Portuguesa de Economia da Sáude.

Assembleia da República. 1976. *Constituição da República Portuguesa*. Lisboa: Assembeia da República.

———. 1979. Lei de bases da saúde. *Diário da República.*

———. 1990. Lei 48/90: Lei de bases da saúde. *Diário da República* 195:3452-9.

———. 1992. *Constituição da República Portuguesa*. Lisboa: Assembeia da República.

———. 2002. Proposta de lei n 17/IX: Aprova o novo regime jurídico da gestão hospitalar e procede à primeira alteração à lei n 48-90, de 24 de Agosto. *Diário da República.*

Atkinson, A. B. 1983. *The Economics of Inequality*. 2nd ed. Oxford: Clarendon Press.

———. 1989. *Measuring Inequality and Differing Social Judgements -Discussion Paper 129*. Edited by A. B. Atkinson, M. A. King and N. H. Stern, *Taxation, Incentives and the Distribution of Income*: Economic and Social Research Council Programme.

Avella, P., S. Benati, L. Canovas-Martinez, K. Dalby, D. Di-Girolamo, B. Dimitrijevic, and G. HGjiani. 1998. Some personal views on the current state and the future of Locational Analysis. *European Journal of Operational Research* 104:269-87.

Barreto, A., ed. 2000. *A Situação Social em Portugal 1960-1999, Volume II, Indicadores Sociais em Portugal e na União Europeia*. Lisboa: Imprensa de Ciências Sociais and Eurostat.

Barros, P. P., and C. Sena. 1999. Quanto maior melhor? Redimensionamento e economias de escala em três hospitais portugueses. *Revista Portuguesa de Saúde Pública* 17 (1):5-18.

Bennet, W. D. 1981. A location-allocation approach to health care facility location: A study of the undoctored population in Lansing, Michigan. *Social Science and Medicine* 15 (D):305-12.

Bentes, M. 1992. DRG-based funding in Portugal: two years after. Paper read at Third EURODRG workshop, May, at Madrid.

———. 1995. Financing hospitals based on case-mix: the situation in Portugal. Paper read at Financement des hospitaux base sur le cout des pathologies en Europe, 18.01.1995, at Centre Hospitalier Intercommunal de Creteil.

Bentes, M., M. L. Gonçalves, E. Pina, and M. Santos. 1994. Case-mix and quality: the utility of DRG information. Paper read at S210: EFMI-WG1 Workshop -Case-mix severity of cases and quality of care, 25 May.

Bentes, M., M. L. Gonçalves, E. Pina, and J. Urbano. 1993. The utility of DRGs to target quality issues. Paper read at 9th International PCS/E Conference, 15-18 September, at Munich.

Bentes, M. E., J. Urbano, M. C. Carvalho, and M. S. Tranquada. 1993. Using DRGs to fund hospitals in Portugal: an evaluation of the experience. In *Diagnostic Related Groups in Europe: Uses and Perspectives*, edited by M. Casas and M. M. Wiley. Germany: Springer-Verlag.

Bentes, M., M. C. Mateus, S. Esteves, M. C. Valente, and J. Veertres. 1996. Towards a more comprehensive financing system for the Portuguese NHS hospitals. Paper read at 12th International PCS/E Working Conference, at Sidney, Australia.

Bentes, M., M. C. Valente, C. Mateus, S. Esteves, and Instituto de Gestão Informática e Financeira da Saúde. 1997. Feedback and audit: ingredients for quality improvement. Paper read at 13th PCS/E International Working Conference, 1-3 October, at Florence, Italy.

Benzeval, M., K. Judge, and M. Whitehead. 1995. The role of the NHS. In *Tackling Inequalities in Health -An Agenda for Action*, edited by M. Benzeval, K. Judge and M. Whitehead. London: King's Fund.

Berki, S. 1972. *Hospital Economics*: Lexington.

Bevan, G. 1988. How should district allocations be altered for changes in cross boundary flows? *British Medical Journal* 296:144-5.

Bevan, G., and R. Ingram. 1987. Variations in admission rates: implications for equitable allocation of resources. *British Medical Journal* 295:1039-42.

Black, D., J. N. Morris, C. Smith, and P. Townsend. 1982. The Black report. In *Inequalities in Health*, edited by P. Townsend and N. Davidson. London: Penguin books.

Black, N., S. Langham, and M. Petticrew. 1995. Coronary revascularisation: Why do rates vary geographically in the UK? *Journal of Epidemiology Community Health* 49:408-12.

Bland, M. 2000. *An Introduction to Medical Statistics*. Third ed, Oxford Medical Publications. Oxford: Oxford University Press.

Blough, D. K., C. W. Madden, and M. C. Hornbrook. 1999. Modeling risk using generalized linear models. *Journal of Health Economics* 18:153-71.

Bond, J., B. Gregson, M. Smith, J. Lecouturier, N. Rousseau, and H. Rodgers. 2000. Predicting place of discharge from hospital for patients with a stroke or hip fracturee on admission. *Journal Health Services Research and Policy* 5 (3):133-9.

Bowling, A. 1997. *Research Methods in Health: Investigating Health and Health Services*. Buckingham: Open University Press.

Brown, H. S. 2001. Income, location, and the demand for health care from public, nonprofit, and for-profit hospitals. *Journal Health Care Finance* 27 (4):24-38.

Bruneau, T. C., P. N. Diamandouros, R. Gunther, A. Lijpart, L. Morlino, and R. A. Brooks. 2001. Democracy, Southern European Style. In *Parties, Politics and Democracy in the New Southern Europe*, edited by P. N. Diamandouros and R. Gunther. Baltimore: John Hopkins.

Butler, J. R. G. 1995. *Hospital Cost Analysis*. Edited by P. Zweifel and H. E. Frech, *Developments in Health Economics and Public Policy*. Dordrecht: Kluwer Academic.

Campos, A. C. 1984. Hospitals and Health Services in Portugal: Setting the Scene - I.H.F. Study Tour of Hospitals. Lisboa.

———. 1991. Public needs, private services. The NHS and public-private mix in Portugal. *Revista Portuguesa de Saúde Pública* 9 (1):5-8.

———. 2002. Despesa e défice na saúde: o percurso financeiro de uma política pública. *Análise Social* XXXVI (161):1079-1104.

Carey, K. 2000. A multilevel modelling approach to analysis of patient costs under managed. *Health Economics* 9:435-46.

Carreira, C. M. G. 1999. *Economias de escala e de gama nos hospitais públicos portugueses: uma aplicação da função de custo variável translog*. Edited by Associação Portuguesa de Economia da Sáude. Vol. 3/99, *Documentos de trabalho*. Lisboa: Associação Portuguesa de Economia da Sáude.

Carr-Hill, R. A., and T. Sheldon. 1992. Rationality and the use of formulae in the allocation of resources to health care. *Journal of Public Health Medicine* 14 (2):117-26.

Carr-Hill, R. A., G. Hardman, S. Martin, S. Peacock, T. A. Sheldon, and P. Smith. 1994. *A Formula for Distributing NHS Revenues Based on Small Area Use of Hospital Beds*. York: Centre for Health Economics of The University of York.

Carr-Hill, R. A., T. A. Sheldon, P. Smith, S. Martin, S. Peacock, and G. Hardman. 1994. Allocating resources for health authorities: development of method for small area analysis of use of inpatient services. *British Medical Journal* 309:1046-9.

Carstairs, V., and R. Morris. 1991. *Deprivation and Health in Scotland*. Aberdeen: Aberdeen University Press.

Carvalho, R. 1998. Comentário. *Revista Portuguesa de Saúde Pública* 16 (2):27-9.

Casasnovas, G. L. 1999. Health care and cost containment in Spain. In *Health Care and Cost Containment in the European Union*, edited by E. Mossialos and J. Le Grand. Hants: Ashgate.

Cho, C.-J. 1998. An equity-efficiency trade-off model for the optimum location of medical care facilities. *Socio-Economic Planning Sciences* 32 (2):99-112.

Church, R. L., and D. J. Eaton. 1987. Hierarchical location analysis using covering objectives. In *Spatial Analysis and Location-Allocation Models*, edited by A. Ghosh and G. Rushton. New York: Van Nostrand Reinhold Company.

Cliquet, R., and K. Thienpont. 1995. *Population and Development -A Message from the Cairo Conference*, *European Studies of Population*. Dordrecht: Kluwer Academic Publishers.

Collins, C. D., A.T. Green, and D. J. Hunter. 1994. International transfers of National Health Service reforms: problems and issues. *The Lancet* 344:248-50.

Colomer, J. M. 1996. Spain and Portugal -Rule by party leadership. In *Political Institutions in Europe*, edited by J. M. Colomer. London: Routledge.

Congdon, P. 2001. The development of gravity models for hospital patient flows under systems change: A Bayesian modelling approach. *Health Care Management Science* 4:289-304.

Cote, M. J. 1999. Patient flow and resource utilisation in an outpatient clinic. *Socio-Economic Planning Sciences* 33:231-45.

Cowell, F. A. 1995. *Measuring inequality.* Edited by F. A. Cowell. 2nd ed, *LSE Handbooks in Economics Series.* Hertfordshire: Prentice Hall, Harvester Wheatsheaf.

Cremieux, P.-Y., and P. Ouellette. 2001. Omitted variable bias and hospital costs. *Journal of Health Economics* (20):271-82.

Culyer, A. J., and A. Wagstaff. 1992. Need, equity and equality in health and health care. York: Centre for Health Economics, Health Economics Consortium, University of York.

———. 1993. Equity and equality in health and health care. *Journal of Health Economics* 12 (5):431-57.

Current, J., H. Min, and D. Schilling. 1990. Multi-objective analysis of facility location decisions. *European Journal of Operational Research* 49:295-307.

Departamento de Estudos e Planeamento da Saúde. 1997/8. *Relatórios do PIDDAC.* Edited by Ministério da Saúde. Lisboa: Departamento de Estudos e Planeamento da Saúde, Ministério da Saúde.

———. 1997a. *Estabelecimentos Hospitalares.* Lisboa: Departamento de Estudos e Planeamento da Saúde, Ministério da Saúde.

———. 1997b. *Inquérito Nacional de Saúde 1995/1996, Continente, Dados Gerais.* Edited by Ministério da Saúde. Lisboa: Departamento de Estudos e Planeamento da Saúde, Ministério da Saúde.

Department of Health (United Kingdom). *Data Sources for the 1996/7 Weighted Capitation Targets* 1996 [cited 2000.

———. 1999. Resource Allocation: Weighted Capitation Formulas. Leeds: Resource Allocation and Funding Team.

Department of Health and Social Security (United Kingdom). 1976. *Sharing Resources for Health in England -Report of the Resource Allocation Working Party.* London: HMSO.

Department of Health. 1998. *Public Health database 1998.* Department of Health, DGA, Direcção Geral do Ambiente. 1998. Mapa por concelhos -Portugal. Lisboa: CNIG -Centro Nacional de Informação Cartográfica.

DGS. 1998. *Elementos Saúde /1996 Estatísticos*. Lisboa: Ministério da Saúde.

———. 2001. Psiquiatria e saúde mental: Rede de referenciação hospitalar. Lisboa: Direcção-Geral da Saúde.

DGS, and Ministério da Saúde. 1996. *Portugal Saúde 1996*. Edited by Direcção Geral da Saúde and Ministério da Saúde. Lisboa.

———. 1998. *Mortalidade infantil perinatal e materna -Estatísticas 1992-1996*. Edited by Ministério da Saúde. Lisboa: Ministério da Saúde.

———. 1998a. *Hospitais Distritais Nível 1 -Estatísticas 1996*. Edited by Ministério da Saúde. Lisboa: DGS, Ministério da Saúde, Portugal.

———. 1998b. *Hospitais Distritais Gerais -Estatísticas 1996*. Edited by Ministério da Saúde. Lisboa: DGS, Ministério da Saúde, Portugal.

———. 1998c. *Hospitais Centrais e Especializados -Estatísticas 1996*. Edited by Ministério da Saúde. Lisboa: DGS, Ministério da Saúde, Portugal.

———. 1998d. *O Hospital Português*: Ministério da Sáude, Portugal.

DGS, and Ministério da Saúde. *Óbitos por Distritos, Sexo e Grupos de Idade*. Ministério da Saúde 1994/5/6/7/8 [cited.

Dias, J. A. 1994. Évora e Beja: vizinhos, mas... diferentes! -Gradientes na morbi-mortalidade e determinantes de risco a nível distrital. *Revista Portuguesa de Saúde Pública* 12(3):33-47.

Diderichsen, f. , E. Varde, and M. Whitehead. 1997. Resource allocation to health authorities: the quest for an equitable formula in Britain and Sweden. *British Medical Journal* 315:875-8.

Dismuke, C. E. 1996. A preliminary analysis of the DRG system in Portugal: hospital response as measured by length of stay. In *As Reformas dos Serviços de Sáude*, edited by A. Vaz, C. G. Pinto, F. Ramos and J. Pereira. Lisboa: Associação Portuguesa de Economia da Sáude.

Dixon, A., and E. Mossialos. 2000. Has the Portuguese NHS achieved its objectives of equity and efficiency? *International Social Security Review* 53 (4):49-78.

Donabedian, A. 1988. The quality of care. How can it be assessed? *Journal of the American Medical Association* 260 (12):1743-8.

Drummond, M. F., B. O'Brien, G. L. Stoddart, and G. W. Torrance. 1999. *Methods for the Economic Evaluation of Health Care Programmes*. 2nd ed. Oxford: Oxford University Press.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse. 1983. A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics* 1 (2):115-126.

Duncan, C., K. Jones, and G. Moon. 1998. Context, composition and heterogeneity: using multilevel models in health research. *Social Science and Medicine* 46 (4):97-117.

El-Darzi, E., C. Vasilakis, T. J. Chaussalet, and P.H. Millard. 2000. A simulation model to evaluate the interaction between acute, rehabilitation, long stay care and the community. In *Decision-Making: Recent Developments and Worldwide Applications*, edited by S. H. Zanakis, G. Doukakis and C. Zopounidis. Dordrecht: Kluwer Academic Publishers.

EOHCS. 1999. *Health Care Systems in Transition -Portugal.* Edited by European Observatory on Health Care Systems, *Health Care Systems in Transition.* Copenhagen: EOHCS, European Observatory on Health Care Systems.

Erkut, E. 1993. Inequality measures for location problems. *Location Science* 1 (3):199-217.

Erlander, S. 1977. Accessibility, entropy and the distribution and assignment of traffic. *Transportation Research* 11:149-53.

Erlenkotter, D. 1983. On the choice of models for public facility location. In *Locational Analysis of Public Facilities*, edited by J.-F. Thisse and H. G. Zoller. New York: North Holland.

Eurostat, INE, and European Commission. 1998. *Portrait of the Regions -Portugal.* Edited by Eurostat, *Portrait of the regions.* Lisboa: Office for Official Publications of the EC.

Evans, R. G., and G. L. Stoddart. 1990. Producing health, consuming health care. *Social Science and Medicine* 31 (12):1347-63.

Eyles, J., and S. Birch. 1993. A population needs-based approach to health-care resource allocation and planning in Ontario: a link between policy goals and practice? *Canadian Journal Public Health* 84 (2):112-7.

Fang, S.-C., J. R. Kajasekera, and H. S. J. Tsao. 1997. *Entropy Optimization and Mathematical Programming.* Edited by F. S. Hillier, *International Series in*

Operations Research and Management Science. Massachusetts: Kluwer Academic Publishers.

Fattore, G. 1999a. Cost containment and reforms in the Italian National Health Service. In *Health Care and Cost Containment in the European Union*, edited by E. Mossialos and J. Le Grand. Hants: Ashgate.

———. 1999b. Cost containment and health care reforms in the British NHS. In *Health care and cost containment in the European Union*, edited by E. Mossialos and J. Le Grand. Hants: Ashgate.

Feldstein, P. J. 1988. *Health Care Economics*. Edited by W. M. Publication. 3rd ed, *Wiley Series in Health Services*. New York: John Wiley & Sons.

Folland, S. T. 1983. Predicting hospital market shares. *Inquiry* 20:34-44.

Folland, S. T., and R. A. Hofler. 2001. How reliable are hospital efficiency estimates? Exploiting the dual homothetic function. *Health Economics* 10:683-98.

Folland, S., and M. Stano. 1990. Small Area Variations: A critical review of propositions, methods, and evidence. *Medical Care Review* 47 (4):419-65.

Folland, S., A. C. Goodman, and M. Stano. 1997. *The Economics of Health and Health Care*. 2 ed. New Jersey: Prentice-Hall.

Forsund, F. R., C. A. K. Lovell, and P. Schmidt. 1980. A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics* 13:5-25.

Gaffey, W. R. 1976. A critique of the standardized mortality ratio. *Journal of Occupational Medicine* 18 (3):157-60.

Gerdtham, U.-G., M. Lothgren, M. Tambour, and C. Rehnberg. 1999. Internal markets and health care efficiency: A multiple-output stochastic frontier analysis. *Health Economics* 8:151-64.

Getzen, T. E. 2000. Health care is an individual necessity and a national luxury: Applying multilevel decision models to the analysis of health care expenditures. *Journal of Health Economics* 19:259-70.

Ghosh, A., and G. Rushton. 1987. Progress in location-allocation modelling. In *Spatial Analysis and Location-Allocation Models*, edited by A. Ghosh and G. Rushton. New York: Van Nostrand Reinhold Company.

Giraldes, M. R. 1988. The equity principle in the allocation of health care expenditure on primary health care services in Portugal: the human capital approach. *International Journal Health Planning and Management* 3 (3):167-83.

———. 1989. Análise crítica da estrutura da despesa em cuidados de sáude primários. *Revista Portuguesa de Sáude Pública* 7 (3):31-40.

———. 1990. The equity and efficiency principle in the financing system of the NHS in Portugal. *Health Policy* 14:13-28.

———. 1995a. Portugal: current distribution and development of the human health resources. *Cahiers Sociologie Demographie and Medicine* 3-4:343-60.

———. 1995b. Caracterização dos principais recursos humanos de sáude em Portugal - sua evolução de curto prazo e desigualdades na respectiva distribuição. *Revista Portuguesa de Sáude Pública* 13 (3):5-20.

———. 1998. Morbilidade declarada no INS 1995/96. Que respostas? -Uma abordagem realizada numa perspectiva de equidade. *Revista Portuguesa de Sáude Pública* 16 (1):43-60.

Giraldes, M. R., and A. C. C. Ribeiro. 1995. Desigualdades sócio-económicas na mortalidade em Portugal no período 1980/82-1990/92. *Revista Portuguesa de Saúde Pública* 13 (4):5-28.

Glazier, R. H., E. M. Badley, J. E. Gilbert, and L. Rothman. 2000. The nature of increased hospital use in poor neighbourhoods: Findings from a Canadian Inner City. *Canadian Journal of Public Health* 91 (4):268-73.

Gonzalez, M. P. L. A. 2001. Modelos multicriterio para la financiación de los Hospitales Comarcales del Servicio Andaluz de Salud. PhD Thesis, Departamento de Economía Aplicada (Matemáticas), Universidad de Málaga, Málaga.

Goodwin, P., and G. Wright. 1998. *Decision Analysis for Management Judgment*. 2nd ed. Chichester: Wiley.

Grannemann, T. W., and R. S. Brown. 1986. Estimating hospital costs -a multiple-output analysis. *Journal of Health Economics* 5:107-27.

Grossman, M. 1972. On the concept of health capital and demand for health. *Journal of Political Economy* 80 (2):223-55.

Grundy, E. M. D. 1996. Populations and population dynamics. In *Oxford Textbook of Public Health*, edited by R. Detels, W. Holland, J. McEwan and G. S. Omenn. Oxford: Oxford University Press.

Gujarati, D. N. 1995. *Basic Econometrics*. 3rd ed. International: McGraw Hill.

Hakkinen, U. 1999. Cost containment in Finnish health care. In *Health Care and Cost Containment in the European Union*, edited by E. Mossialos and J. Le Grand. Hants: Ashgate.

Hallefjord, A., and K. Jornsten. 1984. Gravity models with multiple objectives -Theory and applications. *Transportation Research -B*:19-39.

Hanlon, P., D. Walsh, B. W. Whyte, S. N. Scott, P. Lightbody, and L. M. Gilhooly. 1998. Hospital use by an ageing cohort: An investigation into the association between biological, behavioural and social risk markers and subsequent hospital utilization. *Journal of Public Health Medicine* 20 (4):467-76.

Hansen, P., D. Peeters, and J.-F. Thisse. 1983. Public facility location models: A selective survey. In *Locational Analysis of Public Facilities*, edited by J.-F. Thisse and H. G. Zoller. Amsterdam: North Holland.

Hodgson, M. J., K.E. Rosing, and A. L. G. Storrier. 1996. Applying the flow-capturing location-allocation model to an authentic network: Edmonton, Canada. *European Journal of Operational Research* 90:427-443.

Holland, W. W. 1986. The National Health Service -the RAWP review: pious hopes. *The Lancet* ii:1087-90.

———. 1998. Lessons learnt many years ago have been forgotten. *British Medical Journal* 316:939.

Holland, W., and Working Group on Health Services and Avoidable Death. 1991. *European Community Atlas of Avoidable Death.* Edited by J. Buxton, H. Stein, G. M. Etten and A. E. Baert. 2nd ed, *Commission of the European Communities Health Services Research Series No. 6.* Oxford: Oxford University Press.

Holland, Walter W., and Commission of the European Communities. Working Group on Health Services and Avoidable Death. 1997. *European Community Atlas of Avoidable Death 1985-89.* Edited by J. Buxton, H. Stein, L. Ayres, G. M. Etten, A. E. Baert and Fracchia. 2nd ed, *Commission of the European Communities Health Services Research Series No. 9.* Oxford: Oxford University Press.

Hurst, J. 1996. The NHS reforms in a international context. In *Reforming Health Care Systems -Experiments with the NHS*, edited by A. J. Culyer and A. Waggstaff. Cheltenham: Edward Elgar.

Hutchison, B., J. Hurley, R. Reid, J. Dorland, S. Birch, M. Giacomini, and G. Pizzoferrato. 1999. Capitation formulae for integrated health systems: a policy synthesis. Hamilton: Centre for Health Economics and Policy Analysis, McMaster University.

IGIF. 1997. *Serviço Nacional de Saúde -Receitas e Despesas 1994-95*. Edited by M. d. Saúde. Lisboa: Ministério da Saúde.

———. 1998. *Orçamento do SNS para 1999 (Orçamento Ordinário) -Critérios de Financiamento*. Lisboa: Ministério da Sáude.

———. 1999. *Agrupamento dos Hospitais do Serviço Nacional de Saúde*. Lisboa: Ministério da Saúde.

———. 2000. *Serviço Nacional de Saúde -Contas Globais 1998*. Edited by Ministério da Saúde. Lisboa: Ministério da Saúde.

INE. 1990/1/2/3/4/5/6/7/8. *Estimativas da população residente por NUTS e distritos: estimativas segundo o sexo e por idades*. Lisboa: Ine.

———. 1993a. *Census 91 -XIII Recenseamento Geral da População, Lisboa e Vale do Tejo*. Lisboa: Instituto Nacional de Estatística.

———. 1993b. *Census 91 -XIII Recenseamento Geral da População, Norte*. Lisboa: Instituto Nacional de Estatística.

———. 1993c. *Census 91 -XIII Recenseamento Geral da População, Centro*. Lisboa: Instituto Nacional de Estatística.

———. 1993d. *Census 91 -XIII Recenseamento Geral da População, Alentejo*. Lisboa: Instituto Nacional de Estatística.

———. 1993e. *Census 91 -XIII Recenseamento Geral da População, Algarve*. Lisboa: Instituto Nacional de Estatística.

———. 1998. *Portugal social 1991-1995*. Lisboa: Instituto Nacional Estatística.

———, ed. 2000. *Estatísticas da Saúde 1998*. Edited by Instituto Nacional de Estatística, *Estatísticas da Saúde*. Lisboa: Instituto Nacional de Estatística.

———. 2002. Census 2001 -Resultados definitivos. *Destaque do INE*, 21.10.2002, 34.

INE, Instituto Nacional Estatística. 2000. *Infoline Service*. INE, Instituto Nacional Estatística., 2000 [cited 2000].

INE, Direcção Geral do Centro, and Gabinete de Estudos Regionais. 2000. Estudo sobre o poder de compra concelhio. Coimbra: Núcleo de Estudos Regionais da Direcção Regional do Centro.

INFARMED. 1999. *Informação Estatística -INFARMED*. Edited by INFARMED and Ministério da Saúde, *Informação Estatística*. Lisboa: Ministério da Saúde.

Inskip, H., V. Beral, and P. Fraser. 1983. Methods for age-adjustment of rates. *Statistics in Medicine* 2:455-66.

Intercooled Stata 7.0 for Windows 98/95/NT 15-student Stata for Windows (network). Stata Corporation.

Jacobs, P. 1996. *The Economics of Health and Medical Care*. Maryland: Aspen Publishers.

Johnston, R. J., D. Gregory, and D. M. Smith, eds. 1994. *The Dictionary of Human Geography*. 3rd ed. Surrey: Blackwell.

Jornsten, K. O., and J. T. Lundgren. 1989. An entropy-based model split model. *Transportation Research -B* 23B (5):345-59.

Judge, K., and N. Mays. 1994. Equity in the NHS allocating resources for health and social care in England. *British Medical Journal* 308:1363-6.

Kirkup, B., and D. Forster. 1990. How will health needs be measured in districts? Implications of variations in hospital use. *Journal of Public Health Medicine* 12 (1):45-60.

Kostreva, M. M., and W. Ogryczac. 1999. Equitable approaches to location problems. In *Spatial Multicriteria Decision Making and Analysis*, edited by J.-C. Thill. Aldershot: Ashgate.

Kunst, A. E., F. Groenhof, J. P. Mackenbach, and The EU working group on socioeconomic inequalities in health. 1998. Mortality by occupational class among men 30-64 years in 11 European countries. *Social Science and Medicine* 46 (11):1459-76.

Lamers, L. M. 1998. Risk-adjusted capitation methods: developing a diagnostic cost groups classification for the Dutch situation. *Health Policy* 45:15-32.

Lave, J. R., and B. L. Lave. 1970. Hospital cost functions. *American Economic Review* 60 (3):379-95.

Le Grand, J. 1982. *The Strategy of Equality -Redistribution and the Social Services*. London: George Allen.

———. 1987. Three essays on equity -Essay 1: Equity, health and health care. London: Suntory Toyota International Centre for Economics and Related Disciplines, London School of Economics.

Le Grand, J., and P. Vizard. 1998. The National Health Service: crisis, change, or continuity? In *The State of Welfare: the Economics of Social Spending*, edited by H. Glennerster and J. Hills. Oxford: Oxford University Press.

Leonardi, G. 1983. The use of random-utility theory in building location-allocation models. In *Locational Analysis of Public Facilities*, edited by J.-F. Thisse and H. G. Zoller. New York: North Holland.

Leung, S. F., and S. Yu. 1996. On the choice between sample selection and two-part models. *Journal of Econometrics* 72:197-229.

Li, T., and R. Rosenman. 2001. Estimating hospital costs with a Generalized Leontief function. *Health Economics* 10 (6):523-38.

Lima, E. 1998. The financing of health care: an analysis of the impact of the Portuguese hospital financing system. PhD, Economics, University of Nottingham, Nottingham.

———. 2000. A produção e a estrutura de custos dos hospitais públicos. Lisboa: Associação Portuguesa de Economia da Saúde.

Linna, M., U. Hakkinen, and E. Linnakko. 1998. An econometric study of costs of teaching and research in Finnish hospitals. *Health Economics* 7:291-305.

Long, J.L. 1981. The role of consumer location in the demand for inpatient care. *Inquiry* 18:266-73.

Love, R. F., J. G. Morris, and G. O. Wesolowsky. 1988. *Facilities Location: Models and Methods*. Edited by S. I. Gass, *Publications in Operations Research Series*. New York: North-Holland.

Lucas, J. S. 1986. Portugal: a research note. In *The Health Burden of Social Inequities*, edited by R. Illsley and P. G. Svensson. Copenhagen: WHO Regional Office for Europe.

Lynch, W. D., D. W. Edington, and A. Johnson. 1996. Predicting the demand for health care. *Health Care Forum Journal*:20-4.

Macintyre, S. 1997. The Black report and beyond what are the issues? *Social Science and Medicine* 44 (6):723-45.

Mackenbach, J. P., M. H. Bouvier-Colle, and E. Jougla. 1990. "Avoidable" mortality and health services: a review of aggregate data studies. *Journal of Epidemiology and Community Health* 44 (106-11).

Malmstrom, M., S. Johansson, and J. Sundquist. 2001. A hierarchical analysis of long-term illness and mortality in socially deprived areas. *Social Science Medicine* 53:265-75.

Mandell, M. B. 1991. Modelling effectiveness-equity trade-offs in public service delivery systems. *Management Science* 27 (4):467-82.

Manning, W. G. 1998. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 17:283-295.

Manning, W. G., and J. Mullahy. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* 20:461-94.

Marianov, V., and D. Serra. 2001. Hierarchical location-allocation models for congested systems. *European Journal of Operational Research* 135:195-208.

Marsh, M. T., and D. A. Schilling. 1994. Equity measurement in facility location analysis: a review and framework. *European Journal of Operational Research* 71:1-17.

Martin, J. J., M.P.L. Amo, R. Caballero, T. Gomez, J. Molina, and F. Ruiz. 2000. A goal programming scheme to determine the budget assignment among the hospitals of a sanitary system. In *Decision-Making: Recent Developments and Worldwide Applications*, edited by S. H. Zanakis, G. Doukakis and C. Zopounidis. Dordrecht: Kluwer Academic Publishers.

Matthew, G. K. 1971. Measuring need and evaluating services. In *Portfolio for Health, The role and Programme of the DHSS in Health Services Research.* edited by G. McLachlan. London: Published for the Nuffield Provincial Hospitals Trust by the Oxford University Press:27-46.

Mayhew, L. D., and G. Leonardi. 1982. Equity, efficiency, and accessibility in urban and regional health care systems. *Environment and Planning A* 14:1479-507.

Mayhew, L. D., R. W. Gibberd, and H. Hall. 1986. Predicting patient flows and hospital case-mix. *Environment and Planning A* 18:619-38.

Maynard, A. 1999. Inequalities in health: an introductory editorial. *Health Economics* 8:281-2.

Mays, N. 1987. Reviewing RAWP: measuring morbidity for resource allocation. *British Medical Journal* 295:703-6.

———. 1995. Geographical resource allocation in the English national health service 1971-94: the tension between normative and empirical approaches. *International Journal of Epidemiology* 24:96-102.

Mays, N., and G. Bevan. 1987. *Resource Allocation in the Health Services -A Review of the Methods of the Resource Allocation Party*. Edited by B. S. Press, *Occasional Papers*. Chippenham: Bedford Square Press.

Mayston, D. 1990. Managing capital resources in the NHS. In *Competition in Health Care -Reforming the NHS*, edited by A. J. Culyer, A. K. Maynard and J. W. Posnett. London: MacMillan Press.

McCullagh, P., and J. A. Nelder. 1983. *Generalized Linear Models*. Edited by D. R. Cox, D. V. Hinkley, N. Reid, D. B. Rubin and B. W. Silverman, *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

McGuire, A., and D. Hughes. 2002. The Economics of the Hospital: Issues of asymmetry and uncertainty as they affect hospital reimbursement. In *Advances in Health Economics*, edited by R. Elliot and A. Scott: Wiley.

McGuire, A., J. Henderson, and G. Mooney. 1988. *The Economics of Health Care*. London: Routledge.

McLafferty, S. 1988. Predicting the effect of hospital closure on hospital utilization patterns. *Social Science and Medicine* 27 (3):255-62.

Mills, A., S. Bennet, P. Siriwanarangsun, and V. Tangcharoensathien. 2000. The response of providers to capitation payment: a case-study from Thailand. *Health Policy* 51:163-80.

Ministério da Saúde. 1968. Decreto No. 48358. *Diário da República* I Série (101):612-26.

———. 1971. Decreto-lei no. 413/71: Organização do Ministério da Saúde e Assistência. *Diário da República* I Série (228):1406-34.

———. 1988. Decreto-lei no. 19/88: Aprova a lei da gestão hospitalar. *Diário da República* 17 (248):248(20)-(23).

———. 1993a. Decreto-lei no. 11/93: Aprovação do Estatuto do Serviço Nacional de Sáude. *Diário da República* 12/93 (15.01.93):129-34.

———. 1993b. Decreto-Lei n.° 10/93: Lei orgânica do Ministério da Saúde. *Diário da República* 12/93 (15.01.93).

———. 1997. Despacho normativo 46/97: criação e entrada de funcionamento das cinco agências regionais de contratualização, visando a separação financiamento-prestação. *Diário da República* 45/97 (08.08.97).

———. 1998a. Aviso nº 922/98 (2ªserie): protocolo de acordo entre o Ministério da Saúde, a Portugal Telecom, S.A., e a Portugal Telecom -Associação de Cuidados de Saúde. *Diário da República* 17/98 (21.01.1998):893-4.

———. 1998b. *Carta de Equipamentos da Saúde*. Lisboa: Ministério da Saúde.

———. 1998c. Decreto-lei 97/98: Modelo explícito para gerir as convenções com o sector privado. *Diário da República* 91 (18.04.1998):1709-10.

———. 1998d. Decreto-lei 117/98: Regime remuneratório experimental dos médicos da carreira de clínica geral que exerçam funções nos centros de saúde do Serviço Nacional de Saúde. *Diário da República* 103 (05.05.1998):1991-4.

———. 1998e. Portaria nº348-B/98: Aprova as tabelas de preços a practicar pelo Serviço Nacional de Saúde em relação a todos os subsistemas de saúde, bem como a quaisquer entidades, públicas ou privadas, responsáveis pela assistência prestada. *Diário da República* I-B Serie (138-B/98):2762(8)-(59).

———. 1999a. Decreto-lei no. 157/99: Estabelece o regime de criação, organização e funcionamento dos centros de sáude. *Diário da República* 108/99 (10.05.99):2424-35.

———. 1999b. Decreto-lei no. 156/99: Estabelece o regime dos sistemas locais de saúde. *Diário da República* 108/99 (10.05.99):2421-4.

———. 1999c. Decreto-lei no. 285/99: Definição e implementação do programa de acesso para recuperação de listas de espera. *Diário da República* 172/99 (27.07.1999):4687.

———. 1999d. Despacho nº5804/99 (2ª serie): Programa de Promoção e Acesso. *Diário da República* 68/99 (22.03.1999):4165-6.

———. 1999e. *Sáude um Compromisso -A Estratégia de Sáude para o Virar do Século (1998-2002)*. Revisão 1999 ed. Lisboa: Ministério da Sáude.

MLwiN 1.10.0006 (Multilevel Modelling). Multilevel Models Project, Institute of Education, London.

Mohan, J. 1983. Location-allocation models, social science and health service planning: An example from North East England. *Social Science and Medicine* 17 (8):493-9.

Mooney, G. H. 1983. Equity in health care: confronting the confusion. *Effective health care* 1 (4):179-84.

Mooney, G., and A. McGuire. 1987. Distributive justice with special reference to geographical inequality and health care. In *Health and Economics*, edited by A. Williams. London: Macmillan Press.

Morgan, M., N. Mays, and W. W. Holland. 1987. Can hospital use be a measure of need for health care? *Journal of Epidemiology Community Health* 41 (4):269-74.

Morgenstern, H. 1982. Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health* 72:1336-44.

Mossialos, E., and J. Le Grand. 1999. Cost containment in the E.U.: An overview. In *Health Care and Cost Containment in the European Union*. London: Ashgate.

Mullahy, J. 1998. Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 17:247-81.

Musgrove, P. 1999. Public spending on health care: how are different criteria related? *Health Policy* 47:207-23.

National Centre for Health Outcomes Development, 1999 [cited 1998].

Newbold, K. B., J. Eyles, S. Birch, and A. Spencer. 1998. Allocating resources in health care: alternative approaches to measuring needs in resource allocation formula in Ontario. *Health Place* 4 (1):79-89.

Newhouse, J. P. 1994. Frontier estimation: how useful a tool for health economics? *Journal of Health Economics* 13:317-22.

Newhouse, J. P., W. G. Manning, E. B. Keeler, and E. M. Sloss. 1989. Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review* 10 (3):41-54.

NHS Centre for Reviews and Dissemination. 1997. Concentration and choice in the provision of hospital services: the relationship between concentration, patient accessibility and utilisation of services. York, UK: University of York.

Nijkamp, P. 1978. Compromise choices in spatial interaction and regional planning models. In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars and J. W. Weibull. Netherlands: North Holland.

OECD. 1994. *OECD: The Reform of Health Care Systems: A Review of 17 OECD Countries*. Edited by OECD, *Health Policy Studies*. Paris: OECD.

———. 1998. OECD economic surveys 1997-1998, Portugal, edited by OECD. Paris: OECD.

———. 1999. OECD economic surveys: Portugal -October 1999. Paris: OECD.

———. Several. *OECD Health Data 1998: A Comparative Analysis of 29 Countries.* OECD 1998 [cited Several].

———. Several. *OECD Health Data 2000: A Comparative Analysis of 29 Countries.* OECD 2000 [cited Several].

Ogryczak, W. 2000. Inequality measures and equitable approaches to location problems. *European Journal of Operational Research* 122:374-91.

O'Kelly, M. E. 1987. Spatial-interaction-based location-allocation models. In *Spatial Analysis and Location-Allocation Models*, edited by A. Ghosh and G. Rushton. New York: Van Nostrand Reinhold Company.

Oliveira, M. 2001a. Insights from Portugal's experience with waiting lists. *Euro Observer* 3 (2):6-7.

———. 2001b. Voluntary health insurance in Portugal: a study for the European Commission. London: London School of Economics and Political Science.

———. 2002. Will hospital management reform in Portugal work? *Euro Observer* 4 (4):3-5.

Oliveira, M., and G. Bevan. 2001. Measuring geographic inequalities in the Portuguese health care system: An estimation of hospital care need. Lisboa: Associação Portuguesa de Economia da Saúde.

Oliveira, M., J. Magone, and J. Pereira. 2003. Nondecision-making and inertia in Portuguese health policy. *Paper submitted to the Journal of Health Politics, Policy and Law.*

Oliver, A. O. 1999. *Risk Adjusting Health Care Resource Allocations -Theory and Practice in the United Kingdom, The Netherlands and Germany.* London: Office for Health Economics.

Paiva, R. L. 1993. A medição da eficiência no sector hospitalar -O caso português. PhD thesis, Instituto Superior de Economia e Gestão -Universidade Técnica de Lisboa, Lisboa.

Palmer, S., P. West, D. Patrick, and M. Glynn. 1979. Mortality indices in resource allocation. *Community Medicine* 1:275-81.

Pereira, J. A. 1990. Equity objectives in Portuguese health policy. *Social Science and Medicine* 31 (1):91-4.

———. 1993. What does equity in health mean? *Journal of Social Policy* 22 (1):19-48.

———. 1995. Equity, health and health care: an economic study with reference to Portugal. PhD thesis, Department of Economics and Related Studies, University of York, York.

———. 1996. Who paid for health care, 1980-90? In *As Reformas dos Serviços de Sáude*, edited by A. Vaz, C. G. Pinto, F. Ramos and J. Pereira. Lisboa: Associação Portuguesa de Economia da Sáude.

———. 1998. Equity of health care finance progressivity in Portugal: how robust? *Estudos de Economia* 18 (3):351-72.

Pereira, J., and C. G. Pinto. 1990. Regressivity in an NHS-type system -the financing of the Portuguese health care. Lisboa: Associação Portuguesa de Economia da Saúde.

———. 1993. Portugal. In *Equity in the Finance and Delivery of Health Care: An International Perspective*, edited by E. Van Doorslaer, A. Wagstaff and F. Rutten. Oxford: Oxford University Press.

Pereira, J. A., A. C. Campos, M. F. Cortês, and C. Costa. 1987. *Equidade Geográfica no Sistema de Saúde Português*. Lisboa.

Pereira, J., A. C. Campos, F. Ramos, J. Simões, and V. Reis. 1999. Health care reform and cost containment in Portugal. In *Health Care and Cost Containment in the European Union*, edited by E. Mossialos and J. Le Grand. London: Ashgate.

Persaud, D., and L. Narine. 1999. Cross-national comparison of capitation funding: the American, British and Dutch experience. *Health Services Management Research* 12:121-35.

Pinto, C. G., and M. Oliveira. 2001. The Portuguese health care system: Current organization and perspectives for reform. In *Como está a Economia Portuguesa?*, edited by Centro de Investigação sobre Economia Portuguesa. Lisboa: Europress.

Pinto, M. M., F. Ramos, and J. Pereira. 2000. Health technology assessment in Portugal. *International Journal of Technology Assessment in Health Care* 16 (2):520-31.

Pohlmeier, W., and V. Ulrich. 1994. An econometric model of the two-part decisionmaking process in the demand for health care. *Journal of Human Resources* 30 (2).

Porell, F. W., and E. K. Adams. 1995. Hospital choice models: A review and assessment of their utility for policy impact analysis. *Medical Care Research Review* 52 (2):158-95.

Powell, M., and G. Boyne. 2001. The spatial strategy of equality and the spatial division of welfare. *Social Policy and Administration* 35 (2):181-94.

Presidência do Conselho de Ministros. 1999. Programa do XIV Governo Constitucional. Lisboa.

———. 2002. Programa do XV Governo Constitucional. Lisbon.

Rahman, S., and D. K. Smith. 2000. Use of location-allocation models in health service development planning in developing nations. *European Journal of Operational Research* 123:437-52.

Reis, V. P. , and R. Carvalho. 1994. Portugal. In *OECD: The Reform of Health Care Systems: A Review of 17 OECD Countries*, edited by OECD. Paris: OECD.

Resource Allocation and Funding Team. 2000. *2000/01 Health Authority Revenue Cash Limits Exposition Book*. Edited by Finance and Performance Directorate and NHS Executive. London: NHS Executive.

Rice, N., and A. Jones. 1997. Multilevel models and health economics. *Health Economics* 6:561-75.

Rice, N., and P. Smith. 1999. *Approaches to Capitation and Risk Adjustment in Health Care: An International Survey*. Edited by Centre for Health Economics. York: Centre for Health Economics.

———. 2000. Editorial: capitation and risk adjustment in health care. *Health Care Management Science* 3:73-5.

Ricketts, T. C., R. Randolph, H. A. Howard, D. Pathman, and T. Carey. 2001. Hospitalization rates as indicators of access to primary care. *Health and Place* 7:27-38.

Ridge, J. C., S. K. Jones, M. S. Nielsen, and A. K. Shahani. 1998. Capacity planning for intensive care units. *European Journal of Operational Research* 105:346-55.

Rohrer, J. E. 1990. Supply-induced demand for hospital care. *Health Services Management research* 3 (1):41-8.

Rushton, G. 1987. Selecting the objective function in location-allocation analysis. In *Spatial Analysis and Location-Allocation Models*, edited by A. Ghosh and G. Rushton. New York: Van Nostrand Reinhold Company.

Santana, P. 1993. Acessibilidade e utilização dos serviços de saúde. Ensaio metodológico em economia da saúde. PhD thesis, Faculdade de Letras da Universidade de Coimbra, Universidade de Coimbra, Coimbra.

———. 1996. Utilização dos cuidados hospitalares. Uma abordagem da geografia da saúde. In *As Reformas dos Sistemas de Saúde*, edited by A. Vaz, C. G. Pinto, F. Ramos and J. Pereira. Lisboa: Associação Portuguesa de Economia da Saúde.

———. 1999. Geografia das desigualdades em saúde e no estado de saúde. In *Livro de Homenagem a Augusto Mantas*, edited by P. P. Barros and J. Simões. Lisboa: Associação Portuguesa de Economia da Saúde.

———. 2000. Ageing in Portugal: regional iniquities in health and health care. *Social Science and Medicine* 50:1025-36.

Santerre, R. E., and S. P. Neun. 1996. *Health Economics: Theories, Insights, and Industry Studies*, *The Irwin Series in Economics*. United States: Irwin.

Savas, S., I. Sheiman, E. Tragakes, and H. Maarse. 1998. Contracting models and provider competition. In *Critical Challenges for Health Care in Europe*, edited by J. Figueras, R. Saltman and C. Sakellarides. Buckingham: Open University Press.

Scott, A., and D. Parkin. 1995. Investigating hospital efficiency in the new NHS: the role of the translog cost function. *Health Economics* 4:467-78.

Scottish Office, Executive Health Department. 1999. *Fair Shares for All*. Edinburgh: Scottish Executive Health Department.

Senior, M., H. Williams, and G. Higgs. 2000. Urban-rural mortality differentials: controlling for material deprivation. *Social Science and Medicine* 51:289-305.

Senn, S. J., and W. B. Samson. 1981. Estimating hospital catchment populations. *Statistician* 31 (1):81-96.

Shaughnessy, P. W. 1982. Methodological issues in per capita measurement in health care. *Health Services Research* 17 (1):61-81.

Shaw, R., and P. Smith. 2001. Allocating health care resources to reduce health inequalities. *Health Care UK*:7-13.

Sheldon, T. A. 1997. Formula fever: allocating resources in the NHS. *British Medical Journal* 315:964.

Sheldon, T. A., G. D. Smith, and G. Bevan. 1993. Weighting in the dark: resource allocation in the new NHS. *British Medical Journal* 306:835-9.

Shmueli, A. 1999. Inferring capitation rates from aggregate health plans' costs. *Health Economics* 8:547-52.

Smith, P., T. A. Sheldon, R. A. Carr-Hill, S. Martin, S. Peacock, and G. Hardman. 1994. Allocating resources for health authorities: results and policy implications of small area analysis of use of inpatient services. *British Medical Journal* 309:1050-4.

Snijders, T., and R. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. London: Sage.

Soderlund, N., and R. Jacobs. 2001. Towards panel data specifications of efficiency measures for English acute hospitals. York: Centre for Health Economics, University of York.

Subramanian, S. V., I. Kawachi, and B. P. Kennedy. 2001. Does the state you live in make a difference? Multilevel analysis of self-rates health in the US. *Social Science Medicine* 53:9-19.

Taket, A. R. 1989. Equity and access: exploring the effects of hospital location on the population served -A case study in strategic planning. *Journal of the Operational Research Society* 40 (11):1001-10.

Taket, A., and L. Mayhew. 1981. Interactions between the supply of and demand for hospital services in London. *Omega* 5 (519-26).

Teixeira, A. M. M. , J. A. C. R. Coutinho, and M. V. Morgado. 1999. *Relacionamento entre o Estado e as Entidades Privadas no Exercício da Acção Social, Módulos PROFISS*. Lisboa: Instituto do Emprego e Formação Profissional, Instituto de Gestão Financeira da Segurança Social.

Theil, H. 1967. *Economics and Information Theory*. Edited by H. Theil. Volume 7 vols, *Studies in Mathematical and Managerial Economics*. Amsterdam: North Holland.

Townsend, P. 2001. Targeting poor health: Professor Townsend's report of the Welsh Assembly's National Steering Group on the allocation of NHS resources. Wales: Health and Social Services Committee of the National Assembly for Wales.

Tranquada, M. S. 1998. O modelo de financiamento do sistema nacional de sáude para 1999. Paper read at Encontro sobre financiamento do sistema de saúde em Portugal, 10 and 11.12.1998, at Lisboa.

Tranquada, M. S., M. M. Martins, and M. Sousa. 2000. Critérios de financiamento dos cuidados de saúde primários (a experiência portuguesa). Paper read at VII Encontro Nacional de Economia da Saúde, 16.11.2000, at Lisboa.

Treurniet, H. F., C. W. N. Looman, P. J. van der Maas, and J. P. Mackenbach. 1999. Variations in "avoidable" mortality: a reflection of variations in incidence? *International Journal of Epidemiology* 28:225-32.

Tribunal de Contas. 1999. *Auditoria Horizontal e Temática, no Âmbito da Aquisição de Bens e Serviços e Facturação a Terceiros, ao Hospital de Santo António dos Capuchos -Relatório nº54/99 -2ª Secção*. Lisboa: Tribunal de Contas.

———. 2000. *Auditoria Horizontal e Temática, no Âmbito da Aquisição de Bens e Serviços e Facturação a Terceiros, ao Hospital de Santa Maria -Relatório nº05/00 -2ª Secção*. Lisboa: Tribunal de Contas.

Tsai, S. P., and C. P. Wen. 1986. A review of methodological issues of the standardized mortality ratio (SMR) in occupational cohort studies. *International Journal of Epidemiology* 15 (1):8-21.

Tudor, H. J. 1971. The inverse care law. *The Lancet* i (405-12).

Urbano, J., M. Bentes, and J. C. Vertrees. 1993. Portugal: National commitment and the implementation of DRGs. In *The Migration of Managerial Innovation: Diagnostic-Related Groups and Health Care Administration in Western Europe*, edited by J. R. Kimberly and G. de Pouvourville. San Francisco: Jossey-Bass.

Urbanos, R., and B. González. 2002. Equidad territorial en España. Paper read at Jornadas de Economia de la Salud, Pamplona, 29.05.2002.

Valente, M. C., S. Estevens, C. T. Costa, A. Barardo, and M. Bentes. 1998. Portugal: on the route to APGs. Paper read at 14th PCS/E International Working Conference, at Manchester, United Kingdom.

Van Barneveld, E. M., R. C. J. A. Van Vliet, and W. P. M. M. Van de Ven. 1997. Risk-adjusted capitation payments for catastrophic risks based on multi-year prior costs. *Health Policy* 39:123-35.

Van Doorslaer, E., A. Wagstaff, and F. Rutten, eds. 1993. *Equity in the Finance and Delivery of Health Care: An International Perspective*, *Commission of the European Communities health services research series, N. 8*. Oxford: Oxford University Press.

Van Vliet, R. C. J. A., and W. P. M. M. Van de Ven. 1992. Towards a capitation formula for competing health insurers. An empirical analysis. *Social Science and Medicine* 34 (9):1035-48.

Vita, M. G. 1990. Exploring hospital production relationships with flexible functional forms. *Journal of Health Economics* 9:1-21.

Vitaliano, D. F. 1987. On estimation of hospital cost functions. *Journal of Health Economics* 6:305-18.

Wagstaff, A. 1993. The demand for health: An empirical reformulation of the grossman model. *Health Economics* 2:189-98.

Wagstaff, A., E. van Doorslaer, H. van der Burg, S. Calonge, and T. Christiansen. 1999. Equity in the finance of health care: some further international comparisons. *Journal of Health Economics* 18:263-90.

Webber, M. 1978. Spatial interaction and the form of the city. In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars and J. W. Weibull. Netherlands: North Holland.

Whitehead, M. 1992. The health divide. In *Inequalities in Health*, edited by P. Townsend and N. Davidson. London: Penguin books.

———. 1994. Equity issues in the NHS: who cares about equity in the NHS? *British Medical Journal* 308:1284-7.

———. 1995. Tackling inequalities: a review of policy initiatives. In *Tackling Inequalities in Health -An Agenda for Action*, edited by M. Benzeval, K. Judge and M. Whitehead. London: King's Fund.

WHO. 1994. Information support for new public health action at district level. Geneva: World Health Organisation.

Williams, H. C. W. L., and M. L. Senior. 1978. Accessibility, spatial interaction and the spatial benefit analysis of land use -transportation plans. In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars and J. W. Weibull. Netherlands: North Holland.

Williams, H. P. 1993a. *Model Building in Mathematical Programming.* 3rd ed. Chichester: John Wiley & Sons.

———. 1993b. *Model Solving in Mathematical Programming.* Chichester: John Wiley and Sons.

Wilson, A. G. 1970. *Entropy in Urban and Regional Modelling.* Edited by R. J. Chorley and D. W. Harvey, *Monographs in Spatial and Environmental Systems Analysis*. London: Pion Limited.

Wilson, R. M. 1988. Evaluating models for planning and resource allocation in health care systems. Doctor of Philosophy, University of Newcastle, Newcastle.

Wilson, R. M., and R. W. Gibberd. 1990. Allocating resources between health care regions: Catchment populations and allocating policy. *European Journal of Operational Research* 49:333-47.

Wilson, R., A. Assefa, G. Briscoe, P. Elias, A. E. Green, A. McKnight, J. Stilwell, and Institute for Employment Research. 1996. Labour market forces and NHS provider costs -Final report. Warwick: Institute for Employment Research, University of Warwich.

Wise, P. 2002. World Report -Portugal. *Financial Times*, 21.20.2002.

Wright, J., and R. Williams. 1998. Health needs assessment: development and importance of health needs assessment. *British Medical Journal* 316:1310-1.

Wu, N. 1997. *The Maximum Entropy Method: With 53 Figures*. Edited by T. S. Huang, T. Kohonen, M. R. Schroeder and H. K. V. Lotsch, *Springer Series in Information Sciences*. Verlin: Springer.

Yule, G. U. 1934. On some points relating to vital statistics, more specifically statistics of occupational mortality. *Journal of the Royal Statistical Society* 97(1):1-84.

Zuckerman, S., J. Hadley, and L. Iezonni. 1994. Measuring hospital efficiency with frontier cost functions. *Journal of Health Economics* 13:255-80.

APPENDICES

APPENDIX A - Complete notation by chapter

Table A.1: Chapter 4

Notation	*Interpretation*
r	r is a geographic district unit (district; for Portugal, r =1,2,..18).
cap_index_r	Relative capitation index for district r, accounting for all the selected adjustments of the capitation formula.
P_r, P	Resident population in district r and total resident population.
I_{2r}	Age and additional need index for district r.
I_{3r}	CBFs index for district r.
I_{4r}	UC index for district r.
$District\_share\_1_r$	Share of need for hospital care for district r.
$District\_share\_2_r$	Share of need for hospital care, adjusted by CBFs for district r.
$District\_share\_3_r$	Share of need for hospital care, adjusted by CBFs and UCs for district r.

Table A.2: Chapter 5

Notation	*Interpretation*
a	Age group a.
X_{1a}	Age (and sex) cost for age (and sex) group a.
de_{ar}	Number of deaths in area r from the age group a.
P_{ar}	Resident population of the age group a in area r.
$r_{ar} = \frac{de_{ar}}{P_{ar}}$	Death rate in area r from the age group a, which corresponds to the definition of age specific mortality rates for area r and for age group a (defined below).
$r_a = \sum_r de_{ar} / \sum_r P_{ar}$	National death rate for age group a.
cutoff	Age reference used in the computation of the potential years of life lost index. It is related to life expectancy.
I_a	Mid-age point of age group a (required to compute the potential years of life lost index).
SMR_r	Standardised mortality ratio index for district r.
$ASMR_{ar}$	Age specific mortality ratio index for age group a and for district r.
$PYLL_r$	Potential years of life lost index for district r.
RMI_r	Relative mortality index for district r.
P_r	Defined in Chapter 4.

Table A.3: Chapter 6

Notation	*Interpretation*
h , h'	Hospital identifier ($h \neq h'$).
c	Types of hospital in the administrative (and hierarchical) classification (for Portugal: c = general central, specialised central, district, level I).
k	Geographical place of location.
l	Type of hospital in the costs' statistics classification (for Portugal: l = central, district, level I).
$COutput_h$	Total cost standardised by an index of hospital production. This indicator is referred to as standardised cost.
$TotCost_h$	Total cost.
$OutputIndex_h$	Equivalent patients index.
$Disch_{hl}$	Number of hospital inpatient discharges of hospital h that belongs to hospital group l .
$Outpat_{hl}$	Number of outpatient attendances of hospital h that belongs to hospital group l .
$Emerg_{hl}$	Number of emergency and accident admissions of hospital h that belongs to hospital group l .
a_l , b_l , c_l	Total unit costs from hospitals of type l , for inpatient discharges, outpatient attendances and emergency and accident admissions, respectively.
do_h	Numbers of doctors.
nu_h	Number of nurses.
be_h	Number of beds.
C , C'	Function linking the standardised cost with the covariates; and linear function linking the natural logarithm of standardised cost with the covariates.
α , β , θ	Parameters from the general hierarchical model.
x_h' , x_h'' , x_h	Explanatory variables vector for standardised costs (x_h). x_h' is the sub-set of variables that have a log-linear function relationship with the dependent variable ($x_h' \subset x_h$); and x_h'' is the sub-set of variables with a semi-log function relationship with the dependent variable ($x_h'' \subset x_h$).
e_h	Random error for the general hierarchical model.
α_0 , α_1	Coefficients of the fixed part of the HFEM (excluding the geographical and hospital group related coefficients).
g_{hk}	Dummy variables for the geographical location of hospital h in place k (HFEM and MLM).
α_{2k}	Fixed coefficients for dummies of the geographical area k (geographical related coefficients) (HFEM).
t_{hc}	Dummy variables for the hospital h in the administrative hierarchy c (HFEM).
α_{3c}	Fixed coefficients for dummies of the administrative group c (HFEM).
e_{hck}^{HFEM}	Random error for the HFEM.
$\beta_0 , \beta_1 , \beta_2 , \beta_3$	Coefficients of the fixed part of the cost model (excluding geographical-related and hospital

	group related coefficients) (MLM).
β_{4k}	Fixed coefficients for dummies of the geographical area k (geographical-related coefficients) (MLM).
β_{0c}	Random coefficient of the random intercept of the MLM, defined at the hospital administrative group c.
β_{1c}, β_{2c}	Random coefficients of the random slopes of the MLM, defined at the hospital administrative group c; β_{1c} and β_{2c} are the random coefficients of the nurses to doctors and beds to doctors ratios, respectively.
μ_{0c}	Random component of the random coefficient of the MLM, defined at the hospital administrative group c.
μ_{1c}, μ_{2c}	Random component of the random slopes of the MLM, defined at the hospital administrative group c.
e_{hck}^{MLM}	Random error at the hospital level (MLM).
$\sigma_{\mu 0}^2$, $\sigma_{\mu 1}^2$, $\sigma_{\mu 2}^2$	Variances of the random components of the model at the group level. $\sigma_{\mu 0}^2$ is the variance of the random component of the intercept, while $\sigma_{\mu 1}^2$ and $\sigma_{\mu 1}^2$ is the variance of the random component of the slopes (MLM).
σ_{e0}^2	Variances of the error term at the hospital level (MLM).
$\sigma_{\mu 0 \mu 1}$, $\sigma_{\mu 0 \mu 2}$, $\sigma_{\mu 1 \mu 2}$	Set of covariance between the random components, defined at the group level (MLM).

Table A.4: Chapter 7

Notation	*Interpretation*
i, i', v and q	Population points representing small area population units. Each i, i', v and q belongs to one district r $(i,i',r,q \in r)$ $(i \neq i' \neq r \neq q)$.
n	n is the number of population points
j, w and z	Hospital points representing hospital site geographic units. Each j, w and z belongs to one district r $(j,w,z \in r)$ $(j \neq w \neq z)$.
m	m is the number of hospital points, which is a sub-set of the total number of population points n $(m \subset n)$.
U_{ij}	Utilisation flow between population point i and hospital site j.
D_j	Size of hospital site j.
$\widetilde{D}_{ij}$	Index for alternative supply to hospital site j available for population i.
$other_{ij}^{'}$	A set of other variables related with population and hospital characteristics that explains flows.
$other_i^{''}$	A set of population-related variables that explains flows.
$other_j^{'''}$	A set of hospital-related variables that explains flows.
P_i	Resident population in i.
Dem_i	Demographic characteristics of the population (age and sex) that imply higher need for hospital care for population i.
N_i	Need for hospital care for population i.
X_i	Socio-economic level of population i.
G_{ij}	Accessibility costs for population i to access hospital services in j.
d_{ij}, $d_{ii'}$	Distance between population point i and hospital site j, and between population points i and i' (Euclidean distances as defined in Chapter 3).
A_i	Perceived availability of hospital care to population i.
I_{ij}	Set of institutional characteristics of the hospital system (such as hospitals hierarchy, sites with hospital teaching functions, spatial hospital subsystems, etc), to be specified below. Some of these characteristics relate to population points.
O_{ij}	Set of variables that characterise access to other sectors of health care and non-health care systems (such as welfare system and private supply) and other variables that are expected to influence demand for hospital care –such as spatial variables along the territory.
PC_i	Accessibility to primary care for population located in i.
c_j	Role of hospital j in the hospital hierarchy (for example, dummy variables for central and district hospitals).
$i1_{ij}$	Indicator of whether hospital j is the first hospital of use by population i (dummy variable).
$i2_{ij}$	Indicator of whether hospital j is the second hospital used by population i (dummy variable).

$i3_{ij}$	Indicator of whether hospital j is the central hospital used by population i (dummy variable).
$i4_j$	Vector of hospital variables that characterise hospital j outputs other than inpatient care (such as external consultations and emergencies).
$i5_j$	Vector of variables representing the hospital input mix of hospital j (labour vs. equipment vs. beds).
y	Utilisation variable as a dependent variable.
x	Set of the covariates that are hypothesised as affecting utilisation.
x' and x''	Two sub-sets of covariates of the set x ($x' \subset x$ and $x'' \subset x$).
d'_{ij} , $d''_{ii'}$	Dummy on whether hospital j is within 25 km from population point i, and dummy on whether population point i' is within 25 km from population point i.
β''	Set of coefficients of the econometric model.
e^{ε}	Residuals in the natural scale of the second part of the two-part model.
$\hat{p}_{qw}$	Predicted probability of population point q making use of hospital site w.
$\hat{U}_{qw}$	Predicted level of utilisation flows of population point q to hospital site w, given that the probability of that flow being positive is positive.

Table A.5: Chapter 8

Symbols	*Intuition and explanation of the choices made*
$UCOutput_h$	UC index for hospital h.
I_{1r}	Age adjustment index for district r.
$Catchment_r$	Catchment population of district r.
D_r, D	Discharges from hospitals of district r; total discharges in the system.
O_r	Discharges from the resident population of district r.
W_r, W	Population need for hospital care in district r (resident population weighted by age); total population need.
$W_r^{'}$	Population need for hospital care in district r, scaled so that total need sums up total discharges in the system.
r, I_{2r}, I_{3r}, I_{4r}, cap_index_r, P_r	Defined in Chapter 4.
X_{1a}, P_{ar}, $ASMR_{ar}$	Defined in Chapter 5.
h, do_h	Defined in Chapter 6.
U_{ij}	Defined in Chapter 7.

Table A.6: Chapter 9

Notation	*Interpretation*
$Flow\_1_{ij}$	Dummy **variable** for expressing whether population i is served by hospital j, as a first hospital (DBM) (0 or 1 values).
$Flow\_2_{ij}$	Dummy **variable** for showing whether population i is served by hospital j as a second hospital (DBM) (0 or 1 values).
$Flow\_c_{ij}$	Dummy **variable** for denoting whether population i is served by hospital j as the closest central hospital (DBM) (0 or 1 values).
d_1_i	Distance travelled between population point i and the first hospital of use (non-negative **variable** depending on $Flow\_1_{ij}$) (DBM).
d_2_i	Distance travelled between population point i and the second hospital of use (non-negative **variable** depending on $Flow\_2_{ij}$) (DBM).
d_c_i	Distance travelled between population point i and the closest central hospital of use (non-negative **variable** depending on $Flow\_c_{ij}$) (DBM).
W_i	Needs-weighted population at population point i (DBM). This is derived from weighting resident population per age group by the age weighting index estimated in Chapter 5.
$share\_1_i$	Share (%) of population i that is assumed to go to the first hospital (DBM).
$share\_2_i$	Share (%) of population i that is assumed to go to the second hospital (DBM).
$share\_3_i$	Share (%) of population i that is assumed to go to the closest central hospital (DBM).
U_1_i	Utilisation flow by population i to the closest hospital (DBM).
U_2_i	Utilisation flow by population i to the second closest hospital (DBM).
U_3_i	Utilisation flow by population to the closest central hospital (DBM).
$\overline{U}^N$	National utilisation rate ($\overline{U}^N = U^o / W$) (DBM/UBM).
$\widetilde{p}_{ij}$	Probability of population i using hospital j, as produced by the gravity model, with $\sum_j \widetilde{p}_{ij} = 1, \forall i$ (UBM).
U_i^N	Normative utilisation for population area i depending on total national utilisation rate (non-negative **variable**) (UBM).
D_j^0, D^0, D	Current level of supply of hospital j; total current level of supply; total level of supply, computed within the model (UBM)
U_{ij}^o, U^0	(Past) flows and (past) total level of utilisation (DBM/UBM)
a_i	Auxiliary **variable** used to obtain an absolute value of difference between utilisation and expected utilisation, per population area i (UBM).
$f(\beta_j^{'''}, d_{ij})$	Decay function that relates the effect of distance (accessibility costs) from population i to hospital j (definition in Appendix E). The decay function might differ for hospital type

	and the decay parameter β_j will depend on the level of attraction between hospital j and patients located at different distances from that hospital (UBM).
β_j'''	Parameter that defines the elasticity of utilisation in relation to distance, for hospital j (UBM).
$f\_\min$	Proportion of current level of supply of hospital j to be kept, as a minimum (UBM/UFBM)
$f\_\max$	Proportion of current level of supply of hospital j to be increased, as a maximum (UBM/UFBM)
$\min\_D_j$	Minimum level of supply of hospital j to be maintained (UBM/UFBM).
$\max\_D_j$	Maximum level of supply to be allowed for hospital j (UBM/UFBM).
$\log U_{ij}^r$	Distribution of the natural logarithm of utilisation flows that operates as the target. This target is a distribution formulated in accordance to some type of equity principle (in this case, patients making use of the closest hospital) (UFBM).
b_{ij}	Auxiliary **variable** for defining the difference between variations in the logarithm of utilisation flows (UFBM).
$\log \hat{p}_{ij}$	Logarithm of the probability of use, generated in the first part of the estimated two-part FDM, developed in Chapter 7 (UFBM).
$\log U_{ij}'$	Natural logarithm of the utilisation **variable** between hospital i and hospital j, as defined in the second part of the two-part FDM, developed in Chapter 7 (UFBM).
$DumFirst_{ij}$, $DumSecond_{ij}$ and $DumCentral_{ij}$	Dummy for whether hospital j is the closest hospital to a population i; dummy for whether hospital j is the second closest hospital to a population i; and dummy for whether j is the closest central hospital to a population i (UFBM).
$DumLisboa_{ij}$, $DumPorto_{ij}$ and $DumCoimbra_{ij}$	Dummy for the central hospital site in Lisboa and for populations from the South; dummy for the central hospital site in Coimbra and for populations from the Centre; and Dummy for the central hospital site in Porto and for populations from the North (UFBM).
$oth\hat{e}rs_{ij}$	Parameter capturing the influence on flows of all the factors from the FDM, with the exception of the variables that relate to hospital supply (UFBM).
$\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\alpha}_5$, $\hat{\alpha}_6$, $\hat{\alpha}_7$	Parameters that relate utilisation flows and hospital supply, taken from the estimated flows demand model (estimated in Chapter 7) (UFBM).
$\log U_{ij}$, D_j, $\widetilde{D}_j$, U_i, d_{ij}, m, n	Notation presented in Chapter 7.
W	Notation presented in Chapter 8.

APPENDIX B - Review of Portuguese studies related with hospital costs and performance

Table B.1: a review of Portuguese studies related with hospital costs and performance

Study and objectives	Techniques and data applied	Findings	Strengths	Weaknesses
Paiva (Paiva 1993) Objective: Measure inefficiency in Portuguese hospitals	Techniques: Model mixed two approaches: behavioural and dual (assuming a Cobb-Douglas technology) Model used alternative aggregate output case-mix indices Output decomposed in three effects: beds, occupancy rates and average delay Included vector prices, input prices, capital proxy and labour cost proxies Use of individual and fixed effects, and of random effects Data: Panel data: 46 hospitals; 5 years	High variations in inefficiency between hospitals (explained by size and hospital type); high variations in inefficiency for central large hospitals and for district hospitals. (After introducing adjustments for inefficiencies) cost curve function followed a U-shape format, reached a minimum at the level of 400 beds. Separated estimates of technical and allocative efficiency Technical efficiency implied that costs exceed minimal costs 47%. Implications from allocative efficiency estimates: excessive physicians expenditure, in relation to capital expenditure; higher variability of technical in relation to allocation inefficiency; reallocation of funds between personnel categories would generate improvements in efficiency, as well as reallocation between personnel and capital expenditures.	It does not assume cost minimisation behaviour, and searches for other alternative paradigms. Correction of the model by inefficient levels, before analysis, which introduces realism.	Neglects structural inefficiency. Inefficiency is explained by dimension and centre type: this might be explained by the lack of control of confounding variables; or by an inadequate capture of economies/diseconomies of scale.
Paiva (Paiva 1993) Objective: Analyse of allocative efficiency	Techniques: Translog model, with estimation of input share equations for doctors, nurses, other personnel and capital Computation of substitution elasticities, using the SURE technique	Doctors are complementary with nurses and with other personnel. Excessive costs in personnel, in relation to capital expenditure.	Model does not make assumptions on the structure of hospital costs.	Lack of use of confounding variables. Traditional problems of multicollinearity.

	(Seemingly UnRelated Estimation) Data: Sample: 46 hospitals			
Lima (Lima 1998) Objective: Impact of prospective payment per case on hospital performance	Techniques: Three utility maximising models: one-way random effects, two-way fixed effects and one-way fixed effects Includes control for labour and capital price Data: Sample: 36 district hospitals and 10 years data (1984-1994)	Decline on unit costs per admission and per day, as well as average LOS, along the period; but the effect on the number of admissions is indeterminate.	Behavioural model using a full sample of hospitals.	It is not clear that the period after 1990 was characterised by prospective finance, given high levels of deficits and hospital managers were not accountable for deficits. Inadequate control for case-mix.
Lima (Lima 1998) Objective: analyse the cost structure of Portuguese hospitals	Techniques: Translog and multi-product cost function Testing the model for two types of hospitals: district and central Includes control for labour and capital price Data: Sample: 44 hospitals; 10 years data (1984-1994)	District hospitals with an average size of 241 beds are operating under economies of scale and of joint production of services; central, large and teaching hospitals, with an average size of 869 beds are operating under overall diseconomies of scale and of joint production of services. Elasticity of demand for labour and capital are price elastic. District hospitals are treating less costly cases or being more efficient.	Translog flexible model, making use of panel data.	Problems in first order output measures and regularity conditions. Indirect control for case-mix.
Carreira (Carreira 1999) Objective: Analysis of hospitals cost structure and detection of potential savings from economies of scale and scope	Techniques: Translog flexible cost function with multi-product and multiple inputs Control for labour price and other prices (GDP deflactor) Data: Panel data: 82 hospitals; 5 years data (1991-95 data)	The average hospital is operating with economies of scale in the short-run and with diseconomies of scale in the long-run; a decrease in inpatients and an increase on emergencies and external consultations would imply cost savings. Joint-production has substantial cost savings. Optimal dimension of 215 beds, and strong joint production economies.	Decomposition of the residuals of the model between fixed and variable components, and correction of the model by these coefficients. Testing for different technology by splitting the hospital sample in sub-	High standard variation of hospital variables. Incomplete control for confounding variables, such as for geographic variations and quality.

			samples in accordance with dimension and administrative status. Attention given to short vs. long-run economies.	
Barros and Sena (Barros and Sena 1999) Objective: Impact of opening and re-dimensioning three hospital units on current expenditure and on productivity levels	Function based on production function theory; use of a simplified regression: expenditure as a function of a production indicator, over hospital and over time, and depending on hospital age Use of a composite hospital output indicator –the discharged adjusted patient: hospital outputs weighted by comparative cost ratios Test of a quadratic function on the production indicator, and a dummy differentiating between old and new hospitals Test for quality impact on hospital costs Data: Hospital sample: Abrantes, Almada and Leiria	Diseconomies of scale for both old and new hospitals Expenditure increase is not compensated by productivity gains. Quality not significant in explaining costs.	Single output measure aggregating several hospital outputs. Attempt to test for quality.	Neglects prices of inputs; reduced number of observations in the sample, which implies low degrees of freedom, and constrains the estimation technique; and univariate analysis, with no control for confounding variables, nor for inefficiency.
IGIF (IGIF 1999) Objective: Reformulating the NHS hospital grouping under the principle that similar hospitals, under a set of characteristics, must present similar cost structures and similar efficiency levels	Techniques: Cluster analysis, using multivariate techniques and peer review Statistical model with fixed components (independent from production, services quality and efficiency) Restriction to four main characteristic types to measure hospitals: dimension, diversity, diversity of services provided and resources (variables chosen: number of beds, number of special beds, case-mix, number of doctors and nurses, number of distinct DRGs, percentage of distinct surgical DRGs, types of equipment) Exclusion of variables due to high correlations, dependence from productivity, independent from hospital structure, lack of consensus in the group or representing out-of-date classifications Labour price: personnel costs over inpatient costs Data: National sample of public hospitals under management of the MoH, 1996 data	Grouping of all the acute care hospitals in five groups.	Analysis of structures and production, neglecting costs Reliance on microeconomics literature in the choice of variables (three groups: structure, inputs and outputs).	Special treatment applied to outliers, dependent on peer group decisions. This approach does not give information about how different hospital structures impact on costs. Exclusion of the variable supply of teaching activities – the authors considered insensitive to infer its influence in the institutions cost structure.

Lima (Lima 2000) Objective: Test the productivity of Portuguese public district level hospitals	Techniques: Transcendental logarithmic flexible functional with multi-product and multiple inputs Capital price: capital expenditure over the number of beds Labour price: labour expenditure over number of hospital professionals Data: Panel data: 36 district hospitals; 1984-94 data	Global economies of scale for the average hospital (241 beds). Joint production economies for the specialties: gynaecology/obstetrics, external consultations and emergencies.	Translog flexible model.	Cost minimising behaviour is debatable in the context of the Portuguese health system. Lack of control for confounding variables (such as for quality. High variability of hospital variables.

Bibliography

Barros, P. P., and C. Sena. 1999. Quanto maior melhor? Redimensionamento e economias de escala em três hospitais portugueses. *Revista Portuguesa de Saúde Pública* 17 (1):5-18.

Carreira, C. M. G. 1999. *Economias de escala e de gama nos hospitais públicos portugueses: uma aplicação da função de custo variável translog.* Edited by Associação Portuguesa de Economia da Sáude. Vol. 3/99, *Documentos de trabalho.* Lisboa: Associação Portuguesa de Economia da Sáude.

IGIF. 1999. *Agrupamento dos Hospitais do Serviço Nacional de Saúde.* Lisboa: Ministério da Saúde.

Lima, E. 1998. The financing of health care: an analysis of the impact of the Portuguese hospital financing system. PhD, Economics, University of Nottingham, Nottingham.

Lima, M. E. M. 2000. A produção e a estrutura de custos dos hospitais públicos. Lisboa: Associação Portuguesa de Economia da Saúde.

Paiva, R. L. 1993. A medição da eficiência no sector hospitalar -O caso português. PhD thesis, Instituto Superior de Economia e Gestão - Universidade Técnica de Lisboa, Lisboa.

APPENDIX C - Complementary information to the multilevel model

Table C.1: Geographic classification in use

Regional classification	*Districts*
North coast	Porto, Braga, Viana do Castelo
North interior	Vila Real, Bragança
Centre coast	Aveiro, Coimbra, Leiria
Centre interior	Viseu, Guarda, Castelo Branco
South coast	Lisboa, Setúbal
South interior	Santarém
Alentejo	Beja, Évora, Portalegre
Algarve	Faro

Table C.2: Average and standard deviation (in brackets) of selected variables

	Average	*General*	*Specialised*	*District*	*Level*
Doctors	183.0 (252.2)	538.8 (391.7)	160.1 (111.8)	136.7 (106.8)	29.6 (27.8)
Standardised Cost	410.4 (231.3)	736.1 (295.5)	524.6 (223.0)	323.3 (89.5)	291.8 (92.2)
Case-mix	1.04 (0.3)	1.44 (0.4)	1.15 (0.6)	0.87 (0.1)	1.00 (0.2)
LOS	8.0 (2.5)	10.3 (3.9)	8.0 (2.7)	6.9 (1.1)	8.2 (1.9)
Occupancy rates	71.6 (10.7)	76.2 (8.2)	66.0 (12.1)	72.1 (8.9)	69.8 (13.2)
Consumption costs/total costs	17.7 (8.6)	28.5 (6.6)	16.4 (11.3)	16.8 (6.2)	12.3 (5.4)
Outsourcing/output	73.5 (31.4)	111.1 (47.6)	72.5 (23.3)	61.2 (17.8)	68.7 (16.7)
Personnel costs/total costs	57.3 (8.6)	48.4 (8.2)	61.7 (12.0)	59.2 (6.7)	58.5 (6.6)
Purchaser Power Index	98 (36.9)	136 (36.8)	138 (33.6)	83 (28.0)	83 (18.2)
Nurses/doctors	2.5 (1.5)	1.4 (0.5)	1.7 (0.5)	2.4 (0.9)	3.7 (2.2)
Beds/doctors	2.8 (2.2)	1.7 (1.8)	1.6 (0.7)	2.4 (1.1)	4.6 (3.1)
Employees/doctors	4.3 (3.3)	2.4 (0.9)	3.1 (0.9)	3.4 (1.3)	7.4 (4.8)
Non-NHS revenue/total revenue	13.2 (4.2)	12.7 (3.8)	14.9 (6.3)	13.5 (4.3)	12.5 (3.0)
Growth expenditure last two years	31.8 (25.0)	21.2 (15.4)	26.1 (13.8)	36.6 (32.8)	33.3 (15.2)
Total extra-hour payments per doctor	8.85 (0.2)	8.82 (0.1)	8.94 (0.1)	8.85 (0.2)	8.83 (0.2)

Figure C.1: Residuals at the hospital level

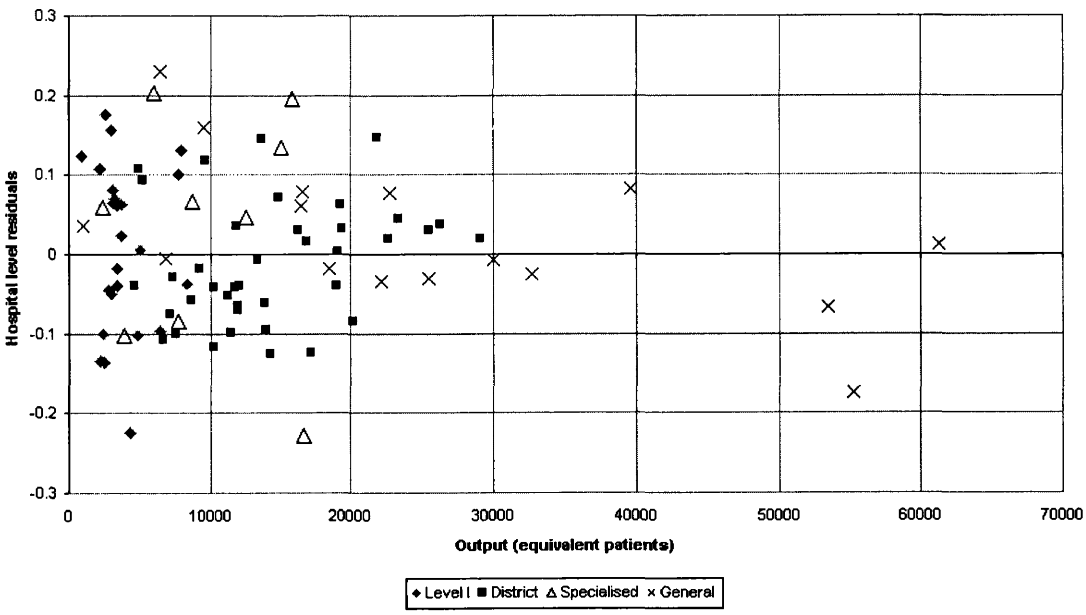
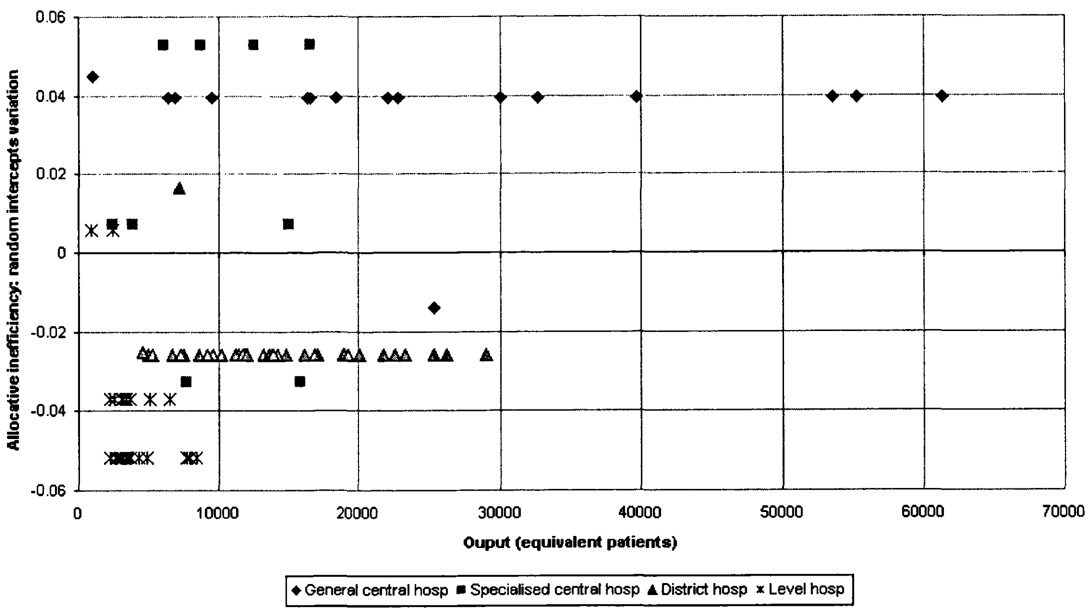


Figure C.2: Allocative inefficiency –random intercepts



APPENDIX D - Descriptive statistics of the application of the flow demand model

Table D.1: descriptive statistics –first part of the TPM

	Average	*Standard deviation*	*Minimum*	*1st quartile*	*2nd quartile*	*3rd quartile*	*Maximum*
Utilisation flows	48.19	732.55	0.00	0.00	0.00	0.00	70,674.00
Population	34,451.16	56,469.20	1,800.00	8,410.00	15,930.00	36,340.00	535,740.00
Needs index	107.52	10.97	85.18	99.85	105.92	114.10	145.17
Discharges	14,070.71	26,142.22	725.00	2,820.25	8,151.50	14,546.75	185,329.00
PC utilisation	95,561.42	173,789.00	9,970.00	25,411.00	44,235.00	92,300.00	2,190,998.00
Private hospital access	14.42	106.50	0.00	0.00	0.00	0.00	1,335.00
PPI	66.58	28.90	33.72	49.68	58.30	74.41	305.19
Illiteracy rates	0.15	0.07	0.03	0.10	0.14	0.19	0.48
Emergency per discharge	12.12	9.33	0.00	5.93	8.58	15.76	51.76
Outpatient per discharge	5.52	1.72	3.09	4.30	5.01	6.40	10.80

Table D.2: descriptive statistics –second part of the TPM

	Average	*Standard deviation*	*Minimum*	*1st quartile*	*2nd quartile*	*3rd quartile*	*Maximum*
Utilisation flows	405.45	2,093.14	0.00	5.00	17.00	136.75	70,674.00
Population	63,532.12	95,491.46	1,800.00	11,677.50	24,500.00	68,560.00	535,740.00
Needs index	104.31	10.69	85.18	96.52	104.03	111.08	145.17
Discharges	43,147.65	56,401.50	725.00	9,316.00	17,546.00	79,174.00	185,329.00
PC utilisation	179,537.42	327,133.75	9,970.00	35,136.00	70,751.00	181,467.00	2,190,998.00
Private hospital access	47.36	218.26	0.00	0.00	0.00	0.00	1,335.00
PPI	78.62	44.99	33.72	52.11	66.94	94.48	305.19
Illiteracy rates	0.13	0.07	0.03	0.08	0.12	0.17	0.48
Emergency per discharge	7.62	5.91	0.00	4.25	5.85	8.58	51.76
Outpatient per discharge	6.31	1.96	3.09	4.55	6.00	7.49	10.80

APPENDIX E - Gravity model formulation and score for perceptions of availability

An attraction-constrained gravity model was chosen (described in Batty (Batty 1976) - pages 39-44; Equations E.1, E.2 and E.3), as hospital production is constrained by hospital capacity (Equation E.2). The main equation is Equation E.1. Notation is presented in Table E.1. The database used to compute the gravity model used the 1999 DRG data. An exponential function was used to estimate the impact of distance on utilisation for each type of hospital.

$$U_{ij} = B_j D_j W_i f(\beta_j^{'''}, d_{ij}) \quad \text{(E.1)}$$

$$\sum_i U_{ij} = D_j \quad \text{(E.2)}$$

$$B_j = \left[\sum_i W_i f(\beta_j^{'''}, d_{ij}) \right]^{-1} \quad \text{(E.3)}$$

Table E.1: Additional notation in use

Symbols	*Intuition and explanation of the choices made*
$\beta_j^{'''}$	Deterrence parameter that should represent the closest match between estimated and observed flows from the gravity model.
$f(\beta_j^{'''}, d_{ij})$	Decay or deterrence function by type of hospital (acts as a spatial discount factor).
g	Constant.
B_j	Balancing factor for hospital site j.
W_i	Population needs in population point i (resident population weighted by need)
U_{ij}, D_j, d_{ij}	Notation from Chapter 7

The score for perceptions on availability (used in Chapter 7) was computed using the following formula:

$$A_i = g\left(\sum_j U_{ij}\right)/W_i = g\sum_j B_j D_j f(\beta_j^{'''}, d_{ij}) = g\sum_j \left(\frac{D_j f(\beta_j^{'''}, d_{ij})}{\sum_i W_i f(\beta_j^{'''}, d_{ij})}\right) \qquad \text{(E.4)}$$

BIBLIOGRAPHY

Batty, M. 1976. *Urban Modelling: Algorithms, Calibrations, Predictions*. Edited by L. Martin and L. March, *Cambridge urban and arquitectural studies*. Cambridge: Cambridge University Press.