Essays in Health Economics:

Volume-Outcome Effect, Upcoding Behaviour and the Evaluation of a Short-Term Intervention Program

By

Gisele Braun

Dissertation

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Economics
in the NOVA School of Business & Economics
at Universidade Nova de Lisboa, 2014

Lisbon, Portugal

Financial support from

Fundação para a Ciência e a Tecnologia - FCT

PTDC/ECO/71867/2006

and
SFRH/BD/71799/2010

To my dad, Cláudio.

"Nothing in life is to be feared,
it is only to be understood.
Now is the time to understand more,
so that we may fear less."
–Marie Curie

Acknowledgments

This thesis is the output of my last five years of research and its completion would not have been possible without the support and assistance of many people and institutions.

If there is anybody in this life that masters the art of teaching by example, it is my advisor Pedro Pita Barros. His research excellence combined with his generosity, provided me with the ideal environment to the development of my research skills and for the writing of this work. He also shared his enthusiasm to embrace the most different roles in the academic life and his patience to beginners (and not only to them). Pedro believed in my research potential when not even I did and I owe him my deepest gratitude. I am forever indebted for his criticisms, advices and help.

The vibrant academic environment of the Faculty of Economics at *Universidade Nova de Lisboa* should also be mentioned. The institution has principles, rules and characteristics that allowed me to grow as a student: a very qualified group of professors committed to the objective of producing high quality work, and an equal opportunity space for students of all race and gender. The faculty is not an impersonal place; it was built by the people that work there. I would like to thank the PhD program directors and the Research Office's manager: João Amaro de Matos, Ana Balcão Reis, Susana Peralta, Pedro Vicente and Silvana Figueiredo.

I cannot find words to express my appreciation to the country of Portugal, to the Portuguese people, its foreign policy and social responsibility to the world. One expression of its generosity is the financial support offered by the Government through *Fundação para a Ciência e a Teconologia*. As a foreign student, I could never have pursued this degree without such help. This is a reflection of the portuguese social values, which I was lucky to benefit from and for which I will always be indebted.

Special thanks are due to Luis Catela Nunes, Alberto Holly and Céu Mateus for the invaluable advise on econometrics and health economics. I stand on the shoulder of giants having you so close to be inspired and taught what really maters on the field. My gratitude to Céu Mateus for her kind attention with me in con-

ferences and meetings, your words made me feel at home. To the recently created Nova Health Care Initiative Group, I thank its members for several suggestions on my work.

Nevertheless, my path in the academia started few years before coming to this University. It goes back to 2004 at *Universidade Federal do Rio Grande do Sul* (UFRGS). A very special acknowledgment is due to professor Giacomo Balbinotto Neto. He was the person that encouraged me to keep studying after college and I am forever grateful for his lessons and friendship. In times that I wanted to give up, he was always there to cheer me up and support me.

Since I left Brazil, many people have come and left. However, the friends have stayed, no matter what or where. A special thanks to the old buddies from Porto Alegre, UFRGS and Banrisul. They know who they are. In the distance, you make me feel special.

Sincere thanks to all my family. Each member had a special role in shaping the person that I became and the opportunities that I have had in life. My grandparents Ataliba, Ana, Alfredo and Celita gave me so much love that living this life became an easier task. I could not think of another dad to teach me how to survive: hard work, forgiving heart and generosity. My mom is the culprit for everything and I thank you for all the courage and strength to be alive. My sister Cibele, that is my reference of equilibrium, my safe place to come back, gives me everyday reasons to keep working for my dreams. As the years passed, I had the luck to witness the enlargement of my family, people that came to add and that helped me in several ways: my step-mother, Diva, with her unconditional faith in me; my stepsisters, Karen and Hellen, with a much funnier view of the world than mine; my brother-in-law, Richard Kalil, a brother overall, and his family; and Arturo's family, Edgar, Anabella and their children, for the pure love offered to me. To all of them, thank you for the hard work of bearing with me in the best and worst.

Tom Jobim said once that *"fundamental é mesmo o amor"*. And my last word goes to Arturo. Our history started during the writing of this thesis and he was of extreme importance for me to conclude it. All the mathematic lessons, the hard comments for the improvement of my work and the patience with my weaknesses are of invaluable price. But the essencial goes beyond these helps: your love and loyalty makes me a better version of myself and I am eternally grateful to share my life with you.

Gisele Braun
May, 2014

Contents

List of Figures

List of Tables

Introduction

The Portuguese health system has made progress in providing better health services to the whole population, specially in the last two decades. Real improvements range from rearrangements of the health care network to specific programs that aim at solving a given downside in the system. Factors aside the health care provision, such as legislation on health and the ways of payment to health providers, which might influence health care, are as much important to determine the quantity and quality of the health system as the provided care. For that reason, they also changed in one or another direction to give right incentives for provision of health care.

This dissertation consolidates the results and conclusions from three empirical studies in the field of health economics. It is composed by three essays that evaluate major recent reforms implemented in Portugal to raise quality on its health care system. Together, they also offer an wide and updated analysis of the services provided though the Portuguese national health service (NHS).

Since the beginning of the new century, a relative big number of health reform initiatives were implemented, covering five areas of intervention: health promotion, long-term care, primary and ambulatory care, hospital management and inpatient care, and the pharmaceutical market.[1] The essays carried on in this work assess changes on health promotion and hospital management and inpatient care.

The need to analyse the impact of hospital merges on quantity and quality of health services emerged from the unification of some public hospital management teams and the change to a more entrepreneurial-like status from 2003 onwards,

[1] Barros P, Machado S, Simões J. 2011. Portugal health system review. *European Observatory on Health Systems and Policies* **13**(4).

plus the implementation of a purchaser-provider relation between Government and hospitals, through contracting. At the same time, some primary care emergency services were closed to better organize the network. In this context, the first essay, named "The Volume-Outcome Effect in a National Health Service: the Portuguese Case", addresses whether a new network distribution of medical procedures, namely the merge of medical activity of two or more hospital units, can lead to any social or economic benefit for the population. I found evidence that a hypothetical alternative arrangement network of the Portuguese public hospitals, based on economies of scale, for some specific diseases, can lead to health benefits for the population, through the argument of "practice-makes-perfect", which claims that the bigger the number of times that a team performs a given medical procedure, the bigger the rate of success.

The cost of services provided by the public health systems is another variable under constant scrutiny as for philosophical reasons or for the scarcity of resources to match demand and supply. In Portugal, as in many European countries, the payment of a medical procedure performed by a hospital is made by the Government (purchaser) based in the "diagnosis-related group" (DRG) classification. The price of each DRG, which corresponds to the provided health care, is established by the Ministry of Health, with infrequent adjustments, usually to update costs. Auditing is the most frequent measure used to control the principal-agent problem. However, this is costly and not always efficient. In such context, the second essay, "Upcoding in a NHS", is an analysis of whether the price setting is or is not inciting upcoding behaviour.[2] The results obtained show that upcoding occurred in the Portuguese NHS, for the investigated DRGs, and a new price setting could discourage such behaviour. The cost of this behaviour was calculated and is relatively small.

The last essay, "Does it last? Effects from a public policy to recover waiting lists", is an evaluation of a direct intervention policy, the ophthalmology intervention programme (PIO), that was active in the years of 2008 and 2009. This was a health promotion initiative reform. The Ministry of Health defined the programme following recommendations of the Portuguese National Health Plan 2004-2010. The waiting time and waiting list for eye-related first consultation

[2]Upcoding is defined as classifying patients in DRG codes associated with larger payments.

and subsequent medical procedures were far from acceptable and a patient could wait several months to receive medical care. The main eye-related diagnosis is the cataract, a medical condition that can reduce drastically the quality of life and lead to blindness if not adequately treated. In spite of improving the universal access to this medical treatment, the intervention programme was completed and post evaluations conducted were mainly descriptive, not considering the randomness and the dynamic problems of waiting lists. Therefore, I add to the literature an economic investigation that consider such problems. I investigated whether the intervention programme was successful in attaining the goals established and the conclusion was that size waiting lists and waiting times reduced in a statistically significant way that allowed me to conclude for the success of the programme. This evidence may help future investment decisions done by policy makers, regarding the allocation of public resources in the health care market.

Aggregating these achievements, it is noteworthy that the Portuguese public health system has been improving since the beginning of the century. Moreover, the investigated recent reforms adopted corroborate with one of the broad goals defined in the National Health Plan, namely the promotion of an effective and efficient health care service.

The decision of these three research questions were not random. On the contrary, I tried to establish a very specific pattern, going from a very broad and concrete subject (the network arrangement of the health care system), passing through the regulatory system (the impact of DRG's price on behaviour of physicians), to a policy intervention (the ophthalmology programme to reduce waiting lists and waiting times). Another highlight to make refers to the data. I used medical records of the Portuguese public hospitals in all three essays, which is another fact to reinforce the connection among the studies. Depending on the subject, I selected medical records from 2001 to 2010. The dataset was provided by *Administração Central do Sistema de Saúde* (ACSS) that consolidates the information of all Portuguese public hospitals. It is primarily designed for administrative purposes, but it also has a relevant number of extra-information regarding characteristics of the patients and medical conditions that allows economic analyses. The advantages of designing randomized controlled trials or applying specific questionnaires are counterbalanced by the size of the sample provided by the ACSS, which accounts

for almost the whole population of Portuguese public medical records. Beside that, the time required and high estimated cost to build more accurate samples leaded me to use the already available dataset.

Having information of individuals allowed me to conduct analyses at the specific level of patient or at a more aggregate level of hospital, for example, when it is more suitable. Although there is no possibility of tracking patients over the years, due to confidentiality reasons, I was able to apply the techniques of panel data when aggregating information across the hospitals. Comparing such technique with the alternative of cross-section analysis, I was able to measure more precisely the impact of one variable over another, isolating no-related influences in the fixed-effects term, an advantage of panel data techniques. Other econometric tools used include descriptive statistics, instrumental variables and simultaneous equation models. In any case, the econometric results obtained were used to make estimations in real terms such as number of deaths avoided, cost in euros and waiting time in days. Moreover, I always brought the results back to the context in which the samples were extracted to explore the potential benefits of the conclusions achieved.

The main contribution of the thesis was to provide a broad innovative and updated economic analysis of the Portuguese public health system and also to offer a measure of the impact of reform initiatives adopted in the last decade. The establishment of these results render evidence for new investment decisions in the public health system. Although there are much work done for the United States health market as well as for the largest countries of the Commonwealth, the results does not carry on automatically due to different characteristics of the samples and the functionality of the health systems. The results offered in this dissertation may be extended for countries other than Portugal, such as other southern European countries, that also have NHSs with similar obstacles to overcome.

In terms of local policy implications, the results suggest that hospital merges should be kept going as specific programs like the PIO are effective to attack problems such as waiting lists for eye-related surgeries. The combination of higher investment associated to behavior incentives lead to solutions in the short-term. In terms of price scheme, the spread of DRG prices, if smaller, could reduce upcoding. However, the estimated upcoding is not high compared to the global expenditure,

meaning that the ongoing auditing is enough to avoid large damage in the health system.

This dissertation is organized into 3 additional chapters. Chapter 2 is the first essay, "The Volume-Outcome Effect in a National Health Service: the Portuguese Case". Chapter 3 presents the second essay, "The Upcdogin in a NHS" and chapter 4 has the third essay of the dissertation, named "Does it last? Effects from a public policy to recover waiting lists".

Chapter 1

Volume-Outcome Effect in a National Health Service: the Portuguese Case

Abstract

Evidence in the literature suggests a negative relationship between volume of medical procedures and mortality rates in the health care sector. In general, high-volume hospitals appear to achieve lower mortality rates, although considerable variation exists. However, most studies focus on US hospitals, which face different incentives than hospitals in a National Health Service (NHS). In order to add to the literature, this study aims to understand what happens in a NHS. Results reveal a statistically significant correlation between volume of procedures and better outcomes for the following medical procedures: cerebral infarction, respiratory infections, circulatory disorders with AMI, bowel procedures, cirrhosis, and hip and femur procedures. The effect is explained with the practice-makes-perfect hypothesis through static effects of scale with little evidence of learning-by-doing. The centralization of those medical procedures is recommended given that this policy would save a considerable number of lives (reduction of 12% in deaths for cerebral infarction).

1.1 Introduction

Over the last decades the Portuguese hospital adult mortality rate has fallen. On the other hand, the volume of activity inside the hospitals has increased each year. In particular, we observe an increasing trend in the number of hospital admissions. Comparing these two facts, can we find any relationship? Is it the case that a reduction in the mortality rate is related to an increase in the volume of activity inside a hospital? This is an important question in Portugal, especially due to new arrangements that have been adopted in the health sector (e.g., creation of specialized units of care).[1]

The analysis of the correlation between number of procedures and outcomes is common in several markets. An application of this sort can be found in the health care context as well. In general, the results indicate that better health outcomes tend to come with a larger number of procedures. However, the studies generally concentrate efforts on the US health market or other private markets, leaving hospitals belonging to the National Health Service (NHS) outside the discussion. This is quite unexpected, because the European health services are basically publicly provided and a NHS exists in many countries (UK, Spain, Italy, etc).

The correlation is called volume-outcome effect and the main hypotheses to explain the mechanism are "practice-makes-perfect" (PMP) and "selective referral" (SR) (Luft et al., 1987). The first hypothesis, which can be explained by scale economies or learning-by-doing, says that a greater volume of procedures leads to improved outcomes (given that the severity level of the cases remains relatively constant); whereas SR supports the notion that high quality centers attract more patients, which leads to a higher volume of procedures.

Considering potential benefits derived from the volume-outcome effect, we try, in a context of NHS, to answer the question "Does centralization of some medical procedures increase the quality of health indicators?" The novelty of this paper relates to the nature of the system analyzed. Private and public markets share

[1] In fact, the concentration of hospital services is one of the recommendations contained in the Portuguese Memorandum of Understanding (MoU), an agreement required for obtaining financial assistance. More details available at http://www.portugal.gov.pt/pt/GC19/Documentos/PCM/1R_MoU_20110901.pdf.

similar features, such as hospital payment system, among others. However, public health markets present characteristics that are idiosyncratic to them, including the fact of not being exposed to competition, and these might make a difference in the quantity and quality of services provided for the patients as well as for the expense of resources needed by the system to provide such services. Arguments such as the fact that size of public hospitals are set according to planning criteria instead of an economic optimal criteria and that closure of hospitals seldom occurs in the NHS, reinforce the possibility of having a different interaction between volume of medical procedures and outcomes than what is observed in the private markets.

Using a dataset from the Portuguese NHS spanning eight years, this study concentrates on the existence, or not, of benefits in terms of better health outcomes coming from increases in volume of medical procedures. These potential benefits are considered at both hospital and patient level. The linear and probit estimation methods show evidence that supports an inverse relation between volume of medical procedures and mortality rate/outcome for some diagnosis-related groups (DRG).

At the hospital level 7 out of 21 DRGs showed evidence of the volume-outcome effect, whereas 10 of them did so at the patient level. Sensitivity analysis excluding observations of transfer cases between hospitals revealed a similar result, suggesting that a transfer to another hospital is likely to occur randomly (in the medical sense and from the perspective of the external observer). The initial restriction of the volume-outcome effect to be explained by the SR hypothesis, due to the normative definition of a catchment area for each NHS hospital, leaves only the PMP hypothesis to be explained.[2] To this purpose, probit models were run and the strongest evidence of the volume-outcome effect, with causality from volume to outcome, was found for the following medical procedures: cerebral infarction, respiratory infections, circulatory disorders with acute myocardial infarction (AMI), bowel procedures, cirrhosis, and hip and femur procedures. Moreover, the mechanism works through a static scale effect, with little evidence for learning-by-doing. Instrumental variables (IV) analysis was provided to conclude for the exogeneity of the *volume* variable. The quadratic specification between volume and outcome rather than the linear one gives consistent evidence of the volume-outcome effect

[2]See details in section 1.4, subsection 1.4.3.

across the different model specifications for 5 out of the 21 DRGs. With regard to policy, the centralization of those medical procedures is recommended given that such a policy, keeping all other variables constant, would save a considerable number of lives.

This study is organized into six additional sections. Section 2 discusses the literature. Section 3 presents the methodology used. Section 4 discusses the results. Section 5 provides some extensions. Section 6 concludes.

1.2 Literature review

The literature reports a relationship between hospital volume or surgeon volume and outcomes through mortality rates and other proxies. In general, there seems to exist a positive correlation between higher volume and better outcomes. Some of the studies seek to identify the causality, mainly summarized as the PMP or SR hypotheses. However, the data and diseases investigated vary across studies. The review of 135 papers from 1980 to 2000 reached such conclusion: the variation in the results found is considerably large (Halm et al., 2002; Hodgson et al., 2005).

The most widely used outcome measure is the in-hospital mortality rate. In many cases it is adjusted for severity of the disease. For some diseases, more demanding measures can also be used as the outcome (e.g., coronary artery bypass graft (CABG) as the outcome for the volume of percutaneous translumial coronary angioplasty (PTCA)). Current volume of procedures in a given hospital prevails over the cumulative volume of procedures, occasionally used to try to capture learning-by-doing effects. Several studies apply the logistic regression, according to the distribution of the data. Furthermore, the PMP hypothesis seems to dominate the SR hypothesis for most of the investigated diseases in the majority of results (Flood et al., 1984; Birkmeyer et al., 2002; Ho, 2002; Gaynor et al., 2005).

One limitation in this vast literature, however, is related to the samples used. Almost all of them are conducted using data from the United States and the results can only be extended to private markets (at first). This work aims to fill a gap in the literature, investigating the volume-outcome effect in the Portuguese health sector, a NHS similar to others in Europe. This might allow the results to be extended to other NHS. We stress the fact that we work in a scenario without

competition, which can affect the behavior of the agents in several ways that can influence the relationship between volume of medical procedures and outcomes. The fact that public hospitals have less pressure to be in the efficient size than private hospitals is an example of how competition can affect the scenario.

1.3 Methodology

1.3.1 Data

The data come from the DRG dataset, which is organized by *Administração Central do Sistema de Saúde, I.P. (ACSS)*. It includes all inpatient discharges at the NHS and the procedures are classified according to the International Classification of Diseases, Clinical Modification (ICD-9-CM). The sample is drawn using observations from 2001 to 2008 (and 1997 to 2008, in some cases).[3] We included hospital information data (case-mix index (*cmi*), *beds*, length of stay (*lstay*), discharged patients (*dp*), and *cost*) from ACSS and hospital reports, available at the ACSS website as well.

The initial sample size including all discharges from 2001 to 2008 is 9,523,301 observations. Private hospitals and observations related to hospitals that appear only in three or fewer years were excluded (4.45% of the dataset, 6 hospitals). Also, observations of some hospitals were combined due to the creation of the health centers.[4]

The DRGs used in the study were selected according to the following criteria, applied cumulatively: all DRGs that represent more than 1% in the whole sample (discharges of all hospitals from 2001 to 2008) were kept; DRGs with no positive mortality rate were excluded; DRGs with small representation in the whole sample but with high relative mortality rates or cost were kept; DRGs studied exhaustively in the earlier literature were included whenever possible.

After applying the selection criteria the sample size becomes smaller, with 880,449 observations (9.67% of the original sample) and 21 workable DRGs. There

[3] Data from 1997 to 2000 were used to build the accumulated volume variable from 2001 to 2004.

[4] *Decretos-lei* and *Portarias* of the Portuguese *Ministério da Saúde* were used to combine the data.

are medical and surgical DRGs in which 7 are cancer-related, which were also analyzed in the literature (Birkmeyer et al., 2002; Schoder and Lichtenberg, 2010).

Table 1.1: Selected DRGs

DRG	Type	Description	Criteria	Obs.	%
14	M	Intracranial hemorrhage or cerebral infarction	1	179,131	20.35
78	M	Pulmonary embolism	3	12,851	1.46
79	M	Respiratory infections and inflammations. age >17, CC	3	28,628	3.25
82	M	Respiratory neoplasms	2	37,074	4.21
89	M	Simple pneumonia and pleurisy	1	158,017	17.95
107	S	Coronary bypass with cardiac catheterism without PTCA	3	2,804	0.32
109	S	Coronary bypass without cardiac catheterism without PTCA	3	12,798	1.45
113	S	Amputation for circulatory system disorders except upper limb and toe	3	17,989	2.04
121	M	Circulatory disorders with AMI	3	60,325	6.85
127	M	Heart failure and shock	1	115,447	13.11
148	S	Major small and large bowel procedures, CC	3	33,605	3.82
172	M	Digestive malignancy. CC	2	36,717	4.17
202	M	Cirrhosis and alcoholic hepatitis	3	51,320	5.83
203	M	Malignancy of hepatobiliary system or pancreas	2	27,271	3.1
210	S	Hip and femur procedures. except major joint. age >17, CC	3	19,964	2.27
274	M	Malignant breast disorders, CC	2	7,586	0.86
318	M	Kidney and urinary tract neoplasms, CC	2	7,083	0.8
366	M	Female reproductive system malignancy, CC	2	7,371	0.84
395	M	Red blood cell disorders, age >17	3	30,317	3.44
403	M	Lymphoma and non-acute leukemia, CC	2	16,123	1.83
416	M	Septicemia	2	18,028	2.05
			880,449	100.00	

Notes:
Type: M is medical and S, surgical.
Criteria: 1 means high significance; 2, high relative mortality; and 3, high relative cost and/or earlier research.
CC: With complication.
DRG 121 = DRGs (121 + 122 + 123)

Some earlier studies performed the analysis of the volume-outcome effect at disease level instead of DRG level, i.e., grouping DRGs that are related to the same disease but differ in the degree of complication, age of patients, among other factors. There are good reasons to work at disease and DRG levels. However, we choose not to aggregate them, otherwise additional control variables (for the differences among the DRGs grouped) such as percentage of more complicated cases should be included in the model.

The measure of volume is the current total number of discharges of each DRG, by hospital and year. The outcome variables vary depending on what level of data we are working with. The mortality rate, measured as

$$mort_{h,t} = \frac{(\text{number of deaths})_{h,t}}{\text{volume}_{h,t}}, \tag{1.1}$$

where h indexes the hospital and t, the year, is the outcome variable when the level of analysis is the hospital. Otherwise, using patient level data, the outcome is a binary variable such that:

$$outcome_{n,h,t} = \begin{cases} 0, & \text{if the patient } n \text{ leaves the hospital alive;} \\ 1, & \text{otherwise;} \end{cases} \tag{1.2}$$

where n indexes the patient and remaining indexes are the same as before.

The measure of the outcome is a constraint imposed by the availability of data. Instead, the utilization of a more accurate measure such as the thirty-days outcome would be more appropriate in some cases. For this reason, the results for some DRGs must be interpreted carefully. To give an example, most of the people do not die immediately after the hip and femur procedure (DRG 210) and thirty-days mortality rate would be the most preferred outcome measure, if available.

The sample characteristics reveal that the mean age of the patients is 70 years, which is an old sample generated by chance. Approximately half of them are women. For the hospitals, the mean *cmi* is 1.06, whereas *beds* is 327, *dp* is 11,837 per year, *lstay* is around 8 days and *cost* is 5,000 euros. Volume of medical procedures and mean mortality rates were fairly stable over the years, with slight variation for some DRGs.[5]

1.3.2 Empirical strategy

The first set of analyses is done at the hospital level, using panel data, mean variables of mortality rates, and patient characteristics. Even when taking into account the potential loss of control for some individual characteristics of the patients, we believe that the benefits of such analysis are worthwhile. It is possible to set fixed effects at hospital level and control for potential unobservable characteristics, as done by Farley and Ozminkowski (1992) and Hamilton and Ho (1998).

[5]Complete data at https://sites.google.com/site/giseletbraun/home/research/appendix.pdf.

For each DRG the estimated equations are:

$$mort_{h,t} = \alpha + \beta \times volume_{h,t} + \delta \times m\_age_{h,t} + \phi \times f\_women_{h,t} + \theta \times \chi_{h,t} + \epsilon_{h,t} \quad (1.3)$$

where the indexes are the same as before. The m_age is the mean age of the patients, f_women is the percentage of women, and χ is a vector of hospital characteristics, namely, *cmi, beds, dp, lstay*, and *cost*. We do not have physician payment information or any other proxy to control for the potential fact that better paid physicians are expected to perform better. However, the fixed effects set at hospital level controls for any variation of this kind. The volume-outcome effect is identified if $\beta < 0$.

The second part deals with individual patient data. The dependent variable is the outcome and mean variables are replaced by individual patient characteristics, increasing the amount of information in the model. The change from panel data to a repeated cross-section one is due to data restrictions.[6] The standard errors are clustered at hospital level to accommodate the fact that we have not only patient information on the regressors but also group averages for hospital characteristics.[7] For each DRG, the estimated equations are:

$$outcome_{n,h,t} = \alpha + \beta \times volume_{h,t} + \delta \times age_n + \phi \times gender_n + \gamma \times tddx_n + \theta \times \chi_{h,t} + \epsilon_{n,h,t} \quad (1.4)$$

where n indexes the patient and the other indexes and variables are the same as before. The $tddx$ variable is the total number of main and secondary diagnoses, which is included in the model to control for the severity of the diseases. It is an imperfect substitute for the fact that our measure of outcome has no severity adjustments. The identification of the volume-outcome effect is again identified using the estimated β in each regression.

The next specification is applied to the learning-by-doing analysis. In order to capture influence of past volume of procedures on outcome, besides the current

[6] The dataset has no patient identification.

[7] The assumption of independence over observations is replaced by the independence among clusters, which are the hospitals.

one, we define the *accumvol*, which stands for accumulated volume and is the sum of the number of discharges in a given hospital in the last four years. The equation, estimated for each DRG, is:

$$outcome_{n,h,t} = \alpha + \beta \times volume_{h,t} + \eta \times accumvol_{h,t} + \delta \times age_n + \phi \times gender_n + \gamma \times tddx_n + \theta \times \chi_{h,2008} + \epsilon_{n,h,t} \quad (1.5)$$

where

$$accumvol_{h,t} = \sum_{t-4}^{t-1} volume_{h,t}, \quad t \in \{2001, \ldots, 2008\} \quad (1.6)$$

The hypotheses tested are: (1) existence of economies of scale: $\beta < 0$ and $\eta = 0$; (2) existence of learning-by-doing: $\beta \leq 0$ and $\eta < 0$.[8]

1.4 Results

1.4.1 Analysis at the hospital level

The natural assumption is that the volume of medical procedures has an inverse relationship with mortality rate. The starting point is the hospital level analysis through a linear regression, considering fixed effects at this level, with *volume* as the only independent variable to explain mortality rates. From these, 6 out of 21 DRGs support evidence for a potential volume-outcome effect, at the 5% level of significance (and 4 more DRGs at 10% level of significance). However, these estimated coefficients might be over or even underestimating the real effects due to the absence of further controls. Adding extra information about average patients (*m_age* and *f_women*) and hospitals (*cmi, beds, dp, lstay*, and *cost*), the regressions show that 7 out of 21 DRGs reveal the volume-outcome effect at the same significance level. The mean age of the patients, the fraction of women, and

[8]Estimation of a quadratic relation between outcome and volume were considered by the addition of the term $\beta_2 \times volume^2$ in equations 3, 1.4 and 1.5. The term $volume \times beds$ is added to equations 1.4 and 1.5 to control for the fact that the volume-outcome effect might be of different magnitude depending on the size of the hospital.

all hospital control variables have, in most of the cases, a non-significant effect.[9]

At this point it is worth giving attention to the predictions of the mortality rates to be between 0 and 1. For the DRGs 78, 203, and 366 only one prediction was greater than one and this can be considered as an error. Moreover, predictions of mortality rates less than 0 were identified for less than 5% of the observations, not for all DRGs, allowing us to interpret the results safely even when the model does not restrict the predictions.

1.4.2 Analysis at the patient level

Overview

The results show that 10 out of 21 DRGs have a significant estimated coefficient for the *volume* variable.[10] For the remaining independent variables, *age* and *tddx* have mainly a significant and positive relationship with mortality. The *gender* variable has a significant coefficient with sign that varies depending on what disease we are dealing with. On the other hand, most hospital control variables are statistically non-significant.[11] A subsample without transfer cases was drawn and it shows results similar to the general sample, with addition of DRG 127, which indicates that the inclusion or exclusion of the transfer cases might modify some results. It can be interpreted as if the most complicated cases are being transferred to the hospitals that have more success when performing this medical procedure, which refers to heart failure and shock.[12]

It is noteworthy to highlight the robustness of our results across the two levels

[9]Subsample without hospital transfer cases (patients that are first admitted in a given hospital but end up being treated in another) was selected to verify if instead of being a random process, it might be the case that transfers occur in the most complicated cases. The results were similar: 6 DRGs support the volume-outcome effect at the 5% level of significance, where DRG 403 is new on the list, suggesting that transfer cases might be relevant. Still, DRGs 89 and 210 do not support the effect anymore. See https://sites.google.com/site/giseletbraun/home/research/appendix.pdf for complete results.

[10]The positive sign of β for DRG 109 signs the opposite of the PMP hypothesis and it may indicate that transfers are not happening as a random process, but instead, selective referral, which is not officially permitted in the Portuguese NHS, is actually taking place.

[11]No relevant interaction between the explanatory variables was found.

[12]Transfers in and transfers out represent, respectively, 11.5% and 5.28% of the total sample. Moreover, it was observed that both transfers are decreasing over time.

of analyses. Furthermore, medical, surgical, and cancer-related DRGs are included among those that show evidence of volume-outcome effect.

Comparing them with results reported in the literature, DRGs 107, 121, 148 and 203 were previously described as medical procedures with volume-outcome effect (Flood et al., 1984; Luft et al., 1987; Birkmeyer et al., 2002; Gowrisankaran et al., 2006; Ramanarayanan, 2006). For DRG 127 it was found the opposite result of selective referral (Farley and Ozminkowski, 1992) and for DRG 203 it was identified no correlation. These comparisons indicate that some diseases might have a similar volume-outcome effect even though we are considering different types of health systems, but still the type of health system might be important for other diseases.

Economies of scale and learning-by-doing

The volume-outcome effect has two main possible explanations: PMP or SR hypotheses. Given that the Portuguese health system design precludes the SR hypothesis, we end up with the need to further explain the PMP hypothesis. It may come from two different effects: scale effect or the learning-by-doing process.

One structure to deal with this question is to use current and accumulated volume as explanatory variables of the outcome. If only today's volume affects today's outcome, then we are facing a static scale effect. On the other hand, if today's volume might affect both today's outcome and future outcomes, than we have the effect via learning-by-doing (Gaynor et al., 2005). Regarding policy implications, the identification of scale effects matters less for the selection of which hospital will concentrate the activities, it is only important to increase volume in any hospital. However, if learning-by-doing is recognized, more careful merging of activities should be considered: given the fact that a merger implies the reduction in volume of activity in one hospital and the increase in the volume of activity in another hospital, it should not be reduced in the hospital that is improving through learning-by-doing.

The lagged accumulated volume's variables appear to be highly correlated among themselves, and some of the earlier studies exclude the correlated variables (Schoder and Lichtenberg, 2010). This pattern clouds the results when describing

the mechanism through which volume affects outcome in the context of the PMP hypothesis. Our sample is not an exception and the current volume has correlation with accumulated volume that is around 0.87.

The regressions show that 6 out of the 10 DRGs, the ones identified in the patient level analysis of the general sample, have a static relationship that characterizes the scale effect. The coefficients for the *accumvol* variable are mainly non-significant, except for DRG 148, indicating that learning-by-doing is not a plausible explanation for the PMP hypothesis.[13]

Similar result in the literature can be found in Gaynor et al. (2005). It shows that only contemporaneous effects were statistically significant when explaining probability of death in terms of current and accumulated volume of medical procedures for the CABG.[14]

1.4.3 Exogeneity

The causality between the *outcome* and *volume* variables are key for policy implications and there is no consensus in the literature. It was reported that some DRGs support the PMP hypothesis while others might support the SR hypothesis. In some cases both hypotheses can be supported for the same DRG.

Portuguese patients in the NHS are directed to one or another hospital according to their area of residence. In principle this would automatically exclude the SR hypothesis. However, different arrangements can occur between patients and physicians. The sensitivity analysis of patient results demonstrated that DRG 127 supports the volume-outcome effect after the exclusion of the transfer cases. This fact may be a consequence of SR or a random process. In any case, we need to argue, more formally, that SR is not occurring.

Most studies support the PMP hypothesis and they exclude the SR hypothesis sometimes using the Hausman-Wu test, as in Gaynor et al. (2005), or the Smith and Blundell exogeneity test, as in Ho (2002), or even using a test to check if there

[13]The subsample without the hospital transfer cases has a weaker scale effect, identified only for 3 DRGs.

[14]These results are not perfectly comparable with ours, given that Gaynor et al. (2005) distinguish past volume using annually lagged variables, i.e, using $\sqrt{vol_t}$, $\sqrt{vol_{t-1}}$, $\sqrt{vol_{t-2}}$, and so on as the explanatory variables.

is a correlation between volume and the residuals of the regressions, as in Farley and Ozminkowski (1992). Other authors explain in plain words the impossibility of SR as a plausible hypothesis, as did in Schoder and Lichtenberg (2010).

The analysis of no SR in our data is performed with the IV probit regression model. The current volume of medical procedures is instrumented with the past year volume of medical procedures. The argument is the following: (1) the reason to suspect the endogeneity of volume of medical procedures comes from the fact that unobservable shocks affecting outcome can also affect volume of medical procedures. Therefore, we treat volume as an endogenous variable, i.e, the exogenous increase in the volume of medical procedures can affect outcome, but the better outcome can also attract more patients to the hospital that is improving the outcomes, affecting volume again; (2) the selection of past year volume of medical procedures as the instrument variable relies on the fact that it is correlated with current volume of medical procedures, the instrumented variable, but it is uncorrelated with outcome, because an exogenous shock in outcome can also affect today's volume of medical procedures but cannot affect past year's volume of the same medical procedure. The results of the Wald test of exogeneity for the *volume* variable implies that we cannot reject the null hypothesis that volume of medical procedures is an exogenous variable, except for DRGs 107 and 416.

1.5 Extensions

1.5.1 Alternative specification

The volume-outcome effect based on a linear relationship between *outcome* and *volume* is more intuitive and the most usually applied specification in the earlier literature. However, it is feasible to think that there is a limit in enhancement of the outcomes due to an increase in the number of medical procedures inside a hospital. An almost empty hospital is expected to have poorer outcomes than the one that performs a higher number of a given medical procedure, but a crowded hospital is also expected to suffer from not being able to offer the best medical care that it could offer in the case that demand were smaller. Therefore, it is adequate to determine what the volume-outcome effect is in the Portuguese NHS if we allow

the relationship between *outcome* and *volume* to be quadratic.

The results for hospital data are less evident than in the linear case and the magnitude of the effects are smaller. The same applies to the patient data: the effect was identified for 6 DRGs out of 21, with scale effect recognized for 5 DRGs. Noteworthy is the concave shape of the curve that relates outcome and volume of medical procedures indicated by the predominant negative sign of the estimated coefficients of the *volume* variable and positive sign of the *volume_sq* variable. Moreover, the variable *volume*beds* is predominantly negative and significant for almost all DRGs, pointing out that the magnitude of the volume-outcome effect varies with the size of the hospital.[15]

1.5.2 Economic impact

The real impact of the volume-outcome effect depends on the volume of activity inside the hospital. Medical activity is affected by different arrangements of the health system. Volume could be leveraged with centralization of medical procedures. A possible exercise is to verify what would happen with outcome when the volume is hypothetically increased through centralization of the two highest-volume hospitals, for each DRG that was identified with volume-outcome effect in the linear model of the hospital level analysis.

We calculated the change in the 2008 mortality rate induced by the hypothetical centralization effect (sum of the volume of medical procedures for both hospitals). Other criteria could be used to reunite activities: reunion based on close-by hospitals would be more expected in reality. However, we stick to the analysis with the largest ones because they might be a good sign of existence of benefits.

We can avoid up to 52 out of 446 deaths with the centralization of the medical care, as calculated for DRG 14. *Ceteris paribus*, the centralization of the two highest-volume hospitals that provide medical care classified as DRG 14 would imply a reduction of 12% in deaths, according to the estimates of our model.

[15]Subsample without transfer cases shows the volume-outcome effect with economies of scale, for DRGs 14, 79, 148, and 202.

Table 1.2: Economic impact of centralization

		(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
DRG	$hosp$	β	vol	$gain$	$mort$	$\widehat{mort}$	$deaths$	$\widehat{deaths}$	POTENTIAL SAVED LIVES	
14	X	-0.00010	1480	-0.117	0.186	0.186	275	275	-60	-52
	Y		1170		0.205	0.146	240	171		
89	Y	-0.00006	1435	-0.056	0.265	0.258	380	371	-32	-31
	W		931		0.199	0.204	185	190		
121	X	-0.00041	635	-0.141	0.055	0.055	35	35	-11	-10
	Z		345		0.125	0.100	43	35		
202	W	-0.00026	300	-0.077	0.137	0.162	41	49	-7	-7
	S		296		0.152	0.138	45	41		
203	S	-0.00083	224	-0.149	0.241	0.307	54	69	-22	-22
	Y		180		0.528	0.452	95	81		
210	Y	-0.00067	205	-0.088	0.049	0.013	10	3	-1	-1
	T		132		0.015	0.052	2	7		
274	U	-0.00229	86	-0.110	0.349	0.318	30	27	-6	-6
	V		48		0.458	0.549	22	26		

Notes:
$III = \beta \times (\sum vol - vol)$;
$VI = \beta \times mort$;
$VII = \beta \times \widehat{mort}$;
$VII = gain \times \sum mort$;
$VIII = gain \times \sum \widehat{mort}$.

The earlier literature reported reduction in the mortality rates goes from 0.5% to 6.8% in response to an increase in the volume of activity. The hypothetical centralization used in our computations implies an increase in volume that sometimes goes up to 100%, which can explain the larger numbers achieved (from 5.53% to 14.67%).

1.6 Final remarks

In the work developed by Gaynor et al. (2005), a positive correlation between greater volume of medical procedures and better health outcomes for CABG in the US health market is reported. A question that we can address is whether it might also occur in a NHS. It is an important issue because several countries' NHS in Europe are nationally supported and so, the incentives and behavior of the economic agents might be different from what is reported in the literature for private markets. We used data from Portugal, where the NHS has some different features. Patients cannot choose the hospital where they are going to receive medical care. They are allocated according to the place of residence.

The volume-outcome effect, defined as the correlation between volume of medical procedures and outcomes, has been identified for some DRGs. The data show strong evidence in favor of the PMP hypothesis for 6 medical procedures: cerebral

infarction, respiratory infections, circulatory disorders with AMI, bowel procedures, cirrhosis, and hip and femur procedures. The gains in the outcomes are mainly caused by scale effects. Weak evidence of learning-by-doing is observed. Moreover, exogeneity of the *volume* variable is verified with IV probit models, where lagged volume variable is used as the instrument. The centralization of those medical procedures is recommended given that such policy, keeping all other variables constant, would save many lives.

Other relevant issues, such as transfer of costs from the State to the patients (like transportation costs to the health center) were not considered here.

Appendices

Partial regression coefficients

	Hospital analysis Unit of observation: hospital Dependent variable: mortality rate			Patient analysis Unit of observation: patient Dependent variable: outcome							
		Fixed effect I	Fixed effect II		Probit I	Probit II			Probit III		
DRG	obs.	volume	volume	obs.	volume	volume	volume_sq	volume×beds	obs.	volume	accumvol
14	463	-0.00009**	-0.00010**	144789	-0.00015*	-0.00050*	0.00000**	-0.00066**	135609	-0.00025*	0.00002
		(0.00002)	(0.00004)		(0.00007)	(0.00020)	(0.00000)	(0.00016)		(0.00012)	(0.00002)
78	429	-0.00104+	-0.00087	10343	-0.00022	-0.00685*	0.00003	0.00136	9846	0.00035	-0.00039
		(0.00055)	(0.00069)		(0.00174)	(0.00296)	(0.00003)	(0.00319)		(0.00176)	(0.00055)
79	455	0.00033*	-0.00013	22721	-0.00109**	-0.00074	0.00000*	-0.00260**	21470	-0.00116*	0.00014
		(0.00014)	(0.00018)		(0.00033)	(0.00076)	(0.00000)	(0.00078)		(0.00047)	(0.00017)
82	468	-0.00024	-0.00042+	28072	0.00003	0.00048	0.00000	-0.00225*	26721	0.00012	-0.00004
		(0.00023)	(0.00023)		(0.00050)	(0.00204)	(0.00001)	(0.00093)		(0.00088)	(0.00021)
89	466	-0.00001	-0.00006*	128048	-0.00002	-0.00013	0.00000	-0.00033*	120622	0.00006	-0.00003+
		(0.00002)	(0.00003)		(0.00007)	(0.00019)	(0.00000)	(0.00015)		(0.00005)	(0.00002)
107	46	-0.00002	0.00019	2260	-0.00550**	0.02234**	-0.00006	-0.01844*	2091	-0.00770+	0.00092
		(0.00029)	(0.00028)		(0.00166)	(0.00544)	(0.00004)	(0.00936)		(0.00414)	(0.00142)
109	48	0.00000	-0.00001	10431	0.00150	0.00827+	-0.00003*	0.00842*	10026	0.00301+	-0.00022
		(0.00002)	(0.00003)		(0.00124)	(0.00441)	(0.00001)	(0.00389)		(0.00171)	(0.00024)
113	441	-0.00000	0.00001	14285	0.00025	-0.00017	0.00001	-0.00332+	13338	0.00078	-0.00047
		(0.00034)	(0.00037)		(0.00058)	(0.00156)	(0.00000)	(0.00187)		(0.00092)	(0.00030)
121	402	-0.00021*	-0.00041**	47626	-0.00127**	-0.00018	-0.00000	-0.00209**	45262	-0.00107**	-0.00001
		(0.00010)	(0.00012)		(0.00029)	(0.00053)	(0.00000)	(0.00043)		(0.00032)	(0.00010)
127	475	0.00000	0.00005	93564	-0.00017	-0.00028	0.00000	-0.00111**	87922	-0.00014	-0.00002
		(0.00005)	(0.00008)		(0.00023)	(0.00055)	(0.00000)	(0.00041)		(0.00023)	(0.00006)
148	414	-0.00011	0.00018	26780	-0.00223*	-0.00701**	0.00002**	-0.00388**	25287	-0.00173*	-0.00050**
		(0.00015)	(0.00018)		(0.00105)	(0.00145)	(0.00000)	(0.00067)		(0.00086)	(0.00017)
172	463	-0.00017	-0.00005	28681	-0.00002	0.00061	0.00000	-0.00254**	27286	0.00035	-0.00012
		(0.00023)	(0.00021)		(0.00044)	(0.00083)	(0.00000)	(0.00043)		(0.00046)	(0.00011)
202	467	-0.00029**	-0.00026*	42160	-0.00109**	-0.00184**	0.00000*	-0.00095	39887	-0.00103**	-0.00003
		(0.00008)	(0.00011)		(0.00029)	(0.00059)	(0.00000)	(0.00073)		(0.00037)	(0.00009)
203	465	-0.00075*	-0.00083*	21939	-0.00139*	-0.00103	-0.00001	0.00164+	20849	-0.00142	0.00014
		(0.00032)	(0.00039)		(0.00066)	(0.00177)	(0.00000)	(0.00088)		(0.00112)	(0.00024)
210	353	-0.00056+	-0.00067*	16368	-0.00563**	-0.00775**	0.00003*	-0.00655**	15290	-0.00573**	0.00013
		(0.00030)	(0.00033)		(0.00080)	(0.00294)	(0.00001)	(0.00233)		(0.00100)	(0.00024)
274	394	-0.00169*	-0.00229*	5530	-0.00326*	-0.00631	0.00000	0.00610+	5297	-0.00244	-0.00031
		(0.00065)	(0.00098)		(0.00144)	(0.00483)	(0.00002)	(0.00322)		(0.00203)	(0.00036)

Continued on next page

Continued from previous page

318	432	-0.00210	-0.00189	5615	-0.00084	0.00709	-0.00010	-0.00088	5335	-0.00428	0.00094
		(0.00147)	(0.00168)		(0.00349)	(0.00648)	(0.00008)	(0.00853)		(0.00496)	(0.00109)
366	431	-0.00163+	-0.00008	5535	-0.00437**	-0.00625	0.00002	-0.00141	5360	-0.00254	-0.00057
		(0.00084)	(0.00091)		(0.00167)	(0.00658)	(0.00004)	(0.00234)		(0.00164)	(0.00053)
395	475	-0.00043+	-0.00034+	24082	-0.00087	-0.00164	0.00000	-0.00085	22703	-0.00109+	0.00023
		(0.00024)	(0.00018)		(0.00058)	(0.00157)	(0.00001)	(0.00172)		(0.00058)	(0.00020)
403	447	-0.00075	-0.00070	12575	-0.00018	-0.00089	0.00001	-0.00217	11969	-0.00031	0.00009
		(0.00047)	(0.00049)		(0.00056)	(0.00262)	(0.00001)	(0.00134)		(0.00089)	(0.00021)
416	434	0.00008	-0.00001	14035	-0.00043	-0.00126	0.00000	-0.00062	12889	-0.00060	0.00014
		(0.00018)	(0.00023)		(0.00027)	(0.00171)	(0.00000)	(0.00103)		(0.00038)	(0.00014)

Notes:
** p<0.01, * p<0.05, + p<0.1
Robust standard errors adjusted for cluster in hospital in parentheses.
Year dummies included in all models, except Fixed effects I.
Fixed effects I - Independent variable: *volume*.
Fixed effects II - Independent variables: *volume*, *mean_age*, *f_women*, *cmi*, *beds*, *dp*, *lstay*, and *cost*.
Fixed effects at hospital level.
Probit I - Independent variables: *volume*, *age*, *gender*, *tddx*, *cmi*, *beds*, *dp*, *lstay*, and *cost*.
Probit II - Independent variables: *volume*, *volume_sq*,*age*, *gender*, *tddx*, *cmi*, *beds*, *dp*, *lstay*, and *cost*.
Probit III - Independent variables: *volume*, *accumvol*, *age*, *gender*, *tddx*, *cmi*, *beds*, *dp*, *lstay*, and *cost*.

Table 1.3: Partial regression coefficients

Bibliography

Barker D, Rosenthal G, Cram P. 2010. Simultaneous relationships between procedure volume and mortality: do they bias studies of mortality at specialty hospitals? *Health Economics* **20**(5) : 1099-1050.

Birkmeyer J, Siewers A, Finlayson E, Stukel T, Lucas F, Batista I, Welch H, Wennberg D. 2002. Hospital volume and surgical mortality in the United States. *New Englad Journal of Medicine* **346**(15) : 1128-1137.

Cooper Z, Gibbons S, Jones S, McGuire A. 2011. Does hospital competition save lives? Evidence from the English NHS patient choice reforms. *The Economic Journal* **121**(554) : F228-F260.

Farley D, Ozminkowski R. 1992. Volume-outcome relationships and in-hospital mortality: the effect of changes in volume over time. *Medical Care* **30**(1) : 77-94.

Flood A, Scott W, Ewy W. 1984. Does practice-make-perfect? I: the relation between hospital volume and outcomes for selected diagnostic categories. *Medical Care* **22**(2) : 98-114.

Flood A, Scott W, Ewy W. 1984. Does practice-make-perfect? II: the relation between hospital volume and outcomes and other hospital characteristics. *Medical Care* **22**(2) : 115-125.

Gaynor M, Seider H, Vogt W. 2005. The volume-outcome effect, scale economies, and learning-by-doing. *The American Economic Review* **95**(2) : 243-247.

Gowrisankaran G, Ho V, Town R. 2006. Causality, learning and forgetting in surgery. *Unpublished manuscript.*

Grumbach K, Anderson G, Luft H, Roos L, Brook R. 1995. Regionalization of cardiac surgery in the United States and Canada. *Journal of the American Medical Association* **274**(16) : 1282-1288.

Halm E, Lee C, Chassin M. 2002. Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature. *Annals of Internal Medicine* **137**(6) : 511-520.

Hamilton B, Hamilton V. 1997. Estimating surgical volume-outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics* **6**(4) : 383-395.

Hamilton B, Ho V. 1998. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. *Medical Care* **36**(6) : 892-903.

Hannan E, Kilburn H, O'Donnell J, Bernard H, Shields E, Lindsey M, Yazici A. 1992. Coronary artery bypass graft surgery: the relationship between in-hospital mortality in New York state and the volume of abdominal aortic aneurysm surgeries performed. *Health Services Research* **27**(4) : 517-542.

Ho V. 2000. Evolution of the volume-outcome relation for hospitals performing coronary angioplasty. *Circulation* **101** : 1806-1811.

Ho V. 2002. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Journal of Health Economics* **21**(5) : 873-885.

Luft H, Hunt S, Maerki S. 1987. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Services Research* **22**(2) : 157-182.

Ramanarayanan S. 2006. Does practice make perfect: an empirical analysis of learning-by-doing in cardiac surgery. *Working paper.*

Schoder J, Lichtenberg F. 2010. Impact of specialization on health outcomes - evidence from U.S. cancer data. *Working paper.*

Hodgson D, Urbach D, Przybysz R, Gong Y, Sullivan T, Rabeneck L. 2005. Analysis of current research related to the impact of low-volume procedures/surgery and care on outcomes of care. *Institute for Clinical Evaluative Sciences - Toronto, Ontario* in Access to Health Services in Ontario: ICES Atlas : 9-42.

Wen S, Simunovic M, Williams J, Johnston K, Naylor C. 1996. Hospital volume, calendar age, and short term outcomes in patients undergoing repair of abdominal aortic areurysms: the Ontario Experience, 1988-92. *Journal of Epidemiology and Community Health* **50** : 207-213.

Chapter 2

Upcoding in a National Health Service

Abstract

Evidence from the US pointed out, over the years, to the existence of upcoding in management practices. Upcoding is defined as classifying patients in DRG codes associated with larger payments. The incentive for upcoding is not particular to private providers of care. Conceptually, any patient classification system that is used for payment purposes may be vulnerable to this sort of strategic behavior by providers. We document here that upcoding occurs in a National Health Service (NHS) where public hospitals have their payment (budget) tied to the classification of treatment episodes. Using DRG data from Portugal we found that the practice of upcoding has being used in the hospitals in a way leading to larger budgets. The effect is quantitatively small.

2.1 Introduction

Upcoding has been documented empirically in US hospitals and their Medicare system.[1] However, no current theory or empirical evidence addresses the issue in the context of national health services.[2] This is somewhat surprising as patient classification systems are now common in many countries, several of them having a National Health Service (NHS) and using DRG-like systems to pay providers of health care.

The use of prospective payment based on patients classification systems has become widespread. A crucial aspect of payments by episode is the coding of each patient. In the US, the first country to have payments to health providers based on diagnosis-related groups (DRG), a concern often raised was that of upcoding. The practice of upcoding consists in shifting the DRG of a patient to another DRG yielding a higher payment from the third-party provider.

As countries with a NHS also adopt patient classification systems and prospective payments to their public hospitals, the very same question can be asked: do we observe upcoding in the public sector? Should it be a concern in the context of the NHS? Our analysis shows that upcoding can also be present in a NHS system, though the link between episode coding and funding is weaker than in private markets. The study was done with data from the Portuguese NHS. A politically-driven change in DRG prices was the exogenous fact used to assess whether public hospitals upcode in response to changes in the relative prices of DRGs.

Portugal has a NHS since 1979, and in the early eighties a version of the US DRGs was introduced. The system has been adjusted over time, both in terms of number and definition of DRGs, as well as the prices paid for each DRG. During the early years, there was no clear link between the budget that each public hospital received and its activity. For a long time, budgets were yearly revised based on historic costs or deficits were run without managers caring about it as sooner or later fresh money from the Government would come in any way. Thus, the soft-budget constraint effect rendered relatively irrelevant the "price" given to each DRG and they were revised sporadically and without a sound underlying cost analysis.

[1] Upcoding is also known as DRG creep in the related literature.

[2] See Silverman and Skinner (2004) and Dafny (2005) for recent accounts of upcoding.

By 2002, a major change occurred in the Portuguese NHS, with new, more entrepreneurial-like, management rules being introduced in nearly half of public hospitals. Different management teams were brought in to the public hospitals. Their performance, either under the new or the old statutes, became under more scrutiny. Even if no public hospital has become bankrupt, it has occurred a clear hardening in the budget constraint. Now the hospitals have to sign a contract with the Ministry of Health that describes the expected activity and overall budget for the next year. The information from the DRGs is actively being used to establish the terms of the contracts. Thus, while under the old system, funding of hospitals was dominated by the soft budget constraint effect and no incentive for upcoding exists, under the new environment, upcoding delivers benefits to the hospital and therefore can become a matter of concern to the purchaser of care (Ministry of Health).

The changes in DRG prices provide an exogenous source of variation, not related to evolution in hospital costs, that may have an impact on hospital decisions. This enables us with variation to identify upcoding behavior within a NHS framework. Since the price's changes are not equally relevant to all DRGs, we trace whether shifts in DRG coding toward higher priced DRGs were stronger in those cases where the relative price of a DRG changed the most. Another issue of interest is whether upcoding is more likely to occur in the set of hospitals that received a new type of statute (entrepreneurial-like) or is common to all hospitals, including pure public hospitals.[3]

In particular, the study analyzes the Portuguese exogenous DRGs price change that occurred in 2006, using public hospital data from 2001 to 2008. We found evidence of upcoding in the more recent years. The hospitals respond to price changes in DRGs including more patients into the DRGs that are more profitable. No distinction across types of hospitals was detected, regarding upcoding behavior.

The paper is organized into 7 sections. Section 2 presents a literature review. Section 3 describes the Portuguese NHS and the exogenous politically-driven change in prices. Section 4 presents a simple model generating our empirical predictions. Section 5 explains the methodology used and section 6 discusses the results ob-

[3]The managers of pure public hospitals were also put under pressure as the transformed hospitals become a benchmark for them.

tained. Section 7 presents the concluding remarks.

2.2 Literature review

A recent theory treatment of upcoding is due to Kuhn and Siciliani (2013). Their focus is on how quality, "manipulative effort" (their upcoding aspect) and auditing interact with the payment rule defined by a third-party payer (which can be a NHS) to a provider. Upcoding takes in Kuhn and Siciliani (2013) a particular role: it increases the number of patients treated and the average DRG weight. The analysis assumes hospital management to be profit maximizer or surplus maximizer within the NHS.[4]

From their model, it follows quite naturally that higher prices generate higher manipulative efforts and higher output. More auditing effort reduces upcoding. The price instrument influences quality provided but auditing does not. Their work nicely explores this difference in the instruments available to the purchaser (third-party payer).

To our empirical purposes, their model is not directly applicable although the implications do carry out. We stick closer to the implicit setup of Dafny (2005), in which demand is exogenous to the hospital and upcoding consists in the classification of patients in a higher price DRG. Exogeneity of demand is, given the existence of well defined catchment areas for the Portuguese NHS hospitals, the reasonable assumption. Although some auditing exists, we do not have information on its results, and we cannot explore this dimension of the problem. Since DRGs have been in place since the mid-eighties and coding accuracy has been audited since the beginning, we do not expect a major role played by this element. The work of Dafny (2005) focused on the US hospitals responses to price changes of the DRGs. She found that hospitals upcoded patients in order to increase profits, by reimbursements, and hospitals upcode more in the DRGs where the price increase was greater. Moreover, it occurred chiefly in the for-profit hospitals, and there was no increase in intensity or quality of care. Another studies found the same conclusion, namely Carter and Ginsburg (1985), Hsia et al. (1988), Steinwald

[4]Kuhn and Siciliani (2013) argue on behalf of their assumption with managerial discretion to use generated surplus even within the NHS.

and Dummit (1989), Carter et al. (1990) and Silverman et al. (1999). Some studies, using a selected group of DRGs, could not prove the upcoding behavior, though they highlighted consistent evidence with it, as Silverman and Skinner (2004).
There is a part of the literature that mainly studied the case mix index (CMI), which measures the complexity of episodes treated in a hospital. It is calculated using DRG weights and a proxy of the total number of patients treated in the hospital. The aim of these studies is to explain the reasons that led CMI to increase since the implementation of the prospective payment system (PPS). Carter and Ginsburg (1985), Hsia et al. (1988), Steinwald and Dummit (1989) and Carter et al. (1990) attempted to explain such increase in the CMI. Carter and Ginsburg (1985) explained that, from the 8.4% accumulated CMI increase from 1981 to 1984, 3.3% could be due to upcoding. Hsia et al. (1988) measured the incorrect coding that occurred under the PPS from 1984 to 1985. Doing a review of medical records, they found an error rate of 20.8% in the DRG codes. Also, they concluded that small hospitals were associated with greater error rate. Steinwald and Dummit (1989) argued that, behind the 20% CMI increase in US from 1983 to 1988, only 8% was due to formal changes in DRG weights, suggesting upcoding as the culprit for the remaining increase. Using a sophisticated code system, Carter et al. (1990) decomposed the 2.4% increase in the CMI of the US hospitals between 1986 and 1987. They found that one third of that value was due to upcoding. Kroneman and Nagy (2001), using a sample of Hungarian episodes, did not find any relation between implementation of DRGs and increasing in the CMI from 1992 to 1995.
Some studies analyzed the association between hospital ownership and upcoding behavior. Silverman et al. (1999) found that for-profit US hospitals presented the greatest evidence, after the adoption of DRGs, suggesting the existence of upcoding behavior in 1989, 1992 and 1995. Silverman and Skinner (2004) studied the hospital ownership connection with upcoding for DRGs associated to pneumonia and respiratory infections, using data from 1989 to 1996. They did not prove the existence of upcoding, though it was suggested, especially for those DRGs for which the severity level is more difficult to distinguish. Xirasagar and Lin (2006) tested the presence of upcoding for hernia operation and cataract surgery using Taiwan National Health Insurance Research Database of 2001. They concluded that large public teaching hospitals are more likely to admit these types of procedures than

for-profit and not-for-profit hospitals.

Other related research topics were investigated, indirectly motivated by the upcoding behavior, after the changes in the US health market's rules. Lave (1985) pointed out that compression was occurring in the US DRG prices, in the sense that high cost DRGs had been set lower to their actual costs and low cost DRGs had been set higher, and this might have important implications on DRG creep. Psaty et al. (1999) calculated the potential upcoding costs per year for heart failure procedures in US from 1986 to 1993, highlighting the size of the problem for the society.

Two other contributions are to be mentioned. Rauner and Schaffhauser-Linzatti (2002) studied the impacts in the health system of Austria after the implementation of the PPS in 1997. They obtained indicators that showed cost reductions and improvements in quality indexes for their hospitals; Steinbusch et al. (2007) conducted a comparative study among American, Australian and Dutch's case mix reimbursement systems, showing the type of market, control system and case mix characteristics as the variables that could motivate or inhibit the upcoding. This is a relevant study in political terms that points out to particular requirements that should be taken into account when planning to implement or improve a health system.

The conclusion in the literature is that there is an agreement among the results of studies for the US, making clear the existence of upcoding after the PPS's implementation. It seems to be a non equal behavior that depends on hospital ownership and the characteristics of the DRGs. This study addresses upcoding behavior in the context of a public hospital system. While it is often believed that public hospitals may behave differently than for-profit or not-for-profit private hospitals, it can be counterargued that incentive is what matters. Our contribution is the identification of upcoding evidence in a public health system, using data from Portugal, a country with a NHS that adopted the use of a DRG system in the mid-eighties.

2.3 The Portuguese health system

The Portuguese health care system is based on a NHS. Under the traditional public service approach to financing NHS hospitals, little incentive for upcoding exists. Although a budget was attributed to the hospital at the beginning of the year, the values were typically below the expenditures of the hospital in the previous year. Hospitals were aware, generally speaking, that budget reinforcements would be received, sooner or later. Thus, the payment system to the public NHS hospitals entailed no relation to DRG values.

The implementation of the DRG system in Portugal started in 1984, with an agreement between the Ministry of Health and Yale University. The two main objectives of the protocol were to test the possibility of making use of the DRGs and to develop an information and financing system. The results of the study were good enough in such a way that in 1987 a patient classification system was starting to be implemented in the Portuguese NHS. The transition period to a system of payments based on DRGs started in 1989 and this modified the incentives to produce and improve health goods. The operationalization was made in 1990, but not reflected in the payment system in a straightforward away.

The budgeting process did allow for some role of predicted costs by the DRG system in its formulae, though it also took into account historical cost data. Initial budgets were seldom respected by hospitals. Regularly, extra budgets were given, making any role of DRGs in the payment system illusive. The year of 2003 witnessed an important change in hospital management rules, bringing roughly half of the public hospitals to a trust-like juridic form and management rules closer to those of the private sector. No fundamental changes in the budgeting process took place. In 2005, the new Government established a budgeting process that brought hospital budgets at the beginning of the year closer to expected costs and management performance became under closer scrutiny. Hospital management teams signed contract programs with regional health authorities, committing to achieve pre-determined activity levels and costs. These changes introduced the financial incentive for upcoding, as higher complexity of cases would bring more funds in the future.

The Portuguese DRG prices changed in 2003 by a Government decree and remained

stable until early 2006, when another set of prices was fixed. Then, an equal change in prices for all DRGs was set in 2007 and no further change in prices occurred in 2008. The price change was exogenously determined and politically-driven. Therefore, the Portuguese NHS provides a natural bed test to assess the extent of upcoding in a NHS.

The policy change of DRG prices was defined in the *Portaria* 567/2006, which followed the previous prices set of the *Portaria* 132/2003 issued by the Portuguese Ministry of Health.[5] Table 2.1 shows the DRG pairs where the modification in the weight or in the price was greater. An increase or decrease in weight or in price acts as an impulse for upcoding responses. For example, from 2003 to 2006, the DRG 32 increased its price and weight respectively 77.3% and 61.2%. It means that DRG 32 become more profitable, given that costs did not increase in the same magnitude. Such increases raised incentives for the occurrence of DRG creep.

Using these kind of changes, we aim to understand how hospitals respond to the price increases or decreases: do they keep on the same behavior pattern, given that the reasons for changes in price are mainly to adjust hospital costs or do the hospitals behave strategically, upcoding, taking into account that control instruments used are not powerful enough to verify every hospital action?

Table 2.1: Top 10 changes in prices and weights of DRGs

		Price (euros)		Weight		Variation	
DRG	**Description**	**2003**	**2006**	**2003**	**2006**	**Δprice**	**Δweight**
32	Concussion age >17 w/o CC	638.83	1132.77	0.3	0.4836	0.773	0.612
34	Other disorders of nervous system w CC	3482.13	1713.92	1.64	0.7317	-0.508	-0.554
151	Peritoneal adhesiolysis w/o CC	2338.86	3738.67	1.1	1.5961	0.599	0.451
155	Stomach esophageal & duodenal procedures age >17 w/o CC	4131.51	7245.92	1.94	3.0934	0.754	0.595
165	Appendectomy w complicated principal diag w/o CC	1827.81	3306.97	0.86	1.4118	0.809	0.642
168	Mouth procedures w CC	6458.66	2451.3	3.03	1.0465	-0.620	-0.655
206	Disorders of liver except malig.cirr.alc hepa w/o CC	1321.87	2279.6	0.62	0.9732	0.725	0.570
257	Dotal mastectomy for malignancy w CC	4667.36	2341.44	2.19	0.9996	-0.498	-0.544
283	Minor skin disorders w CC	3044.28	1247.32	1.43	0.5325	-0.590	-0.628
398	Reticuloendothelial & immunity disorders w CC	4545.15	2003.91	2.13	0.8555	-0.559	-0.598

[5]The changes in DRG weights are equal to the ones in prices, making the correlation between changes in DRG prices and weights roughly equal to one.

2.4 A quick illustrative model

To provide a background to our empirical procedure and the design of the robustness checks on the evidence of upcoding, we use a simple model. There is an exogenous demand, unaffected by coding decisions, at the pair of closely related DRGs: $n = x_1 + x_2$, where x_1 refers to top DRG and x_2, to bottom DRG, which corresponds to the true demand.
The observed treatment is defined as $u = x_1 + g$ and $d = x_2 - g$, where g measures the extent of upcoding (moving patients from bottom DRG to top DRG).
The objective function of the hospital is given by the profits minus the cost of upcoding:

$$V = p_1 u + p_2 d - c_1 x_1 - c_2 x_2 - 0.5 s g^2 \tag{2.1}$$

where the last term aggregates all costs from upcoding into a quadratic cost function (auditing, detection and fines & ethical costs). The cost c_i is independent of coding and upcoding levels depend on price difference between the top and bottom DRGs (this is the basis of the empirical strategy, borrowed in part from Dafny (2005)). Under equation 2.1, the first-order condition for the choice of g depends solely on s (the cost parameter of upcoding) and on $(p_1 - p_2)$, the price difference between the two DRGs of the same pair.
The fraction of top DRG in a pair of DRGs is defined as:

$$f = \frac{x_1}{x_1 + x_2} + \frac{g}{x_1 + x_2} \tag{2.2}$$

Under constant x_1 and x_2, f and g move in the same direction. If x_1 and x_2 move due to other reasons (real effects), then inference is more complicated and it is necessary to accommodate these facts into the empirical analysis. For this reason, the drivers of the ratio $x_1/(x_1 + x_2)$ need to be studied carefully.

2.5 The empirical approach

We follow closely the empirical strategy laid down by Dafny (2005), however, we add to her main equation other relevant variables, namely, age and gender of

patients, characteristics of hospital and fixed effects at the pair of DRGs.[6]
Still relative to Dafny (2005), another difference in our sample is the co-existence of hospitals operating under two distinct juridic regimes within the NHS. One of them is the traditional public sector management system, while the other is an entrepreneurial-like management model (although hospitals remain under complete public ownership).

2.5.1 Data

The data used was the Portuguese DRG database, which is organized by *Administração Central do Sistema de Saúde* (ACSS). Our sample is composed of hospital discharges from 2001 to 2008. We included all hospitals or hospital centers, with the exception of *Hospital Amadora-Sintra*, due to particular management characteristics, not present in the other hospitals.[7] The sample size is 54,593 observations from 104 hospitals and/or hospital centers. Not all of them have been active in all periods (unbalanced panel) as several hospitals merged to originate hospital centers.
We selected 112 pairs of DRGs, each of them corresponding to two similar DRGs that differ only by having or not having complications. For example, the pair number 1 is composed by the DRGs 7, "periph & cranial nerve & other nerv syst proc w CC", and 8, "periph & cranial nerve & other nerv syst proc w/o CC".[8]
The rationale behind the DRGs selection relates to the likelihood of catching the upcoding: the ability to distinguish a patient wrongly coded between two DRGs that only differ in the level of complication of the disease requires a specific knowledge that in general only doctors have. Hence, the purchaser must proceed with auditing and it is costly. In such scenario, the upcoding might be profitable, if not detected.[9]

[6]From Dafny (2005), equation (3): $fraction_{pt} = \alpha + \varsigma pair_p + \delta year_t + \psi \Delta spread_{p,88-87} \times post + \varepsilon_{pt}$.

[7]This hospital has been managed since 1996 by a private consortium under a public-private partnership for management.

[8]Criteria of selection: from description of the DRGs, 145 pairs of DRGs were selected, where the difference between each two DRGs in a pair is only by having or not having complications; 33 pairs were excluded due to lack of observations in our sample.

[9]Auditing does exist and codification of patients is done by especially trained doctors. These factors help to contain the extent of upcoding. However, not much information is available about

2.5.2 Variables and model specification

The basic dependent variable is the fraction of patients in the top DRG, $fraction_{ipt}$, in hospital i, DRG pair p, and year t. Our interest lies in the relationship of the $fraction_{ipt}$ with the change in the relative price of a DRG, the $spread_p$. This latter is the politically-driven exogenous variable, which is constructed in two steps. First, we take the difference between the value paid to the hospital for an episode classified in the DRG with complications minus the value paid for an episode of DRG without complications. Similar to Dafny (2005), p. 1533, equation (2), we define:

$$spreadp_{pT} = \begin{pmatrix} DRG\ price\ in \\ the\ top\ code \end{pmatrix}_{pT} - \begin{pmatrix} DRG\ price\ in \\ the\ bottom\ code \end{pmatrix}_{pT} \tag{2.3}$$

where p indexes the DRG pair and T indexes the year, which is 2003 or 2006, the time of price changes. The DRG price is common to all hospitals as it is published in the official Journal of the Government. Second, given this absolute value, we take the difference over the years.[10] Therefore, it is defined as:

$$spreadp_p = spreadp_{p06} - spreadp_{p03} \tag{2.4}$$

Besides prices, DRGs are also assigned weights to reflect their relative complexity. Analogous to the spread of prices, the $spreadw_p$ is computed using the spread of weights, defined as:

$$spreadw_{pT} = \begin{pmatrix} DRG\ weight\ in \\ the\ top\ code \end{pmatrix}_{pT} - \begin{pmatrix} DRG\ weight\ in \\ the\ bottom\ code \end{pmatrix}_{pT} \tag{2.5}$$

$$spreadw_p = spreadw_{p06} - spreadw_{p03} \tag{2.6}$$

For some pairs of DRGs, price and weight changes are perfectly correlated while in other cases prices may change without a change in the weight. Also, weights may

those audits.

[10] We use the full year as an approximation, though the new prices started to have effect only in the middle of the year.

be changed with price kept constant in other occasions. The DRG weights are used to compute a case-mix index, which is used in the calculation of a hospital budget.

Table 2.2 presents the values of the top 10 variation in spreads. As an example, the pair 32 is composed by DRGs 110 and 111. The value of the price spread in 2003, €5,943.18, is the difference between the DRG prices €12,586.84 (DRG 110) and 6,653.66 (DRG 111).[11] We obtained the $spreadp_p$ for this pair by taking the difference of price spreads in 2006 and 2003, after doing the same calculation for price spread in 2006. For this specific pair of DRGs, the spread of prices reduced by 90.43%. It is noteworthy that, even within this top 10 list, there is a considerable difference in variation of spread across DRG pairs.

Table 2.2: Source data example

Pair	spreadp_p03	spreadp_p06	spreadp_p	spreadw_p03	spreadw_p06	spreadw_p	Δspreadp	Δspreadw
	(€)			(units)			(%)	
32	5943.18	568.73	-5374.45	2.8	0.24	-2.56	-90.43	-91.33
35	1041.06	155.77	-885.29	0.49	0.07	-0.42	-85.04	-86.43
39	1236.18	2480.82	1244.64	0.58	1.06	0.48	100.68	82.60
52	1348.79	2091.04	742.25	0.64	0.89	0.25	55.03	39.48
87	1417.97	21.55	-1396.42	0.66	0.01	-0.65	-98.48	-98.61
88	988.95	56.92	-932.03	0.46	0.02	-0.44	-94.24	-94.72
98	637.54	1050.09	412.55	0.3	0.45	0.15	64.71	49.43
107	184.84	274.29	89.45	0.09	0.12	0.03	48.39	30.11
111	1573.79	186.21	-1387.58	0.74	0.08	-0.66	-88.17	-89.26
140	1777.47	2533.75	756.28	0.84	1.08	0.24	42.55	28.77

Source: *Portaria* 189/2001 of March 9^{th}, effective April 1^{st};
Source: *Portaria* 132/2003 of February 5^{th}, effective March 1^{st};
Source: *Portaria* 567/2006 of June 12^{th}, effective August 1^{st}.

The mean age of patients in the pair of DRGs, m_age_pair, was calculated as the mean age of all patients coded with the DRGs that correspond to each pair, for each pair of DRGs, hospital and year. For the mean age of patients in the DRGs with and without complication of each pair, m_age_cc and m_age_sc was calculated following the same logic. Furthermore, other auxiliary variables that control for the percentage of female and number of days that patients stayed in the hospital were also calculated. The number of discharged patients per year (dp), occupied beds ($beds$), case mix index (cmi), length of stay ($lstay$) and $cost$, the hospital characteristics variables, were obtained from hospitals or ACSS reports.

The dependent variable $fraction_{pit}$ is the percentage of the total patients coded

[11] DRG prices are described in the web appendix at https://sites.google.com/site/giseletbraun/research.

in the DRG with complications in each pair of DRGs, by pair of DRG, hospital, and year:[12]

$$fraction_{pit} \equiv \frac{\#\text{patients in top code DRG}_{pit}}{\#\text{patients in top code DRG} + \#\text{patients in bottom code DRG}_{pit}} \quad (2.7)$$

where p indexes the pair of DRG, i indexes the hospital, and t indexes the year. The relevant equation to be estimated and variants of this equation that will be detailed in the next section, all follow the form:[13]

$$\begin{aligned} fraction_{pit} = \alpha + \psi_1 spreadp_p \times post + \beta_1 m\_age\_cc_{pit} + \beta_2 m\_age\_sc_{pit} \quad (2.8) \\ + \phi_1 p\_female\_cc_{pit} + \phi_2 p\_female\_sc_{pit} + \lambda X_i + \gamma H_i + \delta year_t + \epsilon \end{aligned}$$

where the ψ vector measures the marginal effect of changes in the variation of price spread. The β_1 and β_2 measure the impact of patient's mean age in the DRGs that belongs to a pair (same logic for ϕ). They also work as a control for the (possible) increase in severity of cases over time. The variable *post* is a dummy variable that equals to zero if the observation is from 2003 to 2005 (the new prices are not applicable for these years), and equals to one otherwise (from 2006 to 2008).
Hospital characteristics coefficients (λ) of *dp*, *beds*, *cmi*, *lstay* and *cost* have different relations with the upcoding issue. Discharged patients, length of stay and cost may or may not be directly associated with upcoding. Whenever upcoding does not imply a different course of action but merely a distinct classification code (and the corresponding payment), the observed number of discharged patients, length of stay and costs will be independent of the extent of upcoding. On the other hand, if classification of patients into a different DRG leads immediately to another treatment protocol, yearly average costs per patient treated at the hos-

[12] In the econometric analysis, the $fraction$ is multiplied by 10^5

[13] For the DRG weight analysis (the results are available upon request), the relevant equation is:

$$\begin{aligned} fraction_{pit} = \alpha + \psi_2 spreadw_p \times post + \beta_1 m\_age\_cc_{pit} + \beta_2 m\_age\_sc_{pit} + \phi_1 p\_female\_cc_{pit} \\ + \phi_2 p\_female\_sc_{pit} + \lambda X_i + \gamma H_i + \delta year_t + \epsilon \end{aligned}$$

pital level are positively associated with the extent of upcoding. The number of beds that a hospital has is usually defined in the construction phase, or posterior adjustments take place at spaced intervals of time. It is unlikely to be correlated contemporaneously with upcoding. A different situation occurs for the *cmi*. The computation of this index is based on the DRG episodes of each hospital. Systematic upcoding also creates an upward pressure in the *cmi*. However, the final index value is computed normalizing the average national value to the unit value. Common upgrading to all hospitals does not necessarily change the relative position of each hospital. Nonetheless, it potentially is an endogenous variable.

The δ vector of coefficients can be interpreted as the mean impact of the price changes on all pairs (similar interpretation to γ).

Estimation of the coefficients is done using panel data, considering fixed-effects at the level of DRG pairs (hospital fixed-effects were also computed). To correct for potential autocorrelation and clusters, robust standard errors clustered at DRG pairs were computed.

A different analysis is performed to assess the role of age in the potential upcoding behavior. Suppose that older patients are on average more complex cases and remember that upcoding means moving some of the more complex cases from the lower DRG to the top DRG. Given the association of age with case complexity, this means moving some of the older patients from the bottom DRG to the top DRG, in which they will belong to the youngest within that DRG. Thus, average age of patients in both DRGs decreases and there is upcoding, under the assumption of positive association between case complexity and age of patient. In other words, we want to ask the following question: "do hospitals upcode taking into account the age of the patients?" Hence, we define:

$$d\_age\_cc_{pit} = m\_age\_cc_{pit} - m\_age\_pair_{pit} \tag{2.9}$$

$$d\_age\_sc_{pit} = m\_age\_sc_{pit} - m\_age\_pair_{pit} \tag{2.10}$$

where the dependent variables measure the difference between the mean age of patients coded in the DRG with complications (without complications) and the mean age in the pair of DRGs, for each pair, hospital and year. These two variables

are then used to estimate the coefficients of the following equations:

$$d\_age\_cc_{pit} = \alpha + \lambda X_i + \gamma H_i + \delta year_t + \epsilon \tag{2.11}$$

$$d\_age\_sc_{pit} = \alpha + \lambda X_i + \gamma H_i + \delta year_t + \epsilon \tag{2.12}$$

where the years range from 2001 to 2008.

2.6 Results

2.6.1 Initial results

The estimates show no statistical significance of the spread variation on the share of patients in the top DRG (model 1) at 5% significance level, even though there is statistical significance at the 10% level. The overall average effect cannot be interpreted as clear effect from upcoding, though the dummy variables after 2005, which coincide with the years of price changes, are statistically significant.[14] The estimated coefficient of spread has positive sign, implying that larger price changes are associated with a stronger share of patients in the top DRG.

Table 2.3: Partial estimation results: basic models

Ind. var.	model 1	model 2	model 3	model 4	model 5
spreadp_post	0.643*	0.636*	0.557*	0.829***	0.831***
	(0.335)	(0.335)	(0.317)	(0.303)	(0.303)
m_age_pair	595.9***	601.5***	537.6***		
	(34.29)	(34.23)	(33.08)		
p_female_pair		-3,332***	-2,715**		
		(1,223)	(1,201)		
totdays_pair			826.9***		
			(80.77)		
m_age_cc				90.79***	89.44***
				(18.21)	(18.30)
m_age_sc				142.9***	143.0***
				(22.68)	(23.05)
p_female_cc					466.1

[14] See the complete result in the appendix.

					(546.6)
p_female_sc					103.2
					(789.3)
Constant	21,325***	23,045***	17,617***	32,488***	32,180***
	(5,896)	(6,009)	(5,939)	(7,143)	(7,161)
Obs	29,747	29,747	29,747	23,754	23,754
R-squared	0.166	0.167	0.214	0.152	0.152
Number of pair_drg	112	112	112	112	112

Dependent variable: fraction100.

Robust standard errors in parentheses.

Year dummies and fixed effects at DRG pairs.

*** p<0.01, ** p<0.05, * p<0.1

The number of discharged patients, beds and cost are negatively correlated with the fraction of patients in the top DRG, whereas the opposite is true for case-mix index and length of stay in most of the models.[15] It is worthwhile to highlight the coefficient sign associated with cost, which is negative, meaning that higher-cost hospitals tend to have a smaller fraction of patients coded in the DRG with complications than lower-cost hospitals.

Model 2 is the same as model 1 with the addition of gender control. There is still no evidence of upcoding. The same applies in model 3, where the variable *totdays_pair*, which accounts for the mean number of days stayed in the hospital, is included. This variable has the potential to be endogenous, considering that patients coded in the top DRG, with complications, are supposed to stay more days in the hospital. Taking this potential endogeneity into consideration, plus the fact that the results remain unchanged in model 3, the variable *totdays_pair* will not be used further.

Different from Dafny (2005), the age of the patients has a very important role in the upcoding process. From graphical analyses, which will be done later in the section, it is observed that the mean age of patients in the top and bottom DRGs of a pair have different patterns that must be considered. Model 4 then has the mean age of patients in the top and bottom DRGs, for each pair of DRGs, as explanatory variables. The estimation shows evidence of upcoding behavior as the

[15]Larger hospitals are the ones that have a higher technological differentiation that may not be fully captured by the case-mix index.

coefficient of the variable *spreadp_post* is statistically significant.[16] In model 5, we also add the percentage of female in the top and bottom DRGs of each pair and the results remain roughly equal.[17] Referring back to equation 2.2, it becomes clear that age is an important driver of proportion of patients in the top DRG of a pair.

It is worthwhile to offer an example of the impact of the spread of prices in the share of patients coded in the DRG with complications. Consider the specific pair of DRGs (pair 22) that refers to simple pneumonia and pleurism, with and without medical complications. Suppose that the Government set hypothetical new DRG prices that results in €100 increase in spread, which induces upcoding behavior.[18] In this particular case, it was found that 20 extra episodes would be included in the top DRG, in one year, with associated upcoding cost of €12,045.

Table 2.4: Upcoding size and cost - the case of the DRG pair 22

	t=0	t=1		
Mean number of cases (from sample, by year)	24,835	24,835	Upcoding (number of cases)	20
Marginal effect of price spread	0.831	0.831	Top DRG price (2006)	€1,742
Fraction100	0.1893	0.1901	Bottom DRG price (2006)	€1,139
N of cases in the top DRG	4,701	4,721	Spread of prices	€602
N of cases in the bottom DRG	20,134	20,114	Cost of upcoding	€12,045

Mean number of cases, $24,835$, is the total number of episodes in DRG pair 22, $198,681$, divided by 8, the number of years.

Fraction100 at t=0: 0.1893, the sample value.

Fraction100 at t=1: 0.1901, from $0.1893 + 0.831 \times \frac{10^2}{10^5}$.

€100 (10^2) is the hypothetical increase in spread.

The 10^5 is the number by which the observed fraction was transformed for the regressions.

As previously mentioned, the distinction between pure public hospitals, SPA, and more private-management like hospitals, EPE, is potentially important. We test for it allowing time dummies and the marginal incentive to upcode to be different according to the statute type of the hospital. The two groups of hospitals do not

[16]The estimations were performed considering the fixed effects at the level of DRG pairs. Robustness check altering the fixed effects from the DRG pairs to hospital level reveals the strenght of the results, that remain mainly unchanged. The full set of regressions are available upon request.

[17]The analysis using *spreadw* (the changes in weights of the DRGs) presents results that are similar to the price changes and they are available at https://sites.google.com/site/giseletbraun/.

[18]The €100 was defined in a way to represent an almost 10% increase in the average spread of prices in 2006, which was €1,128 (this average was calculated from the difference between the top and bottom DRGs in all 112 DRG pairs included in the sample).

behave differently in the upcoding matter. In the model, there is an intercept term for EPE hospitals, and the variable *spread_postEPE* allows the measure of marginal incentive for upcoding to be different between hospital statutes. None is statistically significant. Therefore, we conclude that there is no difference in this aspect between the two sets of hospitals. The incentive for upcoding is, according to these results, independent of the juridic nature of hospitals.[19]

Other potential changes in the upcoding results were analyzed. In particular, we considered the possibility of upcoding under the distinction of having medical or surgical pair of DRGs, and also the possibility of upcoding subjected to two different scenarios where there is negative or positive spread of prices. The possibility of asymmetric exposure to DRG price change could be relevant, with upcoding appearing more easily with DRG price increases (the DRG price decrease should be linked to a "downcoding"). There was no modification in the main conclusions. These factors were not statistically significant.[20] They constitute robustness checks.[21]

Another important question relates to the cases of patients' transfer between hospitals. We constructed an index to measure the complexity of the cases transferred in and out of each hospital and the received cases (transfer in) showed to be statistically significant, meaning that the higher the complexity of cases received by a hospital the higher is the share of patients coded in the top DRG, which was naturally expected.[22]

2.6.2 A longer term view

An important issue to be addressed is whether the current upcoding trend is a long term historical trend or it started with the price changes of DRGs in 2003. Previously, we addressed only the 2006 price change, which may just reinforce a

[19] The partial result is presented in the appendix.

[20] The exception is the model that accounts for differences between the type of DRG, being it a medical or surgical one. In this case, the upcoding effect was not identified.

[21] Another model specification, with results not presented here, allows for different hospital upcoding behavior depending on the size of activity. There was no change in the results as the high and low volume hospital regressions show evidence of upcoding without remarkable differences between them.

[22] Results are available upon request.

change that already resulted from a 2003 price change.
Using the simplest regression models, without the impact of DRG price changes and adding two initial years to the previous regressions, we find that years 2001-2003 seem to be somewhat different than the others. The same behavior is observed for the models that include the DRG price changes plus two initial years. The magnitude of the coefficients are bigger after 2003, but much bigger after 2006, coinciding with the time of DRG price changes. Hence, the first price changes that occurred in 2003 may have created some incentives for upcoding, but another shift seems to take place in 2006, with the latest change in DRG relative prices.

Table 2.5: Long term impact: year dummies

Year	model 4 - *spreadp_post*	model 5 - *spreadp_post*	model 4	model 5
2001	reference	reference	reference	reference
2002	559.4	560.0	554.1	554.6
2003	831.6	827.9	834.6	830.8
2004	1,099	1,098	1,099	1,097
2005	1,070	1,069	1,062	1,060
2006	1,755	1,756	2,393	2,393
2007	1,978	1,975	2,618	2,614
2008	2,314	2,313	2,958	2,956

2.6.3 Age effects

Age can be a discriminating variable that is associated with episode complexity, as previously argued. To assess its role in the upcoding, we need to consider the average age in the top DRG and the average age in the bottom DRG.
Figure 2.1 shows the growth rate of the average age in the pairs of DRGs. Even at this aggregate level, a couple of empirical facts are worth noting. First, the evolution before and after 2006 seems to be distinct. In the first years of the sample, average age within each pair of DRGs is increasing but at an intermediate rate between the top DRG growth rate of patients' average age and the bottom DRG growth rate. However, since 2006, we observe a clear shift toward average

age at the DRG-pair level increasing faster than in the bottom and the top DRGs (except for 2008). This can only result from a composition change, characterized by a slower growth in average age of both DRGs caused by older patients in the bottom DRG (which has a lower average age) moving to the top DRG (where they will be relatively young patients).

Figure 2.1: Growth rate of average age

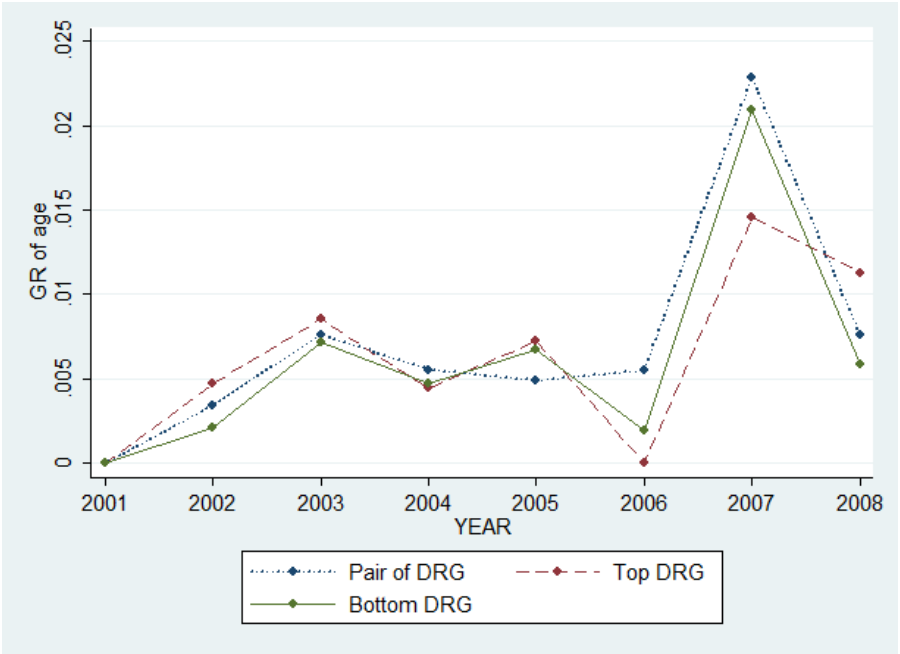


Table 2.6: Average age of patients

Year	DRG pair	Top DRG	Bottom DRG
2001	56.44	62.06	53.56
2002	56.64	62.36	53.68
2003	57.07	62.89	54.06
2004	57.39	63.17	54.32
2005	57.67	63.62	54.68
2006	57.98	63.63	54.79
2007	59.31	64.55	55.93
2008	59.76	65.28	56.26

If age is associated with high complexity and it is used as a lead signal to how severe it will be the case, a change in the criterion for inclusion of younger ages in the top DRG would result in an testable empirical prediction. In case that upcoding takes more patients but less complicated cases to the top DRG, and assuming age to be positively correlated to severity of the case, then we should observe average age decreasing in both DRGs of the pair. The alternative hypothesis is that the patient mix has worsen over time, implying that average age of patients increases over time in both DRGs of each pair, and in a way unrelated to the change in DRG prices.

Thus, the two hypotheses have different implications to the coefficient of interest. We are interested in the overall trend of average age in the top and the bottom DRGs of the pairs, and in the marginal effect associated with the magnitude of the price change. The hypothesis of worst cases leads to a positive trend over time in both top and bottom DRG and statistically non-significant coefficient of price change. The upcoding hypothesis implies a negative trend of the mean age in the top DRG and a positive trend in the bottom DRG, reinforced by price sensitivity - a more pronounced effect is expected when relative prices change more.

In the appendix, table 2.10 shows the estimates related to the role of age. Taking first the results for the top DRG within each pair, we observe that the difference is negative for all years with exception of 2002, though it is statistically non-significant in most of the cases. Statistically significant effects occur only from 2006 onwards, with the average age in top DRG decreasing relative to the average age in pair. For the bottom DRG, a different picture emerges, as the difference over time decreases only in 2002 and 2006, but being statistically non-significant in all years. It means that average age within the bottom DRG is evolving at a smaller pace than average across the pairs of DRGs in 2002 and 2006 (the opposite happens for the remaining years).[23]

Taking both results together, we have support to the basic fact evidence in figure 2.1. Such evidence is consistent with the presence of upcoding.

[23]The regressions including the change in relative prices from 2003 to 2008 do not improve the results. They are available by request.

If the role of age were not to provide conclusive evidence of upcoding, another variables that try to capture the severity of the disease could be investigated in the same way. We could have worked with number of medical diagnoses or procedures. At first, it is expected that patients coded in the bottom DRG present smaller number of medical diagnoses or procedures. However, it could be the case that these variables are taken into consideration when deciding to upcode a patient. This would contribute to hide evidence of upcoding in the data.[24]

2.6.4 Upcoding size and cost

We estimated the total upcoding size and cost, considering the same hypothetical increase of €100 in the DRGs spread of prices, as done for DRG pair 22, this time for the full set of episodes. The results are summarized in the next table.

Table 2.7: Upcoding size and cost - all DRG pairs

	t=0	t=1		
Mean number of cases (by year)	384,615	384,615	Upcding (number of cases)	308
Marginal effect of price spread	0.831	0.831	top DRG price (2006)	€3,366
Fraction100	0.3797	0.3895	bottom DRG price (2006)	€2,138
N of cases in the top DRG	146,038	146,346	spread of price	€1,228
N of cases in the bottom DRG	238,577	238,269	Cost of upcoding	€377,815

Mean number of cases, $384,615$, is the total number of episodes in all DRG pairs, $3,076,921$, divided by 8, the number of years.

Fraction100 at t=0: 0.3797, the sample value.

Fraction100 at t=1: 0.3895, from $0.3797 + 0.831 \times \frac{10^2}{10^5}$.

€100 (10^2) is the hypothetical increase in spread.

The 10^5 is the number by which the observed fraction was transformed for the regressions.

An increase of €100 in the spread of prices would lead to a bigger fraction of patients to be coded in the top DRG. The mean value of the share of patients coded in the top DRG would change from 37.97% to 38.05%. Multiplying the mean number of DRG episodes in the year by these shares and taking the difference between the two values gives the total number of upcoded cases that would be coded in the top DRG, following the incentive of price spread increase.

[24] Another consideration to deepen the investigation of upcoding would be to use the chronic conditions of the patients as an incentive for upcoding. In this case, the unit of observation would need to be the patient and a probabilistic model of whether the patient would be coded in the top DRG or not could be defined. The data used here does not allow for this type of analysis, but we mention it as a suggestion.

Supposing that all hypothetical upcoded cases would have a cost that equals the mean cost of all top code DRG prices, €3,366, the upcoded cases would represent an extra cost of €377,815 in a given year.

2.7 Final remarks

Previously, it was found that US hospitals were quick to react to DRG price changes, and to profit considerably from upcoding (Dafny, 2005). Classifying patients in those DRGs that allowed for a higher payment, and doing so more when the favorable price change was stronger, provided evidence for upcoding.
A natural issue is whether upcoding is specific to the US or it can be found in other health systems as well. This question is particularly important as many countries over the last decade introduced patient classification systems. Purchasers of health care, being them Governments (through NHSs or sickness funds) or health insurance companies, are increasingly using patient classification systems for providers' payment.
We use data from a NHS with the following features: demand to each hospital is basically exogenous, as patients have to comply with Government-defined catchment areas for each hospital; hospitals have to classify patient episodes into a DRG-like system; hospitals are not paid on an episode-by-episode basis but yearly budgets have been increasingly based on the DRGs and the DRG mix the hospital provides; DRG prices are set by Government ruling and have infrequent changes, and the background studies supporting the new prices are not known to hospitals. Thus, price changes can be seen as exogenous from the point of view of hospitals.
Our results provide a mild support for upcoding, which can be described by an increase in the share of top DRG within pairs of DRGs. It has increased over time (2001-2008), and more so in recent years. Moreover, not only upcoding has been occurring above what would be predicted by the simple ageing of population (assumed to be captured by the average age of patients), but has been more important when the price change was stronger. This points to the conclusion that even within national health services, management of hospitals does respond to incentives for upcoding patients.
Given such result, it is recommended another review of prices and weights of the

DRGs to try to avoid such harmful scheme of upcoding, which is not consequence of a sicker population but a way to increase budget from Government to hospitals. In any case, it is not said that the money coming from upcoding is not being invested to increase quantity or quality of services in the health care units. However, a clear necessity of extra budget to cover the expenditure of health care must be negotiated explicitly between hospitals and third-party payers in order to avoid unequal access of the population to health care, which is quite possible (assuming that money from upcoding is been reverted to the population).

Appendices

Basic regression models

Table 2.8: Estimation results: basic models

Ind. Var.	model 1	model 2	model 3	model 4	model 5
spreadp_post	0.643*	0.636*	0.557*	0.829***	0.831***
	(0.335)	(0.335)	(0.317)	(0.303)	(0.303)
m_age_pair	595.9***	601.5***	537.6***		
	(34.29)	(34.23)	(33.08)		
p_female_pair		-3,332***	-2,715**		
		(1,223)	(1,201)		
totdays_pair			826.9***		
			(80.77)		
m_age_cc				90.79***	89.44***
				(18.21)	(18.30)
m_age_sc				142.9***	143.0***
				(22.68)	(23.05)
p_female_cc					466.1
					(546.6)
p_female_sc					103.2
					(789.3)
dp	-0.373***	-0.379***	-0.319**	-0.194*	-0.193*
	(0.128)	(0.128)	(0.127)	(0.116)	(0.116)
beds	-1.513	-1.435	-2.502	0.341	0.327
	(2.340)	(2.320)	(2.157)	(2.084)	(2.082)
cmi	3,431*	3,380*	4,350**	1,692	1,704
	(1,997)	(1,984)	(1,992)	(1,601)	(1,601)
lstay	316.3	301.7	-67.42	-38.71	-35.13
	(311.7)	(310.9)	(313.5)	(221.4)	(221.4)
cost	-535,532*	-549,635*	-402,478	-94,002	-91,031
	(285,172)	(284,688)	(288,626)	(244,791)	(245,609)
yeard4	581.5	615.2	676.8	619.5	618.5
	(483.3)	(483.8)	(498.6)	(384.4)	(384.5)
yeard5	254.9	298.1	362.9	580.7	579.0
	(544.8)	(541.6)	(553.1)	(427.7)	(427.5)
yeard6	1,610***	1,639***	1,661***	1,822***	1,824***
	(607.6)	(606.9)	(602.9)	(499.2)	(499.6)
yeard7	1,609***	1,643***	1,908***	2,188***	2,186***
	(581.9)	(583.1)	(569.5)	(522.8)	(524.3)

Continued on next page...

... table 2.8 continued

yeard8	2,098***	2,137***	2,316***	2,541***	2,536***
	(733.9)	(731.0)	(690.5)	(595.7)	(596.0)
Constant	21,325***	23,045***	17,617***	32,488***	32,180***
	(5,896)	(6,009)	(5,939)	(7,143)	(7,161)
Obs	29,747	29,747	29,747	23,754	23,754
R-squared	0.166	0.167	0.214	0.152	0.152
Number of pair_drg	112	112	112	112	112

Dependent variable: fraction100.
Robust standard errors in parentheses.
Fixed effects at DRG pairs.
*** p<0.01, ** p<0.05, * p<0.1

Alternative regression models

Table 2.9: Estimation results: alternative models

Ind. Var.	EPE - SPA	Med - Surg	Asymmetry
spreadp_post	0.803***	0.598	0.907**
	(0.301)	(0.951)	(0.369)
spreadp_postEPE	0.0301		
	(0.432)		
surgical_spreadp		0.262	
		(0.964)	
upside_spreadp			-0.643
			(1.400)
vol_spreadp			
Constant	30,957***	32,205***	32,191***
	(7,236)	(7,154)	(7,162)
Obs	23,754	23,754	23,754
R-squared	0.153	0.152	0.152
Number of pair_drg	112	112	112

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Year dummies and fixed effects at pair_drg.

The analysis of age

Table 2.10: Estimation results: age effects models

Ind. Var.	d_age_cc	d_age_sc
2002	0.0813	-0.124
	(0.139)	(0.0877)
2003	-0.134	0.00487
	(0.164)	(0.105)
2004	-0.114	0.118
	(0.190)	(0.143)
2005	-0.226	0.0451
	(0.204)	(0.146)
2006	-0.366*	-0.114
	(0.213)	(0.170)
2007	-0.586***	0.00391
	(0.218)	(0.178)
2008	-0.512**	0.0998
	(0.206)	(0.188)
Constant	4.415***	-1.532
	(1.139)	(1.439)
Obs	38,371	41,803
R-squared	0.032	0.034
Number of pair_DRG	112	112

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Bibliography

Carter G, Newhouse J, Relles D. 1990. How much change in the case mix index is DRG creep? *The RAND Corporation*: R-3826-HCFA.

Carter G, Ginsburg P. 1985. The Medicare case mix index increase: medical practice changes, aging and DRG creep. *The RAND Corporation* : R-3292-HCFA.

Dafny L. 2005. How do hospitals respond to price changes? *The American Economic Review* **95**(5) : 1525-1547.

Hsia D, Krushat W, Fagan A, Tebbutt J, Kusserow R. 1988. Accuracy of diagnostic coding for Medicare patients under the prospective payment system. *The New England Journal of Medicine* **318**(6) : 352-355.

Kroneman M, Nagy J. 2001. Introducing DRG-based financing in Hungary: a study into the relationship between supply of hospital beds and use of this beds under changing institutional circumstances. *Health Policy* **55**(1) : 19-36.

Kuhn M, Siciliani L. 2013. Manipulation and auditing of public sector contracts. *European Journal of Political Economy* **32** : 251-267.

Lave J. 1985. Is compression occuring in the DRG prices? *Inquiry* **22**(2) : 142-147.

Psaty B, Boineau R, Kuller L, Luepker R. 1999. The potential costs of upcoding for heart failure in the United States. *American Journal of Cardiology* **84**(1) : 108-09.

Rauner M, Schaffhauser-Linzatti M. 2002. Impact of the new Austrian inpatient payment strategy on hospital behaviour: a system-dynamics model. *Socio-Economic Planning Science* **36**(3) : 161-182.

Silverman E, Skinner J. 2004. Medicare Upcoding and Hospital Ownership. *Journal of Health Economics* **23**(2) : 369-89.

Silverman E, Skinner J, Fisher E. 1999. The Association between for-profit hospital ownership and increased Medicare spending. *The New England Journal of Medicine* **341**(6) : 420-26.

Steinbusch P, Oostenbrink J, Schaepkens F. 2007. The risk of upcoding in casemix systems: a comparative study. *Health Policy* **81**(2-3) : 289-299.

Steinwald B, Dummit L. 1989. Hospital case-mix change: sicker patients or DRG creep? *Health Affairs* **8**(2) : 35-47.

Xirasagar S, Lin H. 2006. Effects of payment incentives, hospitals ownership and competition on hospitalization decisions for ambulatory surgical procedures. *Health Policy* **76**(1) : 26-37.

Chapter 3

Does it last? Effects from a public policy to recover waiting lists

Abstract

Waiting lists generate policy interest in several countries. We often observe specific public programmes directing funds in an attempt to solve the problem. Incentives are a key element. Using data from Portuguese public hospitals we show that temporary funds, in a time and budget-delimited programme, had temporary effects on supply the side of health care.

The number of patients on the waiting list for first ophthalmology appointment and subsequent surgery has increased over the years in the Portuguese public health system. The ophthalmology intervention programme was designed to reduce patients' waiting list and waiting time to receive cataract and other eye-related surgeries. The programme induced an increase in the supply of medical procedures and a reduction in the waiting time to receive care during the implementation of the programme. It also produced a longer term reduction in the waiting time after the intervention policy.

3.1 Introduction

Managing waiting lists is of high concern in several countries with public health service. Waiting time to receive medical care can be quite long. For instance, waiting time for surgical procedures such as hip and knee replacement can be as long as six months in Spain.[1] Waiting lists are important to evaluate a national health system (NHS) as a whole. They are easily judged by the public service users and reflect results from the combination of investment, productivity and level of patients' health. Short-term programmes aimed at reducing waiting lists by funding extra-activities, have been commonly employed in many European countries with NHS. Portuguese examples of this kind are the *Programa Específico de Resolução de Lista de Espera*, *Programa de Promoção de Acesso*, and *Programa Especial de Combate às Listas de Espera Cirúrgicas*. These programmes were implemented in the late 1990s and early in the new century but did not reduce the size of the waiting lists.

Efforts to enhance the Portuguese NHS have continued in more recent years. In 2004 an information system was created, the *Sistema Integrado de Gestão de Inscritos para Cirurgia* (SIGIC), with the goal of improving management of waiting times for consultations and surgical procedures. The indicators related to ophthalmological care in Portugal show that its quality is poor when compared with that of some other developed countries. In some cases, patients wait several months, or even years, to receive health care.

Reasons for shortage of supply derive from several sources: scarcity of physicians, low levels of productivity, and bad management are the main factors on the supply side; whereas ageing population associated with a greater awareness of the technological developments and facilities available to correct sight problems are responsible for the constant increase in the demand for health care.

The "ophthalmology intervention programme" (PIO) was designed by a group of specialized doctors and professors, selected by the Ministry of Health, to implement actions with the goal of reducing the size of waiting list and waiting time for ophthalmological care. It is a funding plan with incentives for extra consultations

[1] Siciliani et al. (2013).

and surgical procedures, as with the programmes mentioned before, inserted into a more informative scenario created with the SIGIC. The results show evidence of the programme's success to increase volume of activity during the time that the PIO was carried out. Waiting times to receive care were also influenced by the PIO, with a reduction observed during and after the programme.

A potential solution to address the excess of demand is to increase the supply (new Portuguese or foreign physicians). Another possibility is the rearrangement of human resources and installed capacity, offering the right incentives to increase performance and ensuring the level of quality in the services provided.

Following an analysis of these alternatives, which have the objective of improving care in the ophthalmology specialty, it was decided and decreed in *Despacho* 20639/2008 to create the PIO, which is the subject of evaluation of this study.[2] The programme seeks to increase public access to ophthalmological first consultations and surgical procedures through agreements between NHS and hospitals to perform extra services that will be paid for with additional PIO funds.

Cataract is an ophthalmology condition that can drastically reduce the patients' quality of life, and can lead to blindness if not adequately treated. According to *Portaria* 1306/2008, the NHS performed around 600,000 medical appointments and 60,000 surgeries in the year 2007, which resulted in a waiting list for cataract surgery of about 30,000 people and mean waiting time of 3.8 months on December 31^{st} of the same year.[3]

There are two publicly available evaluations of PIO's performance. One was conducted by *Administração Central do Sistema de Saúde* (ACSS) and the other by the *Tribunal de Contas* of Portugal.[4] The analysis of ACSS is centred on the final performance of the hospitals (volume of activity) after the PIO's execution. They concluded, roughly speaking, that the programme was successful. However, the *Tribunal de Contas* evaluated the programme as not satisfactory in reaching the goals it established.[5] Even with these assessments, there is insufficient economic

[2] Implementation of the PIO also came in response to public pressure.

[3] *Portaria* is an administrative act that has the objective of explaining a law's application in detail.

[4] *Tribunal de contas* is a court that is responsible for the control of the Government's public accounts.

[5] PIO's objective was 30,000 extra cataract surgeries and 75,000 extra first consultations. The hospitals agreed on performing 21,055 extra surgeries and 48,075 extra first consultations, but

evaluation of the PIO.
Our contribution in this paper is to offer a measure of the programme impact, after controlling for unknown factors that might have also induced its success or failure.
The results reveal a statistically significant impact of the PIO on the volume of first consultations and eye-related surgical procedures and also on the mean waiting time for these health care services. Furthermore, there is a longer term impact of the programme on mean waiting time to receive medical care.
This paper is organized in 6 sections. Section 2 gives the contextualization of the programme in time and the literature review on the subject. Section 3 describes the PIO with details of implementation and accomplishments. Section 4 explains the methodology used. Section 5 shows the results and Section 6 presents the final remarks.

3.2 Portuguese scenario and other countries' experiences

According to the Portuguese National Health Plan 2004-2010, there was an increase in the incidence of eye-related diseases and during this period no adequate measure was taken to reverse this trend.[6] The plan also highlighted the need for strategic orientation and interventions, which were implemented through the PIO after the release of the publication.
In general, intervention programmes can be classified as horizontal or vertical, as defined in Sena et al. (2006). Horizontal programmes usually connect services offered at the same level. The strategies of interventions tend to be designed for the long run and are usually implemented in countries with more stable Governments.[7]
In contrast, vertical programmes are designed to address a specific problem, establishing direct actions to provide health care, and are more frequent in countries

execution rates were 59% for extra surgeries and 41% for extra first consultations. See Auditoria PIO (2010).

[6]The complete report of the plan can be found in PNS (2004).

[7]The connection of family plan and vaccine programme is an example of horizontal programme.

with less stable Governments and more financial problems and, for this reason, have shorter scope and time of execution.[8]

The literature reports ways to match supply and demand of eye-related medical care. On the supply side the main mechanisms applied are increasing productivity of public hospitals, increasing capacity, and the change of incentives. In order to manage demand size, the options include prioritization of patients with greater need and encouragement of private health insurance through subsidization.

There are several problems that should be considered when trying to increase the supply of medical services. If productivity is increased through the funding of extra activity, dynamic problems may arise: the perception of higher productivity could lead to higher demand. Increasing supply by increasing the number of physicians takes time and may not be successful. In fact, some years are needed to educate them, and the success of this measure also depends on the behaviour of the demand during this period. Another, more immediate option, is to expand supply by sending patients to receive medical care in the private sector and paying for their treatment. This measure has the additional advantage of increasing competition between public and private hospitals. Its success in addressing the problem will depend on whether or not supply from public and private hospitals combined is enough to cover total demand.

The strategy most widely used in European countries to shorten waiting times is to increase supply through increased activity inside hospitals, as explained in Hurst and Siciliani (2003) and Siciliani et al. (2013). However, this type of measure does not appear to be effective in the long run and waiting times are still a concern in some NHS.

The case of Spain has similarities to Portugal's situation. Their NHS is publicly funded, past expenditure and case mix index are used to define the financing of the system, and waiting times are more or less equal to those in Portugal. A difference is that patients are free of charge at the point of delivery of care, which reduces the possibility of solving the problem through demand-side. The main strategies adopted in Spain were imposing a "guarantee of maximum waiting time", a measure which had no impact since the limit was over the maximum

[8] An example of vertical plan is the coordination of actions between hospital level and primary care level to control a given disease.

observed in the hospitals for most of the cases; "increase in working hours", which failed partially because of reduced productivity of physicians; "massive referrals to private providers", which increased the expenditure dramatically and could not be sustained in the long run; "specific funding for procedures with highest waiting time", which led policy makers to concentrate efforts on one part of the problem only and did not provide incentives to solve the general system problem; and the development of "out-patient care centres". This latter measure presented visible benefits in reduction of waiting times, but also attracted more patients to the NHS, which increased waiting lists. Another difference is that the political and administrative division of Spain into autonomous communities reduces incentives for the implementation of a national health plan, as in Portugal.

Italy is another European country that is working to reduce waiting lists inside on NHS. Its decentralized system, with 19 regions and 2 autonomous provinces, shows geographical disparities that are reflected in the health care of each part of the country. Hospital remuneration is achieved through a prospective and activity-based system for in-patient and out-patient care. For in-patient care remuneration is determined according to diagnosis-related groups (DRG) and for out-patient care, by a fee-for-service basis. Physicians have a base salary that is augmented with performance-related payments. These characteristics contribute to the problem of long waiting lists and waiting times with the exception of in-patient care, which has waiting lists at acceptable levels. Out-patient care and diagnostic tests, on the other hand, have waiting lists that need to be reduced.

The direct policies adopted in Italy to address these problems were: "maximum waiting time target", which had no real impact because there was no control or penalties for failure; "data collection improvement"; the creation of a "unified booking centre"; measures that could be categorized as "demand-side policies" (priority groups definition, diagnostic and therapeutic pathway, penalties for no-show patients); measures that could be categorized as "supply-side policies" (increase of supply capacity by specific agreements, purchase of extra visits, and tests from private providers). Some of the indirect policies adopted include co-payments, *intramoenia* private practice, and promotion of voluntary health insurance coverage.[9] Finally, the country released its National Health Plan 2010-12, a

[9] *Intramoenia* is the private health care provided in public health structures.

plan similar to the one designed in Portugal. This plan included some practices to reduce waiting lists such as updates on maximum waiting times, levels of priority and respective waiting times, and efforts to improve the collection and publication of waiting times data.

In Portugal, the measures adopted in the specific plan to reduce ophthalmology waiting lists are not an innovation at all, given that they were implemented in one form or another in countries such as Spain and Italy. The PIO could be classified as a vertical programme with supply-side interventions through the use of extra-activity funding, combined with maximum waiting time guarantees (details in the next section).

The importance of evaluating public programmes, such as the PIO, is stressed in the literature, where the elaboration of evaluation reports is recommended . These reports are not always produced in practice.[10] In the case of the PIO, reports were produced but they do not benefit from the economic techniques used in this study.

3.3 The ophthalmology intervention programme

The size of waiting lists and waiting time is a serious problem in the Portuguese NHS. There are some specialties, such as ophthalmology, for which the indicators are not compatible with the quality of service that NHS is committed to provide. The mean waiting time for medical procedures comprehended in the programme fell sharply during the period analysed.[11] Similar data using only information from hospitals that formally adopted the programme were summarized in a report produced by the *Tribunal de Contas* and the same reduction in waiting time is observed.

The waiting time for eye-related surgery in 2006 was, on average, 5 months. By 2010 average waiting time had fallen to less than 2 months. External factors could have influenced this result, but still the intervention programme might have been the main driver of this improvement.

The intervention programme established a contractual agreement of 30,000 surg-

[10] See Sena et al. (2006).

[11] The mean waiting time was 153, 129, 89, 68, and 47 days in 2006 to 2010, respectively. This data is for complete cases only , that is, patients who entered and left the waiting list.

eries in addition to those set annually by hospitals, and of 75,000 extra medical appointments, from July 1^{st}, 2008 through July 1^{st}, 2009. This was stipulated in the *Portaria* 1306/2008.

The PIO is also called *Plano de Acesso à Cirurgia Oftalmológica* (PACO) and hereinafter the two terms are used interchangeably. The measures contemplated in the plan are: a) a contractual agreement to implement extra activities in each hospital; b) to increase the base production of the hospitals according to the installed capacity; c) to define some hospitals as top-performance centres for cataract surgery (*Centro de Elevado Desempenho* (CED)); and d) to include public hospitals in the CED. The medical procedures comprehended in the PIO are those with ICD9-MC codes that relate to ophthalmology procedures.[12].

The limit waiting time for cataract surgery was set at 5 months and 10 days. When the limit is reached, the patient must be admitted for surgery. If no appointment is assigned after 4 months (75% of the maximum time), the patient can be transferred, in the following order: first, for a *Hospital Público de Destino* (HPD) that belongs to the CED; second, to an HPD that does not belong to the CED; third, the surgery is paid for by the Government, to be performed in a private health unit that belongs to the network previously defined.[13] If the patient does not wish to be transferred, he is once more included in the waiting list and the waiting time resets.[14] For first medical appointment, the limit time is 2 months.

Table 3.1 describes the actions taken by the Government to implement the PIO in chronological order, as well as the evaluations conducted by ACSS and *Tribunal de Contas* and the results achieved by the programme.

Table 3.1: Regulation and evaluation of the PIO

Institution	**Type**	**Number**	**Date**	**Year**	**Goals**
Ministry of Health	*Order*	28478	Nov, 5^{th}	2007	assign a group to analyse the alternatives to reduce differences in supply and demand for consultations and surgeries of ophthalmology
Ministry of Health	*Order*	20639	Jul, 25^{th}	2008	create the PIO

continued on next page

[12]ICD9-MC refers to "International Classification of Diseases-Medical Codes"

[13]See an illustration in figure 3.4.

[14]Regulated by Ministry of Health, 2008, *Portarias* 45 and 1306.

Ministry of Health	*Administrative measure*	1306	Oct, 22^{th}	2008	define the dates, procedures, and specific rules of the PIO
ACSS	*Regulatory instruction*	10	Aug, 8^{th}	2008	establish the hospital's actions to record consultations and surgeries in the PIO
ACSS	*Regulatory instruction*	1	Feb, 11^{th}	2009	clarify the articles of the contract between Government and hospitals to perform the PIO
ACSS	Evaluation	1	Sep, 30^{th}	2008	goals reached in the first quarter of the programme
ACSS	Evaluation	2	Dec, 31^{st}	2008	goals reached in the first semester of the programme
ACSS	Evaluation	3	Dec, 16^{th}	2009	goals reached in the whole period of the programme
Audit Court	Report	48	Dec, 16^{th}	2010	evaluate the financial and social impact of the programme

Notes:

(a) extra medical appointment result by the end of the PIO: + 80,940;

(b) extra surgery result by the end of the PIO: + 36,446.

The analyses performed by ACSS can be classified as descriptive statistics and the one conducted by the *Tribunal de Contas* has a more administrative approach. Both are important but insufficient from the economic point of view to assess the health benefits from the public health spending.

The evaluation of the programme will be enriched with our analysis, which uses inferential statistics, given the problem of modelling variables that are subjected to random variations (this is the reason why the descriptive statistics alone could misrepresent the facts obtained from the data in the available reports). Another important contribution of our analysis comes from the fact that it is not an administrative view, which would usually concentrate the auditing effort on a pre-defined set of variables. We believe that policy decisions (such as the decision for a new intervention programme) would be better informed if there were an economic analysis that incorporated the randomness of the data.

3.4 Methodology

3.4.1 Data

The data used in this paper include information of DRG episodes and completed surgical procedures cases.[15] The national dataset of DRGs records medical episodes of all Portuguese public hospitals.[16] It was primarily designed for administrative purposes and the patients were de-identified before being used in the study. The information is provided by the hospitals and is organized by the *Administração Central do Sistema de Saúde, I.P.* (ACSS). Some disadvantages, such as lack of information regarding the waiting time to receive medical care, explain the necessity of using another dataset where this information is available. Completed surgical procedures cases are recorded in the national *Lista de Inscritos para Cirurgia* (LIC), also organized by the ACSS in the *Unidade Central de Gestão de Inscritos para Cirurgia* (UCGIC).

Since no patient identifier was provided, due to data confidentiality reasons, it has not been possible to merge the two datasets. Therefore, two parallel analyses have been implemented. On one hand, the volume of surgeries is evaluated by using information from the DRG dataset, which allows us to use records of surgeries at the disaggregated level of procedure codes. On the other hand, the study of waiting time is performed considering the DRG, a more aggregated level of information.

Since the analysis is performed across a heterogeneous set of hospitals, characteristics of each hospital unit were collected from regular reports from ACSS and/or hospitals. The information collected includes proxies to medical activity, complexity of the cases treated, installed capacity and cost (aggregation is annual for each hospital).

The sample size of the DRG dataset relevant to our objectives is 2,635 monthly observations of 68 hospital units from 2006 to 2010.[17] The selected procedures

[15]Complete cases are those in which there is information of both the date that a patient was indicated for a surgical procedure and the date of the procedure itself.

[16]DRG definition and Portuguese DRGs are discussed in Busse et al. (2011).

[17]This is the number of Portuguese hospital units that recorded at least 1 medical procedure out of the 19 ICD9-MC procedure codes comprehended in the PIO, which does not correspond to the total number of hospital units in the country.

defined in the PIO by *Portaria 1306/2008* of the Ministry of Health are the following: 1311, intracapsular extraction of lens by temporal inferior route; 1319, other intracapsular extraction of lens; 132, extracapsular extraction of lens by linear extraction technique; 133, extracapsular extraction of lens by simple aspiration (and irrigation) technique; 1341, phacoemulsification and aspiration of cataract; 1342, mechanical phacofragmentation and aspiration of cataract by posterior route; 1343, mechanical phacofragmentation and other aspiration of cataract; 1351, extracapsular extraction of lens by temporal inferior route; 1359, other extracapsular extraction of lens; 1364, discission of secondary membrane (after cataract); 1365, excision of secondary membrane (after cataract); 1366, mechanical fragmentation of secondary membrane (after cataract); 1369, other cataract extraction; 1370, insertion of pseudophakos, not otherwise specified; 1371, insertion of intraocular lens prosthesis at time of cataract extraction, one-stage; 1372, secondary insertion of intraocular lens prosthesis; 138, removal of implanted lens; 1390, operation on lens, not elsewhere classified; and 1391, implantation of intraocular telescope prosthesis.

The subsample of the LIC relevant to our study consists of 2,805 monthly observations from 54 hospitals.[18] It is larger than the DRG dataset, as expected, because the sample is selected based on DRGs, which can combine more than one ICD9-MC procedure. We selected all DRGs that belong to the major diagnostic category (MDC) that corresponds to eye-related DRGs, which are: 36, retinal procedures; 37, orbital procedures; 38, primary iris procedures; 39, lens procedures with or without vitrectomy; 40, extraocular proc exc orbit, age > 17; 41, extraocular proc exc orbit, age 0-17; 42, intraocular proc exc retina, iris and lens; 43, hyphema; 44, acute major eye infections; 45, neurological eye disorders; 46, other disorders of the eye, age > 17 with complication; 47, other disorders of the eye, age > 17 without complication; 48, other disorders of the eye, age 0-17; 534, eye procedures with major complication; 535, eye disorders with major complication.

Information on procedures and DRGs is used to generate the two dependent vari-

[18]The number of hospitals in this sample is less than the number in the DRGs dataset because not all Portuguese hospitals recorded a waiting list during the complete period of analysis.

ables. Volume of surgical procedures is a monthly variable defined as:

$$vol\_srg_{h,t} = \sum_{p} \text{medical procedures} \tag{3.1}$$

where h indexes the hospitals; t, the time, which ranges from January, 2006 to December, 2010 (monthly); and p, which corresponds to the procedure codes described above. In other words, the dependent variable measures the volume of activity that corresponds to the procedures comprehended in the PIO, per month in each hospital of the sample.

We have chosen the volume of activity as our dependent variable but we could have selected some other measure, for example, individual quality of treatment induced by the PIO. One could argue that the latter would be a more accurate measure of the outcomes of the programme but the datasets available did not provide enough information to construct such a variable. There is a trade-off between having a very representative sample with less individualized information and a smaller sample with specific information regarding the quality of treatment. The decision made was to use the representative sample of the Portuguese population.

Waiting time is also recorded monthly, with the same indexes as before for hospitals and time, and it measures the mean waiting time in all DRGs:

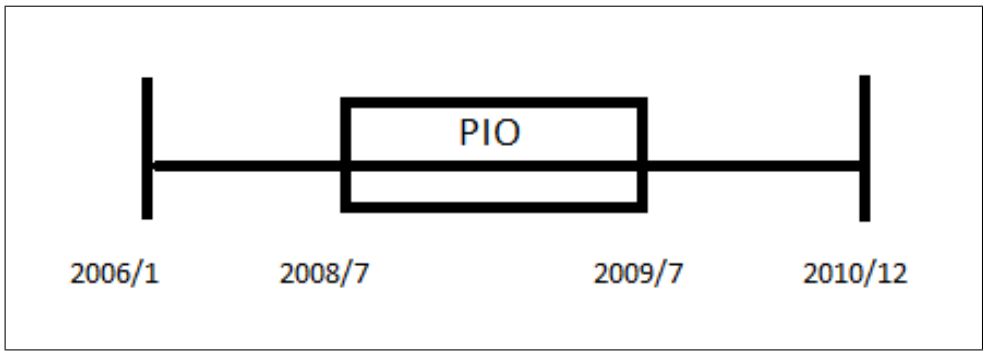
$$mwtime_{h,t} = \text{mean (waiting time in days)}_{\text{all DRGs}} \tag{3.2}$$

The key independent variable, called "PIO", evaluates the effect of the intervention programme, using information from before, during, and after the period when the PIO took place, in order to identify any impact as described in Figure 3.1. It equals 1 for the time when the programme was in place and is defined as follows:

$$PIO_t = \begin{cases} 0, & \text{if } t < 2008m7 \ \vee \ t > 2009m7; \\ 1, & \text{if } t \geq 2008m7 \ \wedge \ t \leq 2009m7. \end{cases} \tag{3.3}$$

In order to check whether the intervention had longer impact on volume of surgeries and waiting time (after the end of the programme), the variable $PIOafter$ is

Figure 3.1: Time's range of observations in dataset



defined, which takes the value 1 after the end of the ophthalmology intervention:

$$PIOafter_t = \begin{cases} 0, & \text{if } t \leq 2009m7; \\ 1, & \text{otherwise.} \end{cases} \tag{3.4}$$

Other explanatory variables include year dummies and control for severity of the disease in the LIC analysis by the mean priority of the cases treated in the hospital. Note that there are some drawbacks to the data. The first is the fact that the LIC dataset is based on DRG information. It may be the case that we included episodes of DRGs in the calculation of mean waiting time that do not have any of the procedures comprehended in the PIO. One of the PIO's rules is that a patient waiting for more than a given time to receive care could be transferred to private hospitals and the Government would pay for the surgery. In this scenario, without information of private complete cases in the dataset of LIC, the results might be underestimated. Another problem is the absence of data from the private hospitals in the DRG dataset, which could alter the results if included.[19]

A description of the variables is available in the appendix (Table 3.6). Briefly, in the DRG dataset mean volume of surgical procedures performed by each hospital in a month is around 200. For the hospitals (annual information) mean number of occupied beds (*beds*) is 451, mean number of discharged patients (*dp*) is 18,023, length of stay (*lstay*) is around 7.5 days, the case-mix index (*cmi*) is 1.05, and

[19]The inclusion of private data in the DRG dataset would increase the volume of activity and magnify any potential effect of the programme.

the mean individual cost (*cost*2) is about 6,000 euros.[20] For each month, the annual information of these variables is repeated, because it is the only variable aggregation available. These variables were included because they might affect the monthly dependent variable even though they are quite stable throughout the year. For the LIC dataset, mean waiting time (*mwtime*) is 95 days, whereas the mean priority of cases (*mpriority*) is 1.31.[21] The hospital characteristic variables have mean values similar to those of the previous sample.

Table 3.2: Summary of DRG dataset

Variable	Obs.	Mean	Sd	Min	Max
vol_srg	2635	200	186	1	1618
beds	2591	451	313	14	1497
dp	2614	18023	11485	364	50128
lstay	2614	7.44	2.76	2.20	33.50
cmi	2508	1.05	0.24	0.65	1.99
cost2 ((totalcost/dp)/10^6)	2532	0.0065193	0.0035865	0.0030216	0.0377743

Table 3.3: Summary of LIC dataset

Variable	Obs.	Mean	Sd	Min	Max
mwtime	2805	95.53	73.13	0	560.14
mpriority	2805	1.31	0.38	1	4
beds	2609	465	315	14	1497
dp	2621	18620	11796	364	50128
lstay	2469	7.58	3.05	2.20	33.50
cmi	2514	1.05	0.23	0.65	1.99
cost2 ((totalcost/dp)/10^6)	2538	0.0065897	0.0036462	0.0032336	0.0377743

[20]The mean individual cost is the result of annual total cost (reported by the hospitals) divided by the annual total number of discharged patients.

[21]Priority ranges from 1 to 4, in decreasing level of urgency.

3.4.2 Estimation procedures

Hospitals are the unit of observation in both analyses. With this data we construct two panels, considering the advantages that panel data offer over cross-section analysis: there are fixed effects at hospital level, because specific characteristics of the hospitals that do not change over time may influence the relationship of the dependent variables with the intervention programme that is being evaluated, the unobserved heterogeneity. Another correction that has to be made in the panel setting is to consider the potential autocorrelation amongst the error terms of a unit (hospital).

The decision regarding the specification of individual-level effects of the model was made based on the Hausman test, which compares the existence of fixed or random effects in the data. Qur results reject the null hypothesis, which states that the difference in the estimated coefficients of the models is not systematic under the assumptions of having fixed or random effects. This result, associated with the particularities of the Portuguese health system, leads us to fit the model using fixed-effects that would capture all temporally constant individual-level effects.

The coefficients of interest are estimated according to the specification of the following equations, in which we correct for autocorrelation and heteroskedasticity and include fixed-effects at hospital level:[22]

$$vol\_srg_{h,t} = \alpha + \beta_1 trend + \beta_2(trend \times PIO_t) + \beta_3(trend \times PIOafter_t) + \delta year_t + \theta \chi_{h,t} + \varepsilon_{h,t} \quad (3.5)$$

$$mwtime_{h,t} = \alpha + \gamma_1 trend + \gamma_2(trend \times PIO_t) + \gamma_3(trend \times PIOafter_t) + \psi mpriority + \phi year_t + \rho \chi_{h,t} + \varepsilon_{h,t} \quad (3.6)$$

We expect the signs of the coefficients β_2 and β_3 to be positive ($\beta_2 > 0$ and $\beta_3 > 0$), meaning that the volume of medical procedures increases during and after the execution of the intervention programme. For the equation that measures the impact of the PIO on the mean waiting time, the sign of the programme

[22]The specification of the model is not a novelty in the literature of programme evaluation, as discussed in Imbers and Wooldridge (2009).

coefficients, γ_2 and γ_3, are expected to be negative, indicating that waiting time diminishes during and after the PIO. The PIO and $PIOafter$ variables were multiplied by the *trend*, which indexes the months in the sample, ranging from 1 to 60.

The differences between the estimated coefficients are relevant for the analysis: if $\beta_3 = 0$, then we have only temporary effect; if $\beta_3 = \beta_2$, the effects are all permanent; if $\beta_3 < \beta_2$, only part of the programme's effects are permanent; and if $\beta_3 > \beta_2$, there are cumulative effects retained even after the end of the programme (the same logic applies for γ.)

3.5 Results

3.5.1 Overview

The null hypothesis was that a) the ophthalmology intervention programme had a positive impact on the volume of surgeries and b) a negative effect on the mean waiting time.

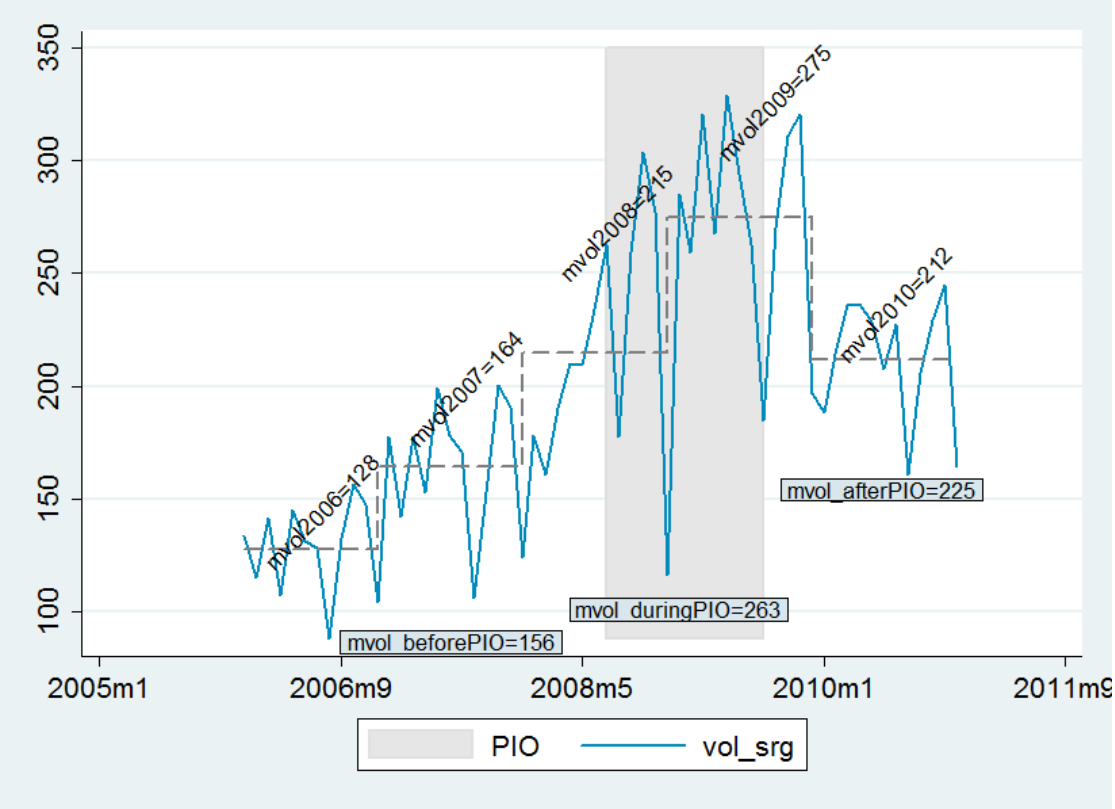
From Figure 3.2 the behaviour of surgeries' volume over time is clear. The average number of surgeries increases from a strip of 100 and 150 to the next, which varies from 150 to 200 surgeries per month. From July, 2008 to July, 2009, we observe the highest mean volume of surgeries. This coincides with the period when the programme was active. This can be interpreted as a first evidence of the intervention programme's positive impact. From the graphic of waiting time (Figure 3.3), it is more difficult to make a visual inference, as there is a trend of decreasing waiting time during the whole period.

3.5.2 Testing the hypotheses

Volume of medical procedures

Although the graph analysis offers insight into the relationship between the intervention programme and the two variables of interest, volume of surgery and waiting time, it is necessary to go further and apply specific techniques that control for other measurable and non-measurable factors that may affect the evaluation of the

Figure 3.2: Surgeries' volume

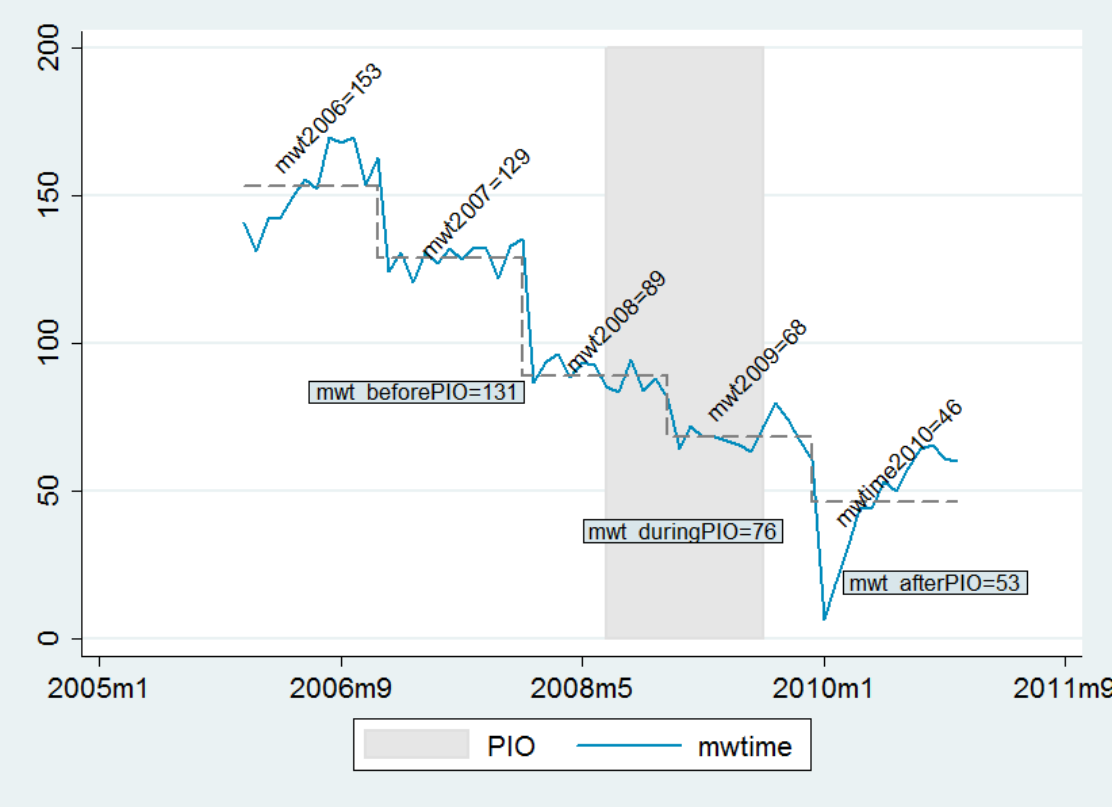


programme in the context of a governmental intervention.

The simplest fixed-effects regression shows that PIO has a positive impact on the mean number of surgeries performed each month by the hospitals (model 1). The results remain valid when we add year dummies (model 2) and hospital characteristic variables (model 3). This shows us that the PIO had a statistically significant impact on the number of surgeries performed by hospital. In other words, it is shown that the number of medical procedures increased in a significant way during the intervention. However, the same cannot be said for the period after the intervention (compared to the pre-intervention period). There is no long-term impact of the programme on the volume of medical procedures.[23]

[23] An extension of model 3, with squared root of volume as the dependent variable, which has a distribution closer to the Normal, shows estimated coefficients that are qualitatively the same as the crude data and the non-normality of the distribution does not affect the consistency of

Figure 3.3: Waiting time



All models were corrected for heteroskedasticity and serial correlation. The Hausman test for fixed and random-effects rejects the null hypothesis that both models are consistent. This, in addition to the characteristics of the Portuguese health system, allows us to conclude and apply the fixed-effects model.

Waiting time

The second set of analyses evaluates the impact of the intervention programme on the mean waiting time of the patients that had a complete case of surgery during the years from 2006 to 2010. Although the results of model A and B indicate a weak effect of the PIO on the waiting time, this result changes when control variables are added (models C and D).

the parameters.

In model D, the estimated coefficients of the PIO have negative signs, as expected, and this means that the waiting time during and after the period of the programme was reduced. The variable *mpriority*, which labels the severity of the disease, has a statistically significant and negative estimated coefficient, corresponding to the fact that more complicated cases have a shorter waiting time for surgeries.
Hausman test results favoured the fixed-effects model as the most suitable one, and corrections of potential heteroskedasticity and serial correlation were applied.

Table 3.4: Linear regression (fixed-effects)

	Unit of observation: hospital						
Dep. var.	Volume			Waiting time			
Model	(1)	(2)	(3)	(A)	(B)	(C)	(D)
trend	1.724***	-0.606	-0.709	-1.990***	-1.873***	2.028***	1.837***
	(0.531)	(0.737)	(0.780)	(0.432)	(0.432)	(0.431)	(0.438)
trendPIO	1.994***	1.421***	1.485***	-0.277	-0.287*	-0.423***	-0.445***
	(0.377)	(0.421)	(0.450)	(0.179)	(0.169)	(0.123)	(0.139)
trendPIOafter	0.175	0.596	0.563	-0.0165	-0.0460	-0.551***	-0.552***
	(0.374)	(0.397)	(0.416)	(0.258)	(0.252)	(0.172)	(0.184)
mpriority					-42.14***	-39.20***	-37.60***
					(7.780)	(7.305)	(8.057)
Constant	128.4***	138.3***	-326.8	160.2***	212.2***	187.1***	240.8**
	(12.93)	(12.58)	(239.3)	(9.337)	(15.38)	(13.20)	(93.96)
Obs	2,635	2,635	2,473	2,805	2,805	2,805	2,350
R2	0.131	0.154	0.173	0.318	0.357	0.395	0.426
N of hosp	68	68	65	54	54	54	46

Notes:

(a) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; robust standard errors in parentheses; fixed effects at hospital level;

(b) models 2 and C include year dummies; models 3 and D include year dummies and hospital characteristics variables.

Simultaneous analysis of volume of medical procedures and waiting time

It is reasonable to suspect that exogenous shocks that affect volume of medical procedures may also have an impact on the waiting time to receive a medical procedure. An example of this is the closure of a surgical unit inside a hospital: besides the fact that the volume of medical procedures would be expected to fall, it is also reasonable to think that waiting time to receive care would increase. In order to verify that the estimation method of the coefficients presented above is correct, we also controlled for these simultaneous shocks, by performing a simul-

taneous equation method estimation. The results are consistent with the previous estimation methods: the programme has a short-term positive impact on volume of medical procedures and it has a short and longer-term impact on waiting time.

Table 3.5: Simultanous equation model

	Unit of observation: hospital					
Model	(1 and B)		(2 and C)		(3 and D)	
Dep. Var.	Volume (supply)	Waiting time (price)	Volume (supply)	Waiting time (price)	Volume (supply)	Waiting time (price)
trend	2.655***	-1.496***	-0.0875	2.042***	-0.441	1.730***
	(0.687)	(0.232)	(1.617)	(0.539)	(1.326)	(0.539)
trendPIO	1.473***	-0.705***	1.219*	-0.553**	1.388**	-0.475**
	(0.505)	(0.171)	(0.673)	(0.224)	(0.553)	(0.225)
trendPIOafter	-0.571	-0.528***	0.393	-0.675**	0.252	-0.577**
	(0.525)	(0.178)	(0.837)	(0.279)	(0.692)	(0.281)
mpriority		-18.02***		-15.43***		-16.79***
		(4.217)		(4.153)		(4.266)
Constant	125.5***	182.8***	206.7***	0	199.4***	19.77
	(12.02)	(6.585)	(68.94)	(0)	(61.02)	(25.72)
Obs	1,958	1,958	1,958	1,958	1,835	1,835
R2	0.065	0.284	0.075	0.312	0.441	0.315

Notes:

(a) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; standard errors in parentheses;

(b) model (2 and C) includes year dummies; model (3 and D) includes year dummies and hospital characteristics variables.

3.6 Final remarks

The PIO was designed to enhance quality of health care in the specialty of ophthalmology. The objective was to address the low performance of some indicators such as waiting lists and waiting times for first consultation and surgery. Earlier evaluations are mainly descriptive and do not take advantage of using econometric techniques to help the analyses.

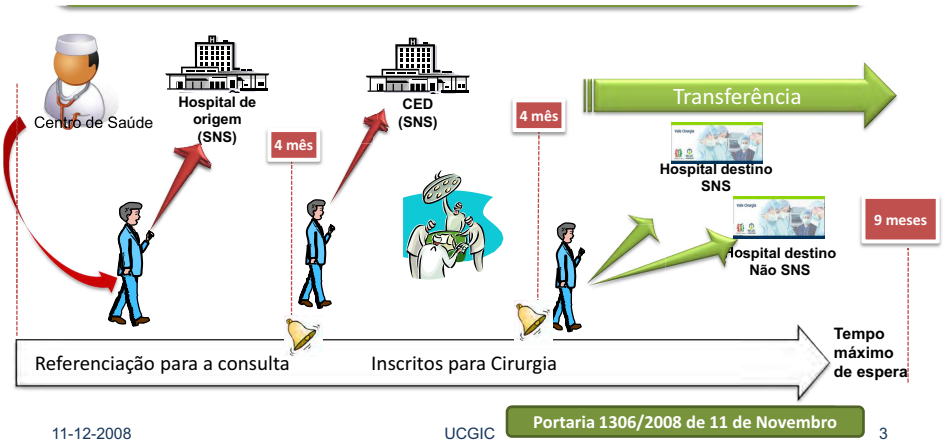
Our study investigated the impact of the PIO on volume of medical procedures and waiting time for eye-related diseases. The PIO was successful in increasing volume of medical procedures and reducing waiting time in the short term. While financial incentives were active, the results on the supply side reached the expectations. We also identified some longer-term effect on waiting time. However, a final conclusion

about the long-term impact of the PIO, in which the results of both increased medical activity and reduced waiting time are incorporated in the structure of the system, should be the object of further research, as more recent data become available.

Appendices

The scheme of the PIO

Figure 3.4: The steps of the PIO[a]



[a]Source: Portal da Saúde - First quarter evaluation of the PIO

Description of the variables

Table 3.6: Complete description of the variables

Variable	**Description**
beds	number of occupied beds (by hospital, by year)
cmi	case-mix index (by hospital, by year)
cost	annual individual cost (by hospital, by year)
dp	number of discharged patients (by hospital, by year)
lstay	length of stay, in days (by hospital, by year)
PIO	dummy of the intervention programme (by time)
PIO_after	dummy of post intervention programme (by time)
*yeard**	dummies of year

Notes:

(a) Case mix index: $cmi_{h,t} = \frac{\sum_{DRG} (\text{number of equivalent patients})_{DRG} \times (\text{DRG\_weight})_{DRG}}{(\text{total number of equivalent patients})}$

(b) $cost$ = (annual total cost/annual discharged patients)/10^6

Bibliography

Anderson G, Black C, Dunn E, Alonso J, Christian-Norregard J, Folmer-Anderson T, Bernth-Peterson P. 1997. Willingness to pay to shorten waiting time for cataract surgery. *Health Affairs* **16**(5) : 181-190.

Busse R, Geissler A, Quentin W, Wiley M (eds.). 2011. Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals. *European Observatory on Health Systems and Policies Series*. Open University Press.

Cullis J, Jones P. 1985. National health service waiting lists: A discussion of competing explanations and a policy proposal. *Journal of Health Economics* **4**(2) : 119-135.

Hacker J, Stanistreet D. 2004. Equity in waiting times for two surgical specialties: a case study at a hospital in the North West of England. *Journal of Public Health* **26**(1) : 56-60.

Halliwell T. 1998. How fair is cataract prioritisation? *New Zealand Medical Journal* **11** : 406-407.

Hanning M, Lundstrom M. 1998. Assessment of the maximum waiting-time guarantee for cataract surgery: the case of a Swedish policy. *International Journal of Technology Assessment in Health Care* **14** : 180-193.

Hurst J, Siciliani L. 2003 Tackling excessive waiting times for elective surgery: a comparison of policies in twelve OECD countries. *OECD Health Working Papers* **6**.

Imbens G, Wooldridge, J. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* **47**(1) : 5-86.

Matos A, Viegas M, Campos A, Damaso I. 2010. Auditoria de resultados ao Programa de Intervenção em Oftalmologia. *Tribunal de Contas de Portugal* : Relatório nº 48/2010 - Processo nº 49/2009 - AUDIT **(I)**.

Sena C, Ferrinho P, Miguel J. 2006. Planos e Programas de Saúde em Portugal: questões metodológicas e macroanálise dos programas nacionais. *Revista Portuguesa de Saúde Pública* **24**(1) : 5-19.

Siciliani L, Borowitz M, Moran V (eds.). 2013. Waiting time policies in the health sector: what works? *OECD Health Policy Studies*. OECD Publishing. http://dx.doi.org/10.1787/9789264179080-en