

From downcoding to upcoding: DRG based payment in hospitals

Carine Milcent

▶ **To cite this version:**

Carine Milcent. From downcoding to upcoding: DRG based payment in hospitals. 2019. halshs-03966671

HAL Id: halshs-03966671
https://shs.hal.science/halshs-03966671

Preprint submitted on 31 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2019 – 53

From downcoding to upcoding: DRG based payment in hospitals

Carine Milcent

JEL Codes:
Keywords: Hospital stays, Diagnosis-related groups (DRGs), Upcoding, manipulated database, heterogeneity in responses

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

From downcoding to upcoding: DRG based payment in hospitals

Carine Milcent

CNRS - PSE

Abstract

A prospective disease group-based payment is a reimbursement rule used in a wide array of countries. It turns to be the hospital's payment rule to imply. The secret of this payment is a fee payment as well as a hospital's activity based payment. There is a consensus to consider this rule of payment as the least likely to be manipulated by the actors. However, the defined fee per group depends on recorded information that is then processed using complex algorithms. What if the data itself can be manipulated? The result would be a fee per group based on manipulated factors that would lead to an inefficient budget allocation between hospitals. Using a unique French longitudinal database with 145 million stays, I unambiguously demonstrate that the implementation of a finer classification led to a up-coding. The end result has been a budget transfer from public non-research hospitals to for-profit hospitals. This budget reallocation did not correspond to any change in the actual production of care.

Résumé

Les établissements hospitaliers français sont tous tarifés par Groupe Homogène de Séjours (GHS) depuis 2008. Ce mode de tarification est aujourd'hui largement adopté par l'ensemble des pays de l'OCDE. Toutefois, une tarification à 100% par forfait par séjour fait aujourd'hui débat en France. Les indicateurs de qualité semblent orthogonaux à la rémunération définie. Afin de prendre en compte plus finement la sévérité des patients, la classification des GHS a été multipliée par 3. A partir d'une base de données originale de 145 millions d'observations, nous sommes à même de comparer le comportement de codage avant la mise en place de cette nouvelle classification et après la mise en place de cette nouvelle classification. Nous montrons que les établissements de santé se sont adaptés à cette nouvelle classification en changeant leur

comportement de codage. Ce changement de codage est intervenu sans modification de la production réel de soins. Suivant le type d'établissement, l'adaptation au codage diffère. La tarification étant à enveloppe fermée, nous montrons une réallocation du budget vers les établissements à but lucratifs au détriment des établissements publics hors établissements de recherche.

key findings/implications of the manuscript

1- implementation of a finer classification led to a up-coding
2- At short term, a budget transfer from public non-research hospitals to for-profit hospitals
3- This budget reallocation did not correspond to any change in the actual production of care

Keywords: Hospital stays, Diagnosis-related groups (DRGs), Upcoding, manipulated database, heterogeneity in responses

From down-coding to up-coding:

DRG based payment in hospitals

Introduction

Manipulating the data from patient's health record during their stay in hospital is one way of shifting from one disease group (called DRG) to a more profitable one. Indeed, raw information manipulation allows coding changes to generate additional payment unrelated to the needs of patients (Steinwald and Dummit, 1989). In this paper, I use a database that has undergone a change in the logic of construction of DRGs, the number of groups increasing from around 800 to 2,200. This shift in classification logic created a natural experiment for testing whether French hospitals tend to enhance profit by taking advantage of regulatory loopholes through a behavior of upcoding (from an initial down-coding context). I also examine the heterogeneity in the response of hospitals to this change in coding structure.

While DRG-based hospital payment systems may provide adequate reimbursement for the average patient within each DRG, they overpay hospitals for patients with below-average resource consumption and underpay for patients with above-average cost. In order to avoid creating perverse incentives, which could lead to patient selection or a priori treatment choices, DRG payments are adjusted to the expected cost of a patient stay. But studies in various countries, including France, have found that DRG classification did not explain all of the variation in cost between different stays (Hakkinen et al., 2012; Mason et al., 2012; Milcent,

2017). While upcoding[i] behavior is widely studied with the US hospitals, we know almost nothing when healthcare is mainly provided by the public sector, as it is the case in France.

To improve the granularity of reimbursement, DRG classifications have often been refined over time. Refinements usually consist in splitting a single DRG category into more DRG categories related to the same primary diagnosis. The current debate in the United States on the potential impact of further refinement of DRGs for certain hospitals is instructive. A study carried out for MedPAC in 2005 demonstrated, for a given DRG, large variations in profitability: DRGs relating to surgery were found to be more profitable than those for medical treatments, notably for orthopaedic and cardiovascular care. On one hand, this led to the idea that a refinement in the classification would improve the payment system by better capturing differences in the severity of treated cases. The logic is straightforward. Highly disaggregated payment systems only offer an incentive to produce the highly disaggregated product at minimum cost. There is no incentive to make efforts to economize on the number or variety of disaggregated services (Newhouse, 2002; Newhouse, 2003; Cash et al., 2003). On the other hand, refinement increased the complexity of the DRG classification, requiring a very detailed and sophisticated information system, with an increased risk of mispayment if the information is not accurate.

This article studies the impact of a decision by French policy makers to refine the DRG classification in 2009 by introducing levels of severity, and, as a consequence multiplying by three the number of DRG groups. For internal purposes, the French health statistics agency used the new 2009-classification of the algorithm to classify stays before and after the implementation of this new classification. With this unique database, I identify changes in hospitals' coding practices without any confounding effects. I was thus able to "go back in time". I assessed the

change in behavior, i.e. changes of the raw information only due to change in the DRG classification. Besides, I distinguished this behavior change according to the type of hospitals.

I find an unambiguous upcoding of the raw information as hospitals' response to a refinement of the DRG classification. This upcoding consists in changes in hospital record-keeping practices in order to increase case-mix indices and reimbursements. The probability of being coded as non-severe decreased by about 2.1%. These results are consistent when controlling for pathology indicators and hospital fixed effects: hospitals are sensitive to the financial incentives created by the DRG classification. I also find that hospitals' response to the refinement of the DRG classification system depended on hospital ownership. For-profit hospitals were more sensitive to the financial incentives created by this change in DRG classification logic. Public non-research hospitals were the least sensitive to the change. An implication of these results is a budget reallocation that went to for-profit hospitals at the expense of public non-research hospitals.

The paper is organized into five sections. Section 2 presents the implications of activity-based pricing (T2A). It introduces a hospital objective function, which provides the theoretical framework for the empirical sections that follows. Section 3 presents the database, and the results in Section 4. Section 5 presents robustness checks. Finally, Section 6 presents concluding remarks.

The implications of activity-based pricing

This research contributes to the literature on providers in healthcare (Ellis and McGuire, 1986; Pope, 1989; Ma, 1994, 1998; Chalkley and Malcomsom, 2000; Mougeot and Naegelen, 2008; Brekke et al., 2011). Models of hospital behaviour predict that hospitals will respond to a change in the refinement of DRG classification by trying to increase revenue per stay. More thorough

coding of secondary diagnoses and procedures is an intended effect of the introduction of DRG-based hospital payment systems, but the risk is that hospitals attempt to increase revenues through fraudulent coding practices, leading to unjustified payments (Simborg, 1981; Steinbusch et al., 2007; Silverman and Skinner, 2004). Kuhn and Sicilliani (2008) assume that providers can increase demand by increasing quality but can also inflate activity through a manipulative effort.

Most of the attention in the empirical literature has focused on how hospitals respond to price changes (Silverman et al., 1999; Silverman and Skinner, 2004 ; Xirasagar, 2006; Brekke et al. 2012). Gilman (2000) investigated the impact of a 1994 reform of Medicaid DRGs for HIV diagnoses in New York. He found that DRG fee changes led to changes in length of stay. These results suggest that hospitals make DRG-specific changes in intensity of care in response to price changes. Researchers have attempted to measure the role of upcoding in increased case-mix indices in US Medicare (Carter and Ginsburg, 1985; Hsia et al., 1988; Carter et al., 1990). Dafny (2005) exploited a 1988 policy reform that generated large price changes in Medicare admissions. She found that the upcoding response was stronger among for-profit hospitals. Barros and Braun (2013), following Dafny (2005), explored the link between upcoding and price increases in the context of a national health service (that of Portugal). Sacarny (2016) uses the change in billing categories to estimate a model of technology adoption. Geruso and Layton (2015) used upcoding data to learn about the game between payers and providers. Using the French natural experiment, I study differences across hospitals in how they respond to the 2009 policy, with the introduction of the severity level in the DRG classification, tripling the number of groups.

Payment system in France

The French national health insurance scheme (Sécurité Sociale) is a single-payer system: this eliminates any concerns about potential cost-shifting behavior by providers, negotiation between providers and payers, or different reimbursement schemes for different patients. Reimbursements cover almost all medical services in hospitals, except an additional fixed fee per day for catering and accommodation in for-profit hospitals.

How is the French hospital system structured? It consists of three types of hospitals: public, private for-profit, and private not-for-profit. Healthcare for in-patient is mainly provided by the public sector. This paper focuses on stays in acute care units. For these, for-profit private hospitals represent 25% of beds, while public hospitals make up two thirds of total beds. The remaining 8% of beds are in private non-profit institutions. In terms of stays, 36% of admissions are to for-profit private hospitals (DREES, 2012).[ii]Until the 2000s, the public sector was largely immune from standard market forces pushing for production efficiency. Since 2008, a DRG-based payment (T2A) has been the sole mode of reimbursement of health care institutions for all acute care stays, regardless of hospital type. However, according to the literature, for a same reimbursement rule, the hospital behavior differs. Dafny (2009, 2005) and Silverman (2004) find that for-profit hospitals or for-profit managers upcode more than hospitals of other ownership forms. It suggests that the upcoding strategies of for-profit hospitals vary from non-profit or government-owned hospitals, a prediction we consider in the empirical work that follows.

Yardstick Competition in practice

The T2A system of activity-based pricing is built on the theoretical model of Schleifer (1983). In this theoretical model, the payment rule is given ex ante and lump-sum transfers are calculated

ex post. Ex ante, the regulator announces to hospitals that they will receive a lump-sum transfer per DRG, whose amount is not yet established but which will correspond to the mean cost. This model thus makes payments dependent on the mean cost for a given DRG. Assuming hospitals and stays to be homogeneous, the cost variable that differs between stays is the hospital level of cost reduction effort. The unique feature of this model is its calculation of the transfer, consisting of the mean cost in all hospitals except the one receiving the payment. The lump-sum transfer received is thus exogenous to the hospital's activity. Rational behavior in the context of payment by lump-sum transfer is to minimize costs in order to capture the rent between the lump sum and the actual cost. Here, this involves making the maximum effort for the minimum cost. All hospitals having the same rational behavior, the cost paid by all hospitals will be the minimum cost. The lump sum thus corresponds ex post to the minimum cost for a given DRG *g*.

The pricing model currently in effect in France is defined for periods of one year. The rule is the following: at the end of the period, each hospital receives an amount of funding for the next period calculated on its activity over the past year. More specifically, each stay is associated to a DRG, which defines a certain quantity of care and a mean cost;[iii] i.e., a certain amount of hospital activity. A lump sum is associated to each stay classified under this DRG. At the end of the period, the hospital thus receives the total of the lump sums corresponding to its activity for the period.

I now present the calculation of lump sums by DRG. Assuming hospitals to be rational agents, I explicitly derive the incentives that result from this financing mechanism. Let s_{hg} denote stay s associated to DRG g in hospital h; S_{hg} denote the total number of admissions associated to DRG g in hospital h; C_{shg} denote the cost of a stay s associated to DRG g in hospital h; h_g denote

hospital *h* having admitted at least one patient for a stay associated to DRG *g*; and H_g denote the number of hospitals having admitted at least one patient for a stay associated to DRG *g*.

The mean cost for a DRG *g* is defined by:

$$\overline{C_g} = \frac{\sum_{h_g}^{H_g} \sum_{s_{h_g}=1}^{S_{h_g}} C_{s_{h_g}}}{\sum_{h_g=1}^{H_g} S_{h_g}}$$

On the basis of the mean cost per DRG, a relative cost scale for the different DRGs can be defined. DRGs are thus ordered on a scale from the least costly to the most costly. DRGs can be expressed as a function of a reference DRG. The choice of reference DRG has no impact on the budget that the hospital receives at the end of the period. Then, a relative index is associated to each DRG. Let $\overline{C_g}$ denote the mean cost of a stay *s* associated to DRG *g*; $\overline{C_{gref}}$ denote the mean cost of a stay *s* associated to the reference DRG, called *gref* . The relative index i_g of a DRG *g* is defined by: $i_g = \overline{C_g} / \overline{C_{gref}}$

Payment per DRG & a fixed annual envelope

Such a pricing system offers no way to control the volume of care. In France, we are in a context where the financial situation of the public system has declined since the 80s. The regulator cannot thus risk an uncontrolled increase in the volume of care. Besides, the period of pricing by global budget that preceded the implementation of PPS was a period of blind rationalization. The extent to which an insufficient supply of hospital care led to a rationing of demand cannot be determined.

As a result, the logic of DRG-based pricing was joined by a fixed budget envelope mechanism. The law on the financing of social security (LFSS) determines the total budget for hospital spending over the course of the year. Thus, the mechanism is characterized by two lump sums: an exogenous, arbitrarily chosen global lump sum *F* for all hospitals, and a lump sum per stay *s* associated to a given DRG *g*, noted as f_g.

A hospital activity corresponds to the sum of the number of relative units for all stays. Total national hospital activity is defined in the same way, as the sum of relative units for all health care institutions in France. The value of the index is calculated as a simple ratio of the total sum of relative units produced by all health care institutions to the size of the overall envelope distributed to all institutions for the period. G represents the number of DRGs in the classification.

$$F = \sum_{g=1}^{G} \sum_{h_g}^{H_g} \sum_{s_{h_g}=1}^{S_{h_g}} i_g v \text{ which implies } v = \frac{F}{\sum_{g=1}^{G} \sum_{h_g}^{H_g} \sum_{s_{h_g}=1}^{S_{h_g}} i_g} \quad (1)$$

According to Equation (1), the greater the amount of care produced by all healthcare institutions over the period, the lower the value of *v*, and thus the lower the reimbursement for a stay, ceteris paribus. The amount of the lump sum per DRG is the relative index multiplied by its monetary value: $f_g = i_g v$. This lump sum per DRG thus depends on the activity of all healthcare institutions combined.

Let g_h denote DRG *g* associated to stays in hospital *h*, and G_h denote the total number of DRGs associated to stays in hospital *h*. On the basis of these definitions, the budget received by hospital *h* is thus:

$$B_h = \sum_{g_h=1}^{G_h} \sum_{s_{h_g}=1}^{S_{h_g}} f_g = \sum_{g_h=1}^{G_h} \sum_{s_{h_g}=1}^{S_{h_g}} i_g v = F * \frac{\sum_{g_h=1}^{G_h} \sum_{s_{h_g}=1}^{S_{h_g}} i_g}{\sum_{g=1}^{G} \sum_{h_g}^{H_g} \sum_{s_{h_g}=1}^{S_{h_g}} i_g} \quad (2)$$

The hospital's objective function (with C_h the total costs of hospital *h* over the period) is:

$$\text{Max } \underbrace{F * \frac{\sum_{g_h=1}^{G_h} \sum_{s_{h_g}=1}^{S_{h_g}} i_g}{\sum_{g=1}^{G} \sum_{h_g}^{H_g} \sum_{s_{h_g}=1}^{S_{h_g}} i_g}}_{B_h} - \underbrace{\sum_{g_h=1}^{G_h} \sum_{s_{g_h}=1}^{S_{g_h}} C_{s_{g_h}}}_{C_h} \text{ with } C_h \leqslant B_h \quad (3)$$

Here, all types of hospitals (state-owned hospitals, for profit hospitals and non-profit hospitals) are considered to have a single objective, revenue maximization. [iv]

Implications

The 2009-shift of classification with the DRG groups' refinement was aiming at reducing the consequences of within-DRG cost heterogeneity, providing an alternative to outlier costs to the hospitals. The more DRGs are in the classification, the greater the distinguishability of the severity of cases. Consequently, the more refined the classification, the more closely relative units reflect the gravity of cases. The hospital, behaving rationally, seeks to maximize the number of relative units per stay. As a result, the more the classification system takes severity into account, the greater the incentive for hospitals to optimize the coding of admissions to associate them to DRGs which present the highest possible level of severity. The refinement issue has been theoretically studied in papers by Siciliani (2006), Hafsteindottir and Siciliani (2009), and Kuhn and Siciliani (2013). They provide a clear analysis of the upcoding phenomenon: Healthcare providers have interest in declaring more diagnoses which lead stays to be associated to high-severity DRGs when low-severity patients are less costly than high-severity patients.

While demand is price-responsive in Shleifer's model, French social security is a system of fully reimbursed insurance for inpatients, which leads to price-inelastic demand. The patient's out-of-pocket cost is negligible and independent of the level of DRG. This inelasticity may induce an increase in the number of stays. To limit this adverse effect, the French regulator has implemented a fixed budget (or envelope) mechanism, as presented in Section 2.2. Because of the fixed budget mechanism, there is no way for total cost to increase. Therefore, another way for a hospital to increase its budget is to increase the number of relative units per stay, that is upcoding.

The objective function outlined above can easily be expanded to include upcoding effects. The number of relative units in hospital *h* can be redefined as an increasing function of the degree of upcoding. Upcoding can then result from maximization of the left hand member of equation (3) (hospital revenue), i.e., maximization of the relative index per stay i_g. The practices of competitors may affect upcoding indirectly through pressure on hospital profits. Returning to equation (3), it is from the numerator of the left-hand member of Equation 3: the number of relative units of hospital *h* increases with the degree of upcoding (indirect channel); Or it may affect directly via the dissemination of upcoding practices. It means, from the denominator of the left-hand member of equation 3, the total number of relative units increases with the degree of dissemination of upcoding practices (direct channel). In the US, one former manager from the largest for-profit hospital chain, Columbia/HCA (now HCA), reported that hospital managers were rewarded for upcoding patients with these diagnoses into the more remunerative "with complications" codes (Lagnado, 1997).

Data

Original data

Whatever the hospital's ownership, for each in-patient, primary diagnosis, the list of all secondary diagnoses, co-morbidities, and procedures performed, as well as all exploratory acts, are registered. This information is then sent to the national health statistics agency[v] providing an exhaustive record of French hospitals. On the basis of this information, the national health statistics agency uses a grouper program to assign each stay to a DRG. Each year there is adaptive change of the grouper program. But, in 2009, there was a political decision of a rethinking of the French DRG classification. This classification switched from around 800 groups to 2,200 groups. For internal purposes, the health statistics agency used the 2009-classification of the algorithm to classify stays over the years 2006-2011.

The database used here covers the period 2006 to 2011 with a total of 145 million stays.

Preliminary statistic on the 2009-logic of the French DRG classification

This new (2009) DRG classification is organized in a nested fashion. DRGs are coded by a series of 6 characters. The sixth character defines the level of severity —4 levels— or the absence of severity, as in the case of exploratory acts (for instance, chemotherapy sessions). [vi]In this paper, we focus on in-patient with level of severity from 1 to 4. The level of severity is independent from the procedures performed on the patient.

The distribution by severity is as follows: 70% in level 1, 20% in level 2, 9% in level 3 and 2% in level 4. Public hospital stays were more spread out among severity levels than those in for-profit institutions (Table 1). Among stays classified by severity level, fewer than two thirds of

stays in public healthcare institutions were of low severity, while in contrast, nearly four fifths of stays in for-profit hospitals were for a level of severity without complications (Table 1).

The proportion of stays coded as low-severity decreased over time (Figure 1). An inflection – an acceleration in the downward tendency – can be observed in 2009. In parallel, the proportion of stays classified as moderate to very high severity increased over time, an increasing trend that became steeper after the reform. This change in the declining slope of 2009 is observed for all types of hospitals (research, other public, private non-profit, and private for-profit).

Insert Table 1 about here

Insert Figure 1 about here

Empirical strategy and results

Econometric models

In this paper, I check whether the refinement of a hospital stay classification system leads to upcoding from a possible initial situation of downcoding. As mentioned in Introduction, the data make it possible to "go back in time" and observe the consequences of the logic change in the DRG classification. For the years before 2009, the data include not only the diagnoses and exploratory procedures reported at the moment of the stay, but also the associated DRG

according to the 2009-shift in classification. Therefore, the analyses here are based on the level of severity that would have been associated to the stay in the 2009-change in the logic of classification, before it actually came into effect. Analysis of data from a 2006-2011 period is used to examine a potential upcoding independently of any changes in treatment.

The explained variable Y_{iht} is an indicator of the level of severity of stay *i* at hospital *h* at time *t*. $Y_{iht} = 1$ if stay *i* is associated to a DRG of severity level 1.

$$Y_{iht} = \alpha + \gamma_1 t + \gamma_2 I(t \geqslant 2009) + \beta_h + \beta_h * t + \beta_p + \epsilon_{iht} \quad (4)$$

- Temporal effects are controlled by a trend variable *t*, coded from 0 to 5 (for years from 2006 to 2011). Any macro changes that are common to hospital stays are picked up by this trend coefficient. The reference year is 2006.
- The effect of the change in classification may also be explained by hospital-specific behaviors. β_h are hospital effects.
- Some hospitals may change in terms of activity or size over the period. To control for this, hospital fixed effects crossed with the trend (β_h * *t*) are also included in the regression model.
- As there may have been changes over the period in the practices recommended by physicians for some pathologies which could affect classification regardless of other incentives, I also control for pathology effects. β_p are pathology effects.
- *I(.)* is an indicator for the post-classification change period, which takes the value 1 for years where the refined DRG classification system was applied (from 2009 onward) and 0 otherwise. *I(.)* is thus the pure impact of the 2009 DRG classification policy obtained

by controlling for any other changes over the period. [vii]This effect may depend on hospital's ownership status. To control for this, an alternative model is to include hospital's ownership crossed with the indicator for the post-classification change period.

- ε_{iht} is random noise.

The linear models do not take into account the correlation between the different levels of severity. In the following, I therefore, used an ordered probit model controlling for the interdependency of coded severity levels. I define a categorical variable $y \in \{1, ..., 4\}$ indicating the observed levels of severity. The discrete probability function of y conditional on all explanatory variables is commonly specified as an ordered probit model. This latent variable y^* is assumed to be generated by a normal regression structure.

$$y*_{iht} = \alpha + \gamma_1 t + \gamma_2 I(t \geqslant 2009) + \beta_h + \beta_h * t + \beta_p + \epsilon_{iht} \quad (5)$$

where y^* is unobserved. What is observable is the coded level of severity y. ε is a normally distributed random term, with variance normalized at 1. Threshold parameters determine the estimates for different observed values of y.

Besides, there are many reasons why upcoding behavior may differ across hospitals. There are a number of theories of the effect of hospital ownership on upcoding, but few consensus predictions (see Silverman and Skinner, 2000 for a comprehensive discussion). Several recent studies document this indirect channel. Duggan (2002) found that for-profit hospitals respond more strongly to financial incentives to treat indigent patients in markets with greater for-profit penetration. In the French context, the DRG classification changes were based on diagnoses rather than interventions. These DRG classification changes were then set up independently from

the intensity of care, by construction. The responses to changes in classification were thus established independently of any changes in treatment. Newhouse (1989) found evidence that private hospitals shifted patients in unprofitable DRGs to public hospitals following the implementation of PPS. Silverman and Skinner (2000) found strong evidence of upcoding between 1989 and 1996. They found that for-profit hospitals upcode the most, and that not-for-profit hospitals are more likely to engage in upcoding when the market share of for-profit hospital in their area is high. Dafny (2005) also found strong evidence of upcoding, a response that was particularly strong among for-profit hospitals.

Public hospitals cover two very different groups of institutions. University hospital centres and regional hospital centres (research hospitals) are distinguished by their research and development activity as well as by their size. Thus, I distinguish between research hospitals and "other public hospitals." As explained in Section 2, the French healthcare market also includes both for-profit and not-for-profit private hospitals. As a result, 4 types of hospitals are considered: Research hospitals, other public hospitals (except the research ones), non-profit hospitals and for-profit hospitals.

Results

Table 3 displays results of Least Square models. The chances that a stay would be coded as low-severity versus higher severity (Y) decreased over time. This indicates either an increase in the severity of the health status of admitted patients over time, or a learning effect: hospitals adapting to the incentives of the DRG-based reimbursement system gradually introduced from 2004 to 2008.

All else being equal, we show a steeper slope of this probability with the 2009-shift of DRG classification. The probability that a stay is associated to a non-severe DRG decreases by about 2.1% over the period. This result is robust: whether or not types of pathologies, hospital fixed effects (or hospital fixed effects multiplied by the trend) are controlled for, I obtain the same result. Therefore, the coding behavior of healthcare actors was modified by the change in DRG classification. [viii]Could the result be explained by more intensive patient care? The answer is no, since severity levels are established exclusively on the basis of diagnoses and not treatment. Thus, what I show here is a coding of more severe patient diagnoses, attributable only to upcoding effects possibly due to a initial situation of downcoding. This change in coding generated no change in hospital production.

I now present the results of the model controlling for interactions between hospital type and I(.), the indicator for the post-classification change period. The effect of the 2009-shift in DRG classification differs according to the type of hospitals. It appears that the policy makers' decision led private for-profit hospitals to more massively manipulate raw information compared with other types (Table 3). The coding behavior of public hospitals changed to a lesser degree. Non-profit institutions fell between private for-profit hospitals and research public hospitals. Regarding public hospitals, "other public hospitals" were disadvantaged in comparison to research hospitals.

Insert Table 3 about here

I now present Model (2), the ordered probit model. From Model (2), I capture the presence of time invariant heterogeneity with a fixed effects procedure. Note that estimating such models with fixed effects typically introduces an incidental parameter problem. Here, because of the large size of the sample, one can estimate these models without having this concern.

Considering the ordered probit model, the results are similar to those of Least Square models with respect to statistical significance.[ix] The reference is severity level 1, the lowest level. Note that a positive sign of any coefficients implies a higher probability of belonging to the higher category, corresponding in principle to greater severity. Controlling for types of pathology (Table 4, Column (2)), fixed hospital effects (Table 4, Column (3)) and fixed hospital effects multiplied by trend (Table 4, Column (4)) did not change the pattern of results. The probability of being coded as higher-severity increased with time. There was a sharp increase due to the 2009 policy. The marginal effects indicate how the probability of level of severity changes with the 2009 change in coding for average individual.[x] I find that the probability that a stay would be coded as severe increased by about 2% after the reform (compared with before the reform), controlling for all of the independent variables (Table 5, Column (4)). Furthermore, as with least squares models, the change in classification led to a greater change in coding behavior in private for-profit healthcare institutions (Table 4, Column (5)). Non-teaching public hospitals changed their coding behavior to a lesser degree.

These results clearly demonstrate a purely exogenous effect of a finer DRG classification leading to a upcoding, whether or not types of pathologies and hospital fixed effects were controlled for. The coding behavior of healthcare actors was thus altered by this DRG-classification shift

independently of any changes in the production of care. Also, these results clearly highlight a differential effect of a refined DRG classification system on the behavior of actors in healthcare provision depending on the type of institution.

Insert Table 4 and Table 5 about here

How to explain these results? As a fact, for-profit hospitals were more active with a very early use of coding optimization softwares in some for-profit institutions. More recently, University hospital centres have made use of such tools. Based on the theoretical and empirical literature, Silverman and Skinner (2004) advocate three potential explanations why for-profit hospitals were more active in upcoding using the Malani et ali (2003)'s simplified taxonomy of theoretical models of not-for-profit hospital behavior: i) the altruism model, ii) the signaling model, iii) the last type of model is based on evidence on the market interaction of hospitals with regard to upcoding. Focusing on the signaling model, if for-profit hospitals were able to increase marginal revenue through upcoding, they could force not-for-profit firms to compete through upcoding (for instance). To some extent, the altruism model may also receive empirical support. If non-for-profit managers instilled a strong ethical norm (Horwitz, 2003), and these standards could extend to conservative billing practices, then it may explain why non-for-profit institutions upcode less than for-profit hospitals. Georgescu and Hartmann (2013) studied the effects of healthcare decision pressure from the hospital's administration and from the professional peer group on physician's inclination to engage in upcoding. They find that the source of pressure is a relevant predictor of physicians' inclination to engage in up coding (from an potential initial situation of downcoding). What can we say about the situation in public hospitals? For these

hospitals, the pro-competitive reform that was introduced gradually from 2004 to 2008 implied a big change in managerial behavior that may be slowed down by the civil servant status of the hospital staff. Moreover, upcoding requires not simply administrators who direct coders to target profitable DRGs, but also physicians responsible for filling in the medical charts with the critical clinical information that can be used to claim the more generous DRG. There was no increase in physicians' revenue that would have created powerful incentives to manipulate the raw information.

This upcoding has direct positive effect on the payment received by the hospital.

Above all, in a fixed-envelope mechanism, the value of the index depends on the total number of index units. A higher level of severity is associated to a higher value of i_g. This situation leads automatically to a decrease in the value v of the index (Equation (1)). This in turn leads to a decrease in the lump-sum transfers associated to each DRG. This result has a number of consequences.

- The "upcoding" led to a public budget reallocation that increased the share of budget allocated to for-profit hospitals, at the expense of non-research public hospitals.
- As this mechanism leads to a tendency for any diagnosis to be coded at a higher level of severity, heterogeneity is shifted to higher severity levels. Thus, heterogeneity is shifted to higher-paying DRGs. This regulatory mechanism led to greater homogeneity in stays coded as not severe (severity level 1).

Financial impact

I do not presume to give the exact financial impact of the data-manipulation on the hospital healthcare budget. data-manipulation means here to switch from a possible downcoding initial situation to a upcoding situation. However, I propose here a simple method to get an approximation of this amount. The additional budget computed is done for the year 2009. **Context A**, I consider the case where there is no fixed annual budget at the national level determined by the regulator. Data-manipulation then leads to an increase in the global hospital healthcare expenditure. **Context B** is a context where a fixed annual envelope is set up. In the French case, a fixed annual budget at the national level is determined by a vote in the French National Assembly.

Being 2 hospitals, Hospital A with a funding x and hospital B with a funding y. With "data-manipulation", Hospital A now claims a funding $x+\alpha$. Hospital B's that does not "manipulate data". We can see the hospitals' funding as a percentage of the total funding.

With **Context A,**

- Without "data-manipulation", Hospital B receives $[y/(x + y)]$ % of the total funding
- With "data-manipulation", Hospital B receives $[y/(x + y + \alpha)]$ % of the total funding

In absolute value, the hospital B receives the same amount whatever the behavior of the hospital A. This latter receives more budget. In relative value, the hospital B receives less in presence of data-manipulation from the hospital A. The global hospital healthcare expenditure can be expressed as Equation (6). With pop DRG-Li: population of patient coded as DRG level i; means fee DRG-Li: mean of fee for patient coded as DRG level i.

$$\text{Additional budget} = 2\% * (\text{pop DRG-L1}) * [\text{mean's fee DRG-L2 - mean's fee DRG-L1}] \quad (6)$$

This data-manipulation is estimated to account for nearly 560 million euros in additional annual hospital healthcare cost, so approximately 1.4% of the annual budget countrywide.

With **Context B,** the fixed-envelope is equal to $x+y$.

- Hospital A gets a funding $[(x+\alpha)/(x+y+\alpha)]*(x+y) = x+[(x+\alpha y)/(x+y+\alpha)] > x$
- Hospital B gets a funding $[y/(x+y+\alpha)]*(x+y) = y - [(\alpha y)/(x+y+\alpha)] < y$

The "data-manipulation" behavior of Hospital A increases the funding of this hospital by decreasing the funding of Hospital B from the same amount. The total budget for acute activities (T2A budget)[xi] in French hospital was around 40 billion euros. For the non-teaching public hospital, the cost of a lesser tendency to data-manipulation was 38 million euros.[xii] This amount appears to be quite not significant. However, here, we discuss on the budget of small size hospitals that already have financially to struggle to survive.

Discussion

The DRG basis prospective payment reform

As the big shift in classification took place in 2009, the database includes periods both before and after the change in classification. I was thus able to study the effect of an exogenous shock on upcoding practices__ from a possible initial situation of downcoding__ with regard to severity. However, the T2A mechanism was applied to all stays in for-profit hospitals beginning in 2005, whereas it was introduced gradually to public sector hospitals between 2004 and 2008.

It was not until 2008 that the T2A mechanism was applied to 100% of acute care stays in all types of hospitals. As a robustness check, I here focus only on the period from the year 2008 (before the 2009 policy change but when financing was 100% based on the T2A in all hospitals) to the year 2010 (after the change). This econometric analysis is thus conducted on a period of time where all hospitals were financed through the same DRG-based mechanism. The results are presented in Appendix A, Tables A1, A2 and A3. The results are unchanged.

I now turn to the formalization presented in this paper. The description of the reimbursement system in Section 2 assumes that all hospitals are priced in the same way, but the formalization presented above is a simplification. DRG-based reimbursement follows the exact same logic for all sectors but there exists differences in the calculation of lump-sum transfers. The national health insurance system yields a fixed price for every DRG,[xiii] but there are two pricing grids: one for the public sector and one for the private sector.[xiv] As a consequence, prices do not converge, although there is competition between health care institutions in the two sectors. In this paper, econometric analysis is conducted controlling for hospital fixed effects crossed with the trend. Then, I control for this difference in pricing grids.

In 2008, the percentage of financing based on the T2A prospective price mechanism reached 100% of acute care stays, but the fee was corrected by a so-called transition coefficient which was specific to each hospital, and which was aimed at correcting inequalities due to the previous global budget mechanism. This coefficient was eliminated in March 2011. In this paper, first the coefficient was applied for the whole period of time of the data used. Second, econometric

analysis is conducted controlling for hospital fixed effects. Because this transition coefficient is specific to each hospital, it is captured by the hospital fixed effect.

In the simple theoretical model used here, the same objective function is applied to all hospitals. This assumption is debatable. A more flexible assumption would be to introduce altruism coefficients: a specific altruism coefficient for each hospital type. Actually, in the econometric model, I opted for a more flexible framework, considering that hospital responses to 2009 policy may differ by hospital type.

Yardstick competition works when the regulator can credibly commit not to cover hospital deficits, which is only partially credible in the French system. While the extreme scenario of closing an hospital due to deficits is not totally credible, getting into a difficult financial situation can have very severe and concrete consequences for hospitals: elimination of units or transformation into a long-term care center. It may also lead to postpone investments, to stop recruitment including not replacing leaving employees.

Why not use a probit or logit model? Least-squares coefficients measure changes in expected levels of severity, while logistic coefficients measure changes in the log odds that the level of severity is equal to 1. The interpretation of a measure of changes in the expected values of the level of severity is direct. However, a probit model was also used to regress the model parameters (results available on request). Of course, the results of least squares and probit models are not directly comparable; however, the corresponding significance tests are. The null

hypotheses for least squares and probit models are identical. As Pohlman and Leitner (2003) pointed out, both models can be used to test relationships with a binary criterion.

To take into account the correlation between the different levels of severity, we used an ordered profit Model. The constancy of threshold parameters is here assumed. However, we have to check whether the model violates the parallel regression assumptions or proportional odds assumption. In such a case, we should consider other models as a generalized ordered logit model. Table A5 presents the BIC's test and the AIC's tests. All of the tests agree that the proportional odds assumption is not violated with such large dataset.

Ambulatory surgery

Thus far, I have considered the four levels of severity. One objective of the French Health Ministry has been to encourage recourse to ambulatory surgery. Patients at a low level of severity, under specific conditions, are eligible for ambulatory surgery. The downward trend observed in admissions for stays coded as low-severity could thus have been due to the rise in ambulatory surgery. To give a general overview, Table 2 presents the distribution of hospital stays by hospital type for stays. In private for-profit establishments, more than 26% of stays were associated to day surgery, and more than 25% were associated to severity level 1. Stays with DRGs coded as ambulatory surgery represent between 5% and 7% of stays in the other types of hospitals. In Appendix A, Table A4 displays the results obtained with a set of stays associated to DRGs of severity levels 1 to 4 and DRGs for ambulatory surgery.[xv]

The inclusion of ambulatory surgery does not change the previously observed pattern of results (Table A4). The change in the classification system negatively affected the probability of being

associated to a DRG of low severity versus a higher level of severity (2, 3 or 4) or day surgery.

Insert Table 2 about here

Conclusion

The analysis here is based on the level of severity that would have been associated to the stay in the new classification, before it actually came into effect. My focus is to assess how far the 2009 classification change in coding offered hospitals opportunities to upcode; I also examined the role of hospital ownership status.

The empirical literature on upcoding is almost exclusively focused on the United States. It is thus natural to ask whether upcoding is a US phenomenon, or whether it can be found in other health systems as well. This question is particularly important given that many countries have opted for a refined DRG classification system over the last decade. This paper examines hospital responses to changes in DRG classification in the context of a DRG payment system.

As the single-payer public health insurance system strengthens financial incentives for hospital efficiency in healthcare production, this question is crucial. The DRG system of fixed prices at the stay level combined with a fixed national envelope for all stays in all hospitals creates incentives for hospitals in both the public and private sectors to increase the volume of admissions in absolute and to increase the number of diagnoses per stay to switch stays into more profitable in DRGs (higher level of severity).

This study illustrates how a simple change in the French DRG classification system in 2009 generated large and exogenous changes in the coded level of severity of in-patient stays. Hospitals responded to this policy by manipulating the coding on patients to get DRGs associated to larger reimbursement. This practice is not an issue when coding practices are uniform across hospitals. Differential upcoding could be seen as attenuation in 'downcoding" caused by more efficient billing systems. Here, the big issue is that the upcoding behavior differs in function of the hospital's ownership whereas hospitals under the same reimbursement system, have the same incentive to manipulate the raw information. Therefore, we find support that the observed effect of a change to the DRG coding system is beyond a reduction in "downcoding" but rather an opportunity to manipulate the coding and hence, the data.

Following the policy change, for-profit facilities availed themselves of this opportunity to the greatest extent. An implication of these results is that upcoding led to a budget reallocation that increased the share of total health spending going to for-profit hospitals at the expense of public non-research hospitals. I also find that the hospitals the least equipped with high-tech equipment—non-teaching public hospitals—were the least likely to engage in upcoding data, with corresponding negative effects on their share of total health spending. Taken together, these findings echo those of studies such as ProPAC (2005) and Scanlon (2006). According to Scanlon, DRG refinements favor research hospitals at the expense of smaller ones, mainly hospitals in rural areas. This may jeopardize the financial situation of the latter, which often struggle to remain financially viable.

A second implication of these results relates to in-patients with low-severity diagnoses. I find that the propensity to code cases as high-severity increased following the implementation of the

policy in 2009. The French National assembly votes, every year, a fixed envelope at the national level. Because of this fixed annual envelope, this led to a decrease in the fee paid for patients with a low-severity diagnosis, for the same intensity of treatment.

A third implication is on the use of number of relative units per DRG to measure the intensity of care. Upcoding practices artificially increase the intensity of care (number of relative units per DRG). As a result, number of relative units per DRG is not per se relevant to judge the intensity of care or services provided by the hospital. The refinement of DRG classification can be said to be useful as a mechanism for cost control, but it does not remove the need for a thorough auditing of hospital coding practices.

As a general conclusion, the 2009 policy lead to upcoding (from an potential initial context of downcoding) disconnected from any changes in production lines. Therefore, these results are free of any effects on healthcare quality but it highlights the financial effects on hospitals and in particular, the negative impact on hospitals in rural areas. This research emphasizes the need to anticipate the impacts of any changes in the DRG classification on public and private-sector actions. Besides, a general comment here is how long French hospitals will be able to absorb the increase in number of admissions with a very short increase in the annual envelope. What are the implication in terms of healthcare and decision of discharge?

Compliance with Ethical Standards:
No funding

Conflict of Interest:
The authors declare that they have no conflict of interest.

REFERENCES

Barros P. and Braun G. 2017. Upcoding in a National Health Service: the evidence from Portugal, 26(5)

Brekke, K. R., Siciliani L., and Straume O. R. 2011. Hospital Competition and Quality with Regulated Prices, *Scandinavian Journal of Economics*, 113(2): 444-469.

Brekke, K. R., Siciliani L., and Straume O. R. 2012. Quality competition with profit constraints, *Journal of Economic Behavior and Organization*, 84(2): 642-659

Carter, G. and Newhouse, J. and Relles, D. 1990. A. How Much Change in the Case Mix Index is DRG Creep? *The RAND Corporation*. R-3826-HCFA.

Carter, G. and Ginsburg, P. 1985. The Medicare Case Mix Index Increase: Medical Practice Changes, Aging and DRG Creep. *The RAND Corporation*. R-3292-HCFA.

Chalkley, M. and Malcomson, J. M. 1998. Contracting for Health Services when Patient Demand Does Not Reflect Quality, *Journal of Health Economics* 17, 1-19.

Currie, J. and MacLeod W. B. 2013. Diagnosis and Unnecessary Procedure Use: Evidence from C-sections, *NBER Working Paper No. 18977*.

Dafny, Leemore S. 2005. How Do Hospitals Respond to Price Changes? *The American*

Economic Review, Vol. 95(5), pp. 1525-1547

DREES (Direction de la Recherche, Etudes, Evaluation et Statistiques) (2012). Le panorama des établissements de santé , Editions 2012

DREES (Direction de la Recherche, Etudes, Evaluation et Statistiques) (2012). Evolution de l'offre et des prises en charge hospitalières entre 2001 et 2009: technicité croissante pour des séjours plus courts, *Dossier Solidarité et Santé*, 25.

Duggan, M. 2002. Hospital Market Structure And The Behavior Of Not- For-Profit Hospitals. *Rand Journal of Economics*, 2002, 33(3)

Ellis, R. P. and McGuire, T. 1986. Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply. *Journal of Health Economics*, 5, 129-151.

Georgescu, I. and Hartmann, F.G.H. 2013. Sources of financial pressure and up coding behavior in French public hospitals. *Health Policy*, 110 (2-3):156-163.

Gilman, Boyd H. 2000. Hospital Response to DRG Refinements: The Impact of Multiple Reimbursement Incentives on Inpatient Length of Stay? *Health Economics*, 9(4): 277-94

Gobillon L. and Milcent C. 2015. The competition effect of a French reform on hospital quality. Working Paper PSE

Hakkinen et al. 2012. Patient classification and hospital costs of care for acute myocardial infarction in nine European countries. Häkkinen U, Chiarello P, Cots F, Peltola M, Rättö H; EuroDRG group. *Health Economics*, 21 Suppl 2:19-29

Hsia, D. C., Krushat, W. Fagan, A. Tebbutt, J. and 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.

Kuhn, M. and Siciliani L. 2008. Upcoding and optimal auditing in health care (or the economics of DRG creep. *CEPR Discussion Paper*, N.6689.

Ma, C. 1994. Health Care Payment Systems: Cost and Quality Incentives? Jou*rnal of Economics and Management Strategy*, 3: 93-112

Ma, C. 1998. Cost and Quality Incentive in Health Care: A Reply. *Journal of Economics and Management Strategy*. 7: 139-142.

Mason A., Zeynep O., Renaud Th., Street A., Thuilliez J., Ward P. 2012. How Well do diagnosis- related groups for appendectomy explain variations in resource use? An analysis of patient-level data from 10 European countries. *Health Economics* (Suppl. 2): 30-40

Milcent, C. 2017. Premier bilan de la T2A sur la variabilité des coûts hospitaliers. *Economie et Prévision*. 210

Mougeot, M. and Naegelen, F. 2008. Supply-side risk adjustment and outlier payment policy. *Journal of Health Economics*. 27(5): 1196-1200

Newhouse, Joseph P. 2002. Pricing the Priceless: A Health Care Conundrum; Cambridge: MIT Press.

Newhouse J.P. 2003. Reimbursing for Health Care Services? *Economie publique* 13: 3-31.

Pohlman and Leitner. 2003. A comparison of Ordinary Least Squares and Logistic Regression. The Ohio Journal of Science, 103(5):118-125.

Pope, G. 1989. Hospital NonPrice Competition and Medicare Reimbursement Policy. *Journal of Health Economics,* 8: 147-172

MedPAC. 2005. A DataBook: Healthcare Spending and the Medicare Program (Washington: Med- PAC, June 2005),102.

Scanlon W.J. 2006. The Future Of Medicare Hospital Payment Modest Proposals in Light of Medicare's Challenges. *Health Affairs* 25(1): 70-80.

Shleifer, A. 1985. A Theory of Yardstick Competition. *The RAND Journal of Economics*, 16(3): 319-327

Siciliani, L. 2006. Selection of treatment under prospective payment systems in the hospital sector? *Journal of Health Economics*, 25(3): 479-499.

Silverman, E., Skinner, J. and Fisher, E. 1999. The Association between for-profit Hospital Ownership and Increased Medicare Spending. *The New England Journal of Medicine*, 1999, 341(6): 420-26.

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

Simborg, D.W. 1981. DRG creep: a new hospital-acquired disease. *New England Journal of Medicine*, 304

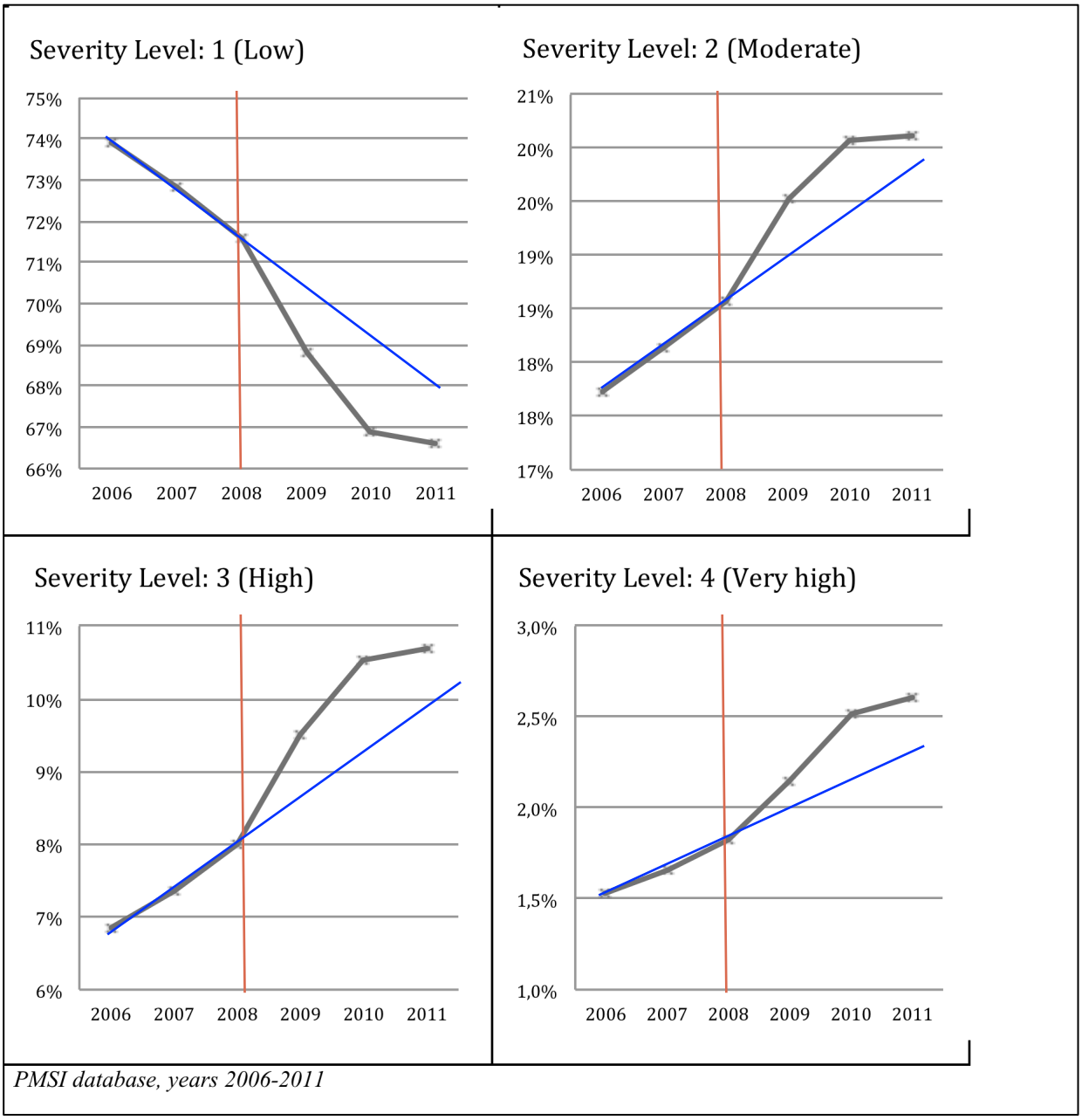
Steinbusch, P.J., Oostenbrink, J.B., Zuurbier, J.J., Schaepkens, F.J. 2007. The risk of up-coding in casemix systems: a comparative study. *Health Policy*, 81

Steinwald, B. and Dummit, L. 1989. Hospital Case-Mix Change: Sicker Patients or DRG Creep? *Health Affairs*. Summer.

Xirasagar, S. and Lin, Herng-Ching. 2006. Effects of Payment Incentives, Hospitals Ownership and Competition on Hospitalization Decisions for Ambulatory Surgical Procedure. *Health*

Policy, 76: 26-37

Figure 1: Trend by level of Severity



PMSI database, years 2006-2011

Table 1: Hospital ownership and severity level (%)

Repartition of hospital ownership by severity level (%)			
	Low Severity Level: 1	Higher Severity Level: 2, 3, 4	All severity
Research hosp.	32,80%	45,26%	36,19%
Other pub. hosp.	20,77%	22,64%	21,64%
NFP hospitals	9,30%	5,89%	7,11%
FP Hospitals	37,13%	25,99%	35,28%
Whole sample	100%	100%	100%
Repartition of severity level by hospital ownership (%)			
	Low Severity Level: 1	Higher Severity Level: 2, 3, 4	All severity
Research hosp.	64,89%	35,11%	100%
Other pub. hosp.	66,23%	33,77%	100%
NFP hospitals	67,72%	32,28%	100%
FP Hospitals	78,40%	21,60%	100%
Whole sample	69,85%	30,15%	100%

PMSI database, years 2006-2011

Notes: Higher level of severity: Moderate (severity level 2), High (severity level 3), Very high (severity level 4)

Table 2: Repartition of DRG's type by hospital's ownership (%)

(%)	Low Severity Level: 1	Higher Severity Level: 2, 3, 4	Death
Research hosp.	30,55%	16,53%	0,23%
Other pub. hosp.	28,07%	13,37%	0,19%
NFP hospitals	17,07%	8,70%	0,06%
FP Hospitals	25,59%	7,05%	0,03%
Whole sample	26,94%	11,62%	0,13%
(%)	Ambulatory surgery	Very short stays	Exploratory stays and sessions
Research hosp.	6,11%	13,99%	32,59%
Other pub. hosp.	5,21%	17,44%	35,72%

NFP hospitals	6,95%	7,32%	59,89%
FP Hospitals	26,87%	4,25%	36,21%
Whole sample	13,71%	10,47%	37,12%

PMSI database, years 2006-2011

Notes: Higher level of severity: Moderate (severity level 2), High (severity level 3), Very high (severity level 4)

Table 3: Least square model

		Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Y : indicator for level of severity 1_Low		**Sample including stays with DRG of Severity level 1 to 4**				
Trend		-0.0098*** (0.0003)	-0.0078*** (0.0001)	-0.0074*** (0.0001)	-	-
DRG classification changes (I(.))		-0.0216*** (0.0003)	-0.0213*** (0.0003)	-0.0213*** (0.0003)	-0.0210*** (0.0003)	-
DRG classification changes (I(.)) *crossed*	Teaching hospital	-	-	-	-	-0.0212*** (0.0004)
	Other Public hosp.	-	-	-	-	-0.0178*** (0.0010)
	Non-Profit hosp.	-	-	-	-	-0.0220*** (0.0005)
	For-Profit hosp.	-	-	-	-	-0.0271*** (0.0005)
Dummies	Pathology	NO	YES	YES	YES	YES
	Hospital	NO	NO	YES	YES	YES
Hospital dummies crossed trend		NO	NO	NO	YES	YES

PMSI database, years 2006-2011

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%.
Model(1): Y= f(Trend, classification DRGs' change)
Model(2): Y= f(Trend, classification DRGs' change, Dummies_pathology)
Model(3): Y= f(Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend)
Models with no dummy pathology include pathology random effect
Models with no hospital fixed effect include hospital random effect

Table 4: Ordered probit model (Coefficients)

		Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Y : indicator for level of severity (from 1 to 4)		**Sample including stays with DRG of Severity level 1 to 4**				
Trend		0.0336*** (0.0002)	0.0309*** (0.0002)	0.0300*** (0.0002)	-	-
DRG classification changes (I(.))		0.0601*** (0.0007)	0.0640*** (0.0007)	0.0653*** (0.0007)	0.0650*** (0.0007)	-
DRG classification changes (I(.)) *crossed*	Teaching hospital	-		-	-	0.0630*** (0.00114)
	Other Public hosp.	-		-	-	0.0575*** (0.0009)
	Non-Profit hosp.	-		-	-	0.0772*** (0.0011)
	For-Profit hosp.	-		-	-	0.1069*** (0.0011)
Dummies	Pathology	NO	YES	YES	YES	YES
	Hospital	NO	NO	YES	YES	YES
Hospital crossed trend		NO		NO	YES	YES
Cut-point		0.6446 (0.0003)	-0.6682 (0.0056)	-0.6922 (0.005619)	-0.7134 (0.0083)	-0.7634 (0.0068)
Cut-point		1.3514 (0.0004)	0.0837 (0.0058)	0.0780 (0.0056)	0.0699 (0.0058)	-0.0640 (0.0068)
Cut-point		2.1657 (0.0005)	0.9414 (0.0056)	0.9531 (0.0056)	0.9138 (0.0055)	0.8804 (0.0068)

PMSI database, years 2006-2011

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%.
Model(1): Y= f(Trend, classification DRGs' change)
Model(2): Y= f(Trend, classification DRGs' change, Dummies_pathology)
Model(3): Y= f(Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend)
Models with no dummy pathology include pathology random effect
Models with no hospital fixed effect include hospital random effect

Table 5: Ordered probit model (Marginal effect)

		Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Y : indicator for level of severity (from 1 to 4)		**Sample including stays with DRG of Severity level 1 to 4**				
DRG classification changes (I(.))		-0.0208*** (0.0002)	-0.0208*** (0.0002)	-0.0207*** (0.0002)	-0.0207*** (0.0002)	-
DRG classification changes (I(.)) *crossed*	Teaching hospital	-	-	-	-	- 0.0209*** (0.0004)
	Other Public hosp.	-	-	-	-	- 0.0177*** (0.0003)
	Non-Profit hosp.	-	-	-	-	-0.0222*** (0.0004)
	For-Profit hosp.	-	-	-	-	- 0.0289*** (0.0007)

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%. ; Standard Error obtained with Delta method in parenthese.

Appendix A

Table A1: Least square model_Sample from year 2008 to year 2010

		Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Y : indicator for level of severity 1_Low		Sample including stays with DRG of Severity level 1 to 4				
Trend		-0.0080*** (0.00007)	-0.0078*** (0.00007)	-0.0075*** (0.0002)	-	-
DRG classification changes (I(.))		-0.0240*** (0.0002)	-0.0229*** (0.0002)	-0.0218*** (0.0002)	-0.0214*** (0.0002)	-
DRG classification changes (I(.)) crossed	Teaching hospital	-	-	-	-	-0.0220*** (0.0005)
	Other Public hosp.	-	-	-	-	-0.0178*** (0.0010)
	Non-Profit hosp.	-	-	-	-	-0.0228*** (0.0005)
	For-Profit hosp.	-	-	-	-	-0.0271*** (0.0005)
Dummies	Pathology	NO	YES	YES	YES	YES
	Hospital	NO	NO	YES	YES	YES
Hospital crossed trend		NO	NO	NO	YES	YES

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%.
Model(1): Y= f(Trend, classification DRGs' change)
Model(2): Y= f(Trend, classification DRGs' change, Dummies_pathology)
Model(3): Y= f(Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend)
Models with no dummy pathology include pathology random effect
Models with no hospital fixed effect include hospital random effect

Table A2: Ordered probit model (Coefficients)_Sample from year 2008 to year 2010

		Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Y : indicator for level of severity (from 1 to 4)		Sample including stays with DRG of Severity level 1 to 4				
Trend		0.0312*** (0.0003)	0.0306*** (0.0003)	0.0300*** (0.0002)	-	-
DRG classification changes (I(.))		0.0610*** (0.0007)	0.0659*** (0.0008)	0.0657*** (0.0007)	0.0652*** (0.0008)	-
DRG classification changes (I(.)) crossed	Teaching hospital	-	-	-	-	0.0648*** (0.0009)
	Other Public hosp.	-	-	-	-	0.0571*** (0.0008)
	Non-Profit hosp.	-	-	-	-	0.0706*** (0.0009)
	For-Profit hosp.	-	-	-	-	0.9180*** (0.0015)
Dummies	Pathology	NO	YES	YES	YES	YES
	Hospital	NO	NO	YES	YES	YES
Hospital crossed trend		NO	NO	NO	YES	YES
Cut-point		0.6389 (0.0007)	-0.7489 (0.0067)	-0.7605 (0.0068)	-0.7624 (0.0068)	-0.7464 (0.0068)
Cut-point		1.3336 (0.0007)	-0.0101 (0.0067)	-0.0035 (0.0068)	-0.0400 (0.0068)	-0.0638 (0.0068)
Cut-point		2.1583 (0.0008)	0.8587 (0.0067)	0.8833 (0.0068)	0.8540 (0.0068)	0.8204 (0.0068)

Notes: *significant at 1% level. *** Significant at 0.1%. Dependent variable = 1 if patient's stay is classified as level of severity j
Model(1): Y= f(Trend, classification DRGs' change)
Model(2): Y= f(Trend, classification DRGs' change, Dummies_pathology)
Model(3): Y= f(Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend)

Models with no dummy pathology include pathology random effect
Models with no hospital fixed effect include hospital random effect

Table A3: Ordered probit model (Marginal effects) _ Sample from year 2008 to year 2010

		Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Y : indicator for level of severity (from 1 to 4)		Sample including stays with DRG of Severity level 1 to 4				
DRG classification changes (I(.))		-0.0215*** (0.0003)	-0.0213*** (0.0002)	-0.0211*** (0.0002)	-0.0211*** (0.0002)	-
DRG classification changes (I(.)) crossed	Teaching hospital	-		-	-	- 0.0220*** (0.0003)
	Other Public hosp.	-	-	-	-	- 0.0173*** (0.0001)
	Non-Profit hosp.	-	-	-	-	- 0.0226*** (0.0003)
	For-Profit hosp.	-	-	-	-	-0.0288*** (0.0007)

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%. ; Standard Error obtained with Delta method in parenthese.

Table A4: Least square model

Y : indicator for level of severity 1_Low		Sample including stays with DRG of Severity level 1 to 4 and day-surgery			
Trend		-0.0080*** (0.0000)	-0.0078*** (0.0001)	-	-
DRG classification changes (I(.))		-0.0228*** (0.0002)	-0.0212*** (0.0002)	-0.0204*** (0.0005)	-
DRG classification changes (I(.)) crossed	Teaching hospital	-	-	-	-0.0220*** (0.0006)
	Other Public hosp.	-	-	-	-0.0178*** (0.0007)
	Non-Profit hosp.	-	-	-	-0.0226*** (0.0005)
	For-Profit hosp.	-	-	-	-0.0271*** (0.0005)
Dummies	Pathology	NO	YES	YES	YES
	Hospital	NO	NO	YES	YES
Y : indicator for level of severity 1_Low		Sample including all stays			
Trend		-0.0079*** (0.0000)	-0.0078*** (0.0000)	-	-
DRG classification changes (I(.))		-0.0218*** (0.0002)	-0.0211*** (0.0003)	-0.0204*** (0.0003)	-
DRG classification changes (I(.)) crossed	Teaching hospital	-	-	-	-0.0217*** (0.0006)
	Other Public hosp.	-	-	-	-0.0176*** (0.0008)
	Non-Profit hosp.	-	-	-	-0.0221*** (0.0005)
	For-Profit hosp.	-	-	-	-0.0271*** (0.0005)
Dummies	Pathology	NO	YES	YES	YES
	Hospital	NO	NO	YES	YES

Notes: *significant at 1% level. *** Significant at 0.1%. Dependent variable = 1 if patient's stay is classified as level of severity j
Model(1): Yj= f(Trend, classification DRGs' change)
Model(2): Yj= f(Trend, classification DRGs' change, Dummies_pathology)
Model(3): Yj= f(Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend)
Models with no dummy pathology and/or hospital include random effect

Table A5: Information criteria

	LL(null)	Akaike's information	Bayesian Information

		criterion	criterion
Ordered Probit Model	- 3.42e+07	6.84e+07	6.84e+07
Generalized ordered Probit Model	- 3.42e+07	6.83e+07	6.83e+07

Notes: Based on Model 2

Appendix B: French DRG 2009-classification and DRG based payment

1. Description of the new French DRG coding system

2009-French DRG classification system is organized in a nested fashion. DRGs are coded by a series of 6 characters. The first two characters define the pathology; the third defines the type of medical practice (medical, surgical, or exploratory); the fourth and fifth characters define the subfields of a given pathology; and the sixth character defines the level of severity of which there are four. The level of severity is independent of the procedures performed on the patient.

The "root" of a DRG consists of the first 5 characters in the corresponding code. There are 617 root DRGs in 2009-classification. Among these roots, 76.2% (470) are subdivided into 4 levels of severity. In addition, certain DRG roots are not subdivided into four severity levels, or are not subdivided by severity at all. Moreover, there exist, in some specific cases, other types of subdivisions that replace or complement subdivision by severity (Manuel des GHM, 2012).

This classification defines four levels of severity

- Level 1: stays without associated complications or morbidity (ACM) belonging to a list validated by consultative and decisional bodies of the national health statistics agency. In what follows, this level is called low or non-severe.

- Level 2, 3 or 4: The list of ACMs contains three sub-lists corresponding to three levels of severity. An ACM always belongs to a single severity level, which does not change. In certain cases, death can also play the role of ACM. In the grouping algorithm, the order of priority is as follows. – Existence of an ACM and its placement at a given level (2, 3, or 4) ; – Patient's age (<2 years, >69 years and >79 years) ; – Death if neither of the two preceding conditions is satisfied; – Duration of stay, which is a mandatory condition.

- Other types of subdivisions
  - The letter "Z" indicates an exploratory stay, with no coding of severity. These stays are almost exclusively chemotherapy sessions and dialysis sessions;
  - The letter "E" indicates stays ending in death;

- The letter ”J” identifies ambulatory surgical stays (no overnight stay) ;
- The letter ”T” corresponds to short stays, between 1 and 3 days. These must be distinguished from stays coded by a ”J”. These are almost exclusively medical stays.

In the primitive 2009-classification of the coding system (before any marginal adaptive changes), used here, stays were classified into 2,192 DRGs.

2. DRG prospective payment system

DRGs are used as a tool to measure the relative resource requirements for hospital care of the population. Hospitals are paid a fixed price for each patient admitted in hospital. At discharge, every patient's stay is assigned to one of the DRGs based on their routinely registered primary diagnosis. At each DRG is associated a weight reflecting the average cost of patients in the given DRG relative to that of the average patient reimbursed by the system (MedPAC, 2007). The price for each patient treated is obtained by multiplying the relevant DRG weight by a fixed monetary value. With the DRG system, hospitals are expected to reduce costs and under certain conditions, to incent increase in activity (Sicilliani, 2012).

They were first adopted in 1983 for the reimbursement of care provided to elderly patients under the US Medicare Program. Since then, the development of DRG systems and DRG-based hospital payments has become an international phenomenon (Kimberly et al., 2008). In Europe, prospective payment based on DRGs was first introduced in Portugal in the late 1980s, and more recently in other European countries such as England, France, and Germany. The principal reason for the popularity of DRG-based hospital payment systems is that they are supposed to lead to eciency: reduced costs for the optimum level of quality.

DRG-based hospital payment creates three main incentives (Cots et al., 2011).

- One of the incentives of DRG-based payment is for hospitals to reduce costs per patient. This can happen through: i) reductions in lengths of stay; ii) reductions in the intensity of services provided; and iii) patient selection. Using French data, the empirical literature showed that in public hospitals both the number of cases treated and case-mix-adjusted production increased significantly between 2004 and 2009. This coincided with a decrease in lengths of stay (DREES (2012), Or et al. (2013), and Gobillon and Milcent (2015)).

- The second incentive is to increase the number of patients: the reduction of the waiting list, splitting care episodes into multiple admissions, admitting patients for unnecessary services (known as induced demand), and improving hospital reputation. According to

DREES (2012), the volume of admissions increased by about 14,6% over the period 2001-2009.

- The third is to increase revenue per stay. Hospitals can achieved this using one of two strategies: (i) changes in coding practices, or (ii) changes in practice patterns. The aim of both strategies is to reclassify patients into higher-severity DRGs with a higher associated payment rate, a phenomenon known as "upcoding". With the data used here, which includes 15 million observations, the effect of changes in coding practices can be isolated and assessed purely independently of all other incentives.

[i] In Health Economic literature, this shift is named upcoding.

[ii] Calculations based on the data used here yield the same results.

[iii] A transformation of the mean cost is actually used

[iv] The French reimbursement system is discussed in Section 5.

[v] Agence Technique de l'Information sur l'Hospitalisation (ATIH)

[vi] Details on the French DRG classification are given in Appendix B

[vii] In the robustness checks (Section 5), the model was run using a more restricted sample from the 2008-2010 period.

[viii] The only potential impact of the 2009 DRG change in coding on the actual production of care could have been through an increase in the length of stay (see Appendix B). The empirical literature does not find support this assumption. A variety of papers have found a decrease of the length of stay over the observed period (DREES (2012), Or et al. (2013), and Gobillon and Milcent (2015)).

[ix] Table 4 for coefficients and Table 5 for the marginal effects. Here, the marginal effect is an estimate of a population-averaged marginal effect. The reference is severity level 1, the lowest level. Note that a positive sign of any coefficients implies a higher probability of belonging to the higher category, corresponding in principle to greater severity.

[x] Here, the marginal effect is an estimate of a population-averaged marginal effect

[xi] For an overview of the methods used, see the IGF report, French Finance Ministry (2011), http://www.igf.finances.gouv.fr/webdav/site/igf/shared/NosRapports/ documents/2012/2011-M-056-01.pdf

[xii] The effect of "upcoding" behavior is standardised on the basis on the non-teaching public hospital "upcoding" behavior. Then, I compute the average "upcoding" behavior of the other hospital's ownership compared with the nonteaching public hospital one. I multiplied this effect by average fee for patient coded as DRG level 2. I then got the average effect of the"upcoding" behavior for one patient coded as DRG level 1. To get the global effect, I multiplied this number by the number of patient coded as DRG level 1 admitted in hospitals except those admitted in the non-teaching public hospital.

[xiii] This is known as the "tarif opposable de l'assurance maladie" [enforceable price to the national health insurance system], and is set for a given diagnosis-related group (DRG)

[xiv] The base for calculating the costs of DRGs differs between the two groups. Doctors' fees, as well as laboratory services, imaging, and function tests are included in the calculation of cost per DRG in the public sector, whereas they are not for the private sector. These prices are national and are published annually by the Ministry of Health.

[xv] As a robustness check, I also ran a least squares model including all admissions to acute care units. The results were unchanged (Table A4).