



HAL
open science

Upcoding and heterogeneity in hospitals' response: A Natural Experiment

Carine Milcent

► **To cite this version:**

Carine Milcent. Upcoding and heterogeneity in hospitals' response: A Natural Experiment . 2016.
halshs-01340557

HAL Id: halshs-01340557

<https://shs.hal.science/halshs-01340557>

Preprint submitted on 1 Jul 2016

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
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2016 – 13

**Upcoding and heterogeneity in hospitals' response:
A Natural Experiment**

Carine Milcent

JEL Codes: I18

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



PARIS-JOURDAN SCIENCES ÉCONOMIQUES

48, Bd JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

Upcoding and heterogeneity in hospitals' response: A Natural Experiment

Carine Milcent *

CNRS, PSE

June 28, 2016

Abstract

How has this administrative change affected the healthcare providers behaviour? Using a unique longitudinal database with 145 million stays, I study the dependence of the severity classification associated with hospital stays on a financial incentive, as well as the resulting budgetary reallocations. The classification of diagnosis-related groups (DRGs) in France changed in 2009. The number of groups was multiplied by 4. Controlling for pathology indicators and hospital fixed effects, I unambiguously demonstrate that a finer classification led to an “upcoding” of stays. Because of a fixed annual budget at the national level, these results directly imply that the upcoding led to a budget reallocation which increased the share of health spending that went to for-profit hospitals, at the expense of public non-research hospitals. This budget reallocation did not correspond to any change in the actual production of care.

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

JEL Classification: I18.

*milcent@pse.ens.fr

1 Introduction

Upcoding phenomenon is almost exclusively studied in the Medicare relationship with the US hospitals. The Diagnosis Related Groups (DRG) prospective payment system (PPS) was firstly introduced in 1983 by the US Medicare Program. From the implementation of the DRGs-based payment, upcoding i.e. shifting from one DRG to a higher profitable one, is one of the major concerns. What can we say when healthcare is mainly provided by the public sector? The 2009 change in DRG coding created a natural experiment for testing whether French hospitals enhance profits by taking advantage of regulatory loopholes with an upcoding behavior over the period 2006-2011. In this paper, I also examine how hospital ownership affected upcoding behavior.

An equitable payment system should encourage more accurate coding, but not allow coding changes to generate additional payments unrelated to patient needs, while being able to recognize and pay for increases in patient care requirements when they occur (Steinwald and Dummit, 1989). 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 costs.

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's 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, 2015). The current debate in the United States on the potential impact of further refinement of DRGs on certain hospitals is instructive. Refinements consist in splitting a single DRG category into more DRG categories relating to the same primary diagnosis. The split may be based either on the type of treatment the patient receives

or on secondary diagnoses. A study carried out for MedPAC in 2005 demonstrated large variations in profitability for a given DRG. In a given clinical context (notably principal diagnoses), DRGs relating to surgery were found to be more profitable than those for medical treatments, notably for orthopedic and cardiovascular care. 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. On the other hand, refinement increased the complexity of the DRG classification, requiring a very detailed and sophisticated information base, with an increased risk of mispayment if the information is not accurate. According to the theory, as the level of aggregation increases, the power of the contracting method increases. The logic is straightforward. The manager's incentive to exert unobserved effort to economize on production costs only extends to the product that is priced; hence, 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).

This research contributes to the literature on providers in healthcare (Ellis and McGuire, 1986; Pope, 1989; Ma, 1994, 1998; Chalkley and Malcomson, 2000; Mougeot and Naegelen, 2008; Brekke et al., 2011). Models of hospital behavior 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 evidently hospitals' attempts to increase revenues through fraudulent coding practices, leading to unjustified payments, are not (Simborg, 1981; Steinbusch et al., 2007; Silverman and Skinner, 2004). Kuhn and Sicilliani (2008) recently offered a theory of upcoding. They "assume that providers can increase demand by increasing quality but can also inflate activity through a manipulative effort (upcoding or

DRG creep).” They compare the optimal price and audit policy for the purchaser under two scenarios : commitment or no commitment to a given audit policy.

Most of the attention in the empirical literature has been 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. Dafny (2005) exploited a 1988 policy reform that generated large price changes in Medicare admissions. He found that the upcoding response was stronger among for-profit hospitals. While there is a large empirical literature focusing on US hospitals in the context of the Medicare system, there are almost no empirical studies in the context of a national public healthcare sector. Barros and Braun (2013), following Dafny (2005), explored the link between upcoding and price increases in the context of such a national health service (that of Portugal). Researchers have attempted to measure the role of upcoding in increased case-mix indices following the implementation of a PPS in US Medicare (Carter and Ginsburg, 1985; Hsia et al., 1988; Carter et al., 1990). As explained by Carter et al. (1990), along with true change and upcoding, the measured case mix can also be affected by changes in the “grouper” program, which assigns stays to DRGs. Here, I bypass this limitation on previous assessments of upcoding effects: because the same grouper program is used for the whole period, the only change is in the coding practices of hospitals.

In France, since 2000, for each stay in each hospital, the primary diagnosis, all secondary diagnoses, co-morbidities, and procedures performed, as well as all exploratory acts, have been registered. Each hospital database is sent to the national health statistics agency, which then uses a grouper program to assign each stay to a DRG, on the basis of this information. In 2009 the French DRG classification system

changed. This change in coding is a true exogenous shock. In 2012, for internal purposes,¹ the health statistics agency ²used the new 2009-version of the algorithm to classify stays over the years 2006-2011. This database offers the unique advantage of containing information on the classification of stays both before and after a change of classification coded in a single version of the system, the 2009-version. I was thus able to “go back in time”. It can then be used to identify changes in hospitals’ coding practices without a confounding effect from changes in the algorithm of the grouper program. Here, for the first time, this database is used for scientific research purposes. My focus is to assess how far the 2009 change in coding offered hospitals opportunities to upcode; I also examined the role of hospital ownership status.

I found an unambiguous upcoding effect as hospitals’ response to a refinement of the DRG classification. Upcoding is defined here as 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 hold after controlling for pathology indicators, hospital fixed effects: hospitals are sensitive to the financial incentives created by the DRG classification. Because of the fixed national budget, this practice is not really an issue when coding practices are uniform across hospitals, but when this is not the case, it results in budgetary reallocations. I then found 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. Public non-research hospitals were the least sensitive to the change. An implication of these results is that the upcoding led to a budget reallocation which increased the share of health spending that went to for-profit hospitals at the expense of public non-research hospitals.

¹See Section 3.1 for details

²Agence Technique de l’Information sur l’Hospitalisation (ATIH)

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. The data are presented in Section 3, and the results in Section 4. Section 5 presents robustness checks. Finally, Section 6 presents concluding remarks.

2 The implications of activity-based pricing

2.1 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 hospital, except an additional fixed fee per day for catering and accommodation in for-profit hospitals.

Healthcare is mainly provided by the public sector. Until the 1980s, the public sector was largely immune from standard market forces pushing for production efficiency. Beginning in 1983, a global budget was adopted for public sector hospitals, but without the benefit of information on their real activity. In 2004, a prospective payment system based on hospitals' activity (diagnosis-related groups, or DRGs) was introduced to promote competition and efficiency. This pro-competitive hospital reform, called Tarification à l'Activité (T2A) was progressively introduced during the period 2004-2008. It was implemented with different timing in different types of hospitals. While the T2A prospective payment system has been applied since 2005 in for-profit hospitals, in 2006 the system only accounted for 35% of financing in public hospitals, which had previously been financed under a global budget mechanism; this figure rose to 50% in 2007. In 2008, this percentage reached 100%, regardless of hospital type. Since

then, in France, DRG-based payment (T2A) has been the sole mode of reimbursement of health care institutions for all medical-surgical-obstetric (MSO) stays.

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 may differ from nonprofit or government-owned hospitals, a prediction we consider in the empirical work that follows. But, how is the French hospital system structured? It consists of three types of hospitals: public, private for-profit, and private not-for-profit. This paper focuses on stays in acute care units. 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).³

2.2 Yardstick Competition

The T2A system of activity-based pricing is based 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 hospitals' level of cost reduction effort. The original 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

³Calculations based on the data used here yield the same results.

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 cost⁴; 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 ; $C_{s_{hg}}$ 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 s associated to DRG g ; and H_g denote the number of hospitals having admitted at least one patient for a stay s associated to DRG g .

The mean cost for a DRG g is defined by:
$$\bar{C}_g = \frac{\sum_{h_g} \sum_{s_{hg}=1}^{S_{hg}} C_{s_{hg}}}{\sum_{h_g=1}^{H_g} S_{hg}}$$

On the basis of the mean cost per DRG, a relative cost scale for the different DRGs can be defined. Hospital pricing rests on a comparison of the relative costs of different DRGs and not on their absolute cost. The lump sums per DRG are evaluated relative to one another. One DRG serves as a reference (the DRG for vaginal delivery without complications). The other DRGs are situated with respect to this reference DRG on the basis of their cost. DRGs are thus ordered on a scale from the least costly to the

⁴A transformation of the mean cost is actually used

most costly. DRGs can be expressed as a function of this reference DRG. For example: the ratio of the cost of the DRG “hepatic transplants, level 1” to the cost of the reference DRG is calculated, yielding a relative index. The relative index for the DRG “hepatic transplants, level 1,” for example, might be equal to 3 times that of the DRG “vaginal delivery without complications.” A relative index is associated to each DRG. The choice of reference DRG has no impact on the budget that the hospital receives at the end of the period.

Let \overline{C}_g denote the mean cost of a stay s associated to DRG g ; $\overline{C}_{g_{ref}}$ denote the mean cost of a stay s associated to the reference DRG, called g_{ref} . The relative index i_g of a DRG g is defined by: $i_g = \frac{\overline{C}_g}{\overline{C}_{g_{ref}}}$

2.3 Payment per DRG in the French context

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. Healthcare authorities thus cannot anticipate what the volume of demand would be under a pricing system that did not constrain it. Once again, the regulator cannot risk an uncontrolled increase in the volume of care.

As a result, the logic of DRG-based pricing was accompanied 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 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 .

The key to the determination of this payment (the lump sums f_g) is the fixed budget (or envelope) defined by the law on the financing of social security (LFSS) and the activity of all hospitals.

Since hospital activity is a set of stays and each stay is associated to a DRG, the hospital's total activity can be quantified as a number of relative units. A hospital's activity corresponds to the sum of the number of relative units for all stays. Total national hospital activity can be defined in the same way, as the sum of relative units for all health care institutions in France. The value of the index can be 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. This envelope F is determined by a vote in the French National Assembly. It is exogenous. G represents the number of DRGs in the classification.

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

According to Equation (1), the greater the amount of care produced by all health care 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:

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

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

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

Here, all types of hospitals (state-owned hospitals, for profit hospitals and non-profit hospitals) are considered to have a single objective, revenue maximization. During the period 2008-2010, the T2A mechanism had been completely implemented for all types of hospitals.⁵

2.4 Implications: Upcoding

There are some limitations on the applicability of Shleifer's model to the French context. Yardstick competition solves the moral hazard problem, but not the adverse selection problem due to patient heterogeneity. In 2009, the French DRG classification system changed. In the shift from version 10 to version 11 of T2A, the number of groups increased from around 800 to 2200⁶.

This change was also accompanied by a change in the logic of the construction of DRGs. The refinement of the classification introduced four levels of severity. This refinement was a way to reduce the consequences of within-DRG cost heterogeneity, providing an alternative to outlier costs to the hospitals.

⁵The period considered in the paper is 2006-2010. In 2006 and 2007, public sector hospitals were reimbursed through a mixed system, with both T2A and global budget components. This change of reimbursement system is discussed in Section 5.

⁶This number slightly fluctuated across previous versions. The exact number of DRGs in version 11c is 2,192.

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 recent 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. These points explain why the upcoding may occur in the French context.

The objective function outlined above can easily be expanded to include upcoding effects. Let $U_{S_{hg}}$ denote the degree of upcoding for a stay s associated to DRG g in hospital h . The number of relative units in hospital h can be redefined as an increasing function of the degree of upcoding. The 2009 policy change studied here involves DRGs that are particularly susceptible to upcoding because the coding of

complications results in a substantially higher price. Upcoding can 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 greater the relative index i_g of treated DRG g , the greater the hospital's budget. 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).

The French DRG-based payment system gradually introduced from 2004 onward is a pro-competition reform. The practices of competitors may also affect upcoding indirectly through pressure on hospital profits, or directly via the dissemination of upcoding practices. Returning to equation (3), *i*) 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); *ii*) 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).

Traditional simple models incorporate the intensity of care into the hospital objective function (Dafny, 2005). Because in these models the objective function is separable into these two arguments (budget and patient's intensity of care), this does not change the relationship between upcoding and the relative index i_g .

Similarly, in the budget part (the left-hand member of equation (3)), the intensity of care may be taken into account. However, in the French T2A system, the switch from a low-severity DRG to a higher-severity DRG is based only on secondary diagnoses, and not on intensity of care.⁷ The relative index i_g thus does not depend on intensity of care. Hafsteinsdottir and Siciliani (2009) studied the

⁷See Section 3 for more details.

refinement of a DRG classification. In their model, they assume that hospitals provide two alternative treatments for a given diagnosis: a less intensive (medical) treatment and a more intensive (surgical) treatment. The tariff can be based only on the diagnosis, or it can be based on both the diagnosis and the treatment. In the 2009 policy changes in France, the refinement of the DRG classification is based only on diagnoses. To conclude, a more complex model integrating intensity of care will not change the following result: according to the simple theoretical model presented here, severity should increase following such a change, *ceteris paribus*.⁸

There are many reasons why upcoding behavior may differ across hospitals and DRGs. 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 of 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 area market share of for-profit hospital is higher. Dafny (2005) also found strong evidence of upcoding, a response that was particularly strong among for-profit

⁸Dafny (2005) considers both nominal and real responses to policy changes, where "nominal" refers to hospital coding practices and "real" refers to admission volumes and intensity of actually provided care. Here, the term "real" refers only to admission volumes.

hospitals. Like Dafny (2005), I study an abrupt change in upcoding incentives which should be followed by a similarly abrupt change in upcoding if hospitals are responsive to these incentives. I also assess the dependence of the effects of the policy change on hospital ownership.

3 Data

3.1 Original data

Each year, after the fixed budget envelope for all hospitals is approved by the French Parliament, the tariff for each DRG is published. Tariffs for each DRG are based on a retrospective calculation of the value of the relative index, which itself depends on costs. Theoretically, costs from the past year are used. In reality, costs from the last two to four years are used, because of the limited sample size of the data on cost per stay. In order to maintain a longitudinal database using the current version of the DRG classification system, the health statistics agency (ATIH) uses a single version of the grouper program for different years. In 2012, they did the same thing for a longer period of time, including the years both before and after the reform.

The PMSI database used here thus contains all information on hospital stays over the period 2006-2011 classified according to the new version (version 11). This data is an exhaustive record of French hospital stays for the years 2006 to 2011, a total of 145 million stays.

3.2 A new French DRG classification

This new (2009) version of the 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. The level of severity is **independent** from the procedures

performed on the patient.⁹

The subsample of DRGs that are subdivided into four levels of severity represents approximately 40% of the database. In this subsample, 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).

A general overview of the coding of hospital stays shows that a quarter of stays in private for-profit institutions were associated to DRGs coded as ambulatory surgery. Stays of severity level 1 made up an additional quarter of admissions. In private non-profit institutions a high level of admissions for exploratory stays is also observed (59.89%). A portion of these institutions are cancer centers, which mainly perform chemotherapy sessions, which are coded as exploratory acts.

⁹Details on the French DRG classification are given in Appendix B

4 Empirical strategy and results

4.1 Econometric models

In this paper, I check whether the refinement of a hospital stay classification system leads to upcoding. As mentioned in the Introduction, these data make it possible to “go back in time” and observe the consequences of the 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 new (2009) version of the classification. Therefore, the analyses here are based on the level of severity that would have been associated to the stay in the new classification, before it actually came into effect. Analysis of data from this period is used to determine the trend in the percentage of low-severity stays independently of changes in the behavior of health care providers due to changes in the DRG classification. I am then able to examine whether, after the reform, the percentage of low-severity stays is lower than the previous trend would predict.

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.

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

- 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 ($\beta_h * t$) 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, which can again be either fixed or random.
- $I(\cdot)$ 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. $I(\cdot)$ is thus the pure impact of the 2009 DRG classification policy obtained by controlling for any other changes over the period.¹⁰ 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. I therefore used an ordered probit model controlling for the interdependency of coded severity levels. I define a categorical variable $y \in \{1, \dots, 4\}$ 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.

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

where y^* is unobserved. What is observable is the coded level of severity y . ϵ is a normally distributed

¹⁰In the robustness checks (Section 5), the model was run using a more restricted sample from the 2008-2010 period.

random term, with variance normalized at 1. Threshold parameters determine the estimates for different observed values of y .

4.2 Results

Table 3 displays results of Least Square models. The chances that a stay would be coded as low-severity (Y) decreased over time. This may indicate either increases in the severity of the health status of admitted patients over time, or a learning effect: hospitals adapting to the incentives of the T2A system. After the reform, they coded comorbidities and diagnoses as a whole in a more exhaustive fashion. All else being equal, the probability that a stay would be associated to a low-severity DRG strongly decreased with the 2009 reform of the DRG classification. This result is robust: the same result is obtained whether or not types of pathologies, hospital fixed effects, and hospital fixed effects multiplied by the trend are controlled for. Therefore, the coding behavior of health care actors was modified by the change in DRG classification independently of any effect on the actual production of care.¹¹ 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 and not to changes in hospital production. As a result of the upcoding, the probability of a stay being coded as non-severe decreased by about 2.1%.

The effect of a change in classification may differ in different types of hospitals. The category of public hospitals includes two very different groups of institutions. University hospital centers and regional

¹¹The only potential impact of the 2009 DRG change in coding on the actual production of care could have been through a 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)).

hospital centers (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. I now present the results of the model controlling for interactions between hospital type and $I(\cdot)$, the indicator for the post-classification change period.

It appears that the 2009 policy change led to a greater change in coding behavior in private for-profit healthcare institutions (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 in terms of coding behavior change induced by the 2009 policy change. Regarding public hospitals, the shift to a finer classification system led to a greater degree of upcoding in research hospitals than in other public institutions. Other public hospitals were disadvantaged in comparison to research hospitals.

Table 4 presents the ordered probit model and Table 5 displays the marginal effects obtained from the ordered probit model. 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. The results are generally similar to those in Table 3 with respect to statistical significance. From Model (2), we capture the presence of time invariant heterogeneity with a fixed effects procedures. 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. 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.¹² The refinement of the DRG classification system thus led to a significant and positive upcoding effect, with the probability that a stay would be coded as severe increasing by about 2% after the reform (compared with before the reform), controlling for all of the independent variables (Table 5, Column (4)). Our results clearly demonstrate a purely exogenous effect of the change in classification system on coding at medium to very high levels of severity relative to the lowest severity level. The probability of being coded at a higher severity level was shown to increase solely due to the change in the DRG classification system, whether or not types of pathologies and hospital fixed effects were controlled for. The coding behavior of healthcare actors was thus altered by the change in classification independently of any effect on the actual production of care.

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. Non-profit institutions fell between private for-profit hospitals and non-teaching public hospitals in terms of coding behavior. 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. One of the reasons for this is probably the very early use of coding optimization software in for-profit institutions. University hospital centers have also made use of these tools, but this seems to be a much more recent phenomenon. 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 al (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

¹²Here, the marginal effect is an estimate of a population-averaged marginal effect

is based on evidence on the market interaction of hospitals with regard to upcoding. Here, 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 health care decision pressure from the hospital's administration and from the professional peer group on physician's inclination to engage in up coding. They find that the source of pressure is a relevant predictor of physicians' inclination to engage in data-manipulation. What can we say about the 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 in upcoding.

Besides, an exogenous shock in classification led to a general increase in the coded severity of cases. A higher level of severity is associated to a higher value of i_g . In a fixed-envelope mechanism, the value of the index depends on the total number of index units. 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.

- This regulatory mechanism led to greater homogeneity in stays coded as not severe (severity level 1). 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.

4.3 Financial impact

I do not presume to give the exact financial impact of the upcoding on the hospital healthcare budget. However, I propose here a simple method to get an approximation of this amount. The additional budget computed are done for the year 2009. First, I consider the case where there is no fixed annual budget at the national level determined by the regulator. The “upcoding” then leads to an increase in the global hospital healthcare expenditure.

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

With pop DRG-L1: population of patient coded as DRG level 1; mean's fee DRG-L1: mean of fee for patient coded as DRG level 1; mean's fee DRG-L2: mean of fee for patient coded as DRG level 2.

I estimated “upcoding” accounted for nearly euros 560 million euros in additional annual hospital healthcare costs, so approximately 1.4% of the fixed national annual budget.

Now, considering the real context where a fixed annual budget at the national level determined by a vote in the French National Assembly. Considering 2 hospitals, Hospital A (with a funding x) and hospital B (with a funding y), the global budget is equal to $x + y$. With “upcoding”, Hospital A claims a funding $x + \alpha$ but the global budget remains equal to $x + y$. We can see the hospitals' funding as percent of the total funding. Considering hospital B's that does not “up-code”, for sake of simplicity,

- Without “upcoding”, Hospital B receives $y/(x + y)$ % of the total funding

- With “upcoding”, Hospital B receives $y/(x + y + \alpha)$ % of the total funding

Hence, in a context of a fixed annual budget and with “upcoding”, the fundings are the following:

- 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 “upcoding” behaviour of Hospital A increase the funding of this hospital by decreasing the funding of Hospital B from the same amount. The total budget for acute activities (T2A budget)¹³ in French hospital were around euros 40 billions. For the non-teaching public hospital, the cost of a lesser tendency to upcode was 38 million euros.¹⁴

The “upcoding” behaviour may also depend on the local market structure. Because of the national budget and the same rule of reimbursement whatever the ownership, each hospital has the same incentive that is to maximise the number of patients and the level of severity for each patient. Therefore, there is no reason to think that the local market share may have any impact on their “upcoding” behaviour.

¹³For 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>

¹⁴The effect of “upcoding” behaviour is standardised on the basis on the non-teaching public hospital “upcoding” behaviour. Then, I compute the average “upcoding” behaviour of the other hospital’s ownership compared with the non-teaching 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” behaviour 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.

5 Robustness Checks

5.1 The DRG basis prospective payment reform

As the change 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 coding practices 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 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 were unchanged.

In this study, I focus on the new version of the DRG. It might be thought that it would be interesting to assess the effect of this change on hospital behavior using the old version of the DRG with data from the 2009-post period. In fact, this is impossible: in the old DRG classification, there is no way to identify the level of severity.

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 some differences in the calculation of lump-sum transfers. For instance, the relative units calculated for for-profit institutions differ from those for the public sector. The national health insurance system yields

a fixed price for every DRG,¹⁵ but there are two pricing grids: one for the public sector and one for the private sector. 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. As a consequence, prices do not converge, although there is competition between health care institutions in the two sectors. In this paper, I focus on the behavior of all health care institutions, disregarding the question of the differences in lump-sum pricing between the sectors. I assume that the logic of pricing is the same for all health care institutions. Since 2008, the logic of pricing has in fact been the same for all French hospitals. As mentioned above, a similar pattern of results was obtained with the sample from 2008 to 2010 (Table A1 and A2).

Here is another simplification of the formalization. In 2008, the percentage of financing based on the T2A prospective price mechanism reached 100% of acute care stays, but the tariff 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

¹⁵This 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)

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. Part of the cost can be covered by the so called “forfait journalier hospitalier”, which may greatly differ among hospitals¹⁶. Nevertheless, currently a deficit in a hospital care unit can lead to anything from reduction in the number of beds in the unit to its elimination. While the extreme scenario of closing a hospital due to deficits is not totally credible, getting into a difficult financial situation can have very severe and concrete consequences for a hospital: the elimination of units or transformation into a long-term care center. It may also lead to postpone investments, to stop recruitment including not to replace employees leaving.

5.2 Effect of ambulatory surgery

Thus far, I have considered the four levels of severity. One of the objectives of recent reforms has been to encourage recourse to ambulatory surgery. Patients at a low level of severity, under specific conditions, may be 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. It can be observed that stays associated to a DRG coded as ambulatory surgery were mainly admissions to private for-profit establishments. 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. Ambulatory surgery was more frequent in for-profit hospitals. The differences between types of public hospitals seem to have

¹⁶Here, I chose to simplify a complex mechanism between the “forfait journalier hospitalier”, the “prix de journée” and the “ticket modérateur”

been relatively modest compared to the differences between public and private 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.¹⁷

The inclusion of ambulatory surgery also did 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.

5.3 The Econometric model

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. It seems to me easier to interpret a measure of changes in the expected values of the level of severity. 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 probit 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

¹⁷As a robustness check, I also ran a least squares model including all admissions to acute care units. The results were unchanged (Table A4).

the AIC's tests. All of the tests agree that the proportional odds assumption is not violated with these very large dataset.

6 Conclusion

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 global budget 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 upcoding patients to 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. Actually, the upcoding behavior differs in function of the hospital's ownership whereas hospitals under the same reimbursement system, have the same incentive to upcode. 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 upcode, at least

partially.

Some hospital types proved quite sophisticated in their upcoding strategies, upcoding more in diagnoses where the reward for doing so was greater. Following the policy change, for-profit facilities availed themselves of this opportunity to the greatest extent. These results are in the line of those of Dafny (2005). An implication of these results is that upcoding led to a budget reallocation which increased the share of total health spending that went to for-profit hospitals at the expense of public non-research hospitals.

A second possible implication of these results relates to in-patients with low-severity diagnoses. I found 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 budget at the national level. Because of this fixed annual budget, this led to a decrease in the fee paid for patients with a low-severity diagnoses, for the same intensity of treatment. I also found that the hospitals the least equipped with high-tech equipment—non-teaching public hospitals—were the least likely to engage in upcoding, 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 cost 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 similar conclusion can be drawn here in the French context.

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 judging 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 is disconnected from any changes in production lines. A change in treatments performed for a given diagnosis has no effect on coded level of severity. Therefore, these results are free of any quality effects.

A straightforward extension of this study would be to assess hospitals' response to changes in DRG prices in the absence of any changes in the DRG classification system. This would allow assessment of changes in practice patterns: the provision of services which lead to the reclassification of stays into higher-paying DRGs. More generally, this research highlights the need to anticipate the impacts of any changes in the DRG classification on public and private-sector actions in this important industry. 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 budget. This year, for the first time, the annual budget is appeared not enough to take in charge the number of admissions. Another possible extension of this study could be to assess how hospitals deal with the lack of financial support? What are the implication in terms of healthcare and decision of discharge?

7 Bibliography

Agence Technique de l'Information sur l'Hospitalisation —ATI— (2011) “Rapport sur les dépenses d'assurance maladie relatifs aux établissements de santé en 2011”, Annexe 1.3

Barros and Braun (2013), “Upcoding in a NHS”, *Working Paper*

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

Brekke, K. R., L. Siciliani, and O. R. Straume (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.

Chandra, A., D. Cutler, and Z. Song (2011). “Who ordered that? the economics of treatment choices in medical care”, in T. G. M. Mark V. Pauly and P. P. Barros (Eds.), *Handbook of Health Economics*, Chapter six, Volume 2 :397-432.

Currie, J. and W. B. MacLeod (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

Delattre E., B. Dormont, M. McClellan et C. Milcent (2002). “Systèmes de tarification et Évolutions de la variabilité des coûts hospitaliers en France et aux Etats-Unis”, dans *La tarification à la pathologie : les leçons de l’expérience Étrangère*. Dossiers Solidarité et Santé, Hors série, La Documentation française.

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.

Ellis, R. and Vidal-Fernandez, M. (2007). “Activity-based payments and reforms of the English hospital payment system”, *Health Economics, Policy and Law*, Volume 2.

Garber, A. M. and Skinner J. (2008). “Is american health care uniquely inefficient” *The Journal of Economic Perspectives*, Volume 22 (4).

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, B.H. (1999). Measuring hospital cost-sharing incentives under refined prospective payment. *Journal of Economics and Management Strategy*, 8(3):433-52.

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

Gobillon and Milcent (2015) “The competition effect of a French reform on hospital quality”, Working Paper PSE

Hafsteinsdottir, E.J.G. and Siciliani, L. (2009). “DRG prospective payment systems: refine or not refine?” *Health Economics*, 19(10):1226-39.

Hakkinen et al (2012). “Patient classification and hospital costs of care for acute myocardial infarction in nine European countries. Häkkinen U1, Chiarello P, Cots F, Peltola M, Rättö H; EuroDRG group”, *Health Economics*, 21 Suppl 2:19-29

Hsia, David C.; Krushat, W. Mark; Fagan, Ann B.; Tebbutt, Jane A. and Kusserow, Richard P. (1988), “Accuracy of Diagnostic Coding for Medicare Patients Under the Prospective Payment System”. *The New England Journal of Medicine*, February, 318(6), pp. 352-355.

Kimberly, J.R., de Pouvourville, G., D’Aunno, T., eds. (2008). *The Globalization of Managerial Innovation in Health Care*. Cambridge: Cambridge University Press.

Kuhn, Michael and Luigi Siciliani, (2008), “Upcoding and optimal auditing in health care (or the economics of DRG creep)”, *CEPR Discussion Paper*, N.6689.

Ma, Ching-to Albert, (1994), “Health Care Payment Systems: Cost and Quality Incentives?” *Journal of Economics and Management Strategy*, 3, 93-112

Ma, Ching-to Albert (1998), “Cost and Quality Incentive in Health Care: A Reply”, *Journal of Economics and Management Strategy*, 7, 139-142, 1998.

Manuel des GHM —groupes Homogènes de Malades— (2012). 11ème version de la classification, 3ème révision (11d) Version 13.11d de la fonction groupage, *Bulletin Officiel*, Fascicule spécial, n2012/5bis

Mason A., O. Zeynep, 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

McClellan, M. (1997), “Hospital Reimbursement Incentives: An Empirical Analysis”, *Journal of Economics and Management Strategy* 6, 91:128.

Medicare Payment Advisory Commission (2008), “Report to the Congress: Reforming the Delivery System” (Washington, D.C.: MedPAC, June 2008).

Milcent, C. (2014), “Coût des soins hospitaliers: Effet de la concurrence par la qualité”, Working Paper PSE, 2014.

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

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

Newhouse J.P. (2003), “Reimbursing for Health Care Services?”, *Economie publique*, 13: 3-31.

Zeynep Or, Julia Bonastre, Florence Journeau, Clément Nestrigue “Production, Productivity and Care Quality in French Hospitals Before and After the Introduction of Activity Based Payment”, Document de travail n° 56 - Irdes - Avril 2013

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*, vol. 8: 147-172 *Journal of Economics and Management Strategy*

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

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

Shleifer, A. (1985). "A Theory of Yardstick Competition", *The RAND Journal of Economics*, Volume 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. and Skinner, J. (2004), "Medicare Upcoding and Hospital Ownership". *Journal of Health Economics*, 23(2), pp. 369-89.

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), pp. 420-26.

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

Simborg, D.W. (1981). DRG creep: a new hospital-acquired disease. *New England Journal of Medicine*, 304:1602-4.

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*, 81:289-99.

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

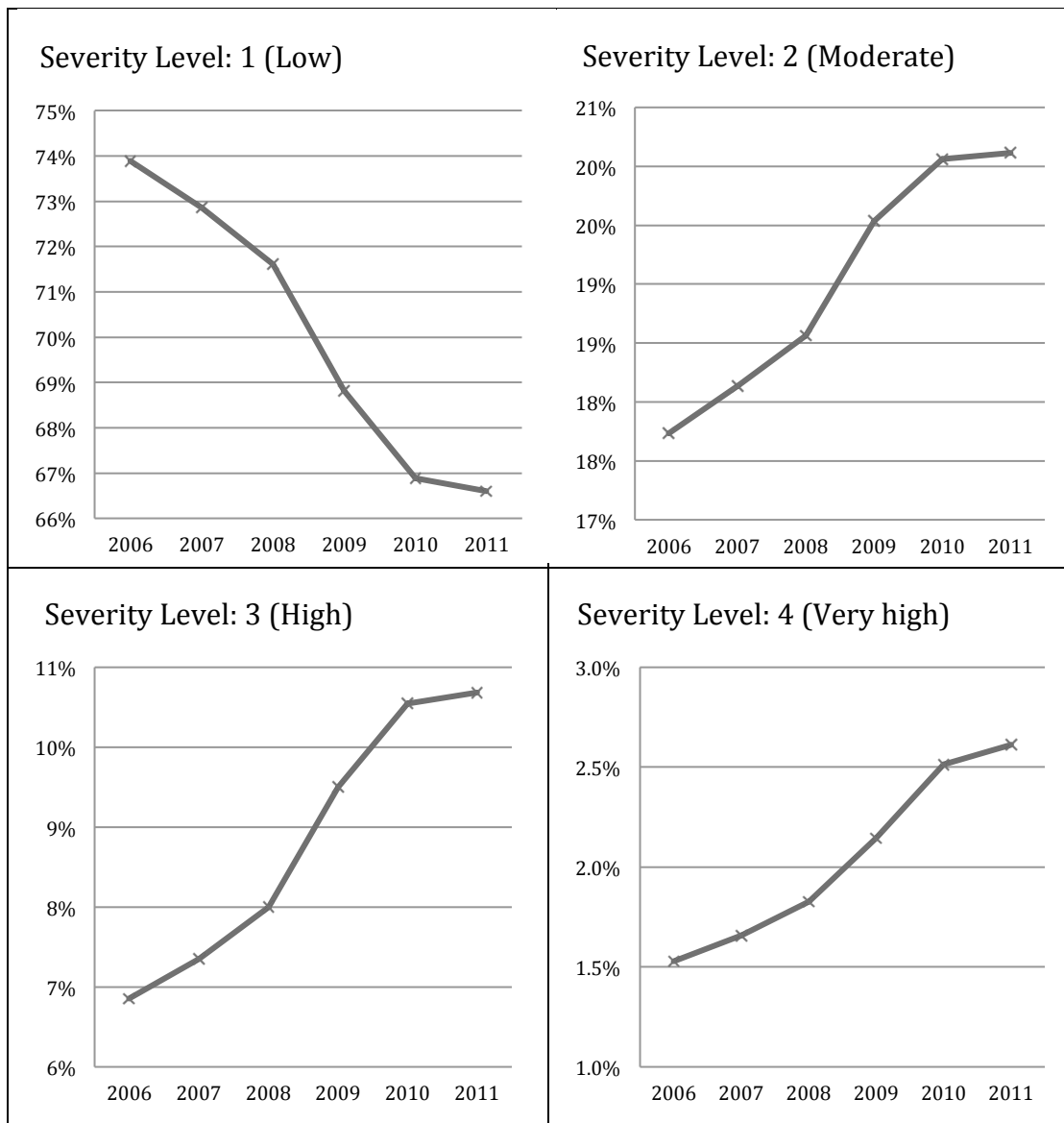
Street, A. and Maynard, A. (2007), "Activity based financing in England: the need for continual refinement of payment by results?", *Health Economics, Policy and Law*, Volume 2: 419-427.

Street A, Kobel C, Renaud T, Thuilliez J. (2012). "How well do Diagnosis Related Groups explain

variation in costs and length of stay among patients and across hospitals? Methods for analysing routine patient data”, *Health Economics* 21 (Suppl. 2): 6-18.

Xirasagar, S. and Lin, Herng-Ching. (2006), “Effects of Payment Incentives, Hospitals Ownership and Competition on Hospitalization Decisions for Ambulatory Surgical Procedure.” *Health Policy*, 2006, 76, pp. 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(.)) <i>crossed</i>	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

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%.

Model(1): $Y = f(\text{Trend, classification DRGs' change})$

Model(2): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology})$

Model(3): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend})$

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(.)) <i>crossed</i>	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)

Notes: *significant at 1% level, ** significant at 0.1%, * significant at 0.01%.

Model(1): $Y = f(\text{Trend, classification DRGs' change})$

Model(2): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology})$

Model(3): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend})$

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(.)) <i>crossed</i>	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 parentheses.

Appendix

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(.)) <i>crossed</i>	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): $Y = f(\text{Trend, classification DRGs' change})$

Model(2): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology})$

Model(3): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend})$

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(.)) <i>crossed</i>	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): $Y = f(\text{Trend, classification DRGs' change})$

Model(2): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology})$

Model(3): $Y = f(\text{Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend})$

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(.)) <i>crossed</i>	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 parentheses.

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(.)) <i>crossed</i>	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(.)) <i>crossed</i>	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): $Y_j = f(\text{Trend, classification DRGs' change})$

Model(2): $Y_j = f(\text{Trend, classification DRGs' change, Dummies_pathology})$

Model(3): $Y_j = f(\text{Trend, classification DRGs' change, Dummies_pathology, Dummies_hospital, Dummies_hospital crossed trend})$

Models with no dummy pathology and/ or hospital include random effect

Table A6: Information criteria

	LL(null)	Akaike's information criterion	Bayesian Information 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

8 Appendix B: French DRG classification version 11 and DRG based payment

8.1 Description of the new French DRG coding system

Version 11 of the 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 “root” of a DRG consists of the first 5 characters in the corresponding code. There are 617 root DRGs in version 11c. Among these roots, 76.2% (470) are subdivided into 4 levels of severity.

This classification defines four levels of severity (Manuel des GHM, 2012). The level of severity is independent of the procedures performed on the patient.

- 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. For a given version of the classification, 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.

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.

- 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". According to the health statistics agency, these are almost exclusively medical stays.

In version 11 of the coding system, used here, stays were classified into 2,192 DRGs.

8.2 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 increase 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 efficiency: 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 achieve 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.