



**HAL**  
open science

# Bias due to re-used databases: Coding in hospital for extremely vulnerable patients

Carine Milcent

► **To cite this version:**

Carine Milcent. Bias due to re-used databases: Coding in hospital for extremely vulnerable patients. 2023. hal-03960584v1

**HAL Id: hal-03960584**

**<https://hal.science/hal-03960584v1>**

Preprint submitted on 27 Jan 2023 (v1), last revised 6 Mar 2024 (v2)

**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.

Bias due to re-used databases:  
*Coding in hospital for extremely vulnerable patients*

**Running title**

Bias due to re-used EHR databases

**Author (corresponding author)**

Prof. Carine Milcent  
Health Economics  
Econometrics

**Institution names :**

Center for National Scientific Research - CNRS  
Paris School of Economics - PSE

**Country :**

France

**Email address**

carine.milcent@psemail.eu

**Full postal address**

PSE  
48 Bd Jourdan  
75014 PARIS

**Title**

Bias due to re-used databases: Coding in hospital for extremely vulnerable patients

**Acknowledgements**

We would like to thank the team of the ATIH French Statistical Agency for Hospital database for the technical support. We also like to thank Borelli seminar participants,

**Data Protection Authority**

Data used for this study are reported to the National Data Protection Authority (CNIL number 2019-100001-166-140).

**Bias due to re-used databases:  
*Coding in hospital for extremely vulnerable patients***

## **Public interest abstract**

Electronic health records (EHRs) are intended to reduce healthcare costs and improve the quality of care. Nevertheless, usability issues common to EHRs have been identified. In this paper, we investigate these usability issues for social vulnerability codes. Using the acute care EHR and the rehabilitation care EHR databases, hospital stays of 800'000 patients are studied. This article highlights the differences in coding processes between public and private institutions observed when there are different incentives to code. Furthermore, it shows that the differences in coding are not random but depend on the coding strategy. This article emphasises that the reuse of data leads to biases in interpretation. Using the example of social vulnerability alerts policymakers to the need to take into account these differences in coding processes when decisions are based on EHR information. Otherwise, this process of coding differences in social vulnerability may exacerbate social inequalities rather than reduce them.

## **Highlights**

- The EHR is fully used for all patients admitted in acute care unit and rehabilitation unit care.
- Heterogeneity in coding social vulnerability between public and private sector
  - During the rehabilitation stays
  - During the acute care stays.
- Heterogeneity in coding is not random but rather calculated.
- Opportunities to optimize reused of EHR databases for policymaker's reform should be further explored.

## Bias due to re-used databases: *Coding in hospital for extremely vulnerable patients*

### Introduction

Researchers and professionals often apply data not only for its primary goal but also for multiple alternative purposes. In this paper, we exhibit limits in re-used databases for alternative objectives taking the example of administrative healthcare databases.

Most health care systems currently have their information systems, the forerunner initially developed in the United States in the 1980s. The first step is to record the information in the patient record (electronic or paper patient record). That information is then coded by a hospital staff member specifically trained for the effect using ICD10 for diagnosis and country-specific other classification for procedures. This electronic health record EHR data is a digital version of a patient's paper chart. Menachemi and Collum [1] defined EHRs as « a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. » Finally, an algorithm leads to classifying each patient in one DRG with the same rules being applied to all patients. Combining a principal diagnosis, secondary diagnoses, and procedures leads to a different DRG. In some countries, DRGs are linked to a fee. Since the DRG algorithm qualifies the codes entered differently, optimizing coding activity has a direct financial impact.

The coding behavior is defined as the incentives to code and the way to code. This coding behavior determines the accuracy and the viability of the information contained in the database [2]. According to Dafny [3], coding behavior proved very responsive to financial incentives. Studies on the use of secondary diagnosis showed changes in coding and insidious incentives for upcoding ([4], [5]). Research focusing on exogenous shifts in prices showed perverse upcoding effects ([3], [6]-[9]). A study on French data analyzed the DRG classification reform and exhibited changes in coding practices ([10]). However, the consistency of coding for hospitals is judged, in this literature, for a specific primary purpose, a payment system.

At the same time, there is a growing common practice of re-using of existing EHR databases for population health. The need to define standards for EHR data quality is emerging. Kahn et al [11] published the harmonized data quality assessment terminology and framework for secondary

use of electronic health record data. However, the data is usually considered as an accurate and homogeneous description of a situation, without concern regarding any concept of miscoding or upcoding. The issue of consistency in hospital coding for alternative use has so far escaped this debate.

Implicit bias in coding practice arises between healthcare facilities when this variability is disconnected from any effect on the primary purpose. For instance, to define the DRG payment, an algorithm mobilizes both ICD-10 and procedures codes. Some of these codes are necessary to optimize fees associated with the DRG attributed to the stay; however, other codes are useless for the DRG algorithm. The time needed for qualified health staff to perform coding activity is costly in itself. When secondary diagnoses unused by the algorithm are not coded, no perverse upcoding effects or downcoding effects impact DRG-based payments. That is what we define as implicit bias.

Social vulnerability is an excellent example of such phenomena. Interest in social determinants of health among clinicians, researchers, and policy-makers has increased in recent years, driven both by a recognition of their role as major contributors to health outcomes and by interest in improving health equity ([12], [13]). However, the codes of social vulnerability are so far rarely used in healthcare outcome literature ([14], [15]). The social vulnerability code has been introduced in the classification since the ICD-9 version. In the next version (ICD-10), the number of codes has been expanded to more refined information on housing, social, behavioral, and economic needs reflect circumstances that influence the receipt of health care services ([16]).

In this paper, we work to illustrate implicit bias caused by heterogeneity in coding processes by exploring the coding practices of healthcare facilities. We do so by taking French facilities as a specific case study. By exploiting the acute care database and rehabilitation care database, we select patients coded as socially vulnerable during their rehabilitation stay. We then examine their ICD-10 socially vulnerable codes when they were previously admitted to an acute care unit. For groups of patients ICD-10 coded as socially vulnerable in a rehabilitation unit, the probability of being coded as such during the previous stay (acute care) varies according to the hospital's ownership. Considering the difference in incentives, we conclude by arguing that heterogeneity in coding is not random but rather calculated. Database re-use for epidemiologic studies or health economics projects has the potential to lead to bias in results.

## Materials and methods

### *The French context*

Administrative and medical data are recorded at discharge from the hospital for all patient stays. Diagnoses are coded using the International Classification of Diseases, 10th revision (ICD-10), and procedures performed during hospitalization are coded according to the Common French Classification of Medical Acts. This hospital activity database – PMSI database – is an extensive, exhaustive, national database that has gathered data transmitted monthly by all public and private hospitals in France.

In the French-DRG algorithm, most individual social vulnerability ICD-10 codes are not taken into account for stays in acute care, giving little incentive for exhaustive coding. However, patients with social vulnerability most likely represent an extra cost that public hospitals have to bear, as they cannot select patients, unlike their private counterparts. Public sector hospitals are then keen on coding the information as leverage in the negotiation with the regulator for an amendment in the DRG algorithm or additional payment. As a matter of fact, a global reform of healthcare facilities reimbursement targets to include social determinants in activity-based payment.

The social vulnerability information is also marginally used for reimbursement in rehabilitation centers. However, admitting patients in the context of vulnerability is one mission of the rehabilitation center. The French system is a publicly funded healthcare system. All expenses need to be justified by the French health authorities, including admissions to a rehabilitation care center. To register vulnerability (ICD-10 codes) allows providing statistics on patient's vulnerability rate that justify the patient's admission to these centers. Thus, even if the vulnerability ICD-codes are not explicitly used in the algorithm for defining the DRGs' fees, these codes may be a patients' admission rationale for control from the regulator. That is a solid and direct incentive for coding. In addition, recent grey literature suggests improving the inclusion of vulnerable aspects with additional subsidies ([17]), and from 2022, new Z-codes will impact the rehabilitation DRG algorithm.

Beyond, the PMSI database is available for researchers. Economic research or epidemiologic investigation is performed using this database. Missing codes unused by the DRG algorithm may create bias in these researches. From the research point of view, the incentive to code

exhaustively is extremely high. It is equivalent for health economic researchers or IA Health researchers. Some of these researchers work in academic research units embedded within healthcare facilities. They promote, from the inside, exhaustive coding activity.

Therefore, for reasons external to the DRG-based payment, individual social vulnerability should be more coded in rehabilitation centers. However, it implies bias in coding due to variability in incentives and motives to code. In this paper, we use these social vulnerability ICD-10 codes to show heterogeneity in coding.

## ***The methodology***

The methodology is as follows (**Figure 1**): we collect information on patients admitted in a rehabilitation centre. We isolate the coding practice for vulnerability as defined by the ICD-10 during the rehabilitation stay. Then, we split the group in two sub-groups according to the presence of a vulnerable ICD-10 code. For each group, we eventually compare facilities in their coding practices by statuses. To do so, we compute the probability to previously have vulnerability code(s) registered during the acute care stay.

From the group of patients described above, we distinguished patients admitted in acute care in public sector facilities from those admitted in private sector ones. We computed t-tests to exhibit the significance of coding behavior differences in public healthcare units compared to the coding behavior in private healthcare centers.

The information mobilized is at the stay level. Some patients have completed an “acute care stay-rehabilitation stay” pathway several times during 2018. However, the vast majority commuted between their acute care facility and their rehabilitation facility. A very marginal proportion was hospitalized twice with an “acute care stay-rehabilitation stay” pathway unrelated to their first pathway. In addition, an even neglectful number (0.002%) were admitted during their "acute care stay-rehabilitation stay" pathway to facilities with different ownership (public versus private). Thus, it is impossible to study two distinct “acute care stay-rehabilitation stay” trajectories for the same patient. After a stay in rehabilitation, some patients return for another stay in rehabilitation for the same pathology. They are excluded from the observations since we select patients with an “acute care stay-rehabilitation stay” pathway afterward.

We use a multilevel Logit model in which we account for clustering at the facility level. As sensitive analysis, we use a multilevel Logit model in which we account for clustering of the same patients repeated within the dataset. The results are totally equivalent.

### ***Coded information at patient's admission***

Data comes from two French national hospital discharge databases: PMSI MCO for Acute Care and PMSI-SSR for Rehabilitation care. The information is available at the stay level. From these databases, patients are followed throughout their hospital stays using an anonymous identification number that is unique to each patient. The analysis of the merged databases for scientific research has been conducted with permission from the Hospital Statistics Agency (ATIH), which is the responsible authority for this data. Data used for this study are reported to the National Data Protection Authority (CNIL number 2019-100001-166-140). Informed consent is not required since the study was based on routinely collected de-identified administrative data, as regulated by French law.

Data selection is patients who were admitted to an acute care center between January 1 and December 31, 2018 and discharged to a rehabilitation center following the acute care stay. The data contains 848'567 stays and 835'086 individuals. The stays were first identified in the acute care database and then searched for in the 2018\_Rehabilitation database using the anonymous identification number. The search was extended to the first two weeks of the year 2019. The two databases (database for acute care and database for rehabilitation care) contain information regarding the patient's demographic characteristics (age, gender) as well as the main diagnosis and comorbid factors based on the International Classification of Disease (ICD), 10th revision. The severity of the patient can change from the acute care stay to the rehabilitation stay. Therefore, we do not compare coding activities based on this information. Information on dramatic socio-economic context and severe psycho-socio conditions are usually constant over the two stays \_ from an acute care stay to consecutively a rehabilitation stay \_. Therefore, we concentrated the study on these codes.

When a patient is admitted to a hospital (for acute care or rehabilitation care), individual information is coded. The medical staff codes patient information into three blocks: 1- general information, 2- acts and procedures performed during patient's stay information, and 3- pathology, diagnosis, and co-morbidities information. The first block of information (1-) is



mainly composed of demographic information, length of stay, where the patient comes from, and where he is then discharged. The second block of information (2-) is coded according to a French national classification called CCAM. Each act or procedure is associated with a code. Some procedures are cut into several codes for accuracy purposes. The third block of information (3-) is coded according to the International Classification Diseases \_ ICD. This classification allows hospital staff to record clinical information as well as dramatic socio-economic context and very severe psycho-socio-environmental issues. These are classified in Z-codes according to the ICD-10 classification.

We then collected this information during the acute care stay of the patient and during the following rehabilitation care stay. For each type of vulnerability, we selected the following ICD-10 Z-code: childhood severe issue (ICD-10 Z61), social environmental issue (ICD-10 Z60), employment severe issue (ICD-10 Z56), physical environment (ICD-10 Z58), severe housing issue (ICD10-Z59), severe education issue (ICD-10 Z55), psycho-socio issue (ICD-10 Z64-Z65)<sup>1</sup>. The hospital staff codes this information during the hospital stay with CMD-10, only when the patient is in a highly critical situation.

From these dummy variables, we computed a global dummy called *Vulnerable*. This variable is equal to 1 if at least one dummy variable in this list of dummy variables is equal to 1.

Knowing that the patient is coded as vulnerable in the rehabilitation center of admission, we assess the probability of being coded as vulnerable during acute care. In doing so, we explore the possible heterogeneity in coding activity between types of hospitals.

## Results

**Table 1** presents the baseline characteristics of patients. Most patients were female (59%), and most were aged 85 and over (48%). In the acute care facilities, 13% of patients were coded as socially vulnerable. These same patients were 26% to be coded as socially vulnerable during their rehabilitation stay.

---

<sup>1</sup> The codes Z61.6 and Z61.5, potentially affecting the algorithm, are excluded.

Among the rehabilitation centers, the share of private ones was 28%, and they admitted 37% of patients. The proportion of private acute care facilities was 36%, and they treated 28% of the inpatients.

**Figure 2** displays the share of patients with vulnerability according to healthcare facilities.

In the socially vulnerable group of rehabilitation stay patients, only around one-fifth was coded as vulnerable during their acute care stay (26.96%). This result is quite intuitive given the incentive differences between acute care unit and rehabilitation care unit. Moreover, since acute care centers and rehabilitation centers are fundamentally different in their goals, this discrepancy has little impact on downstream research.

Second, we shed light on some heterogeneity within acute care centres. In order to do so, two groups of patients are considered: patients coded as vulnerable during their rehabilitation stay and the ones not coded as vulnerable during their rehabilitation stay. For each of these groups of patients, we compare patients previously admitted in an acute care public unit with patients admitted in a private one. First of all, results reveal heterogeneity in coding vulnerability between different types of hospitals. Staff in public acute care centres tend to code more frequently patients with vulnerability than staff in private acute care units. For these patients, the share coded as vulnerable is 29.5% in public rehabilitation units and it falls to 18.6% for patients admitted in a private acute care facility. A similar comment can be made for groups of patients not coded as vulnerable during their rehabilitation stays. Differences between the mean are validated using a T-test.

**Table 2** presents results for probability to be coded vulnerable in the acute care facility. We controlled for the sex, groups of age: between 75 and 84 years old, 85 years old and over and, the facility's status.<sup>2</sup> As expected, from 75 years old, the probability to be coded as vulnerable increases. And, this is even more emphasized over 85 years old. To be a female patient decreases the probability to be vulnerable.

Turning on the acute care facility status, to be admitted in a private acute care facility compared to public one, decreases the probability to be coded as vulnerable. Computing the marginal effect, we found that being admitted in a private acute care facility decreases the probability to be coded as vulnerable by 11.4% (std. err. 0.0008, Delta-method).

---

<sup>2</sup> See Table 2, Note.

**Table 2** exhibits also the results for regression model run on the database of patients **without** vulnerability code(s) during their rehabilitation stays. We controlled for sex, group of age (defined as previously). From 75 years old, in acute care facilities, the probability to be coded as a vulnerable patient, increases. About the type of healthcare centre of admission for acute care, we also found differences. In the private sector, the probability to be coded vulnerable, decreases. Computing the marginal effect, we found a decrease in coding activity of 4.70% (std. err. 0.0006, Delta-method) compared to public acute care hospitals.

We here clearly found differences in coding practices that may impact studies on vulnerability based on the re-use of these databases.

Until now, we have assumed that the way to code is similar in rehabilitation centres whatever their status. This assumption is based on the reimbursement incentive for these centres. To deal with patient vulnerability is one mission of these rehabilitation centres. A direct incentive is associated with the reimbursement fee. The French healthcare authority (ATIH) reviews periodically coding guidelines to improve this incentive. A recent work suggested that to improve the inclusion of vulnerable aspects by additional subsidies (ATIH, 2021). And from 2022, new Z-codes will impact the rehabilitation DRG-algorithm.

To test this assumption, we then started the coding approach within public rehabilitation centres. Public rehabilitation centers have a lower rate of inpatients coded as socially vulnerable (24.31%) than that found for all rehabilitation centers (25.86%). This finding may be due to coding practices or patient mix differences. This group of patients admitted in public rehabilitation centres is divided into two groups: one with Z-codes for social vulnerability and one without these codes. Then, we compare the coding approach between statuses when they were firstly admitted in an acute care facility. We also find that being treated in an acute care private facility decreases the probability of being coded as socially vulnerable. It can be interpreted as the public sector up-codes vulnerability compared with private acute care facilities or vice versa, private sector down-codes vulnerability compared with public acute care facilities. This difference in coding approach is revealed for both, the vulnerable sub-group of patients (by 16.62%) and the non-coded as vulnerable sub-group of patients (by 4.99%).<sup>3</sup>

---

<sup>3</sup> A marginal effect: 16.62%, the std. err. 0.00152, (Delta-method)

A marginal effect: 4.99%, the std. err. 0.00096, (Delta-method)

**Table 3**, we run a regression model where patients were firstly admitted in a public sector acute care facilities and then admitted in a rehabilitation centre. On the database with vulnerable patients (as coded during their acute care stay), we find that being admitted in a private rehabilitation centre increases by 3.27% the probability to be coded as vulnerable (std. err. 0.00116, Delta-method) compared to a public one. With the sub-group of non-vulnerable patients, to be admitted in a private rehabilitation centre leads to an increase of 10.41% (std. err. 0.0005, Delta-method). This upcoding behaviour deriving from direct financial incentive is consistent with the existing literature on the topic ([18]-[20]).

If we focus on a subset of patients admitted in public rehabilitation centres, the results remain very stable.

So, in summary, we observe that comparing public rehabilitation units with private rehabilitation centres, there are more stays coded as vulnerable in private facilities. Similarly, for acute care unit, the share of patients coded as vulnerable is higher in public sector facilities. The bias in coding activity depends on differences in incentives.

## Discussion

This paper highlights the difference in coding activity between healthcare facilities. As the results show, differences in coding behavior within acute care facilities (respectively rehabilitation unit) depend on their status. Medical staff do not code vulnerability according to the same criteria depending on the status of the facility. Specifically, the study is conducted in two stages. First, the patients who were coded as socially vulnerable during their care in rehabilitation were isolated. We then expect to find them homogeneously coded as socially vulnerable during their previous stay in acute care. The results show that this is not the case: 27% were coded as socially vulnerable. It is also expected that their “socially vulnerable” coding would be independent of admission hospital status. This is not what we get either. The results show considerable heterogeneity by facility status. As a sensitivity test, we further restrict the database by selecting only socially vulnerable patients admitted to public rehabilitation centers. We then eliminate any differences in coding between public and private rehabilitation units. We then look at how these patients were coded during their stay in acute care. The results persist: during their acute care stay, heterogeneity of coding between sectors is shown.

This work is also done by starting from the set of patients coded as socially vulnerable during their stay in acute care and studying the variability of this coding between sectors during their stay in rehabilitation unit. Heterogeneity in coding according to the sector is also obtained. In addition, there is a crossed effect in coding approach, in function of i) the financial incentives of the unit – acute care unit versus rehabilitation centre – ; ii) the sector – public or private – . When there is a financial incentive, private sector codes more intensively compared to the public one. The French healthcare system is a publicly funded healthcare system. In consequences, a published aim of the French healthcare authority is to limit unnecessary cost that is questioned by this result. When there is a weak financial incentive, private sector down-codes compared to the public one.

## Conclusions

These results are shown in a context where coding differences are not related to financial incentives created by DRGs. Whether or not vulnerability codes are coded does not change the assignment of the stay to a specific DRG.

With the emergence of data science, database is now valuable for itself, meaning beyond its primary goal. When a coding team is set up, the additional time (with a monetary assessment) to presently code what appears to be useless information can be compared with the potential value of setting up the database for other goals. In regard to the regulation of data confidentiality, the database created for a specific target may be optimized to fit a large set of alternative goals. The opportunity to sell the created database is conditioned by the possibility to re-use it for alternative objectives.<sup>4</sup> Re-using databases is definitely a way to optimise the return on investment in setting-up databases. However, it may not mean to neglect limits intrinsic to the re-use of an existing database, due to the gap from the initial purpose to the alternative one(s).

Part of the role of researchers and any data users is to know whether the database they are using is fit for purpose and to be aware of the limitations and potential biases it may introduce into the results. The administrative healthcare databases in France offer a helpful example of gaps caused

---

<sup>4</sup> The term "database" used here includes all aggregated forms of the original database in function of the client's needs and the regulation's rules.

by the re-use of databases. This paper illustrates one such bias to highlight the need for this preliminary work.

Above the market price of the database, coding decisions may also impact political decision-making. Indeed, to guide healthcare policies and support upcoming reforms, policymakers ask for forward-looking evaluations mobilizing the existing database. For instance, a funding for coordinated healthcare that is emerging implies integrating not only clinical aspects but also socio-economic components and life condition factors into costs. These determinants were previously taken into account; however, they are considered drivers for patient healthcare pathways today. In the current system, healthcare payers are required to anticipate expenses caused by non-clinical aspects, and then use the available existing healthcare database. This means that today's coding activity may influence tomorrow's healthcare reforms. This paper raises the political question on defining funding to support patient's social vulnerability when this funding is based on a re-used of existing database. The limits and risks of such recycling of existing re-use of databases for policy governance have seldom been explored so far.

This result can be extended to other fields where decisions are based on databases whose primary purpose of data collection differs from those they are used to support decision by policy makers.

## References

- [1] Menachemi N., Collum T. Benefits and drawbacks of electronic health record systems. *Risk Management Health Policy*. 2011. 4: p. 47–55.
- [2] Fisher, E.S., et al., The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *American Journal of Public Health*, 1992. 82(2): p. 243-248.
- [3] Dafny, L.S., How Do Hospitals Respond to Price Changes? *The American Economic Review*, 2005. 95(5): p. 1525-1547.
- [4] Steinbusch, P.J., et al., The risk of upcoding in casemix systems: a comparative study. *Health Policy*, 2007. 81(2): p. 289-299.
- [5] Luo, W. and M. Gallagher. Unsupervised DRG Upcoding Detection in Healthcare Databases. in *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops*. 2010. IEEE Computer Society.
- [6] Anthun, K.S., J.H. Bjørngaard, and J. Magnussen, Economic incentives and diagnostic coding in a public healthcare system. *International Journal of Health Economics and Management*, 2017.
- [7] Barros, P. and G. Braun, Upcoding in a National Health Service: the evidence from Portugal. *Health Economics*, 2016. 26(5): p. 600-618.
- [8] Silverman, E. and J. Skinner, Medicare upcoding and hospital ownership. *Journal of health economics*, 2004. 23(2): p. 369-389.
- [9] Barros, P. and G. Braun, Upcoding in a National Health Service: the evidence from Portugal. *Health Economics*, 2016. 26(5): p. 600-618.
- [10] Milcent, C., From downcoding to upcoding: DRG based payment in hospitals. *International Journal of Health Economics and Management*, 2021. 21(1): p. 1-26.
- [11] Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)* 2016; 4 (1): 1244.
- [12] Cook LA, Sachs J, Weiskopf NG. The quality of social determinants data in the electronic health record: a systematic review. *Journal of the American Medical Information Assoc.* 2021 Dec 28;29(1):187-196.
- [13] Bradley E Iott, Matthew S Pantell, Julia Adler-Milstein, Laura M Gottlieb, Physician awareness of social determinants of health documentation capability in the electronic health record, *Journal of the American Medical Informatics Association*, 2022;, ocac154
- [14] Pantell MS, Kaiser SV, Torres JM, Gottlieb LM, Adler NE. Associations Between Social Factor Documentation and Hospital Length of Stay and Readmission Among Children. *Hosp Pediatr*. 2020 Jan;10(1):12-19.
- [15] Milcent, C., Rehabilitation Care: Performance and Ownership. 2021. halshs-03010949 working Paper.
- [16] Torres JM, Lawlor J, Colvin JD, Sills MR, Bettenhausen JL, Davidson A, Cutler GJ, Hall M, Gottlieb LM. ICD Social Codes: An Underutilized Resource for Tracking Social Needs. *Med Care*. 2017 Sep;55(9):810-816.
- [17] ATIH, 2021 Guide méthodologique du PMSI SSR 2021 (version June 2021 – Chapitre 6.13), visited 20/12/2021  
[https://www.atih.sante.fr/sites/default/files/public/content/3971/guide\\_methodologique\\_ssr\\_2021\\_versionp](https://www.atih.sante.fr/sites/default/files/public/content/3971/guide_methodologique_ssr_2021_versionp)

[rovisoire3.pdf](#)

[18] Anthun KS. Predicting diagnostic coding in hospitals: individual level effects of price incentives. *International Journal of Health Econ Manag.* 2021 Oct 6. doi: 10.1007/s10754-021-09314-5.

[19] Neby, S., et al., Bending the Rules to Play the Game: Accountability, DRG and Waiting List Scandals in Norway and Germany. *European Policy Analysis*, 2015. 1(1): p. 127-148.

[20] Steinbusch, P.J., et al., The risk of upcoding in casemix systems: a comparative study. *Health Policy*, 2007. 81(2): p. 289-299.



**Table 1: Basic statistics**

	Mean	Std.	N
Social precarity rate / acute care stays	.13	.33	848'567
Social precarity rate / rehabilitation stay	.26	.44	848'567
Females	.59	.49	848'567
Age			
74 and less	.27	.45	848'567
75 to 84	.22	.40	848'567
85 and over	.48	.50	848'567
	% of facilities	% of stays	
% of private acute care facilities	36%	28%	
% of private rehabilitation facilities	28%	37%	
Total number of acute care facilities			1,929
Total number of rehabilitation centers			1,635

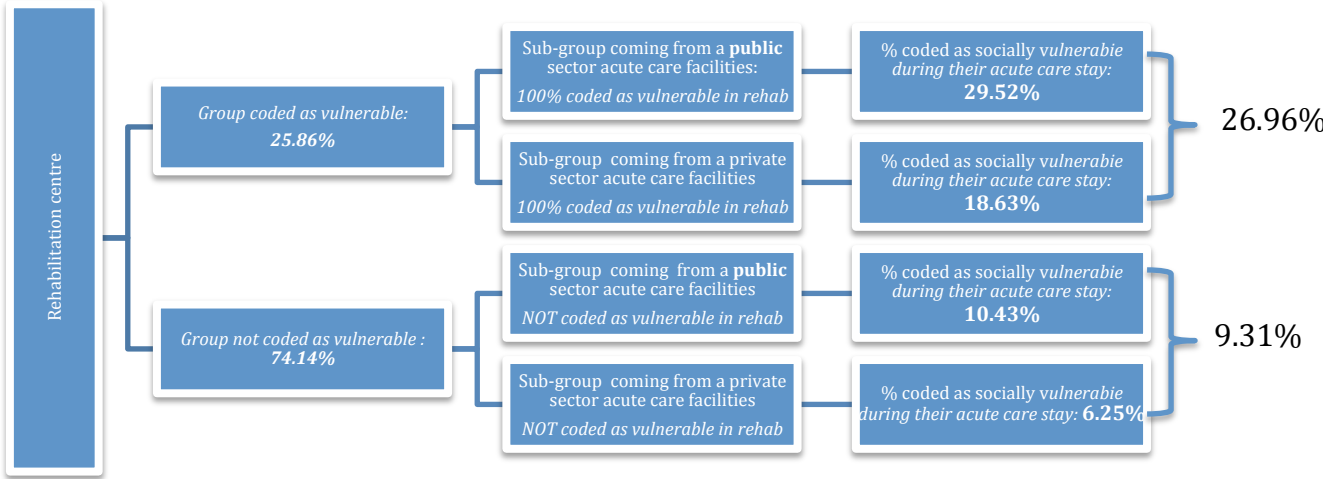
*Source: PMSI MCO and PMSI SSR, years 2018*

Note: Each row reports regression coefficients from a logit regression model, in year 2018. Statistics at the stays level. Patients' stays admitted in SSR subsequent to a acute care stay

**Figure 1:** The comparison groups



**Figure 2:** For group of patients (not) coded as vulnerable in a rehabilitation center: Share in patients (not) coded as vulnerable in the acute care facilities by status



Source: PMSI MCO and PMSI SSR, years 2018

**Table 2** \_ Probability to be coded as vulnerable during the acute care stays (odd ratios)

Dependent variable: Prob.(coded as vulnerable during the acute care stay)	Odds Ratio	Std. Err.	P>z	[95%Conf.	Interval]
<i>Group of patients <b>with</b> vulnerability code(s) admitted in a rehabilitation center</i>					
Gender	.899	(.003)	.000	[.893	.906]
Age: 75 to 84	1.115	(.005)	.000	[1.105	1.124]
Age: 85 and over	1.594	(.006)	.000	[1.581	1.606]
Private acute care facility	.559	(.002)	.000	[.553	.563]
<i>Group of patients <b>without</b> vulnerability code(s) admitted in a rehabilitation center</i>					
Gender	1.302	.006	.000	[1.289	1.314]
Age: 75 to 84	1.143	.007	.000	[1.129	1.156]
Age: 85 and over	1.580	.009	.000	[1.563	1.598]
Private acute care facility	.607	.004	.000	[.600	.615]

Source: PMSI MCO and PMSI SSR, years 2018

Note: Each row reports regression coefficients from a logit regression model, in year 2018. Standard errors are in parentheses. Standard errors are clustered at the hospital level. The model is run at the stay level.

\*: p<.1; \*\*: p<.05; \*\*\*: p<.01

As a sensitivity analysis, we ran models with different combinations where we added information on the length of stay, entrance type, discharge type, the Charlson index during the acute care stay. The results are equivalent. As an additional sensitivity analysis, we ran the model by ICD groups. The results are consistent.

**Table 3** \_ Probability to be coded as vulnerable during the rehabilitation stays (odd ratios)

Dependent variable: Prob. (coded as vulnerable during the rehabilitation stay)	Odds Ratio	Std. Err.	P>z	[95%Conf.	Interval]
<i>Group of patients <b>with</b> vulnerability code(s) admitted in a rehabilitation center</i>					
Gender	1.082	.006	.000	[1.070	1.094]
Age: 75 to 84	.9135	.006	.000	[.901	.926]
Age: 85 and over	1.202	.008	.000	[1.186	1.217]
Private rehabilitation center	1.177	.007	.000	[1.164	1.191]
<i>Group of patients <b>without</b> vulnerability code(s) admitted in a rehabilitation center</i>					
Gender	1.508	.004	.000	[1.50	1.515]
Age: 75 to 84	.941	.003	.000	[.936	.946]
Age: 85 and over	1.235	.004	.000	[1.228	1.242]
Private rehabilitation center	1.541	.004	.000	[1.533	1.548]

*Source: PMSI MCO and PMSI SSR, years 2018*

Note: Each row reports regression coefficients from a logit regression model, in year 2018. Standard errors are in parentheses. Standard errors are clustered at the hospital level. The model is run at the stay level.

\*: p<.1; \*\*: p<.05; \*\*\*: p<.01

As a sensitivity analysis, the length of stay, entrance type, discharge type, the Charlson index during the acute care stay were included in the model. We also ran models with different combinations of these variables. The results are equivalent. As an additional sensitivity analysis, we ran the model by ICD groups. The results are consistent.