Physician Workload and Hospital Reimbursement: Overworked Physicians Generate Less Revenue Per Patient

Adam Powell
8 Garrison St. #101, Boston, MA 02116, adampowell@adampowell.com

Sergei Savin
The Wharton School, University of Pennsylvania,
Philadelphia, PA 19104, savin@wharton.upenn.edu

Nicos Savva
London Business School, Regent's Park,
London NW1 4SA, United Kingdom, nsavva@london.edu

We study the impact of physician workload on hospital reimbursement utilizing a detailed data set from the trauma department of a major urban hospital. We find that the proportion of patients assigned a "high-severity" status for reimbursement purposes, which maps, on average, to a 47.8% higher payment for the hospital, is substantially reduced as the workload of the discharging physician increases. This effect persists after we control for a number of systematic differences in patient characteristics, condition and time of discharge. Furthermore, we show that it is unlikely to be caused by selection bias or endogeneity in either discharge timing or allocation of discharges to physicians. We attribute this phenomenon to a workload-induced reduction in diligence of paperwork execution. We estimate the associated monetary loss to be approximately 1.1% (95% Confidence Interval 0.4% − 1.9%) of the department's annual revenue.

Key words: empirical; hospital operations; healthcare reimbursement; workload management

History: February 4, 2012

---

1. Introduction

Hospitals in the developed world are facing a growing set of challenges. Not only are they treating an aging population that places a higher demand on hospital resources, but they are also under constant pressure by public as well as private payers to substantially reduce their operating costs. The combination of these effects has led to a substantial and sustained increase in the daily workload of medical practitioners. One suggested response to this workload increase has been for physicians to spend less time doing paperwork, such as post-discharge record-keeping, in order to free up time to devote to treating patients (Gottschalk and Flocke (2005)).[1] From an operations management

[1] McCormick et al. (2004) found that the level of paperwork-related stress today is so great that two thirds of physicians would be willing to give up 10% of their income for a substantial reduction in paperwork.

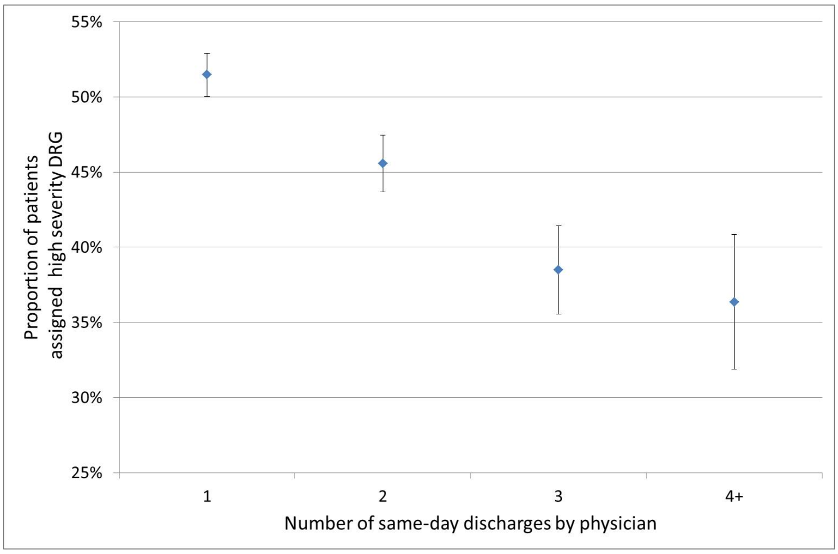

Figure 1 **Proportion of patients assigned a high-severity Diagnostic Related Group (DRG) code (and 90% confidence interval) vs. number of same-day discharges performed by discharging physician for the trauma department of a major non-profit urban hospital, 01/01/2006 to 09/28/2010. A high-severity DRG code assignment corresponds, on average, to a 47.8% higher reimbursement payment than a low-severity DRG code assignment.**

perspective, reducing the time that highly skilled and expensive servers spend performing secondary tasks is desirable. However, reducing the time physicians devote to ancillary activities such as paperwork might have unanticipated, yet significant clinical and financial consequences. Empirical studies have demonstrated that an increase in physician workload has an adverse impact on the quality of residents' discharge summaries (Coit et al. (2010)). Workload-induced degradation in the quality of medical notes and, in particular, of the discharge note, can have implications for follow-up care. Furthermore, since the discharge note is the primary input into the billing process, it may also have implications for hospital reimbursement, which is the focus of this paper.

Utilizing detailed reimbursement data from the trauma department of a major urban hospital, we study the impact of patient-generated workload on hospital reimbursement. Our main finding is that the proportion of patients who are assigned a high- (as opposed to a low-) severity Diagnostic Related Group (DRG) code, which maps, on average, to a 47.8% higher reimbursement payment for the department, is significantly reduced when physicians experience higher than average workloads. As a preview of our results, Figure 1 shows the proportion of patients assigned a high-severity DRG code as a function of the number of same-day discharges completed by the discharging physician. As evident in Figure 1, a physician who discharges a single patient on any given day, and therefore has only one discharge note to write, has a 15% higher chance of getting a high-severity assignment for his patient as compared with a physician who has to do three or more same-day discharges.

The significant reduction in high-severity assignment as workload increases persists even after we control for differences in patient characteristics (gender and age), treatment characteristics (length-of-stay, condition fixed effect, physician fixed effect) and differences in discharge day (day-of-the-week, calendar-month fixed effects). We find that this result is robust to alternative model specifications, and we perform several robustness checks to rule out endogeneity in discharge timing, endogeneity in the allocations of discharges to physicians, and selection bias based on observable factors as the root causes of this observation. We also find that besides the number of same-day discharges a physician performs, the number of inpatients under a physician's care also has an impact on the probability that a discharged patient is assigned a high-severity DRG code. Interestingly, we find that department-level workload does not seem to have an impact on severity assignment, which provides support for the hypothesis that physicians do not share their paperwork load.

We attribute this significant reduction in high-severity assignments to workload-related degradation in the quality of discharge notes. Notes compiled under high-workload conditions are more likely to miss information on patient comorbidities and complications, leading these patients being assigned to a lower severity DRG code than the one medically warranted, a phenomenon one could call "undercoding".[2] This explanation is consistent with a commonly held opinion regarding the value of diligent paperwork execution during the discharge process: Esposito et al. (2006) report that 63% of general surgeons surveyed do not believe that paying closer attention to discharge notes will increase reimbursement. Nonetheless, Reed et al. (2003) demonstrate that increased verification and collaboration in the trauma billing process can increase payments.

Since a high-severity DRG code assignment is associated with an average extra payment of 47.8%, we estimate the relative loss of revenue to the department due to workload-related undercoding to be approximately 1.1% (95% C.I. 0.4% − 1.9%). This amount, even at the lower end of the confidence interval, is substantial, and it would more than cover the cost of reasonable changes in the hospital's discharge process designed to prevent such prevalent undercoding. For example, even at the lower end it would be enough to cover the salaries of two full-time nurses.

Furthermore, we find that the impact of workload on the probability that a patient is undercoded is moderated by the frequency with which the trauma department is handling the patient's condition. The more frequent a condition is to the department, the smaller the impact an extra same-day discharge will have on the probability that a patient is assigned a high-severity DRG code. This reflects the likely underlying causes of DRG undercoding. Physicians, in general, do their best to produce high-fidelity discharge notes that include all relevant information regarding

[2] We would like to thank the anonymous reviewer for suggesting this term.

the patient, the treatment, and any complications or comorbidities. Such complete documentation would lead to high-severity patients being correctly assigned the high-severity DRG code. However, when other activities compete for their time and attention, physicians may not put as much time and/or effort into writing the discharge note. This is less problematic for frequent cases for which the department has developed organizational routines and therefore physicians know what needs to be written in the discharge note. It is, however, a problem for less common cases in which physicians would have to do more extensive research on what information should be recorded.

Beyond hospital reimbursement, our paper has wider implications for operations management. It points to a behavioral impact on system performance that, to the best of our knowledge, has not received substantial attention in the operations management literature. The increase in workload not only affects the speed and the quality of the primary service (see KC and Terwiesch (2009)), but also compromises the ancillary activities that are secondary to the quality of outcomes while essential to generating income. Moreover, this has implications for the optimal design of prospective reimbursement systems in healthcare and elsewhere as it shows that the ability to accurately observe the system, a critical factor in the successful implementation of prospective reimbursements, may depend on the workload level the system is subjected to.

The rest of the paper is organized as follows. In Section 2 we provide a detailed account of hospital reimbursement procedures and of the operations of the trauma department we studied. We then present a literature review in Section 3, followed by a discussion of our main hypotheses in Section 4. We present data and results in Sections 5 and 6. We conclude with a discussion of our main findings and of implications for hospital operations along with suggestions on mitigating measures in Section 7.

2. Hospital Reimbursement and the Trauma Department

2.1. Hospital Reimbursement and DRG Codes

The Centers for Medicare and Medicaid Services (CMS) operate a Prospective Payment System (PPS) to reimburse hospitals for the care of Medicare patients (Mayes (2007)).[3] Under this system, CMS reimburses hospitals for the treatment of eligible patients by predetermined amounts based on patient diagnosis (summarized by a single 3-digit DRG code) rather than on actual procedures performed. The difference between the costs a hospital incurs and the Medicare reimbursement amount is the responsibility of the hospital (Goldsmith (1984)). Thus, while the treatment provided to a patient determines the cost of rendered services, it does not directly influence the revenues that the hospital receives under the PPS, which are determined by diagnosis and patient characteristics,

[3] Many private insurers have chosen to follow suit and use the diagnosis categories provided by CMS as the basis of their payment systems (Reinhardt (2006)).

rather than by physician treatment decisions. CMS deliberately uses prospective payments rather than cost-based payments in order to eliminate the incentive for physicians to over-treat patients in order to increase revenue. This system has long been recognized as promoting "yardstick" competition (Shleifer (1985)).

In order to more accurately reflect the costs of providing services, as well as to reduce the potential incentive for hospitals to selectively treat more profitable patients, CMS enables differential payments to be made, according to the severity of a patient's condition. Thus, several diagnoses have more than one DRG codes associated with them (Cleverley et al. (2010), Ch. 3). For instance, there are three different DRG codes associated with the diagnosis "Traumatic stupor and coma of greater than one hour" (MS-DRG version 25).[4] Patients without Complications or Comorbidities (CCs) are classified as DRG 084, while patients with CCs are classified as DRG 083 and patients with major CCs are classified as DRG 082.

The actual DRG code assigned to a patient is determined by a three-step process (Cleverley et al. (2010), Ch. 2). First, during the patients' care process but mostly during patient discharge, physicians record notes on the patients. After the patient is discharged from the hospital, a professional coder translates the physicians' notes into diagnostic and treatment codes. Once coding is complete, the diagnosis and procedure codes, along with patient characteristics (age, sex, discharge status, complications and comorbidities) are passed to a "DRG grouper" software application, which determines the ultimate DRG code assignment (Department of Health and Human Services (2007)). Note that although a patient might be assigned multiple diagnosis and procedure codes, the patient is ultimately assigned a single DRG code for the duration of their stay at the hospital. The professional coders in charge of transcribing the discharge notes come from a central pool shared by the hospital. As the billing processing does not need to be done urgently, a variable time lag may occur between the time a patient is discharged and the time a patient's discharge note is processed by a coder. Furthermore, coders are assigned to transcribe notes from different parts of the hospital in no systematic fashion. While coding is considered to be accurate enough to be used as the basis for most US hospital payments, Fisher et al. (1992) found that there exists a variation in coding accuracy across medical conditions. This implies that while the accuracy of coding may be related to the nature of the content being coded, for any given condition there is no substantial variation in accuracy. It is worth noting that due to the time lag between discharge and coding, and due to coders being a pooled resource shared by a number of departments across the hospital, it is unlikely that coding errors (if any) made by professional coders would be correlated with physician discharge workload.

[4] For a complete list see http://www.cms.hhs.gov/AcuteInpatientPPS/downloads/FY_2010_FR_Table_5.zip last accessed June 16, 2011.

There is an ongoing debate in economic literature as to whether hospitals engage in "upcoding" behavior, where patients are switched from low-paying to high-paying DRG codes on grounds other than medical. Early studies (Carter et al. (1990)) find little evidence of upcoding, while more recent studies (Silverman and Skinner (2004), Dafny (2005)) demonstrate that hospitals did engage in upcoding for selective DRGs where such behavior is profitable. They show that for-profit hospitals engage in these activities more actively than non-profit or government hospitals. These studies are based in data from the late 1980's and early 1990's, when coding practices where not audited as rigorously as they are today (in particular, under Section 302 of the Tax Relief and Healthcare Act of 2006).[5] Due to the fact that the hospital we study is a not-for-profit institution and because of the rigorous auditing regime which hospitals are subject to, it is unlikely that upcoding is a significant problem at the hospital we investigate. Therefore, any reduction in DRG severity assignment associated with physician discharge workload is unlikely to be due to workload preventing physicians from attempting to "upcode" patients.

2.2. The Operations of the Trauma Department

We have decided to focus on the trauma department since we believe that econometrically it offers the cleanest test case for our enquiry, as a number of confounding factors are less pronounced in the trauma treatment environment.

The first advantage of studying the trauma department has to do with the nature of the admission process. Trauma department patients, due to the fact that they are admitted as a result of accidents or violence,[6] have a restricted ability to exercise choice, as do the physicians working for the trauma department. More specifically, neither trauma patients nor their physicians are able to select the location or timing of their admission. Patients are allocated to the admitting physician who happens to be on duty at the time they arrive to the hospital and have neither the time nor the ability to research or select their care providers or even the hospital they are admitted to. For the same reason, trauma physicians are also restricted in both the number and the case mix of patients they treat. They treat the patients that happen to be admitted while they are on admission duty. This inability to load-balance admissions creates greater workload variability than is present in departments in which patient admissions may be scheduled in advance.

The second reason we chose the trauma department has to do with the fact that, during the patient's entire hospital stay, his/her care is coordinated mainly by one physician, known as the "practitioner of record". The practitioner of record in the vast majority of cases is the admitting

[5] see http://www.cms.gov/recovery-audit-program/ last accessed June 16, 2011.

[6] Patients arrive at a typical trauma center both through ground transportation and by helicopter, traveling a distance of up to forty miles (Cocanour et al. (1997)).

physician. While this physician may be supported by a team of residents and nurses, the provided treatment is ultimately his or her responsibility. Similarly, although the team contributes to writing the discharge notes, it is the practitioner of record who is responsible for their contents.

Third, trauma department physicians have less flexibility than their colleagues in other departments in determining when to discharge patients under their care. As the patient recovers, the trauma department monitors the patient's health through the case management team, which is responsible along with the practitioner of record for planning the patient's eventual discharge. Due to the complexity of most trauma cases, the trauma department remains in charge of patients even if the patient is transferred to another department within the hospital. The case management team meets every weekday to discuss when patients will be ready to leave the hospital and what services they will need once they depart (Curtis (2007)). When the case management time decides that a patient is ready to leave, the discharge process is initiated by the practitioner of record.

The fourth advantage of the trauma department for the purposes of studying the impact of workload on severity assignment and thus on income, has to do with the proportion of patients allocated to a high-severity DRG code in the trauma department. As is shown in Figure 1, this proportion is near 50%, which is relatively high compared to the other departments within the hospital that we studied, which makes identification of any workload-induced systematic changes in severity assignment easier to observe.

3. Literature Review

Our research relates to the growing body of empirical research on healthcare operations in general and to the literature on the impact of workload in particular. On the one hand, there is evidence that higher physician volume, i.e. higher number of patients treated, leads to higher-quality care for a number of conditions (e.g. cancer treatment (Hillner et al. (2000)), coronary angioplasty (Jollis et al. (1997)), and pneumonia treatment (Lin et al. (2008))). Similarly, there is also evidence that higher volume at the hospital level lead to higher-quality care as well (Luft et al. (1979), Jollis et al. (1997), Hillner et al. (2000), Birkmeyer et al. (2002), Macias et al. (2009)). Both of these effects can be explained by individual and organizational learning effects as well as the benefits of specialization. On the other hand, such an increase in volume, when coupled with an increase in provider workload, i.e. an increase in the number of patients seen by individual nurses, physicians, or the department per unit of time, has unintended consequences. KC and Terwiesch (2009) find that workload affects both service times and mortality rates. They show that as the workload increases, procedures take less time to complete, but that this excess speed comes at the expense of patient safety as mortality rates increase during high-workload periods. Furthermore, they show that the decrease in completion time is not sustainable, since medical practitioners eventually tire and slow down. Kuntz et al. (2011) find that organizational workload has a nonlinear impact on quality of care. They show that outcomes deteriorate when workload increases from already high workload levels, while they improve if workload is increased from low workload levels. Green et al. (2010) study a different effect of workload on the healthcare system. They show that workload is linked to nurse absenteeism and find that absenteeism is exacerbated on days when the workload is anticipated to be higher due to insufficient staffing levels. Workload can also inhibit organizations from learning from past mistakes. For example, Tucker and Edmondson (2003) note, based on in-depth field research, that heavy workload is in part responsible for nurses' focus on providing "locally optimal" solutions for individual process failures at the expense of spending time identifying and rectifying the root cases of such failures. Such behavior does not create a conducive environment for long-term and system-wide changes that can improve hospital efficiency. Our work complements this line of research by demonstrating an unintended consequence of increased workload on hospital reimbursement rates. Thus, our paper also contributes to the debate on how human factors affect the productivity of a system (Schultz et al. (1998), Schultz et al. (1999), Oliva and Sterman (2001), Boudreau et al. (2003), Mas and Moretti (2009)).

When organizations perform tasks repeatedly they learn and develop routines adapted to their needs and environments (Nelson and Winter (1982)). Routines reduce the need for organizations to discover solutions every time they face a problem, and these routines evolve over time through experiential learning. While repetition allows organization to develop routines, Zollo and Winter (2002) argue that the frequency with which an organization executes specific tasks has a dramatic impact on the development of such routines. They argue that tasks performed at low frequency suffer significant losses in their capability-building power. Furthermore, they argue that for such tasks, explicit learning mechanisms, such as knowledge codification, will be more effective in building organizational routines as opposed to relying on tacit experience accumulation. Although our research does not explicitly tests the impact of frequency on learning to develop organizational routines, by investigating the moderating effect of task frequency on workload-related undercoding we indirectly examine one of its consequences. Our findings are consistent with the hypothesis that developing organizational routines is more effective for high-frequency tasks.

Finally, our work is related to the literature on incentives in hospital reimbursement (see a review by Newhouse (1996)). For example, Shleifer (1985) discusses how the Medicare DRG-related reimbursement policy provides incentives for monopolist providers to reduce costs. Fuloria and Zenios (2001) present a principal-agent payment system and show that social welfare is maximized when reimbursements are outcome-adjusted. In general, this literature assumes that even if the treatment is private information for the provider, the payer is able to verify the diagnosis and health outcome of a patient. As our research shows, this assumption might not be valid, since the documentation generated by physicians is sensitive to operational conditions such as system workload. Therefore any reimbursement plan that aims to align incentives between payers and providers needs to take this endogeneity into account.

4. Hypotheses Development

Broadly speaking, there are three main activities that compete for the time and attention of healthcare providers. These are the admission of new patients, the monitoring and continuation of care for already admitted patients, and the discharging of sufficiently recovered patients. The admissions process at the trauma department we study is decoupled from care and discharge which are treated more as business as usual. Admitting physicians serve at the admissions bay in shifts, and while they are working on admissions, they are not treating or discharging any of their previously admitted patients. For the development of our hypotheses, we focus on the last two activities.

Each individual discharge generates a substantial amount of physician-related workload, and, as a consequence, physicians have a limited amount of time to devote to each discharge note. Similarly, the time a physician spends attending to inpatients competes with the time he/she devotes to writing discharge notes. Consistent with previous research in this area (Coit et al. (2010)), we conjecture that the quality and the level of detail of each discharge note will decrease as the workload, defined as the number of same-day discharges performed by, or the number of in-patients under the direct of, a physician increases. More specifically, we conjecture that a discharge note compiled under high-workload conditions will have a less detailed record of the patient's comorbidities and complications. This conjecture reflects the fact that compiling such evidence takes time, effort and attention to detail; all of which are likely to be in short supply when the discharging physician is under high-workload pressure. While the adverse impact of workload on the quality of the discharge note might be symmetric across patients of different severities, we expect it to have an asymmetric impact on severity assignment, since for a patient to be awarded the higher-severity status the discharge note needs to provide detailed evidence for comorbidities and complications. If such evidence is missing or is not complete, then a patient whose condition should have classified as high-severity will be given a low-severity assignment. Conversely, if such evidence is missing from the discharge note of a patient whose condition should be classified as low-severity, then the patient is not more likely to be assigned the high-severity status. Therefore, as the discharge workload increases we would expect to see fewer patients assigned to a high-severity DRG code.

Hypothesis 1. *As the number of same-day discharges performed by a physician increases, the probability that each of these discharged patients is awarded a high-severity DRG code decreases.*

Hypothesis 2. *As the number of inpatients under the care of the discharging physician increases, the probability of a discharged patient being awarded a high-severity DRG code decreases.*

The benefits of learning from experience have long been a part of the management literature, from Smith's arguments for the benefits of specialization in pin production, to Weber's description of bureaucracies undergoing experiential learning, to Cyert and March's Behavioral Theory of the Firm (Smith (1776), Weber (1922/1978), Cyert and March (1963)). When organizations perform tasks repeatedly, they learn and develop routines adapted to their needs and environments (Nelson and Winter (1982)). Routines reduce the need for organizations to discover solutions every time they face a problem, and these routines evolve over time through experiential learning. It is likely that trauma departments have more developed routines for handling patients with common conditions, than those with rare conditions, since common conditions provide more opportunities for experiential learning. Such frequent interactions are well suited for the development of a "capability-building mechanism based on tacit accumulation in the minds of "expert" personnel" (Zollo and Winter (2002)). When physicians at the trauma department are under high workload pressure, they can fall back on optimized organizational routines when handling patients with familiar conditions, but they must expend more effort on discovering the care and documentation process when handling patients with rarer conditions for which such routines are not as well established. As these routines are established at the level of the department as opposed to the individual physician, our third hypothesis is that the impact of discharge and inpatient workload on the probability that a patient is undercoded is moderated by the frequency with which the trauma department treats a specific condition.

Hypothesis 3. *The negative relationship between physician workload on discharge day and probability of high-severity DRG code assignment is moderated by the frequency with which the patient's condition is treated by the trauma department.*

5. Data Description

To empirically test the hypotheses articulated in the previous section, we utilize a detailed data set from the trauma department of a large urban hospital. Our data spans a period of 45 months, starting on January 1, 2006 and ending on September 28, 2010, and represents a complete description of treatment activities and patient characteristics for all $7,100$ patients admitted to the trauma department for the reported period. For each admission we know the admission date, discharge date, patient characteristics (gender, age, race, insurance plan), treatment characteristics (assigned DRG, days stayed in hospital, physician identifier) and billing data. During the period for which we collect data, CMS moved from version 24 of the DRG code through to version 27. The differences between versions 25 to 27 were minor, while the differences between versions 24 and 25, which was implemented on October 1, 2007, were more drastic. In version 24, many DRG classifications were "paired" to reflect the presence of complications or comorbidities (CCs), while in version 25 the pairing was, in many instances, replaced with a three-tiered system: absence of CCs, presence of CCs, and a higher level of presence of major CCs (see Department of Health and Human Services (2007)). To allow a like-to-like comparison the hospital re-assigned all the claims in our data set using the official software for the DRG version 25 system.

In order to be able to examine how severity assignments are made, we have lumped related DRGs that are only differentiated by severity into the same "Base-DRG". For those diagnoses with two or three severity levels, we created a separate severity variable, which was set to 0 for the lowest severity and 1 for the other severity level(s). For the example of "Traumatic stupor and coma of greater than one hour", which has three different DRGs associated with it - 084, 083 and 082 for patients without CCs, with CCs and with major CCs, respectively - all of these DRGs would fall under the same base-DRG code, while patients with DRG 084 would be mapped to a severity level of 0, and patients with DRG 082 and 083 would be mapped to a severity level of 1.

To construct workload measures, such as number of in-patients treated at the trauma department, we rely on the admission and discharge data. For the first few days of the data set we do not have an accurate picture of how many patients the department is treating because these patients were admitted before our data set starts. To avoid this left-censoring problem, we exclude from the workload regressions the 59 discharges that took place in the first 20 days of the data set. Since the length of stay is less than 20 days for about 96% of patients, the choice of 20 days as a cutoff is a conservative one. Similarly, we cannot construct workload measures for patients discharged after September 28, 2010 when we no longer have admissions data. To avoid this right-censoring issue, we exclude the 11 discharges that occurred after the last admission date.

While we use all remaining patients to construct workload measures, for the severity assignment regressions we also exclude any base-DRG code that does not have a high-severity assignment (this excludes 664 observations) and any base-DRG codes that have only high- or only low-severity assignment in our data sample (190 data points). For example, one such excluded DRG is that for chest pain, DRG 313, which has no high-severity counterpart. We also drop an additional 11 observations where the practitioner of record was either not available or was not a trauma department regular, as well as 9 patients whose age was recorded to be between 124 and 128 years, which we consider a typo. This leaves $6,165$ observations.

Our billing data includes the actual payment the hospital received and not just the amount charged. The actual payment can be very different from the amount charged as patients might not pay because they are uninsured, or because insurers negotiate preferential treatment or volume discounts, etc. Having the actual payment allows us to examine the impact of workload-induced undercoding on actual financial performance as opposed to estimated financial performance based on charges or prospective payments. To fully reflect the true severity-related differences in reimbursement, we use reimbursement data starting on October 1, 2007, when DRG version 25 went into effect. We also exclude any observations in which the payment amount is not reported, mainly because the bill has not yet been settled. This happens mostly in the last 3 months of data. This leads to a final count of 3793 patient observations for the reimbursement regression. Of the 3793 patients we observe, 257 (or 6.78%) did not pay for the delivered care.

Table 1 shows descriptive statistics for the dependent and independent workload-related variables and their correlations. The unit of analysis is individual patient discharges. The dependent variable "High-Severity Assignment" takes the value 1 if a patient receives a high-severity DRG code and 0 otherwise. The "Physician's Discharge Workload" variable stands for the total number of same-day discharges performed by the discharging physician, while the "Total Patient Discharges" designates the respective number of patients discharged by the entire trauma department. The rest of the variables are similarly defined. Note that some of the pairwise correlation coefficients are significant, however we discovered no significant multi-collinearity problems. Descriptive statistics for hospital payments and for patient age and length of stay appear in Table 2. All analysis and econometric tests presented were implemented in STATA/IC 10.0 for windows

Table 1 Descriptive statistics and correlation table

	Descriptive Statistics					Correlation Table					
Variable	Obs	Mean	S.D.	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)
(1) High-Severity Assignment	6165	0.47	0.50	0	1	1.00					
(2) Physician's Discharge Workload	6165	1.70	0.91	1	6	-0.10***	1.00				
(3) Physician's Admission Workload	6165	0.26	0.83	0	11	0.00	0.02	1.00			
(4) Physician's Inpatient Workload	6165	3.69	2.50	0	15	-0.11***	0.19***	-0.00	1.00		
(5) Total Patient Discharges	6165	5.48	2.48	1	15	0.01	0.39***	0.02*	0.06***	1.00	
(6) Total Inpatients	6165	25.41	7.48	0	50	-0.01	0.06***	0.03**	0.38***	0.16***	1.00
(7) Total Patient Admissions	6165	4.34	2.30	0	13	-0.00	0.07***	0.11***	0.19***	0.20***	0.44***

***, **, * denote statistical significance of the correlation coefficient at 1%, 5%, and 10% confidence levels, respectively.

Table 2 Descriptive statistics of hospital payment and control variables.

Variable	Obs	Mean	Std. Dev.	Min	Max
Payment ($)	3793	15,266	20,744	0	347,386
Patient Age (Years)	6165	43.3	19.8	12	106
Length of Stay (Days)	6165	5.49	5.49	0	140

6. Econometric Analysis

Our first model focuses on testing our hypotheses on the influence of the discharge and inpatient workloads on the probability of high-severity DRG code assignments. In particular, we estimate the following logistic regression:

$$\begin{aligned} &\ln\left(\frac{\text{Prob}(\text{SevAssign}_{i,s,t}=1)}{\text{Prob}(\text{SevAssign}_{i,s,t}=0)}\right)=\beta_0+\beta_{\mathbf{c}}\text{Controls}_{i,s,t} \\ &+\beta_1\text{PhysDisLoad}_{s,t}+\beta_2\text{PhysAdmLoad}_{s,t}+\beta_3\text{PhysInpLoad}_{s,t} \\ &+\beta_4\text{DepDis}_t+\beta_5\text{DepAdm}_t+\beta_6\text{DepInp}_t, \end{aligned} \tag{1}$$

where the unit of analysis is patient i, discharged by physician s on date t. $\text{SevAssign}_{i,s,t}$ is the indicator for patient severity assignment. It takes the value 1 when the patient receives a high-severity assignment and 0 otherwise. The variables $\text{PhysDisLoad}_{s,t}$, $\text{PhysAdmLoad}_{s,t}$ and $\text{PhysInpLoad}_{s,t}$ denote the number of discharges, admissions, and inpatients looked after by the discharging physician s on the discharge date t. The variables DepDis_t , DepAdm_t and DepInp_t denote the number of discharges, admissions, and inpatients of the trauma department on the discharge day t.

Models I and II, presented in Table 3, have different controls $\text{Controls}_{i,s,t}$. Model I controls for patient characteristics (age, gender and length of stay), physician and time fixed effects (calendar month and day of the week), while Model II adds base-DRG fixed effects. This last control allows us to model heterogeneity in the propensity of specific conditions to generate high-severity assignments. However, this is computationally and data intensive, as it requires an additional 111 controls. To address potential correlation of error terms, we cluster standard errors in both models on physicians.

As evident from Table 3, we find strong support for both hypotheses 1 and 2. Both models show that the discharging physician's discharge workload as well as the number of inpatients assigned to the discharging physician have a significant and negative impact on the probability that a patient is assigned high severity. According to Model I (Model II), the coefficient of the physician discharge workload is -0.0938, p-value<0.00 (-0.0853, p-value$=0.02$) while the coefficient of the physicians inpatient workload is -0.0289, p-value$=0.02$ (-0.0261 p-value$=0.03$). Furthermore, Model I (Model II) suggests that the average marginal effect of an extra same-day discharge by the discharging physician on the probability of high-severity assignment is -2.3%, p-value<0.000, (-2.1%, p-value$=0.02$) while the marginal effect of an extra inpatient under the care of the discharging physician is -0.7%, p-value$=0.015$ (-0.6%, p-value$=0.03$). The physician's admission workload and all the department workload measures are not significantly different from zero, confirming that admissions are decoupled from the rest of the discharge workload and that physicians do not share workload with one another. Turning to goodness-of-fit measures, the Hosmer-Lemeshow statistic

	Dependent variable: Was the patient assigned high severity?	
	Model I	Model II
VARIABLES	Logit Coefficient	Logit Coefficient
Physician's Discharge Workload	-0.0938***	-0.0853**
	(0.0208)	(0.0351)
Physician's Admission Workload	0.0696	0.0415
	(0.0470)	(0.0607)
Physician's Inpatient Workload	-0.0289**	-0.0261**
	(0.0118)	(0.0120)
Total Patient Discharges	0.0097	0.0193
	(0.0125)	(0.0126)
Total Inpatients	-0.0031	0.00111
	(0.00885)	(0.00948)
Total Patient Admissions	0.0099	-0.00209
	(0.0114)	(0.0142)
Constant	-1.89***	-2.11
	(0.439)	(1.82)
Fixed Effects		
Calendar Month	Yes	Yes
Day of the Week	Yes	Yes
Patient Gender	Yes	Yes
Patient Age (5 categorical variables)	Yes	Yes
Length of Stay (5 categorical variables)	Yes	Yes
Physician	Yes	Yes
Base-DRG	No	Yes
Observations	6165	6165
Log Likelihood	-3273	-2506
Pseudo R-Squared	0.233	0.412

***, **, * denote statistical significance at 1%, 5%, and 10% confidence levels, respectively.
Errors shown in parentheses are clustered on physicians.

Table 3 Results of Models I and II.

(with 10 and 20 groups) rejects the hypothesis that the logit model is not the correct model for both Models I and II. As a robustness check we also ran two additional models. First, we estimated a regression model with all the controls of Model I and with the physician level workload variables, but without the unit-level workload variables. Second, we estimated a regression model with all the controls of Model I and with the unit-level workload variables but without the physician-level workload variables. The former model produced results very similar to the ones presented in Table 3, while the latter produced no significant results. This confirms that physician-level discharge and inpatient workload matter when it comes to severity assignment while unit level workloads have no explanatory power in predicting severity assignment.

The linear specification used in Models I and II (Table 3) is, to a large extent, *ad hoc*. After all, it seems plausible that the impact of an increase in workload only "kicks in" after a certain threshold has been exceeded. To this end, we present a non-linear model specification for the three physician-related explanatory variables in Table 4. The non-linear specification breaks the physician's discharge workload into three categories (with 1, 2 and 3+ discharges, containing 53.5%, 30.0% and 15.5% of the observations, respectively), the physician's admission workload into two categories (0 and 1+, containing 87.2% and 12.8% of observations, respectively) and the rest of the variables in quartiles. For each category, we define a dummy variable that takes the value 1 when the explanatory variable is in that category and 0 otherwise. For each explanatory variable, we omit the first category so that the coefficients should be interpreted as difference from the lightest workload state. As evident from Table 4, the main finding is that physician discharge workload has a substantial and significant effect once the discharge workload is equal to or greater than 3 (marginal effect $= -5.3\%$, coefficient$= -0.217$, p-value$= 0.003$). Similarly, when the number of patients assigned to the same physician is in the 3rd or 4th quartile, then the probability that a discharged patient is assigned the high-severity DRG code is reduced significantly in the case of the 3rd quartile (marginal effect$= -3.2\%$, coefficient$= -0.129$, p-value$= 0.08$), but not significantly in the case of the 4th (marginal effect$= -3.4\%$, coefficient$= -0.139$, p-value$= 0.19$). The lack of significance in the case of the 4th quartile, which is observed despite the fact that the coefficient is larger in absolute terms than the coefficient of the 3rd quartile, might be indicative of the variability in the workload generated by inpatients as the standard error of the coefficient is quite high. Individual physician admissions workload does not play a significant role in severity assignment. Looking at the department-level variables, there seems to be a positive effect when the department processes a smaller number of discharges than the median (marginal effect$= 5.6\%$, coefficient$= 0.228$, p-value$= 0.01$). The rest of the variables are not significant.

We next examine the impact of the frequency with which the trauma department treats specific conditions on severity assignment. We measure the frequency by counting the number of instances a base-DRG code appears in the 45 months for which we have data. Models IV and V, presented in Table 5, include the interaction of DRG frequency with the continuous workload variables of Model II and the dummy workload variables of Model III that were statistically significant. We find that frequency has a significant moderating impact on the effect of discharge workload on severity assignment (coefficient$= 0.000659$, p-value$= 0.02$). The average marginal effect of an extra discharge on the probability a patient is assigned the high-severity DRG code[7] is increased (i.e. becomes less negative) and is significant for most patients as the number of times the trauma department sees a condition increases. Averaged over all patients, an increase in the frequency with which the trauma department sees a condition from just under 1 per month (43 instances in our data-set) to 3.8 per month (i.e. 170 instances in our data-set) halves the marginal effect of discharge workload (from -2.2% to -1.1%). The frequency with which the trauma department treats specific conditions does not seem to play a role in severity assignment when the number

[7] To calculate the marginal effect of an extra discharge as a function of the frequency we follow Ai and Norton (2003). Despite the fact that the interaction coefficient is positive and significant, as Ai and Norton (2003) point out, the non-linear nature of the models we use implies that for some patients the impact of a change in the frequency on the marginal effect of an extra discharge could be negative. Although this change of sign is due to a "mechanistic saturation effect" that occurs as the probability of high-DRG code assignment approaches either 0 or 1 (Kolasinski and Siegel (2010)), one needs to take into account that the interaction effect is different for different patients.

VARIABLES	Dependent variable: Was the patient assigned high severity? Model III Logit Coefficient
DV=1 if Physician's Discharge Workload=2	-.0576
	(0.0871)
DV=1 if Physician's Discharge Workload=3	-0.217***
	(0.0722)
DV=1 if Physician's Admission Workload>0	0.0875
	(0.120)
DV=1 if Physician's Inpatient Workload in 2nd Qrtle	.0643
	(0.105)
DV=1 if Physician's Inpatient Workload in 3rd Qrtle	-.129*
	(0.0738)
DV=1 if Physician's Inpatient Workload in 4th Qrtle	-.139
	(0.106)
DV=1 if Total Patient Discharges in 2nd Qrtle	.224***
	(0.0781)
DV=1 if Total Patient Discharges in 3rd Qrtle	.0736
	(0.110)
DV=1 if Total Patient Discharges in 4th Qrtle	0.1534
	(0.109)
DV=1 if Total Inpatients in 2nd Qrtle	0.153
	(0.112)
DV=1 if Total Inpatients in 3rd Qrtle	.140
	(0.144)
DV=1 if Total Inpatients in 4th Qrtle	.0614
	(0.160)
DV=1 if Total Patient Admissions in 2nd Qrtle	.0537
	(0.139)
DV=1 if Total Patient Admissions in 3rd Qrtle	.0271
	(0.0851)
DV=1 if Total Patient Admissions in 4th Qrtle	.121
	(0.118)
Constant	-2.21
	(1.05)
Fixed Effects	
Calendar Month	Yes
Day of the Week	Yes
Patient Gender	Yes
Patient Age (5 categorical variables)	Yes
Length of Stay (5 categorical variables)	Yes
Physician	Yes
Base-DRG	Yes
Observations	6165
Log Likelihood	-2502
Pseudo R-Squared	0.413

DV stands for dummy variable. ***, **, * denote statistical significance at 1%, 5% and 10% confidence levels, respectively. Errors shown in parentheses are clustered on physicians.

Table 4 Results of Model III.

of patients a physician is in charge of is higher than the median, although this could be due to substantial variability in the workload generated by inpatients. Therefore, we find partial support for our third hypothesis.

6.1. Robustness Checks

One possible concern with respect to our results is that patients may be less likely to be awarded the high-severity assignment on a busy day not due to undercoding, but rather because physicians *choose* to discharge patients of lower severity when they are busy, perhaps in order to free up capacity Chan et al. (2011). Although one of the reasons for choosing the trauma department, is that each case is monitored by the case management team, and, therefore, the individual physician has

	Dependent variable: Was the patient assigned high severity?	
	Model IV	Model V
VARIABLES	Logit Coefficient	Logit Coefficient
DV=1 if Physician's Discharge Workload=2		-0.152
		(0.167)
DV=1 if Physician's Discharge Workload=3		-0.474***
		(0.153)
Base-DRG Freq x DV=1 if Physician's Discharge Workload=2		0.000629
		(0.000857)
Base-DRG Freq x DV=1 if Physician's Discharge Workload=3		0.00171**
		(0.000761)
DV=1 if Physician's Admission Workload>0		0.0912
		(0.119)
DV=1 if Physician's Inpatient Workload in 2nd Qrtle		0.0709
		(0.104)
DV=1 if Physician's Inpatient Workload in 3rd Qrtle		-0.118
		(0.0777)
DV=1 if Physician's Inpatient Workload in 4th Qrtle		-0.131
		(0.113)
Base-DRG Freq x DV=1 if Physician's Inpatient Workload in 2nd Qrtle		2.73e-05
		(0.000761)
Base-DRG Freq x DV=1 if Physician's Inpatient Workload in 3rd Qrtle		-0.000142
		(0.000877)
Base-DRG Freq x DV=1 if Physician's Inpatient Workload in 4th Qrtle		-0.000221
		(0.000942)
Total Patient Discharges	0.0202	0.0202*
	(0.0125)	(0.0119)
Total Inpatients	0.00123	0.000792
	(0.00955)	(0.00941)
Total Patient Admissions	-0.00118	-0.00110
	(0.0142)	(0.0146)
Physician's Discharge Workload	-0.183***	
	(0.0561)	
Base-DRG Freq x Physician's Discharges	0.000657**	
	(0.000279)	
Physician's Admission Workload	0.0418	
	(0.0597)	
Physician's Inpatient Workload	0.0102	
	(0.0185)	
Base-DRG Freq x Physician's Inpatients	-0.000241**	
	(0.000105)	
Constant	-2.01	-1.439
	(1.87)	(1.206)
Fixed Effects		
Calendar Month	Yes	Yes
Day of the Week	Yes	Yes
Patient Gender	Yes	Yes
Patient Age (5 categorical variables)	Yes	Yes
Length of Stay (5 categorical variables)	Yes	Yes
Physician	Yes	Yes
Base-DRG	Yes	Yes
Observations	6165	6165
Log Likelihood	-2503	-2503
Pseudo R-Squared	0.413	0.413

DV stands for dummy variable. ***, **, * denote statistical significance at 1%, 5% and 10% confidence levels, respectively.
Errors shown in parentheses are clustered on physicians.

Table 5 Results of Models IV and V.

less flexibility as opposed to other hospital departments, the problem remains that the physicians to some extent can choose when to discharge a patient. It is plausible that when a physician is busy, he/she might decide to reduce the threshold of recovery at which they discharge a patient (see KC and Terwiesch (2009)). If they do so selectively, that is, if they choose to reduce the discharge threshold for patients of low-severity more than they reduce the threshold for patients of high-severity, that would also generate results similar to ours. However, if physicians do selectively discharge easier cases during busy times, one would expect that patients remaining in hospital after a busy day would be, on average, of higher severity. Therefore, if the day(s) before the current discharge day were busy, one would expect to see an increase in the proportion of high-severity assignments for the patients released on the current day. An alternative, but related mechanism which could produce results similar to ours is the following. When a physician is busy, he/she might choose not to discharge fully recovered high-severity patients but still go ahead and discharge any fully recovered low-severity patients who would generate little administrative burden. However, if this were the case, then one would expect that these high-severity fully recovered patients would be discharged on the very next day or in the next few days. Therefore, if the day(s) before the current discharge day were busier than normal, we would observe an increase in the proportion of high-severity assignments.

We test this possibility by explicitly including lagged workload in our model as an explanatory variable. In Model VI, presented in Table 6, we include previous-day (lag 1) physician workload variables (i.e. the number of discharges as well as the number of inpatients assigned to a physician on the previous day). Since for roughly two thirds of our sample, any given physician does not discharge patients on two consecutive days, this reduces the data points on which we can fit the model to $2,611$. Nevertheless, the impact of contemporaneous discharge workload remains significant (coefficient$= -0.108$, p-value$= 0.05$) and the impact of the lagged discharge workload variable is non-positive and statistically indistinguishable from 0 (coefficient$= -0.0412$, p-value$= 0.46$). The impact of inpatient workload is no longer significant (coefficient$= -0.0305$, p-value$= 0.29$), but neither is the impact of the previous-day inpatient workload (coefficient$= 0.0256$, p-value$= 0.47$). The second test we perform is to include lags from the previous 7 days. To avoid the problem of further reducing the sample size with the introduction of lags, we set the lagged workload equal to 0 when it is missing and we include 7 dummy variables in the model (1 for each lagged variable) that take the value 1 when we have a missing data point and 0 otherwise. To deal with left-censoring of the lagged variables, we add another 7 days to the first 20 days we exclude from the regression. The estimated coefficients of the lags for the discharge and inpatient workload variables are reported in Figure 2, along with their 95% confidence intervals. The first observation is that there is no systematic pattern in the lags. The second is that none of the lagged discharge workloads has a significant positive effect (even at the 90% confidence level), while the contemporaneous discharge workload remains negative and significant. Similarly, all the inpatient lagged workload variables are statistically indistinguishable from 0. This is consistent with the hypothesis that when physicians chose to discharge patients "early" due to increased workload, it seems that they do not discriminate, at least not in a statistically significant way, between "nearly" fully recovered patients that were of high severity or "nearly" fully recovered patients that were of low severity. This provides evidence that the measured reduction in high-severity DRG code assignments is not due to endogeneity but to undercoding.

	Dependent variable: Was the patient assigned high severity?	
	Model VI	Model VII
VARIABLES	Logit Coefficient	Logit Coefficient
St. Dev. Discharges x Physician's Discharge Workload		0.0223
		(0.0709)
St. Dev. Discharges		0.184
		(0.172)
Physician's Discharge Workload	-0.108**	-0.157*
	(0.0550)	(0.0941)
Physician's Discharge Workload Previous Day	-.0411	
	(.0554)	
Physician's Admission Workload	-0.000340	0.0426
	(0.0736)	(0.0611)
Physician's Inpatient Workload	-0.0305	-0.0254**
	(0.0291)	(0.0124)
Physician's Inpatient Workload Previous Day	.0256	
	(0.0353)	
Total Patient Discharges	0.0175	0.00219
	(0.0193)	(0.0186)
Total Inpatients	-0.00720	0.000613
	(0.0118)	(0.00969)
Total Patient Admissions	0.0290*	-0.00104
	(0.0160)	(0.0146)
Constant	-2.222***	-2.795**
	(0.573)	(1.394)
Fixed Effects		
Calendar Month	Yes	Yes
Day of the Week	Yes	Yes
Patient Gender	Yes	Yes
Patient Age (5 categorical variables)	Yes	Yes
Length of Stay (5 categorical variables)	Yes	Yes
Physician	Yes	Yes
Base-DRG	No	Yes
Observations	2611	6165
Log Likelihood	-1374	-2503
Pseudo R-Squared	0.233	0.413

***, **, * denote statistical significance at 1%, 5% and 10% confidence levels, respectively. Errors shown in parentheses are clustered on physicians.

Table 6 Results of Models VI and VII.

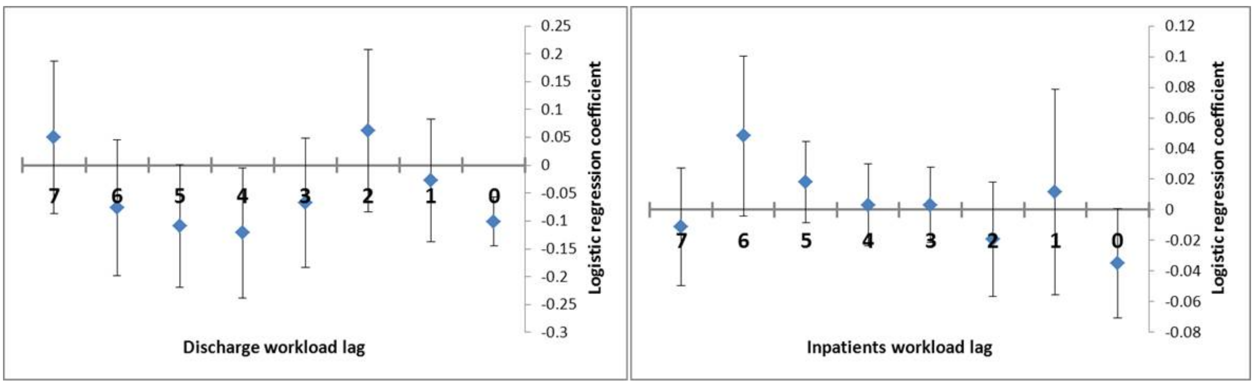

Figure 2 Coefficients of lagged discharge and inpatient workload variables (and 95% confidence intervals). The coefficients were estimated by including the lags in the logistic regression presented in Model I along with a matrix of dummy variables that take the value 1 when a lagged variable is missing and 0 otherwise. The dummy variables for the missing values were not significant (at the 5% level) except for lag 5, which was significant and negative.

Although patient care at the trauma department is coordinated by the "practitioner of record" and handovers are rare, another possible, even though unlikely explanation that would generate results consistent with ours is that physicians coordinate in deciding who gets to discharge which patients in order to equate their workload. More specifically, physicians could "allocate" patient discharges so that one physician does fewer but more time-consuming (i.e. more severe) cases and another physician does a larger number but less-time consuming (i.e. less severe) cases. To test whether this is the case, we measure the standard deviation of the number of discharges done by physicians on any given day and use this as a measure of heterogeneity in the number of discharged patients. If the hypothesis that physicians coordinate their discharge allocations holds true, we would expect the standard deviation to have an amplifying effect on the impact of workload on severity assignments. That is, on days when there is a greater amount of heterogeneity in the number of discharges performed (i.e. some physicians discharge very few but more severe cases and others discharge a greater number of less severe cases), we would expect to find the probability of a high-severity assignment for those physicians with a high workload to be even lower compared to days with less heterogeneity. As evident from Model VII, presented in Table 6, this is not the case. The coefficient of the interaction term between standard deviation and number of discharges is positive instead of being negative and is statistically indistinguishable from zero (coefficient= 0.0223, p-value= 0.76).

The third robustness test we performed is designed to address a potential selection bias problem. Clearly, it is impossible to observe the severity assignment for the same patient under both high- and low-discharge workload conditions. Therefore, one potential concern is that patients discharged under high workload conditions are, for one reason or another, systematically different from patients discharged under low workload conditions. This could be problematic for our conclusions, as any difference in severity allocation might not be due to a causal link between severity assignment and workload but due to this selection bias. To partially address this problem, we included several control variables in all logistic regressions (see, for example, equation (1)). However, the question remains as to whether the selection bias is appropriately and fully addressed by the control variables and by the chosen linear specification. An alternative approach is to match patients discharged under high workload conditions to patients discharged under low workload conditions, so that these individuals are almost identical in all observable characteristics except the workload conditions on discharge. Although this does not alleviate the concern that the patients are heterogenous based on a variable not unobservable to us, the researchers, but observable to physicians, it does provide some extra support as it relaxes the linearity restriction imposed by the logistic regression.

To do the matching, we define a categorical variable that takes the value 1 if the patient is discharged under high workload conditions (i.e. the physician has discharged 3 or more patients on that day or is looking after more patients than the median) and 0 otherwise. We use this as the treatment variable. Since we have a number of observable factors (patient age and gender, length of stay, physician, day-of-the-week, month and condition fixed effects), several of which are continuous variables, finding an exact match for every patient is not advisable due to the curse of dimensionality. Instead, we rely on propensity score-matching methods, which provide a natural weighting scheme that can yield unbiased estimates of the impact of workload on severity allocation (Dehejia and Wahba (2002)). We estimate the propensity score by running a logistic regression of the treatment variable on the above mentioned observable characteristics. We exclude 12 observations that fall outside the common support (Caliendo and Kopeinig (2008)). We use the propensity score to generate a sample of low-workload discharges that matches as closely as possible the high-workload discharge sample. To do so we use the nearest-neighbor (with replacement), two-nearest-neighbors (with replacement) and a non-linear kernel (Epanechnikov kernel) matching algorithms. Using the significance test on pseudo-R^2 before and after matching (Sianesi (2004)), we find that the two-nearest-neighbors approach provides a good match quality (the observable factors are not jointly significant even at the 10% level in predicting the treatment category), while the nearest-neighbor or the kernel approach are not entirely successful (the observable factors were jointly significant at the 5% level in predicting the treatment category). However, for completeness we report results for all three methods in Table 7.

In Table 7, the first row reports the difference in average severity assignment between patients discharged under high-workload and low-workload conditions using all of the observations in our data set. The second row reports the marginal effect (averaged over all observations) of workload treatment after estimating a logit model where all observable factors (patient age and gender, length of stay, physician of record, day of the week, month and condition fixed effects) are used as controls. The last three rows present the difference in average severity assignment between high- and low-workload patients after running a propensity-score matching algorithm (nearest-neighbor, two-nearest-neighbors and Epanechnikov kernel). For the matched samples, the standard errors reported are based on the approximation by Lechner (2001).[8] As the average effect of being discharged on a high-workload day in the unmatched sample (coefficient$= -0.122$, std. err.$= 0.0127$) is significantly lower as compared to all other models presented in Table 7 we can conclude that patients discharged during high workloads are indeed systematically different to patients discharged under low workloads. However, controlling by using observable factors with propensity-score matching or with a logistic regression specification yields very similar results: the probability a patient discharged under high-workload conditions receives the high-severity assignment is about 6% (std.

[8] Errors based on bootstrapping were also estimated and they were essentially identical to the ones reported in Table 7.

Sample	Impact of high workload on severity assignment		
	Average Effect	Standard Error	T-Stat
Unmatched	-.122	.0127	-9.60
Unmatched with controls	-.0616	.0191	-3.22
Matched-nearest neighbor	-.0683	.0228	-3.00
Matched-nearest 2 neighbors	-.0603	.0201	-3.00
Matched-Epanechnikov kernel	-.0683	.0228	-3.00

Table 7 Propensity score-matching results. Patients are divided into two treatments, those discharged under high-workload conditions (i.e. the physician has discharged 3 or more patients on that day or is looking after more patients than the median) and the rest. The column "Average Effect" presents the difference in the probability a patient is assigned a high-severity code between the two treatments for five different models. The standard error and the T-stat of the average effect are presented in the next two columns. The row "Unmatched" presents the difference in probability of high-severity assignment without controlling for any difference between the two treatments. The row "Unmatched with controls" presents the difference in probability of high-severity assignment after controlling for systematic differences based on patient age, gender, length of stay, physician, day-of-the-week, month and condition fixed effects, using a logistic regression. The next three rows present the difference in probability of high-severity assignment using the "nearest neighbor", "nearest 2 neighbors" and "Epanechnikov kernel" matching algorithms.

err. 2%) lower than a matched patient discharged under low-workload conditions. More importantly, this observation provides additional evidence that our finding regarding the impact of workload on the probability of a patient being assigned a high-severity DRG code is unlikely to be due to selection bias on observable factors.

As a last robustness test, we turn to the finding that the frequency with which the trauma department treats specific conditions has a moderating impact on workload. One alternative explantation for what we find is that less-frequent conditions require inherently longer and more complex discharge notes which are more time-consuming for the discharging physicians than more-frequent conditions. Therefore, the moderating impact of frequency is not due to frequency fostering the development of workload-mitigating organizational routines, but rather due to frequency being negatively correlated with the inherent complexity of the discharge note. To investigate if this is a plausible correlated omitted variable problem we check directly whether the frequency of the condition is related to the complexity of the discharge note. To do so we use the total number of diagnosis and procedure codes the professional coder assigned to each discharged patient based on the discharge note as a proxy for complexity (i.e. higher paperwork complexity would generate, on average, more diagnosis and procedure codes). These codes, along with patient characteristics, are the inputs that determine the DRG-code assigned to a patient. For all conditions (DRGs codes) treated by the trauma department we calculate the average number of codes recorded for patients discharged during non-busy times (i.e. the physician has discharged 2 or fewer patients on that day or is looking after fewer patients than the median). We restrict our attention to non-busy times as the physicians might systematically miss treatment and diagnostic codes during busy times. We find that, on average, a low-severity condition generates 6 codes (95% C.I. $5.22-6.79$) while a high-severity condition generates a further 6.4 more codes (95% C.I. $5.01-7.85$). The frequency with which the department treats a condition has no significant relationship with the number of codes assigned irrespective of the condition's severity. This suggests that conditions treated more frequently are not inherently more complex in terms of paperwork (i.e. they are not associated with more diagnosis and procedure codes) than less-frequently treated conditions.

We conclude this section by noting that the moderating impact of the frequency of a condition on the probability a patient is assigned the high-severity DRG code provides an additional piece of evidence that supports the hypothesis that workload leads to undercoding rather than the alternative hypothesis of this being a side-effect of endogeneity in discharge timing or coordination or selection bias. If it was one of the latter, then it is not clear why frequency would have such a moderating effect on the probability of a high-severity DRG code assignment.

6.2. The Impact of Workload on Hospital Reimbursement: A Counterfactual Study

Having established that discharge-day workload conditions reduce the probability a patient is classified as high severity, we now assess the impact of this effect on hospital reimbursement. To do this we would ideally like to know which patients were undercoded and how much revenue is lost per undercoded patient. However, we only have a probabilistic assessment on who was undercoded. Furthermore, we do not always know the counterfactual payment as the high severity assignment commands different premiums for different diagnoses and for different payers. This is further complicated by the fact that 6.8% (std. err. $=0.41\%$) of the patients did not pay mainly because they were uninsured.[9] To overcome these difficulties we proceed in two different ways.

The first and more direct approach to estimating the financial impact of workload-induced undercoding involves estimating, for each patient, the extra revenue generated by high-severity assignment, as well as the probability of undercoding, and use these figures to estimate the lost revenue. To estimate the extra revenue generated by high severity assignment we run a least-squares regression of the logarithm of payments (for those patients that do pay) on the severity assignment with base-DRG, insurance plan and month fixed effects as controls. We use month fixed effects to control for any systematic changes to reimbursements over time. The model estimated is given by

$$\ln(\text{payment}_i) = a_0 + a_1 \text{Severity}_i + a_c \text{Controls}_i + \text{error}_i, \tag{2}$$

and the results are summarized in the first column of Table 8. We find that, conditional on paying, a high-severity patient generates, on average, 47.8% ($=\exp(0.391)-1$) more revenue per patient

[9] After controlling for patient age, neither severity nor workload conditions predict whether a patient will pay or not.

than a low-severity patient. The 95% confidence interval suggests that the average revenue lost for each undercoded patient is between 37.0% and 58.6%.[10] We next calculate the probability an individual patient is undercoded. To do this we first use Model II to estimate the probability of high severity assignment given the workload conditions at discharge and all other patient specific information. Then we calculate the probability of high severity assignment assuming the patient was discharged under low workload conditions (i.e. the physician did no more than one discharge and the number of inpatients is no greater than the one). The probability a patient is undercoded is the difference between the latter and the former. For each patient we also estimate the probability of paying for their bill. Combining these estimated parameters (and their errors) we can estimate that the hospital is losing, on average, what could be approximately an extra 1.1% of revenue (with a 95% C.I. $0.4\% - 1.9\%$).

For the second, and less direct, approach we run a least-squares regression on the logarithm of payments against a dummy variable called "Busy" that takes the value 1 if a patient is discharged under high-workload conditions (where physicians either look after more patients than the median or perform 3+ discharges or both) with base-DRG, insurance plan and month fixed effects as controls. The model estimated is given by

$$\ln(\text{payment}_i) = d_0 + d_1 \text{Busy}_i + d_c \text{Controls}_i + \text{error}_i, \tag{3}$$

and the results are summarized in the second column of Table 8. We find the coefficient of the workload dummy variable to be negative and significant (coefficient$= -0.0748$, p-value$= 0.005$), indicating that conditional on getting paid, the hospital on average (averaging over all base-DRGs and insurance plans), receives a $7.2\% = (1 - \exp(-0.0748))$ lower payment (95% C.I. $2.4\% - 12.0\%$) for patients discharged under high-workload conditions compared to patients discharged under low-workload conditions. Considering that 53.6% of the patients are discharged under high-workload conditions and taking into account the probability of non-payment, we can estimate the overall lost revenue to the hospital to be 3.6% (95% C.I. $1.2\% - 6.0\%$). It is not surprising that the second method produces a higher and more noisy estimate of the impact of workload on revenues than the first method. This is most likely due to the fact that, unlike the first method, the second method does not control for any heterogeneity in the patients discharged under high-workload conditions, which was shown to overestimate the impact of workload on the probability a patient is undercoded (see for example the propensity score matching analysis of Section 6). Note that in the regression of equation (3) we did not control for patient severity. If we do include the severity

[10] A similar analysis using the official CMS (version 25) prospective rates (also known as DRG weights) suggests that high-severity patients generate 78% more revenue per patient for the hospital (95% C.I. 75%-81%) than low-severity patients. This suggests that CMS is more aggressive in adjusting payments for patient severity than other payers.

VARIABLES	Linear Regression Dependent variable: ln(Payment)	
High-Severity Assignment	0.391***	
	(0.0293)	
Busy Dummy Variable		-0.0748***
		(0.0260)
Fixed Effects		
Calendar Month	Yes	Yes
Base-DRG	Yes	Yes
Insurance Plan	Yes	Yes
Observations	3536	3536
R-Squared	0.643	0.625

***, **, * denote statistical significance at 1% 5% and 10% confidence levels, respectively. Errors shown in parentheses are clustered on Base-DRG.

Table 8 Linear regression of the logarithm of hospital payments.

dummy variable as well as the interaction between severity and workload conditions, the workload conditions at discharge have no further impact on hospital payments, i.e. given that a discharged patient is assigned the high (or low) severity DRG code, how busy the discharging physician is does not make any difference to revenues. This provides further confirmation that discharge workload conditions affect hospital revenue only through severity assignment.

While the two methods result in slightly different estimates, they both suggest that the financial impact of workload-induced undercoating is substantial. To protect confidential data we refrain from reporting the absolute numbers, but we can point out that even at the lowest end of our estimation, the lost revenue would cover the salary of two full-time nurses whose sole responsibility could be to do discharges.

7. Discussion and Conclusions

By examining detailed reimbursement data at a major trauma department, we are able to show that the proportion of patients assigned a high-severity DRG code linked to workload. In particular, we establish that the probability of a patient receiving a high-severity assignment decreases with the discharge and the inpatient workload performed by the patient's physician of record. Consistent with the literature, we attribute this probability reduction to workload-induced degradation in the quality of the discharge notes and would be interesting for further research to identify the exact mechanism of how workload affects the quality of the discharge note. Our main finding suggests that the quality of the discharge note, besides having an impact on continuity of care, also has adverse financial implications. We also find that this reduction is moderated by the frequency with which the trauma treats specific conditions. This is an example of human behavior having an impact on performance in a systematic manner that, to the best of our knowledge, has not been studied before. More importantly, we show that the impact of this understandable human deficiency is financially important. We are able to estimate that the hospital loses approximately 1.1% (95% Confidence Interval $0.4\% - 1.9\%$) of its annual revenue due to this effect.

To the extent that hospitals are profit maximizers, our finding suggests that an increase in physician utilization is only guaranteed to increase profits when complemented by operational changes that mitigate the adverse impact of such an increase on the ability of the system to keep reliable patient and service records. Just like in the case of (the more serious) medical errors (see Insitute of Medicine (1999)) we believe coding errors are not due to “bad people” but due to “system failures” (Leape and Berwick (2005)). System interventions such as automation of certain clinical functions, complemented by training of clinical and non-clinical staff in quality management have been shown to reduce medical errors (Aron et al. (Forthcoming)), and we believe that similar interventions can help with workload-induced coding errors. More specifically, information systems already deployed in hospitals that automatically sense and record the actions of clinical agents (Ball (2003)) can provide the discharging physician with relevant content that could make writing the discharge summaries easier, faster and serve as props for presenting a more complete account of the diagnosis and the treatment. Moreover, managerial review systems that generate reports and summary statistics of deviations from prescribed or expected procedures (Aron et al. (Forthcoming)), can be programmed to provide feedback to clinicians and hospital management about individual coding performance, particularly when the discharging physician is under high workload. Complementing automation with training that reinforces to clinical staff the importance of complete discharge summaries both for ensuring quality of patient follow-up care (Kripalani et al. (2007)) as well as for correct reimbursement, along with providing specific templates and guidelines (Rao et al. (2005)) for generating high-quality discharge summaries, might alleviate the under-reimbursement documented by our study. Such deliberate investment in human capital is shown to be effective in creating organizational routines even for low-frequency events (Zollo and Winter (2002)). Furthermore, a more fundamental reorganization, where a fully-trained nurse contributes to the discharge process, particularly in non-routine cases during busy times, would be another possibility. Whether and which of these interventions are effective would be an interesting follow-up research question.

Our work has implications for the design of reimbursement systems such as the prospective system in use by healthcare payers, or any other service system where payments are contingent on system-generated reports. The prospective system is designed to reimburse service providers based on the average cost of providing a specific service. When tending to the needs of undercoded patients, the service provider incurs a cost that is likely to be higher than the reimbursement received. Therefore, the prospective payment system, which is designed to allow the average hospital to break even, may not always be adequate: service providers may incur losses which have nothing to do with operational efficiency, but instead are due to workload-induced undercoding. Such systems need to take into account that the quality of system-generated reports, and, therefore, system verifiability, are endogenous to the workload the service system is subjected to and should specify operational procedures that explicitly decouple, as far as possible, system workload from information collection.

Further research needs to investigate whether other settings, in healthcare and beyond, exhibit a similar response to workload. Any setting subject to stochastic variability with reports generated during (or at the end of) normal operations could in principle display similar behavior to the system we study. For example, relying on managerial reports to compile near-miss statistics (Dillon and Tinsley (2008)) might be problematic if near-misses are more likely when the system is operating at a higher-than-average workload *and* the quality of the data collection process is compromised at higher workloads. Similarly, sophisticated knowledge management systems, an important driver of performance for knowledge-based organizations such as consulting or financial services firms (see Ofek and Sarvary (2001)), might be less effective in helping organizations deal with high-workload conditions if the quality (and quantity) of content that goes into the system is compromised under high-workload conditions. Our empirical findings open up opportunities for further empirical and analytical work in these areas.

8. Acknowledgment

The authors would like to thank the traumatologists and administrators of the hospital examined in this paper and Chris P. Lee for assistance in obtaining the data used in this paper. They are also grateful to Gerard Cachon, Serguei Netessine, Chris Parker, Kamalini Ramdas, Stefan Scholtes, Stefanos Zenios, the Associate Editor and three anonymous reviewers for comments that have greatly improved the paper.

References

Ai, C., E.C. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* **80**(1) 123–129.

Aron, R., S. Dutta, R. Janakiraman, P.A. Pathak. Forthcoming. The impact of automation of systems on medical errors: Evidence from field research. *Information Systems Research* .

Ball, M.J. 2003. Hospital information systems: perspectives on problems and prospects, 1979 and 2002. *International journal of medical informatics* **69**(2-3) 83–89.

Birkmeyer, J.D., A.E. Siewers, E.V.A. Finlayson, T.A. Stukel, F.L. Lucas, I. Batista, H.G. Welch, D.E. Wennberg. 2002. Hospital volume and surgical mortality in the united states. *New England Journal of Medicine* **346** 1128–1137.

Boudreau, J., W. Hopp, J.O. McClain, L.J. Thomas. 2003. Commissioned paper: On the interface between operations and human resources management. *Manufacturing & Service Operations Management* **5** 179–202.

Caliendo, M., S. Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* **22**(1) 31–72.

Carter, G.M., J.P. Newhouse, D.A. Relles. 1990. How much change in the case mix index is DRG creep? *Journal of Health Economics* **9** 441–428.

Chan, C., V. Farias, N. Bambos, G. Escobar. 2011. Optimizing ICU discharge decisions with patient readmissions. *Columbia GSB, Working paper* .

Cleverley, W.O., J.O. Cleverley, P.H. Song. 2010. *Essentials of health care finance*. Jones & Bartlett Publishers.

Cocanour, C.S., R.P. Fischer, C.M. Ursic. 1997. Are scene flights for penetrating trauma justified? *The Journal of Trauma* **43**(1) 83–88.

Coit, M., J. Katz, G. McMahon. 2010. The effect of workload reduction on the quality of residents discharge summaries. *Journal of General Internal Medicine* 1–5.

Curtis, K. 2007. Trauma nurse case management. *Available at http://www.trauma.org/index.php/main/article/532. Accessed December 12, 2010.* .

Cyert, R., J. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ.

Dafny, L. S. 2005. How do hospitals respond to price changes? *The American Economic Review* **95**(5) pp. 1525–1547.

Dehejia, R.H., S. Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics* **84**(1) pp. 151–161.

Department of Health and Human Services. 2007. Medicare program; changes to the hospital inpatient prospective payment systems and fiscal year 2008 rates. Also known as CMS-1533 72(162): 47,130-48,175, Federal Register.

Dillon, R.L., C.H. Tinsley. 2008. How near-misses influence decision making under risk: A missed opportunity for learning. *Management Science* **54**(8) 1425–1440.

Esposito, T.J., M. Rotondo, P.S. Barie, P. Reilly, M.D. Pasquale. 2006. Making the case for a paradigm shift in trauma surgery. *Journal of the American College of Surgeons* **202**(4) 655–666.

Fisher, E.S., F.S. Whaley, M. Krushat, D.J. Malenka, C. Fleming, J.A. Baron, D.C. Hsia. 1992. The accuracy of Medicare's hospital claims data: Progress has been made, but problems remain. *American Journal of Public Health* **82**(2) 243–248.

Fuloria, P.C., S.A. Zenios. 2001. Outcomes-adjusted reimbursement in a health-care delivery system. *Management Science* **47**(6) 735–751.

Goldsmith, J. 1984. Death of a paradigm: The challenge of competition. *Health Affairs* **3** 5–19.

Gottschalk, A., S.A. Flocke. 2005. Time spent in face-to-face patient care and work outside the examination room. *Annals of Family Medicine* **3**(6) 488–493.

Green, L.V., S. Savin, N. Savva. 2010. Nursevendor problem: Personnel staffing in the presence of endogenous absenteeism. *Working Paper* .

Hillner, B.E., T.J. Smith, C.E. Desch. 2000. Hospital and physician volume or specialization and outcomes in cancer treatment: Importance in quality of cancer care. *Journal of Clinical Oncology* **18**(11) 2327–2340.

Insitute of Medicine. 1999. To err is human: building a safer health system. *Committee on quality health Care in America* .

Jollis, J.G., E.D. Peterson, C.L. Nelson, J.A. Stafford, E.R. DeLong, L.H. Muhlbaier, D.B. Mark. 1997. Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. *Circulation* **95**(11) 2467–70.

KC, D., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486–1498.

Kolasinski, A., S. Siegel. 2010. On the economic meaning of interaction term coefficients in non-linear binary response regression models. Tech. rep., Working paper, University of Washington and National Bureau of Economic Research.

Kripalani, S., A.T. Jackson, J.L. Schnipper, E.A. Coleman. 2007. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. *Journal of Hospital Medicine* **2**(5) 314–323.

Kuntz, L., R. Mennicken, S. Scholtes. 2011. Stress on the ward: Linking individual performance to system service quality. *Working paper* .

Leape, L.L., D.M. Berwick. 2005. Five years after to err is human. *JAMA: The journal of the American Medical Association* **293**(19) 2384.

Lechner, M. 2001. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. M. Lechner, F. Pfeiffer, eds., *Econometric evaluation of labour market policies*. Physica-Verlag, 1–18.

Lin, H.-C., S. Xirasagar, C.-H. Chen, Y.-T. Hwang. 2008. Physician's case volume of intensive care unit pneumonia admissions and in-hospital mortality. *American Journal of Respiratory Critical Care Medicine* **177** 989–994.

Luft, H.S., J.P. Bunker, A.C. Enthoven. 1979. Should operations be regionalized? The empirical relation between surgical volume and mortality. *New England Journal of Medicine* **301** 1364–1269.

Macias, C.A., M.R. Rosengart, J.-C. Puyana, W.T. Linde-Zwirble, W. Smith, A. Peitzman, D.C. Angus. 2009. The effects of trauma center care, admission volume, and surgical volume on paralysis after traumatic spinal cord injury. *Annals of Surgery* **249**(1) 10–17.

Mas, A., E. Moretti. 2009. Peers at work. *The American Economic Review* **99** 112–145(34).

Mayes, R. 2007. The origins, development, and passage of medicare's revolutionary prospective payment system. *Journal of the History of Medicine and Allied Sciences* **62**(1) 21–55.

McCormick, D., D.U. Himmelstein, S. Woolhandler, D.H. Bor. 2004. Single-payer national health insurance: Physicians' views. *Arch Intern Med* **164** 300–304.

Nelson, R., S. Winter. 1982. *An Evolutionary Theory of Economic Change*. Belknap Press, Cambridge, MA.

Newhouse, J.P. 1996. Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature* **34**(3) pp. 1236–1263.

Ofek, E., M. Sarvary. 2001. Leveraging the customer base: Creating competitive advantage through knowledge management. *Management Science* **47**(11) pp. 1441–1456.

Oliva, R., J.D. Sterman. 2001. Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* 894–914.

Rao, P., A. Andrei, A. Fried, D. Gonzalez, D. Shine. 2005. Assessing quality and efficiency of discharge summaries. *American Journal of Medical Quality* **20**(6) 337–343.

Reed, R. L. II, K.A. Davis, G.M. Silver, T.J. Esposito, V. Tsitlik, T. O'Hern, R. L. Gamelli. 2003. Reducing trauma payment denials with computerized collaborative billing. *The Journal of Trauma* **55** 762–770.

Reinhardt, U.E. 2006. The pricing of U.S. hospital services: Chaos behind a veil of secrecy. *Health Affairs* **25**(1) 57–69.

Schultz, K.L., D.C. Juran, J.W. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Management Science* **45** 1664–1678.

Schultz, K.L., D.C. Juran, J.W. Boudreau, J.O. McClain, L.J. Thomas. 1998. Modeling and worker motivation in JIT production systems. *Management Science* **44** 1595–1607.

Shleifer, A. 1985. A theory of yardstick competition. *The RAND Journal of Economics* **16**(3) pp. 319–327.

Sianesi, B. 2004. An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics* **86**(1) 133–155.

Silverman, E., J. Skinner. 2004. Medicare upcoding and hospital ownership. *Journal of Health Economics* **23**(2) 369–389.

Smith, A. 1776. *An Inquiry into the Nature and Causes of The Wealth of Nations*. London.

Tucker, A.L., A.C. Edmondson. 2003. Why hospitals don't learn from failures: Organizational and psychological dynamics that inhibit system change. *California management review* **45**(2) 55–72.

Weber, M. 1922/1978. *Economy and Society: An Outline of Interpretive Sociology*. University of California Press, Berkeley, CA.

Zollo, M., S.G. Winter. 2002. Deliberate learning and the evolution of dynamic capabilities. *Organization science* 339–351.