Short-term Fluctuations in Hospital Demand: Implications for Admission, Discharge, and Discriminatory Behavior

For Presentation to

The Applied Microeconomics Workshop
Department of Economics/School of Management
Yale University

April 6, 2006

Draft: Please do not cite or quote without permission
(Comments & Suggestions are welcome)

Abstract

This paper develops a framework for examining how short-term fluctuations in demand affect hospitals' admission and discharge behavior. We develop a new test for discriminatory admissions practices from utilization data that, under certain circumstances, does not require control for underlying differences in treatment seeking behavior and health of different patient groups. We further show that regression or other analyses of differences in mean effects can contain logical flaws and are ill suited to detect discrimination when incentives to discriminate are created by limited capacity. Analysis of Oregon inpatient data indicates that patients are discharged earlier than expected on high demand days relative to low demand days. Consistent with discriminatory behavior, admissions and discharges for Medicaid patients are affected by capacity constraints.

JEL Classification: I1; L2
Kew Words: Hospital behavior; Stochastic demand; Discrimination in treatment.

Rajiv Sharma,[*] Assistant Professor, Department of Economics, Portland State University, Portland, OR 97207.

Miron Stano, Professor, Department of Economics, School of Business Administration, Oakland University, Rochester, MI 48309.

Renu Gehring, Managing Partner, Ace-Cube, LLP, Beaverton, OR 97006

[*] Corresponding author. Department of Economics, Portland State University, PO Box 751, Portland, OR 97207. Email: sharmar@pdx.edu; TEL: (503) 725-3933; FAX: (503) 725-3945.

1. Introduction

The considerable day-to-day variance in hospital utilization and its implications for capacity and costs have long been recognized. In contrast, less attention has been directed to the effects of variance in demand on hospital behavior regarding admissions and discharges of patients requiring different types of treatment or with different types of insurance. This article addresses these neglected issues by introducing a theoretical framework to analyze the impact of short-term fluctuations in inpatient demand on hospital decisions. The model emphasizes hospital decisions when it becomes capacity constrained on high demand days.

The most far reaching, of various predictions of admissions and discharges behavior, is one that provides a test for discriminatory behavior. If a hospital discriminates against patients from one group, a binding capacity constraint will impact patients from that group differently than patients from favored groups. When comparisons are made on otherwise similar days, this differential impact can be observed in the cumulative distribution functions of patients admitted on high and low demand days. Because the difference is observable in the distribution of patients actually admitted, our proposed test eliminates the difficult challenge of controlling for differences across groups in underlying distributions over health conditions and care seeking behavior. With empirical data regarding those admitted to hospitals more readily available than data regarding those who seek treatment, the ability to detect and measure discrimination from such widely accessible utilization data would represent a significant advantage over current methods.[1] Although this paper examines patients with different insurance types, and finds

[1] Reduction of treatment disparities has a high priority on the national health care agenda. The influential Institute of Medicine (IOM) report, initiated by Congress, concluded that "… no single study adequately controlled for all potential confounding factors (e.g., patient preferences, racial differences in disease severity, geographic availability of specific services or procedures) simultaneously" (Smedley, Stith, and Nelson, 2002, p. 39). Aside from the difficulty of controlling for these factors, we show that, when capacity constraints create or exacerbate discriminatory behavior, regression methods or other analyses of differences in mean effects are logically flawed and poorly suited to detect such behavior.

evidence consistent with discrimination against those with Medicaid coverage, the methodology we develop can be adapted to detect discrimination by patient gender, race, ethnicity, or other characteristics.

The paper is organized as follows. After a brief review of the literature on stochastic hospital demand in the next section, Section 3 develops a model of hospital decision-making when patients differ both in treatment requirements and in the payment that their insurance plans provide to the hospital. Section 4 analyzes how the hospital's admissions behavior is affected by fluctuations in demand that may cause it to be capacity constrained. Appendix A contains formal proofs of the main results developed in this section. Section 5 contains a brief discussion of how capacity constraints impact discharge behavior and develops a simple proxy measure for the additional expected stay of current inpatients. This measure is later used to empirically test for discriminatory discharge behavior.

Section 6 uses Oregon hospital discharge data to show that hospitals' discharge behavior is affected by fluctuations in demand. Patients discharged on days when hospitals face high demand are discharged earlier than expected when compared to those discharged on days when demand is low. We also find evidence of discrimination in admissions and discharges against patients covered by the Oregon Health Plan (OHP)—which is Oregon's expanded Medicaid program.[2] Section 7 concludes the paper and summarizes its policy implications.

[2] Hospitals typically receive lower payments, often below reported cost, for Medicaid/OHP patients than for privately insured or Medicare patients. (MedPAC 2003) The effects of the lower Medicaid reimbursement levels and changes in payment systems have been extensively studied. Frank and Lave (1989) showed that hospital length- of-stay is significantly reduced for Medicaid psychiatric patients under prospective payment as compared to cost-based reimbursement. Dor and Farley (1996) examined how hospitals may vary the intensity of treatment they provide, i.e., cost-adjusting behavior, in response to payer generosity. Medicaid marginal costs were substantially below those of Medicare and private insurance. Dranove and White (1998) found that cuts in reimbursement have a relatively large impact, in terms of reduced services per day and per admission, on Medicaid patients especially those in Medicaid-dependent hospitals.

2. Literature Review

The growing economic literature on fluctuations in hospital demand attests to the importance of this phenomenon and the need to better understand its many potential effects. Harris (1977) describes the decision rules—including rules of thumb, contingency plans, bargaining, and sometimes shouting and screaming— that hospitals and physicians employ to ration capacity when faced with excess demand. Joskow (1980) examines how hospitals that compete on quality make capacity decisions when facing stochastic demand. He finds that if a hospital wishes to keep the probability of turning patients away below a threshold level, then the capacity it requires will depend in part on variability in demand. Friedman and Pauly (1981, 1983) develop a model where hospital costs include a latent penalty due to deterioration in quality when demand is unexpectedly high. Gal-Or (1994) models hospital decisions to invest in capacity, and finds that uncoordinated actions by hospitals can lead to excess capacity and low occupancy rates. Gaynor and Anderson (1995) examine how uncertain demand affects cost structures and estimate the cost of empty hospital beds. Using an alternative econometric approach, Keeler and Ying (1996) estimate the impact of excess bed capacity on hospital costs. Carey (1998) develops a framework for estimating the optimal level of spare capacity when hospital demand is stochastic. More recently, Baker et. al. (2004) analyze California data to estimate how variability in demand affects hospital costs.[3]

Despite the growing scholarly attention, many of the potential consequences of fluctuations in short-term hospital demand have yet to be determined. The following sections introduce a model that produces predictions on resource utilization and admissions and discharges behavior that are unique to the literature. Of particular relevance to current policy

[3] In addition to economic analyses, a distinct operations research literature has focused on: the factors underlying variance in demand for inpatient treatment (e.g., Bagust et. al., 1999); models of hospital occupancy (e.g., Millard et. al., 2001); and methods for forecasting hospital demand (e.g., Sterk and Shyrock, 1987; Jones, Joy, and Pearson, 2002). Discharge decisions under congestion are also prominent in that literature (e.g., Berk and Moinzadeh, 1998).

and health care delivery, we show that comparisons of hospital behavior with and without a binding capacity constraint produce a new test of discrimination in admissions and discharges.

3. The Model

Hospitals are required to make admission and discharge decisions each day. From among patients presenting themselves for possible treatment, hospitals have to decide which ones they will admit. From among current inpatients, hospitals have to determine which ones will be discharged. Admission and discharge decisions may be affected by short-term fluctuations in demand—when hospitals are capacity constrained, they may respond by diverting ambulances, rescheduling elective procedures, and discharging some patients early. We model admission and discharge decisions at a representative hospital. While decisions regarding specialization, quality, and investments in technology have a major impact on the treatment hospitals provide, such long-term decisions can be viewed as predetermined in the short-term context that is our focus.

Patients presenting for potential admission differ in the resources their treatment requires in terms of the number of days they will spend in the hospital and in terms of their intensity of treatment. Let $t > 0$ denote the expected hospital resources used in the treatment of the condition for which a specific patient is admitted. Comprising both the number of days spent in hospital as well as the intensity of resource use in treatment during the stay, t is thus a measure similar to those employed by prospective payment systems such as Medicare's DRG relative-weights. Such systems seek to group patients according to the average resources required in their care (Shen, 1996). However, the actual resource use of patients grouped in a single DRG can vary due to differences in physician practice patterns, and to within-group variance in severity of sickness (McMahon and Newbold, 1986; Dranove, 1988).[4] Often, the patients with greater resource requirement are those who are more severely sick, with more complications, and are more difficult to treat. We assume without loss of generality that the hospital restricts admissions to patients with $t \in [t_{\min}, t_{\max}]$, and refer to the interval $[t_{\min}, t_{\max}]$ as the hospital's service range.[5]

The hospital serves patients belonging to health insurance plans x and y, and obtains marginal revenue $r_j t$ from plan $j \in \{x, y\}$ for each admitted patient belonging to that plan with resource requirement t. Marginal revenue from treating patients belonging to plan x is greater than from those belonging to plan y (i.e., $r_x > r_y$). While the true costs of treating a specific patient are only realized at the end of the hospital stay, we assume that, on average, the hospital incurs marginal cost ct when it admits one patient with resource requirement t. To ensure that it is possible for total revenue at the hospital to equal or exceed total costs, we assume that marginal cost per unit of treatment is less than marginal revenue per unit of treatment for patients belonging to plan x (i.e., $c < r_x$). Let $f_{ij}(t) > 0$ for all $t \in [t_{\min}, t_{\max}]$ denote the probability density function, and let $s_{ij} > 0$ denote the number of patients—both scheduled and unscheduled—belonging to plan j who present themselves for possible admission on day i. Let $F_{ij}(t)$ be the cumulative distribution function (CDF) corresponding to $f_{ij}(t)$.

[4] We abstract away from issues arising due to residual variance in resource use within groups, and refer to the typical resource use of a patient within each group as that patient's "resource requirement" in treatment. We also abstract away from issues related to the optimal resource use in treatment by assuming that, subject to capacity constraints, the hospital aims to provide all patients with the typical treatment for their condition. Newhouse (1996) contains a thorough survey of the impacts of payment systems on providers' incentives to optimally choose the quantity and quality of treatment.

[5] Although we assume the service range to be uniform across the hospital for simplicity, the maximum and minimum resource requirements of patients it treats may differ by specialty, and there may be leeway due to physician discretion. While hospital administrators typically determine the range and availability of services in a hospital, it is physicians enjoying considerable autonomy from the administrators who make actual admission decisions (Dranove and White, 1994). Indeed, Harris (1977) describes a hospital as two separate firms comprising the medical staff and the administration.

Following the work of Ellis and McGuire (1986) and Hodgkin and McGuire (1994),[6] we assume that the hospital maximizes an objective function that depends on the quantity and type of treatment it provides to new admissions, on the profit it makes, as well as on the quantity and type of treatment it provides to current inpatients. Specifically, assume that the hospital has a service objective for new patients and let $\beta_j B(t)$ denote the benefit that the hospital derives due to the fulfillment of this objective when it admits one plan *j* patient with resource requirement t. In this formulation, β_x and β_y such that $\beta_x \geq \beta_y > 0$, reflect the respective weights in the hospital's objective function on services provided to patients from plans *x* and *y*, while $B(t)$ is a continuous, twice differentiable function such that $B(t) > 0$ for all $t \in [t_{\min}, t_{\max}]$. We assume that the contribution to the hospital's service objective is larger when it admits for treatment a patient requiring more intensive use of hospital resources than when it admits a patient requiring less intensive use of hospital services (i.e., $B'(t) > 0$ for all $t \in [t_{\min}, t_{\max}]$). There can be several reasons for this. Patients requiring more intensive use of hospital resources are often those who are more severely sick and difficult to treat. They may be viewed as being in greater medical need, and therefore as more deserving of treatment. Serving such patients may also add more to the objective function because hospitals providing highly resource intensive treatments such as multi-organ transplants enjoy higher prestige among patients, and particularly among physicians, than hospitals providing possibly equally effective but less resource intensive treatments such as smoking cessation and cholesterol control regimens. Hospitals often compete for physicians—and the patients they bring—through acquisition of facilities and equipment that permit the most

[6] These and many other contributions (e.g., Newhouse 1996) examine the effects of prospective reimbursement. Although some predictions have been supported, the principal observed impact of prospective payment has been the increasing intensity (case-mix) of admissions (McClellan 1997).

technically sophisticated and resource intensive treatment (Dranove and Satterthwaite, 2000). A focus on resource intensive cases may also increase payment from Medicare and other sources.

The hospital may also be concerned about the impact that admitting additional patients will have on its profits. Let α and $1-\alpha$ respectively denote the weight of profits and service in the hospital's evaluation of potential admissions where $0 \le \alpha \le 1$. Then, the net increase in the hospital's objective function when it admits one person belonging to plan *j* and with resource requirements t is given by

$$\alpha(r_j - c)t + (1-\alpha)\beta_j B(t). \tag{1}$$

In this framework, the hospital prefers admitting a patient from plan *x* rather than one from plan *y* with identical resource requirements if it is either concerned with the impact admission decisions have on profits or if $\beta_x > \beta_y$. We assume that, for both plans, the expression in (1) increases when t increases. Since $B'(t) > 0$, this is clearly true when $r_j \ge c$. When $r_j < c$, the expression will increase with t if the weight the hospital places on profits is small relative to the weight it places on its service mission, or if the contribution to the hospital's service objective of admitting a patient requiring more intensive resource use is sufficiently large to offset the greater financial loss incurred in treating such a patient. The key impact of this assumption is that if the hospital wishes to admit only some of the patients from an insurance plan that does not cover the marginal cost of treating patients, then the hospital will choose the ones with greater resource requirements. Intuitively, this may arise because the medical needs of those with greater resource requirement are often more severe, more immediately life-threatening, and therefore may be viewed as sufficiently more deserving of treatment to offset the additional financial loss the hospital incurs in providing treatment.

Finally, the hospital is concerned with the welfare and costs of current inpatients. Current inpatients differ in the marginal benefit that they derive from remaining in the hospital an additional day. Let $v \in [v_{\min}, v_{\max}]$ denote the marginal patient benefit when one current inpatient remains for an extra day. We assume that v is an element of the hospital's objective function, and that it decreases in the days immediately preceding a patient's discharge from hospital. We further assume that the hospital may assign different weights to the benefit received by patients belonging to different insurance plans, and that these weights are given by β_x and β_y for patients from plans x and y, respectively.[7] Let $h_{ij}(v) > 0$ for all $v \in [v_{\min}, v_{\max}]$ denote the probability density function, and let $n_{ij} > 0$ denote the number of inpatients belonging to plan j who are in the hospital on at the beginning of day i. Let $H_{ij}(v)$ be the CDF corresponding to $h_{ij}(v)$. We also assume for simplicity that the marginal costs the hospital incurs when an inpatient stays an additional day are the same for all inpatients and equal c_v. While this assumption may be problematic for patients with substantial remaining treatment, the assumption of constant marginal costs for an additional day in the hospital is likely a reasonable approximation of reality for patients who are the focus of our analysis—patients who are candidates for discharge on a specific day. Many such patients have already undergone necessary tests and procedures, and remain in hospital for supervision during recovery.

The timing of events and the resolution of uncertainty in our model occur as follows. The hospital learns the number and type of its current inpatients at the beginning of each day (i.e., n_{ij} and $h_{ij}(v)$ are realized at the beginning of day i). The hospital then learns the number and type of potential new patients for that day (i.e., s_{ij} and $f_{ij}(t)$ are realized for day i). Upon

[7] This may arise in a prospective payment framework such as this one because the hospital may be more concerned about its reputation among patients from higher paying plans due to implications for future profits.

learning of the number and distribution of current as well as potential patients, the hospital simultaneously determines which current inpatients will be discharged, and which potential patients will be admitted.

Formally, the hospital's problem on day i is to determine which patients it will admit and which current inpatients it will discharge to maximize the value of the following objective function.

$$\max_{t_{ix},t_{iy},v_{ix},v_{iy}} \quad s_{ix}\int_{t_{ix}}^{t_{\max}}[\alpha(r_x-c)t+(1-\alpha)\beta_x B(t)]f_{ix}(t)dt+s_{iy}\int_{t_{iy}}^{t_{\max}}[\alpha(r_y-c)t+(1-\alpha)\beta_y B(t)]f_{iy}(t)dt$$
$$+n_{ix}\int_{v_{ix}}^{v_{\max}}(\beta_x v-\alpha c_v)h_{ix}(v)dv+n_{iy}\int_{v_{iy}}^{v_{\max}}(\beta_y v-\alpha c_v)h_{iy}(v)dv. \tag{2}$$

The hospital's maximization of (2) is subject to the capacity constraint that

$$C-(s_{ix}\int_{t_{ix}}^{t_{\max}}f_{ix}(t)dt+s_{iy}\int_{t_{iy}}^{t_{\max}}f_{iy}(t)dt+n_{ix}\int_{v_{ix}}^{v_{\max}}h_{ix}(v)dv+n_{iy}\int_{v_{iy}}^{v_{\max}}h_{iy}(v)dv)\geq 0 \tag{3}$$

where C denotes hospital capacity.

If the hospital has sufficient capacity to admit patients with $t=t_{\max}$, and to retain current inpatients with $v=v_{\max}$ from both plans, the Lagrangian for the hospital's problem on day i is

$$\mathrm{L}=s_{ix}\int_{t_{ix}}^{t_{\max}}[\alpha(r_x-c)t+(1-\alpha)\beta_x B(t)]f_{ix}(t)dt+s_{iy}\int_{t_{iy}}^{t_{\max}}[\alpha(r_y-c)t+(1-\alpha)\beta_y B(t)]f_{iy}(t)dt$$
$$+n_{ix}\int_{v_{ix}}^{v_{\max}}(\beta_x v-\alpha c_v)h_{ix}(v)dv+n_{iy}\int_{v_{iy}}^{v_{\max}}(\beta_y v-\alpha c_v)h_{iy}(v)dv$$
$$+\lambda[C-(s_{ix}\int_{t_{ix}}^{t_{\max}}f_{ix}(t)dt+s_{iy}\int_{t_{iy}}^{t_{\max}}f_{iy}(t)dt+n_{ix}\int_{v_{ix}}^{v_{\max}}h_{ix}(v)dv+n_{iy}\int_{v_{iy}}^{v_{\max}}h_{iy}(v)dv)]$$
$$+\mu_x(t_{ix}-t_{\min})+\mu_y(t_{iy}-t_{\min})+\gamma_x(v_{ix}-v_{\min})+\gamma_y(v_{iy}-v_{\min}). \tag{4}$$

where $\lambda \geq 0$ is the multiplier associated with the capacity constraint, $\mu_j \geq 0$ are multipliers associated with the constraint that the hospital only admits patients with $t \geq t_{\min}$, and $\gamma_j \geq 0$ are multipliers associated with the constraint that the hospital only retains inpatients with $v \geq v_{\min}$.

In the analyses that follow, we consider hospital admission behavior discriminatory against plan *y* patients if there exists *t* such that the hospital admits plan *x* patients with resource requirement *t*, but not plan *y* patients with resource requirement *t*. We consider hospital discharge behavior discriminatory against plan *y* patients if there exists *v* such that the hospital discharges plan *y* inpatients who derive marginal benefit *v* from remaining hospitalized an additional day, but retains plan *x* patients who derive marginal benefit *v*.

Causes of variance in hospital demand

An analysis of hospital behavior under capacity constraints must also examine the factors—aside from random ones—that distinguish days when the hospital is capacity constrained from those when it has ample spare capacity.[8] This is especially important because factors that systematically distinguish one day from another may also influence patients' health status and care seeking behavior.

Clearly, patient and staff preferences regarding the scheduling of treatment will influence the days on which the hospital is likely to be capacity constrained. Such preferences could result in both weekly (e.g., weekend vs. weekday, beginning of workweek vs. end of workweek) and seasonal (e.g., summer vs. winter, holiday vs. non-holiday season) variance in hospital demand. Physicians' offices and outpatient clinics may be more likely to be open or keep longer hours on some days of the week than on others, and thus contribute to weekly variance in referrals to

[8] Since we model the capacity constraint as a daily phenomenon, it suffices to consider factors that distinguish one day from another. If capacity constraints were, for example, weekly or hourly phenomena, we would have to consider the factors that distinguish one week from another or one hour from another.

hospitals for admission. Seasonal variations in the incidence or severity of certain diseases, as well as in the propensity to seek health care (e.g., due to variance in the availability of work, due to open enrollment in health insurance, or as a result of a post holiday resolve to improve one's health) may similarly contribute to seasonal variation in hospital demand. Whether a day is at the beginning, middle or end of a month may also potentially affect the likelihood that the hospital will find itself capacity constrained if, for example, treatment seeking behavior is influenced by ready availability of money which may be greatest at the beginning of the month for some patients. In the theoretical analysis that follows, we assume that days when the hospital is capacity constrained are otherwise similar to days when it has ample capacity, so that patients' health status and care seeking behavior does not differ according to whether or not the hospital is capacity constrained. In the empirical portions of this paper, we make efforts to account for possible differences between days we identify as high demand (denoting a possibly binding capacity constraint) and low demand (denoting ample capacity) for the hospital. Generally, since characteristics of days (day of week; month; beginning, middle, or end of month; holiday vs. workday) are far more readily observable than characteristics of patient groups or individual patients (health status, care seeking behavior), they are also likely to be easier to account for in analyses.

4. Short-term fluctuations in demand, and admissions behavior.

If considerations of profit do not affect decisions regarding which patients the hospital will admit (i.e., if $\alpha = 0$) and if the hospital assigns equal weight to services provided to patients from the two health plans (i.e., if $\beta_x = \beta_y$), then its admissions behavior toward patients with plan *x* will be identical to its behavior towards those belonging to plan *y*. In the remainder of this section, we examine the hospital's behavior when it does consider the impact admission decisions will have on profits and, therefore, may want to discriminate between patients from the two plans but assigns equal weights to services provided to patients from the two plans (i.e., $\alpha > 0$ and $\beta_x = \beta_y$).[9] If the hospital does not face a binding capacity constraint on day *i*, it follows that: (1) the hospital admits all patients from plan *x* whose treatment requirements are within its service range; (2) if the contribution to its objective function is non-negative for all plan *y* patients, then the hospital will admit all patients from plan *y*; and (3) if the contribution to its objective function is negative for some patients then, even if it has ample spare capacity, the hospital will admit only those plan *y* patients whose contribution to its objective function equals or exceeds zero.

An examination of the resource requirements of those admitted to the hospital will also yield systematic differences between patients belonging to the two plans. If patients from both plans *x* and *y* who present themselves for possible admission to the hospital have identical distributions over resource requirements in treatment, marginal revenue from plan *y* patients is less than the marginal cost of treating them, and the hospital is sufficiently concerned about profit so that some plan *y* patients will detract from the hospital's objective function, then resource requirements will be smaller for admitted plan *x* patients in the sense of strict first order stochastic dominance than for admitted plan *y* patients.

Proposition 1 summarizes the hospital's behavior towards patients from the two plans when it faces a binding capacity constraint. Let t_{ij}^{c} and t_{ij}^{*} denote the hospital's choice of $t \in [t_{\min}, t_{\max})$ for plan *j* patients that maximizes the hospital's objective function, respectively, with and without a binding capacity constraint on day i.

[9] Our qualitative results would remain unchanged if, instead, $\alpha = 0$ and $\beta_x > \beta_y$. Detailed derivations of the results in this section are available from the authors on request.

Proposition 1: *Suppose* $C-(s_{ix}\int_{t_{ix}^*}^{t_{\max}} f_{ix}(t)dt+s_{iy}\int_{t_{iy}^*}^{t_{\max}} f_{iy}(t)dt+n_{ix}\int_{v_{ix}}^{v_{\max}} h_{ix}(v)dv+n_{iy}\int_{v_{iy}}^{v_{\max}} h_{iy}(v)dv)<0$

where $t_{ix}^*=t_{\min}$ and $t_{iy}^*\geq t_{\min}$. *Then:*

(i) *if* $0<\lambda\leq\alpha(r_x-c)t_{\min}+(1-\alpha)\beta_x B(t_{\min})$, *then* $t_{iy}^c>t_{iy}^*\geq t_{\min}$ *and* $t_{ix}^c=t_{ix}^*=t_{\min}$;

(ii) *if* $\lambda>\alpha(r_x-c)t_{\min}+(1-\alpha)\beta_x B(t_{\min})$, *then* $t_{iy}^c>t_{iy}^*\geq t_{\min}$ *and* $t_{ix}^c>t_{ix}^*=t_{\min}$; *and*

(iii) $t_{iy}^c>t_{ix}^c$.

When a binding capacity constraint, prevents the hospital from admitting all the patients it would otherwise admit, important differences in the hospital's behavior toward patients from the two plans arise because plan *x* patients contribute more to the hospital's objective function than otherwise identical patients from plan *y*. When the marginal impact of the capacity constraint on its objective function is relatively small, the hospital responds by restricting admissions for patients who contribute least to its objective function (plan *y* patients with relatively low resource requirements), leaving admissions of others unaffected by the capacity constraint. When the marginal impact of the binding capacity constraint is larger, its effects extend to patients from plan *x*. Even when patients from both plans are affected by the capacity constraint, the portion of the hospital's service range where admissions are affected is larger for plan *y* patients than for patients from plan *x*.

Let $g_{ij}(t)$ be the probability density function over $t\in[t_{\min},t_{\max}]$ of patients belonging to plan $j\in\{x,y\}$ that the hospital admits on day i when the capacity constraint does not affect admissions on that day. Similarly, let $g_{ij}^c(t)$ be the probability density function over $t\in[t_{\min},t_{\max}]$ of patients belonging to plan $j\in\{x,y\}$ that the hospital admits on day i when a binding capacity constraint does affect admissions. Let $G_{ij}(t)$ and $G_{ij}^c(t)$ be the CDFs corresponding to $g_{ij}(t)$ and

$g_{ij}^c(t)$, respectively. Proposition 1 implies that resource requirements of plan *y* patients admitted when the capacity constraint does not bind will be lower in the sense of strict first order stochastic dominance than resource requirements of plan *y* patients admitted when the capacity constraint binds (i.e., $G_{iy}^c(t) < G_{iy}(t)$ *for all* $t \in (t_{\min}, t_{\max})$, and $\int_{t_{\min}}^{t_{\max}} t g_{iy}(t) dt < \int_{t_{\min}}^{t_{\max}} t g_{iy}^c(t) dt$). Admissions of plan *x* patients are unaffected by the binding capacity constraint when it has a relatively small marginal impact on the hospital's objective function. However, if the capacity constraint has a large marginal impact, then resource requirements of plan *x* patients admitted when the capacity constraint does not bind will be lower in the sense of strict first order stochastic dominance than resource requirements of plan *x* patients admitted when the capacity constraint binds (i.e., $G_{ix}^c(t) < G_{ix}(t)$ *for all* $t \in (t_{\min}, t_{\max})$, and $\int_{t_{\min}}^{t_{\max}} t g_{ix}(t) dt < \int_{t_{\min}}^{t_{\max}} t g_{ix}^c(t) dt$).

When patients from plans *x* and *y* who present for possible admission have identical distribution over resource requirements, then we obtain the additional results contained in Corollary 1.

Corollary 1: *Suppose that* $f_{ix}(t) = f_{iy}(t)$ *for all* $t \in [t_{\min}, t_{\max}]$), *and that the hospital faces a binding capacity constraint on day i. Then*:

(i) $\int_{t_{\min}}^{t_{\max}} t g_{ix}^c(t) dt < \int_{t_{\min}}^{t_{\max}} t g_{iy}^c(t) dt$); *and,*

(ii) $G_{ix}^c(t) \geq G_{iy}^c(t)$ *for all* $t \in (t_{\min}, t_{\max})$.

Although a binding capacity constraint can cause the hospital to restrict admission of patients from the better paying plan, Corollary 1 reports that the constraint will have a more severe impact on admissions of patients belonging to the plan that pays less well. If patients from plans *x* and *y* who are identical with respect to resource requirements seek treatment from the hospital when a capacity constraint causes the hospital to restrict admissions from both plans, then the proportion of patients with relatively low resource requirements will be larger from plan *x* than from plan *y*. Furthermore, mean resource requirement will be higher for patients from plan *y*.

Evidence of differential admissions when patients from different plans are not identical

Differences in hospital admission behavior towards patients belonging to different insurance plans are difficult to isolate empirically from utilization data because unobserved underlying differences between patients from different health plans may exist. For example, we have shown above that both with and without a binding capacity constraint, when payments from plan *y* satisfy certain conditions, the hospital will admit only the relatively more resource intensive patients from plan *y* while admitting plan *x* patients in a larger portion of its service range. However, the mere observation that the hospital admits few patients who have relatively low resource requirements from a lower paying plan while admitting a broader range of patients from a better paying plan is not evidence of discrimination by the hospital. Such an observation constitutes evidence of discrimination only if patients from the different plans who present for possible admission are identically distributed with respect to the resources required in their treatment. Therefore, empirical tests of differential admissions usually require that either patients from different plans who seek treatment are identical with respect to treatment requirements or that analysis of patients from different plans admitted to hospital be controlled for differences across plans in the characteristics of those seeking treatment. We cannot ascertain whether patients belonging to different plans are identical, and we cannot control for differences in resource requirements of those who seek treatment as our data include information only on those actually admitted. Even when feasible, controls have their own limitations. To address these limitations, we develop a test for differential admissions policies that remains valid even when patients from different plans who seek hospital treatment are not identically distributed with respect to resource requirements.

Corollary 2 provides the basis for such a test by comparing hospital behavior when it is capacity constrained to behavior on otherwise similar days when it is not capacity constrained.

Corollary 2: *Let* $D_{ij}(t) \equiv G_{ij}(t) - G_{ij}^{c}(t)$ *for* $j \in \{x, y\}$ *and all* $t \in [t_{\min}, t_{\max}]$.

(i) *If* $0 < \lambda \leq \alpha(r_x - c)t_{\min} + (1-\alpha)\beta_x B(t_{\min})$, *then*

(i-a) $G_{ix}(t) = G_{ix}^{c}(t)$ for *all* $t \in [t_{\min}, t_{\max}]$, $D_{ix}(t)$=0 for *all* $t \in [t_{\min}, t_{\max}]$, *and*

$\int_{t_{\min}}^{t_{\max}} t g_{ix}(t)dt = \int_{t_{\min}}^{t_{\max}} t g_{ix}^{c}(t)dt$. *And,*

(i-b) $G_{iy}(t) > G_{iy}^{c}(t)$ for *all* $t \in (t_{\min}, t_{\max})$, $D_{iy}(t)$>0 for *all* $t \in (t_{\min}, t_{\max})$, *and*

$\int_{t_{\min}}^{t_{\max}} t g_{iy}(t)dt < \int_{t_{\min}}^{t_{\max}} t g_{iy}^{c}(t)dt$.

(ii) *If* $\lambda > \alpha(r_x - c)t_{\min} + (1-\alpha)\beta_x B(t_{\min})$, *then*

(ii-a) *for* $j \in \{x, y\}$, $G_{ij}(t) > G_{ij}^{c}(t)$ for *all* $t \in (t_{\min}, t_{\max})$, $D_{ij}(t)$>0 for *all* $t \in (t_{\min}, t_{\max})$,

and $\int_{t_{\min}}^{t_{\max}} t g_{ij}(t)dt < \int_{t_{\min}}^{t_{\max}} t g_{ij}^{c}(t)dt$. *And,*

(ii-b) $D_{ij}(t)$ *is maximized when* $t = t_{ij}^{c}$ *for* $j \in \{x, y\}$. *If the hospital discriminates against patients from plan y in admissions, then the value of* t *that maximizes* $D_{iy}(t)$ *will be greater than the value of* t *that maximizes* $D_{ix}(t)$.

Corollary 2 reports two results showing that an examination of the differences across plans in the impact that a binding capacity constraint has on the distributions of admitted patients can provide evidence of discriminatory behavior by the hospital. First, if the marginal impact of the binding budget constraint is small, then the distribution and mean resource requirement of only plan *y* patients will be affected by the constraint. Second, if the marginal impact of the binding capacity constraint is large, then the distribution and mean resource requirement of patients from both plans will be affected, but the portion of the service range affected will be larger for plan *y* patients. Intuitively, if the hospital discriminates against plan *y* patients, then plan *y* patients will be the only ones to feel the effects of a binding capacity constraint when the marginal impact of such a constraint is small. When the marginal impact of a binding capacity constraint is large, patients from both plans will suffer its effects but the portion of the service range affected will be larger for plan *y* patients than for plan *x* patients: the maximum difference in the cumulative distribution of admissions without and with a binding capacity constraint will arise a higher level of resource requirement for plan *y* patients than for plan *x* patients.

If the hospital does not discriminate between patients from the two plans, then a binding budget constraint with even a small marginal impact will affect patients from both plans. Furthermore, the portion of its service range over which admissions are affected by a binding capacity constraint will be the same for patients from both plans. The maximum difference in the cumulative distribution of admissions without and with a binding capacity constraint will therefore arise at the same level of resource requirement for patients from the two plans.

A crucial aspect of Corollary 2 is that, provided the analysis is based on comparisons of hospital behavior with and without binding capacity constraints on otherwise similar days, the results it reports depend entirely on distributions over resource requirements of patients actually admitted to the hospital (i.e., on $G_{ij}(t)$ and $G_{ij}^{c}(t)$). It does not depend on distributions of patients who present themselves for potential admission to hospital (i.e., not on $F_{ij}(t)$) that can be affected by underlying differences in health and treatment seeking behavior of patients with different types of insurance. Therefore, the key implication of Corollary 2 is that it permits detection of discrimination in hospital behavior towards patients from different health plans by examining distributions over resource requirements of those admitted without the need to control for underlying differences between different patients groups. Because empirical data regarding those admitted to hospital are more readily available than data regarding those who seek treatment, and because characteristics of days (day of week; month; beginning, middle, or end of month; holiday vs. workday) are far more readily observable than characteristics of individual patients (health status, care seeking behavior), Corollary 2 can dramatically simplify the detection of discriminatory behavior.

Analyses of differences in mean effects as evidence of differential admissions

Analyses of differences in mean effects, such as those using regression techniques, are ill-suited as methods for detecting differential admissions policies when the incentive to discriminate is created or exacerbated by binding capacity constraints. In such analyses, a larger increase in mean resource requirement among plan *y* patients than among plan *x* patients due to a binding capacity constraint would be regarded as evidence of discrimination against plan *y* patients.

Table 1: Discrimination and difference in mean effects—Illustrative Scenarios

Resource requirement in treatment	Scenario 1		Scenario 2	
	$f_{ix}(t)$	$f_{iy}(t)$	$f_{ix}(t)$	$f_{iy}(t)$
1	0.1	0.2	0.1	0.05
2	0.4	0.3	0.1	0.05
3	0.4	0.3	0.2	0.45
4	0.1	0.2	0.6	0.45

To see why such analyses can yield erroneous results, consider a hospital that does not discriminate between patients from the two plans: it admits all patients with $t > 0$ who seek care when the capacity constraint does not bind, and admits all patients with $t > 1$ when the capacity constraint does bind. Assume that values of $f_{ij}(t)$ for patients from the two plans are those shown in Scenario 1 in Table 1. Without a binding capacity constraint, the mean resource requirement for admitted patients from both plans is 2.5. When the capacity constraint binds, the mean resource requirement for admitted patients from plans *x* and *y* is, respectively, 2.67 and 2.88—the impact on plan *y* patients is larger even though hospital admission policies were not discriminatory. In a situation such as this, an analysis of differences in mean effects using regression techniques would yield a false positive result.

Now consider a hospital that discriminates between patients from the two plans. Assume that it admits all patients with $t > 0$ who seek care when the capacity constraint does not bind. When the capacity constraint binds, assume that it admits only those plan *x* patients with $t > 1$ and only those plan *y* patients with $t > 2$. Let values of $f_{ij}(t)$ be those shown in Scenario 2 in Table 1. Then, the mean resource requirement for admitted patients from both plans is 3.3 when the capacity constraint does not bind. When the capacity constraint binds, the mean resource requirement for admitted patients from plans *x* and *y* is, respectively, 3.56 and 3.5—the impact on plan *y* patients is smaller despite discriminatory admissions policies. An analysis of differences in mean effects using regression techniques would yield a false negative result under such circumstances.

5. Impact on discharges.

Even though our formal model abstracts away from issues related to the optimal treatment that patients should receive, some related issues can be addressed by examining the impact of short-term changes in hospital demand on the characteristics of patients discharged by the hospital. Hospitals may have greater influence over how much treatment patients receive before they are discharged than over the resource requirements of patients who are admitted.

Therefore, analysis of discharge behavior may provide us with additional insights about how hospital behavior responds to fluctuations in demand.

Specifically, a capacity-constrained hospital may discharge current inpatients earlier than were capacity constraints not binding. In terms of our model, when the marginal cost of a binding capacity constraint is large (i.e., when λ is large), the hospital derives a large additional benefit if it can increase available capacity. One way of relieving a capacity constraint is to discharge additional patients who would receive additional care in the absence of a capacity constraint. In that case, short-term fluctuations in demand may also have implications for the quality and quantity of care that patients receive.

Analysis of hospital discharge behavior in our model provides results analogous to those contained for admissions behavior in Section 4. In particular, we find a result that is *mutatis mutandis* identical to the one in Corollary 2: If the hospital does not discriminate against plan *y* patients in discharge, then a binding capacity constraint with even a small marginal impact will affect patients from both plans. The type of current inpatients—as denoted by the marginal benefit of an additional day in the hospital—subject to early discharge will be identical across plans. The maximum difference in the CDF (over *v*) of discharges with and without a binding capacity constraint will arise at the same level for patients from both insurance plans.

If the hospital discriminates against plan *y* patients (i.e., if $\beta_x > \beta_y$), and the marginal impact of the binding capacity constraint is small, then discharges of only plan *y* patients will be affected by the constraint. If the marginal impact of the binding capacity constraint is large, then discharges of patients from both plans will be affected, and the maximum difference in the CDF of discharges with and without a binding constraint will arise at a higher level of marginal benefit for patients from plan *y* than for those from plan *x*.

Since current inpatients constitute the potential pool of candidates for possible discharge, we first develop a proxy measure for the expected additional hospital stay of a current inpatient in order to examine how fluctuations in demand affect hospitals' discharge behavior. The expected additional hospital stay of a current inpatient will increase with the marginal benefit the patient derives from additional stay (v in the model in Section 3). Any such proxy measure must account for both the type and severity of illness necessitating hospital admission, as well as for the progression of illness and treatment since admission. We do this by considering patients' DRGs and by accounting for how long they have already been in the hospital.

Specifically, consider patients in DRG D for which the maximum length of stay for any patient in year Y is n. Let NDY_k denote the number of these patients who remain hospitalized at the end of day k as inpatients where the day of admission is defined as day 0. Then, the expected remaining length of stay for a patient in DRG D in year Y who remains an inpatient on day k is

$$ERLOS_{DYk} \equiv \sum_{l=k}^{n} (l-k)\left[\frac{NDY_{l-1} - NDY_l}{NDY_k}\right]. \tag{5}$$

The term in square brackets in equation (5) is the probability that a patient who remains hospitalized on day k as an inpatient will be discharged on day l. When $k = 0$, then $ERLOS_{DYk}$ is simply the mean length of stay for patients in DRG D who obtained treatment in year y. When $k > 0$, $ERLOS_{DYk}$ is the mean remaining length of stay of patients who remained hospitalized on day k of their hospital stay. This measure thus accounts for the fact that the subset of patients who remain in hospital on day k of their stay often respond more slowly, suffer more setbacks in treatment, and were more severely sick at the time of admission than patients discharged prior to day k. Therefore, the treatment of the cohort of patients who remain in hospital on day k of their stay often requires more hospital resources than of those discharged prior to day k.

A potential shortcoming of this measure is that while $ERLOS_{DYk} \geq 0$ for all D, Y, and k, the realized remaining length of stay on day k of all patients discharged from hospital on day k is zero. Therefore, if this measure is employed to predict the additional resources used in treatment of patients discharged on a specific day, then it will provide an overestimate. However, since patients with high *ERLOS* are those who would ordinarily be expected to receive substantial additional hospital treatment, analyses of differences in *ERLOS* between patients discharged when the hospital does and does not face a binding capacity constraint may provide evidence of how its discharge behavior responds to such constraints.

6. Oregon hospitals' behavior in response to fluctuations in demand.

We now examine hospital data for evidence of short-term fluctuations in demand that may cause hospitals to be capacity constrained, and the impact that binding capacity constraints may have on hospital admissions and discharge behavior. We use Medicare DRG relative weights as empirical proxy for patient's resource requirement in treatment (t in the model in Section 3), and ERLOS as a proxy for current inpatients' marginal benefit of additional hospital stay (v in Section 3).

The Data

We began with the 381,499 records that comprise inpatient discharges from Oregon hospitals between December 1, 1997 and November 30, 1998. Our initial intent was to analyze inpatients who had received treatment in calendar year 1998. Since our data include only patients discharged in the sample period, patients in the hospital during 1998 but discharged after December 31, 1998 were not counted. This leads to an undercount in the days leading up to the end of calendar year 1998, so we dropped December 1998 from our analysis.[10] Data on patients discharged in December 1997 was included instead.

Our data do not include “day-surgery” or “short-stay” patients who obtain treatment in a hospital, but are never formally admitted. We excluded records associated with lengths of stay of over one year, as well as records where our computation of length of stay based on admission and discharge dates was not in concert with the value contained in length of stay variable included in the raw data. Together these exclusions amounted to less than 0.2% of records. These data are publicly available from the Oregon Office for Health Policy and Research, and have been stripped of elements such as names and addresses that would permit patients to be individually identified.

The data contain discharge records from 65 hospitals. The single hospital with the largest number of patients accounted for about 8% of discharges, whereas the smallest accounted for less than .02%. The 7 largest hospitals comprised 41% of all discharges. Each record contained patient diagnosis, date of birth, date of admission, date of discharge, and information on payer characteristics. Using discharge date, admission date, and length of stay variables, we compute daily occupancy by hospital. Our measure does not include patients who are admitted and discharged on the same day because they are not in the hospital at the time of any midnight census.

The literature on stochastic hospital demand has recognized regular weekly and seasonal patterns in admissions and discharges (e.g. Baker et. al. 2004). Our analysis focuses on Thursdays as the day with the highest mean occupancy to avoid the confounding effects of weekly patterns. We excluded Thanksgiving, Christmas and New Years Day, which all

---

[10] Baker et. al. (2004) eliminate the last month in their discharge dataset from calculations of utilization patterns for similar reasons. Their analysis of California data covers the period from 1983-1995, and they exclude all of 1995 from cost analyses.

happened to be Thursdays in our data, from analysis due to their special nature as important holidays. We excluded patients transferred to other hospitals from our analysis of discharges since such discharges did not represent a potentially early end to treatment. The results reported in this section are based on an analysis of 13,440 admissions and 11,412 discharges on days when hospitals may have been capacity constrained, and of 11,416 admissions and 8,944 discharges on days when hospitals were likely not capacity constrained.

Hospital inputs and capacity constraints

A hospital becomes capacity constrained when the available quantity of any input necessary in the provision of treatment is inadequate to meet realized demand. In a theoretical analysis, Gal-Or (1994) permits hospital capacity to be variously interpreted as the number of rooms available for hospitalization, quantity of diagnostic equipment, laboratory space, or the number of physicians in different specialties. Several empirical studies have classified hospital inputs as fixed, quasi-fixed, and variable (e.g., Gaynor and Anderson, 1995; Carey, 1998). The number of beds is usually used to represent fixed inputs for hospitals. Quasi-fixed inputs may include some types of labor that are obtained on long-term contracts. Only variable inputs such as supplies and some types of labor obtained in spot markets can be altered in response to short-term fluctuations in demand. It is generally not possible to directly observe whether a hospital has adequate quantities of each necessary input to meet realized demand. Therefore, identifying the times when it faces binding capacity constraints raises several important issues.

The number of beds available for inpatient admissions is a common measure of hospital size and capacity. While it is possible to observe if hospital occupancy equals or exceeds hospital bed capacity, Baker et. al. (2004) find that such events are rare. In our own analysis of occupancy for the more than 20,000 hospital-days between December 1997 and November 1998, we find that occupancy equaled or exceeded available beds for 18 hospital-days. For several reasons, a hospital may face binding capacity constraints even if occupancy is strictly lower than bed capacity. First, occupancy measured by a midnight census of patients underestimates demand because it does not account for those treated as outpatients without formal admission. It also misses the patients who are admitted and discharged on the same day. Second, hospital beds must be staffed in order to constitute available capacity on a specific day, and hospitals may not be able to adequately alter staffing levels in response to short-term fluctuations in demand. Finally, because of limited substitutability in facilities and equipment between various hospital departments—available beds in the obstetric or neonatal units may be of limited use to cardiac patients, for example—a hospital may be capacity constrained even when the number of patients is strictly below the hospital's bed capacity.

Another issue has to do with the extent to which capacity may be a public good among hospitals in a market. In the Gal-Or (1994) model, a hospital can transfer some patients to a neighboring hospital that has available capacity if its own realized demand exceeds available capacity. This effect may be especially important in markets dominated by hospital chains or systems because the transfer of patients among affiliated hospitals does not involve loss of revenue to rival organizations. Because Oregon—and particularly the preponderant Portland metropolitan area—is dominated by hospital systems such as Providence and Legacy, capacity constraints at the hospital level may not be as important as market or system-wide capacity constraints. Furthermore, physicians often have admitting privileges at several hospitals. Rizzo and Goddeeris (1998) estimate from a nationally representative sample of self-employed physicians that, on average, specialist physicians have admitting privileges at 2.78 hospitals and primary care physicians have privileges at 2.17 hospitals. When physicians with admitting privileges at multiple hospitals find that one hospital is capacity constrained, they may direct patients to an alternative hospital. Finally, providers who advise patients to seek care in hospital emergency rooms may also direct patients to alternative hospitals if they are aware that a specific hospital is very busy on a specific day. If hospital capacity is indeed shared among neighboring or affiliated hospitals to a significant extent, then the largest impact on a hospital's admissions and discharge behavior is likely to occur not when the hospital itself appears busiest, but when hospital system or market-wide demand is greatest. The occasions when market-wide demand is greatest are also likely to be those when nursing and other specialized labor is the scarcest in spot markets, thus leaving hospitals capacity constrained even when they have an ample number of available beds.

For this paper, we conducted analyses using three different ways of identifying Thursdays when specific hospitals could be capacity constrained. In the first, we assume that each hospital serves a market comprising all hospitals within a fifteen mile radius. We identify the 20% of Wednesdays with the greatest and lowest cumulative DRG relative weight counts (i.e., sum of all inpatients' DRG relative weights) in each hospital's market as preceding high and low demand Thursdays for that hospital. In the second approach, we identify the 20% of Wednesdays with the greatest and lowest cumulative DRG relative weight counts in each hospital as preceding high and low demand Thursdays for that hospital. In the third approach, we restrict our analysis to hospitals in the Portland metropolitan area which we assume to be a single market. We identify the 20% of Wednesdays with the greatest and lowest cumulative DRG relative weight counts in the Portland area as preceding high and low demand Thursdays for hospitals in the area. In each approach, we postulate that a hospital faces no capacity constraints on low-demand days, but may face binding capacity constraints on high-demand days. Our empirical analysis of hospital behavior in response to short-term changes in demand is based on differences between observed admissions and discharges on high and low demand days in pooled data from all hospitals in our data set for the first two approaches, and from all hospitals in the Portland metropolitan area for the third approach. The results from the analysis using the first approach to identifying days when capacity may be constrained are reported in the text of the paper. Appendix B (not included in current draft) contains summaries of results using the other two approaches.

Statistical Analysis

We conducted F-tests of differences in means as well as Kolmogorov-Smirnov (KS) tests to identify differences in distribution functions between high and low demand days for admissions and discharges of patients with each of the three types of insurance that we analyze (e.g., H_0: The distribution over DRG relative weights of Medicare patients admitted on high demand days is the same as that for Medicare patients admitted on low demand days). We also tested for differences in means and in distribution functions between patients with different types of insurance for admissions as well as discharges on both high and low demand days (e.g., H_0: The distribution over DRG relative weights of Medicare patients admitted on high demand days is the same as that for OHP patients admitted on high demand days). Since the results in Section 4 (particularly Corollary 2) imply differences in both CDFs and means, we conclude that admission or discharge patterns are systematically different across insurance types, or between high and low demand days for the same insurance type, only if both the K-S test and the F-test for difference in means are statistically significant.

Admissions on high and low demand days

Panels A and B of Figure 1 show the distribution of admissions on low and high demand days, respectively, over Medicare's DRG relative weights for patients covered by private non-HMO insurance, Medicare, and OHP. Since all CDFs have a value of 0 at the beginning of the range and a value of 1 at the end of the range, differences in resource requirements can be observed in the divergence of the CDFs for intermediate weights as well as in divergence in means.

As both panels illustrate, there are important differences across insurance types in proportion of patients admitted with high resource requirements. The proportion of OHP patients with low resource requirements was higher, and the proportion of OHP patients with high resource requirements was lower than for either Medicare or privately insured patients. This is also reflected in OHP patients' lower mean DRG relative weights of 0.907 and 0.854 on high and low demand days, respectively. By comparison, the mean weight of privately insured non-HMO patients was 1.002 for those admitted on high demand days and 1.033 for those admitted on low demand days. Medicare patients were the most resource intensive per admission on average of the three groups on the measure we employed with mean DRG relative weights of 1.452 and 1.438 on high and low demand days, respectively. Across all patients, the mean DRG relative weight for patients admitted on high demand days was 1.129, and 1.128 for those admitted on low demand days.

The value of the CDF on low demand days was greater than on high demand days for nearly the entire range of resource requirements in treatment for OHP patients (P-value: 0.012). This was not the case for Medicare or privately insured patients, even though the K-S test found the CDFs on high and low demand days for such patients to be different (P-values: 0.008 for Medicare, and 0.012 for the privately insured). Mean DRG relative weights were not statistically significantly different on high and low demand days for Medicare and privately insured patients (P-values: 0.63 and 0.22, respectively), while for OHP patients the difference was significant at the 10% level (P-value: 0.087). Across all patients, the difference in mean DRG relative weights was not statistically significant (P-value: 0.923), even though the K-S test found the two CDFs statistically significantly different (P-value: 0.005).

These results suggest that, to the extent that there was an impact on hospitals' admission behavior due to high demand that could be measured using DRG relative weights, this impact was felt by OHP patients. A cautious reading of these results does not permit us to conclude that Oregon hospitals discriminate against OHP patients in admissions.

Discharges on high and low demand days

To examine how the discharge behavior of Oregon hospitals differs on high demand days when compared to low demand days, we first compute probability distributions of how long people in each DRG stay in the hospital. Using these distributions we calculate, on each day of each inpatient's hospital stay, the conditional probability of discharge on every subsequent day. These conditional probabilities are used to compute *ERLOS* conditional on the number of days s/he has already spent in the hospital as defined in equation (5).

Figure 3 shows the *ERLOS* for patients admitted to hospital with conditions classified under DRGs 121 (circulatory disorders with acute myocardial infarction and major complications who were discharged alive; N=1,625) and 122 (circulatory disorders with acute myocardial infarction without major complications who were discharged alive; N=1,527). *ERLOS* on the day of admission was 4.94 for patients in DRG 121, and 3.74 for patients in DRG 122. In a pattern seen consistently in almost all DRGs, *ERLOS* for the patients shown in Figure 3 initially decreased with the number of inpatient days as the relatively less resource intensive patients neared discharge, but then increased as the discharge of such patients left a remaining cohort of more resource intensive inpatients. For patients in both DRGs 121 and 122, this occurred on day 5. Peak *ERLOS* occurred on day 12 for patients in DRG 121 on day 11 for patients in DRG 122.

To ensure that the following results are due to systematic differences between high and low demand days, we adjusted each patients' *ERLOS* at discharge to account for within DRG variation due to month of year, and whether a day occurred at the beginning, middle, or end of a month. Specifically, each person's ERLOS at discharge was adjusted to DRG-specific mean values of independent variables using the following regression:

$$\ln ERLOS^{d} = \alpha_{E} + D \times Month + D \times WOM + \varepsilon , \tag{6}$$

where $ERLOS^{d}$ is *ERLOS* at discharge. WOM is an index of the weeks within a month (1 for the first seven days, 2 for the next seven days, and so on), while *D* and *Month*, are indices of DRGs and months respectively; ε is an error term.

Panels A and B of Figure 4 show the distribution over adjusted *ERLOS* of patients discharged on low and high demand days, respectively. On both high and low demand days, the proportion of Medicare patients discharged early relative to expectation (i.e., high adjusted *ERLOS* at discharge) was higher than for either OHP or privately insured patients. Over most of the range of *ERLOS*, a larger proportion of OHP patients were discharged early relative to expectation than privately insured patients. These differences in discharge patterns across insurance types are also reflected in mean adjusted *ERLOS* at discharge: 3.563 and 3.473, respectively, on high and low demand days for Medicare patients; 3.047 and 2.853 for OHP patients; 2.637 and 2.602 for privately insured patients. Across all patients, the mean adjusted *ERLOS* at discharge was 3.046 for those discharged on high demand days, and 2.964 for those discharged on low demand days.

Since adjusted *ERLOS* is a measure of how much longer a specific patient can be expected to remain hospitalized, differences in adjusted *ERLOS* of those discharged on high and low demand days can provide evidence of how hospitals' discharge behavior responds to capacity constraints. Panel A of Figure 5 shows the differences between cumulative distribution of discharges on low and high demand days for Medicare and OHP patients, while panel B shows the differences for privately insured and OHP patients. For patients with OHP or with private insurance, the value of the CDF on low demand days was greater than on high demand days almost the entire range of adjusted *ERLOS*. However, this was not the case for patients covered by Medicare. The K-S test was significant for patients with OHP, Medicare, and private insurance (P-values: 0.032, 0.013, and 0.015, respectively). Mean adjusted *ERLOS* at discharge was not statistically significantly different on high and low demand days for Medicare and privately insured patients (P-values: 0.138 and 0.501, respectively). However, for OHP patients the difference in means between high and low demand days was statistically significant (P-value: 0.032). Across all patients, the difference in means was statistically significant (P-value: 0.019), as was the difference between the overall distribution of patients admitted on high and low demand days (P-value: 0.013).

These results show that, as predicted by our model, hospitals tend to discharge patients earlier relative to expectation on high demand days when compared to low demand days. Our results also suggest that there are important differences in Oregon hospitals' discharge practices towards OHP patients when compared to either Medicare or privately insured patients. While discharges of privately insured and Medicare patients are not systematically affected by high demand that may cause hospitals to be capacity constrained, OHP patients tend to be discharged earlier relative to expectation when the hospital may be capacity constrained.[11]

7. Conclusion.

Modeling variations in hospital occupancy, and their effects for scheduling and excess demand has a long history in the field of operations research (e.g., Shonick 1970; Harrison, Shafer, and McKay 2005). With considerable anecdotal evidence that hospitals cope with inadequate capacity during periods of congestion by restricting admissions and hastening discharges, there is a

[11] If within DRG severity is greater for OHP patients, the degree of discharge discrimination is understated.

need for more formal economic analyses. This paper develops a model to examine the impact of short-term fluctuations in demand for inpatient care on hospitals' admission and discharge behavior. The model predicts that when admission decisions are affected by inadequate capacity due to high demand, a hospital will only admit patients whose requirements in treatment exceed a certain threshold. When discharge decisions are affected by a binding capacity constraint, the hospital will discharge patients earlier relative to expectation than when it has ample spare capacity. We also develop a simple proxy measure for the expected additional stay of current inpatients. When incentives to discriminate are created by limited capacity, we show that regression or other analyses of differences in mean effects can be logically flawed as methods for detecting such discrimination. Perhaps most significantly, we develop a test for discriminatory behavior that eliminates the need to control for underlying differences in health and treatment seeking behavior of different patient groups provided the analysis is based on comparison of hospital behavior when it is capacity constrained to behavior on otherwise similar days when it is not capacity constrained. Since characteristics of days (day of week; month; beginning, middle, or end of month; holiday vs. workday) are far more readily observable than characteristics of individual patients or patient groups (health status, care seeking behavior), our test can dramatically simplify the detection of discriminatory behavior.

Because of our focus on short-term changes in hospital behavior, our model differs significantly from earlier models of hospital behavior. Specifically, the Newhouse (1970) model with its explicit choice by the hospital of the quality of care it will provide, and a budget constraint are not contained in our approach. This decision is not addressed because it is unlikely that a hospital will either be able or wish to choose a new level of quality on each day. Similarly, it is unlikely that the hospital would be required to meet a specific budgetary target each day. These longer-term issues can be viewed as predetermined parameters in a short-term model. Choices of different quality levels would alter marginal costs and demand in our model. A stringent budget constraint would force the hospital to place greater weight on profit in its objective function, and create an interdependency between admissions of profitable and unprofitable patients as the former enable provision of services to the latter. This can again be reflected in a greater weight on profits in a short-term model such as ours. None of these would alter the qualitative results presented in this paper.

In our analysis of hospital discharge data from Oregon, we found that patients were discharged earlier relative to expectations on high demand days than on low demand days. The impact of high demand on admissions behavior was not statistically significant across all patients. This may indicate that a majority of hospitals' response to capacity constraints takes the form of early discharge of current inpatients rather than that of denying admission to potential patients with low resource requirements. It may additionally indicate that while hospitals can be selective about which patients they discharge early, the techniques used to limit admissions in the short term (e.g., diverting ambulances to other hospitals) may not be amenable to patient selection.   Our comparison of hospital discharge behavior towards patients with different types of insurance shows that, on both high and low demand days, Medicare patients are discharged earlier relative to expectation than patients covered by either private insurance or Medicaid. If the Medicare payments system is more effectively prospective than payment systems employed by private insurers or Medicaid, then this result may indicate that hospitals respond very strongly to the incentive to discharge early that is created by prospective reimbursement. Our results also suggest important differences in Oregon hospitals discharge practices towards Medicaid patients when compared to Medicare or privately insured patients. While discharges of Medicare and privately insured patients were not systematically affected by high demand, Medicaid patients tended to earlier relative to expectations on high demand days.

Future research can refine and extend our work in several important ways. Our analysis of admissions decisions is based on DRG relative weights that preclude adjustment for seasonal and other factors because they are identical for all patients within a DRG. Use of measures that take better account of within DRG variance in resource requirement in treatment and the urgency of medical need would address this issue. Our methods for identifying capacity constraints can also be refined. Capacity constraints may be shorter or longer in duration than the daily phenomena we model, or they may be more relevant for individual hospital departments than for the hospital as a whole.

Perhaps most importantly, our model and the results it yields have wider applications for future research in attempts to detect discrimination in medical treatment using widely available data. Typical analyses of data regarding the utilization of health care services yield information about the treatment that is provided to different groups, but cannot detect discrimination because of the inability to account for possible underlying differences in treatment requirements and treatment seeking behavior of different groups. Even when controls for these differences are possible, such controls are imperfect. Our analysis finds that discrimination in treatment can be detected from utilization data by examining the differential impact that inadequate capacity has on the treatment of individuals with different types of insurance. If hospital behavior under capacity constraints is compared to behavior on otherwise similar days when the hospital is not capacity constrained, we show that such discrimination can be detected even if individuals with different types of insurance have different treatment requirements and treatment seeking behavior. Most crucially, our technique for detecting discrimination does not require us to control for health and other differences that also influence the treatment that patients from different insurance plans will receive. With minor modifications, the test can also be used to detect gender, race, and income-based discrimination in hospital treatment. It may also be possible to modify the test to detect discrimination in outpatient settings, as well in areas beyond health care such as the responsiveness of fire or police departments.

There is an extensive health services literature on disparities in healthcare utilization,[12] and the need to reduce disparities has emerged as a policy priority (Smedley, Stith, and Nelson, 2002; Putsch and Pololi, 2004). As a result of this priority, economists have also recognized the need to better understand the sources of the disparities.[13] However, the extent to which discrimination, as opposed to other factors, contributes to the disparities remains a major challenge for research. The ability to detect and isolate discrimination from other potential contributors, using widely available utilization data, provides investigators with a powerful tool to support this effort.

[12] Weinick, Zuvekas, and Cohen (2000) and Mayberry, Mili, and Ofili (2000) provide comprehensive reviews.
[13] See, for example, Balsa and McGuire (2001, 2003), and Balsa, McGuire and Meredith (2005).

Appendix A

Proof of Proposition 1: If the hospital faces a binding capacity constraint on day *i*, then $\lambda > 0$.

(i) if $0 < \lambda \leq \alpha(r_x - c)t_{\min} + (1-\alpha)\beta_x B(t_{\min})$, then $\mu_x \geq 0$ and $t_{ix}^c = t_{ix}^* = t_{\min}$. Since the capacity constrained hospital cannot admit all patients it would otherwise admit and since patients from plan y with $t = t_{iy}^*$ add least to the hospital's objective function, a capacity constrained hospital will restrict admissions of such patients. Therefore, $t_{iy}^c > t_{iy}^* \geq t_{\min}$.

(ii) if $\lambda > \alpha(r_x - c)t_{\min} + (1-\alpha)\beta_x B(t_{\min})$, then $t_{ix}^c > t_{ix}^* = t_{\min}$. It has already been shown that $t_{iy}^c > t_{iy}^* \geq t_{\min}$ when $\lambda > 0$ as it must be if $\lambda > \alpha(r_x - c)t_{\min} + (1-\alpha)\beta_x B(t_{\min})$.

(iii) if $\lambda > \alpha(r_x - c)t_{\min} + (1-\alpha)\beta_x B(t_{\min})$ then, $\mu_x = 0$ and $\lambda = \alpha(r_x - c)t_{ix}^c + (1-\alpha)\beta_x B(t_{ix}^c)$. Then, $\mu_y = 0$ and $\lambda = \alpha(r_y - c)t_{iy}^c + (1-\alpha)\beta_y B(t_{iy}^c)$. Since $r_x > r_y$ and $\beta_x = \beta_y$ by assumption, this yields that $t_{iy}^c > t_{ix}^c$.

Q.E.D.

Proof of Corollary 1: From Proposition 1, $t_{iy}^c > t_{ix}^c$. Then $f_{ix}(t) = f_{iy}(t)$ *for all* $t \in [t_{\min}, t_{\max}]$, implies that

(i) $g_{ix}^c(t) = g_{iy}^c(t)$ for $t \in [t_{\min}, t_{ix}^c)$, $g_{ix}^c(t) > g_{iy}^c(t)$ for $t \in [t_{ix}^c, t_{iy}^c)$, and $g_{iy}^c(t) = \dfrac{g_{ix}^c(t)}{1 - G_{ix}^c(t_{iy}^c)}$ for $t \in [t_{iy}^c, t_{\max}]$. Therefore, $\int_{t_{\min}}^{t_{\max}} t g_{ix}^c(t)dt < \int_{t_{\min}}^{t_{\max}} t g_{iy}^c(t)dt$.

(ii) $G_{ix}^c(t) \geq G_{iy}^c(t)$ for all $t \in (t_{\min}, t_{\max})$ follows from the argument in part (i) of this proof.

Q.E.D.

Proof of Corollary 2:

(i) This follows from the argument in part (i) of Proposition 1.

(ii-a) This follows from the argument in part (ii) of Proposition 1.

(ii-b) Since $g_{ij}^c(t) = g_{ij}^*(t)$ for $t \in [t_{\min}, t_{ij}^*)$, $g_{ij}^c(t) \leq g_{ij}^*(t)$ for $t \in [t_{ij}^*, t_{ij}^c)$, and $g_{ij}^c(t) \geq g_{ij}^*(t)$ for $t \in [t_{ij}^c, t_{\max}]$, $D_{ij}(t)$ is at its maximum for $t = t_{ij}^c$. Since $t_{iy}^c > t_{ix}^c$ from Proposition 2 when $\alpha > 0$, the value of t that maximizes $D_{iy}(t)$ is greater than the value of t that maximizes $D_{ix}(t)$ when the hospital discriminates against patients from plan *y* in admissions.

Q.E.D.

References

Bagust, A., Place, M. and Posnett, J.W. "Dynamics of Bed Use in Accommodating Emergency Admission: Stochastic Simulation Model." British Medical Journal, Vol. 319 (1999), pp. 155-158.

Baker, L.C., Phibbs, C.S., Guarino, C., Supina, D. and Reynolds, J.L. "Within-Year Variation in Hospital Utilization and its Implications for Hospital Costs." Journal of Health Economics, Vol. 23 (2004), pp.191-211.

Balsa, A.I. and McGuire, T.G. "Statistical Discrimination in Health Care." Journal of Health Economics, Vol. 20 (2001), pp. 881-907.

Balsa, A.I. and McGuire, T.G. "Prejudice, Clinical Uncertainty and Stereotyping as Sources of Health Disparities." Journal of Health Economics, Vol. 22 (2003), pp. 89-116.

Balsa, A.I., McGuire, T.G., and Meredith, L. "Testing for Statistical Discrimination in Health Care." Health Services Research, Vol. 40 (2005), pp. 227-252.

Berk, E, and Moinzadeh, K., "The Impact of Discharge Decisions on Health Care Quality." Mangement Science, Vol. 44(1998), pp. 400-415.

Carey, K. "Stochastic Demand for Hospitals and Optimizing "Excess" Bed Capacity." Journal of Regulatory Economics, Vol. 14 (1998), pp. 165-187.

Dor, A. and Farley, D.E. "Payment Source and the Cost of Hospital Care: Evidence from a Multiproduct Cost Function with Multiple Payers." Journal of Health Economics, Vol. 15 (1996), pp. 1-21.

Dranove, D. "Pricing by Non-Profit Institutions: The Case of Hospital Cost-Shifting." Journal of Health Economics, Vol. 7 (1988), pp. 47-57.

Dranove, D. and Satterthwaite, M.A. "The Industrial Organization of Health Care Markets." In: Culyer, A. and Newhouse, J.P. (Eds.), Handbook of Health Economics, Vol. 1B, Elsevier, Amsterdam, 2000, pp. 1093-1139 (Chapter 20).

Dranove, D. and White, W.D. "Medicaid-Dependent Hospitals and their Patients: How have they fared?" Health Services Research, Vol. 33 (1998) pp. 163-186.

Dranove, D. and White, W.D. "Recent Theory and Evidence on Competition in Hospital Markets." Journal of Economics and Management Strategy, Vol. 3(1994), pp. 169-209.

Ellis, R.P., and McGuire, T.G., "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply," Journal of Health Economics, Vol. 5(1986), pp. 129-151.

Frank, R.G. and Lave, J.R. "A Comparison of Hospital Responses to Reimbursement Policies for Medicaid Psychiatric Patients." RAND Journal of Economics, Vol. 20 (1989), pp. 588-600.

Friedman, B. and Pauly, M.V. “Cost Functions for a Service Firm with Variable Quality and Stochastic Demand.” Review of Economics and Statistics, Vol. 63 (1981), pp. 610-624.

Friedman, B. and Pauly, M.V. “A New Approach to Hospital Cost Functions and Some Issues in Revenue Regulation.” Health Care Financing Review, Vol. 4 (1983), pp. 105-114.

Harris, J.E. “The Internal Organization of Hospitals: Some Economic Implications.” Bell Journal of Economics, Vol. 8 (1977), pp. 467-482.

Harrison, G.W., Shafer, A. and Mackay, M. "Modelling Variability in Hospital Bed Occupancy." Health Care Management Science, Vol. 8(2005), pp. 325-334.

Gal-Or, E. “Excessive Investment in Hospital Capacities.” Journal of Economics and Management Strategy, Vol. 3 (1994), pp. 53-70.

Gaynor, M. and Anderson, G.F. “Uncertain Demand, the Structure of Hospital Costs, and the Cost of Empty Hospital Beds.” Journal of Health Economics, Vol.14 (1995), pp. 291-317.

Hodgkin, D., and McGuire, T.G., “Payment Levels and Hospital Response to Prospective Payment.” Journal of Health Economics, Vol. 13(1994), pp. 1-29.

Jones, S.A., Joy, M.P. and Pearson, J. “Forecasting Demand for Emergency Care.” Health Care Management Science, Vol. 5 (2002), pp. 297-305.

Joskow, P. “The Effects of Competition and Regulation on Hospital Bed Supply and the Reservation Quality of the Hospital.” Bell Journal of Economics, Vol.11 (1980), pp. 421-447.

Keeler, T.E. and Ying, J.S. “Hospital Costs and Excess Bed Capacity: A Statistical Analysis. Review of Economics and Statistics, Vol. 78 (1996), pp. 470-481.

Mayberry, R.M., Mili, F. and Ofili, E. “Racial and Ethnic Differences in Access to Medical Care.” Medical Care Research and Review, Vol. 57 Supplement (2000), pp. 108-145.

McMahon, L.F. and Newbold, R. “Variation in Resource Use Within Diagnosis-Related Groups: The Effect of Severity of Illness and Physician Practice.” Medical Care, Vol. 24 (1986), pp. 388-397.

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

Medicare Payment Advisory Commission (MedPAC), Report to Congress: Medicare Payment Policy, March 2003. Available at: www.medpac.gov/publications/generic_report_display.

Millard, P.H., Christodoulou, G., Jagger, C., Harrison, G.W. and McClean, S.I. “Modeling Hospital and Social Care Bed Occupancy and Use by Elderly People in an English Health District.” Health Care Management Science, Vol. 4 (2001), pp. 57-62.

Newhouse, J.P. “Reimbursing Health Plans and Health Providers: Efficiency in Productions versus Selection.” Journal of Economic Literature, Vol. 34 (1996), pp. 1236-1263.

Newhouse, J.P. “Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital.” American Economic Review, Vol. 60 (1970), pp. 64-74.

Putsch, R.W. and Pololi, L. “Distributive Justice in American Healthcare: Institutions, Power, and the Equitable Care of Patients.” American Journal of Managed Care, Vol. 10 Special Issue (2004), pp. SP45-SP53.

Rizzo, J.A. and Goddeeris, J.H. “The Economic Returns to Hospital Admitting Privileges.” Journal of Health Politics, Policy and Law, Vol. 23 (1998), pp. 483-516.

Shonick, W., "A Stochastic Model for Occupancy-Related Random Variables in General-Acute Hospitals." Journal of the American Statistical Association, Vol. 65(1970), pp. 1474-1500.

Shen, J.Z. “DRG Relative Weights.” In: Shen, J.Z., Textbook of DRG-based Payment System Greenwich Village Press, New York, 1996, pp. 172-188 (Chapter 6).

Smedley, B.D., Stith, A.Y. and Nelson, A.R. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Institute of Medicine, National Academy Press, Washington, DC, 2002.

Sterk, W.E. and Shryock, F.G. “Modern Methods Improve Hospital Forecasting.” Healthcare Finance Manager, Vol. 41 (1987), pp. 96-98.

Weinick, R.M., Zuvekas, S.H. and Cohen, J.W. “Racial and Ethnic Differences in Access and Use of Health Care Services, 1977 to 1996.” Medical Care Research and Review, Vol. 57 Supplement (2000), pp. 36-54.

Figure 1A: Distribution of Admissions on Low Demand Thursdays

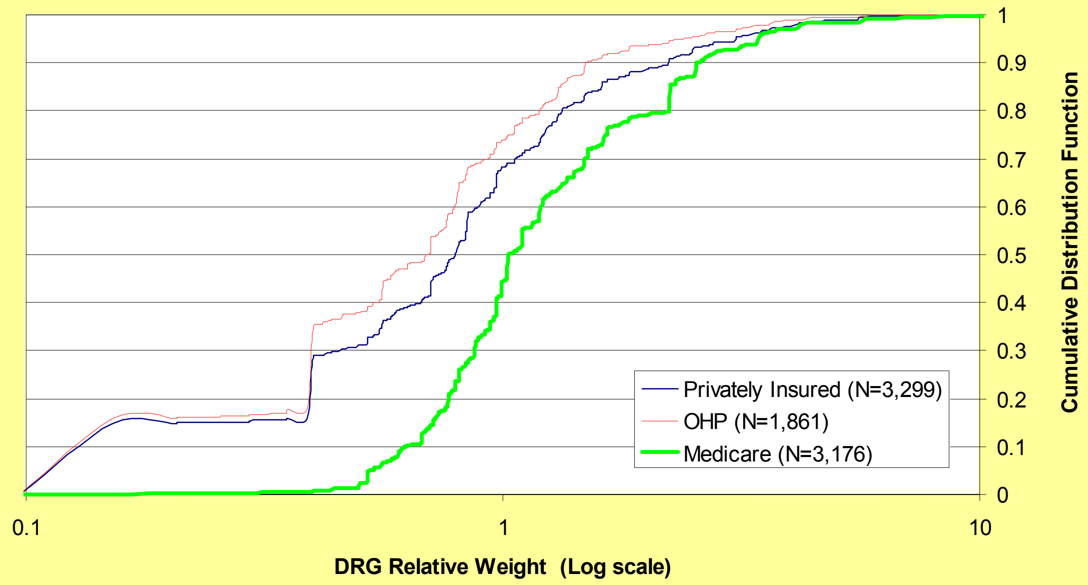
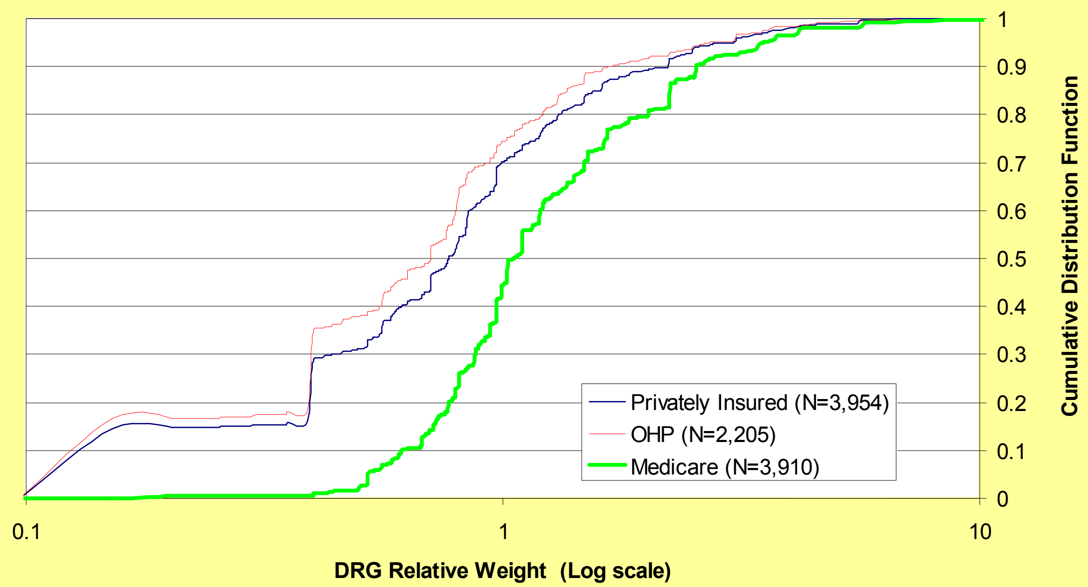


Figure 1B: Distribution of Admissions on High Demand Thursdays



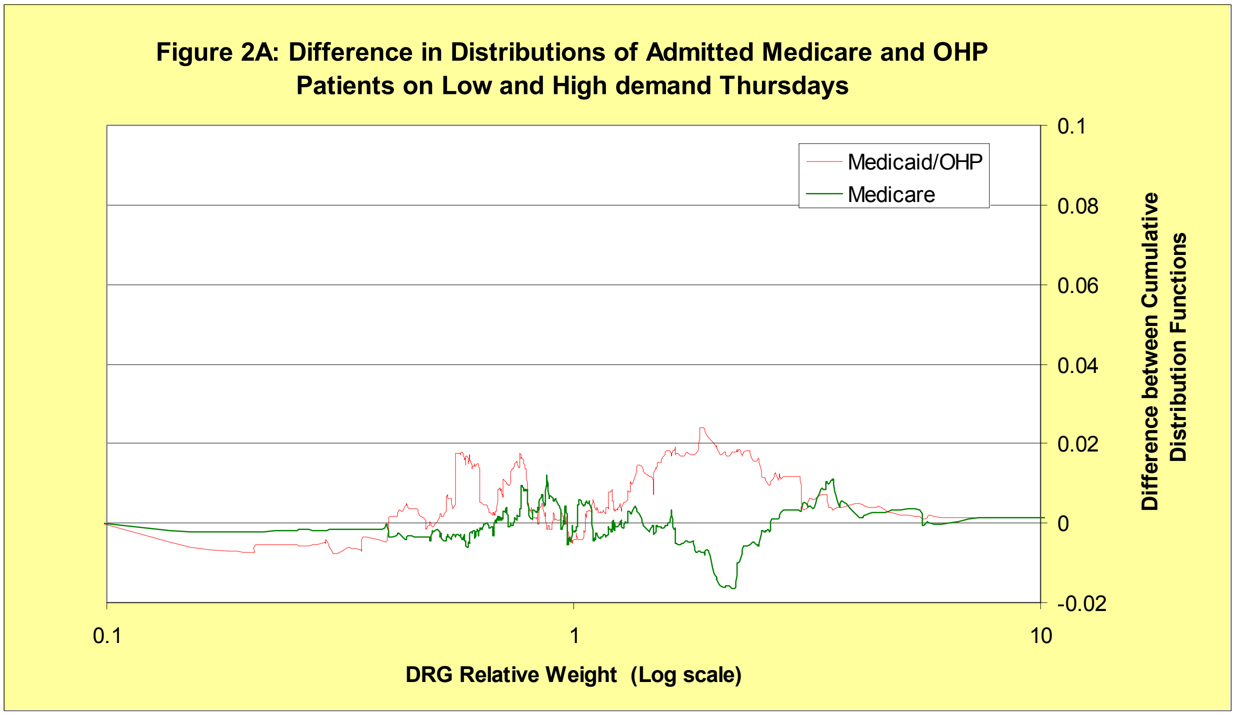


Figure 2A: Difference in Distributions of Admitted Medicare and OHP Patients on Low and High demand Thursdays
Medicaid/OHP
Medicare
0.1
0.08
0.06
0.04
0.02
0
-0.02
Difference between Cumulative Distribution Functions
0.1
1
10
DRG Relative Weight (Log scale)

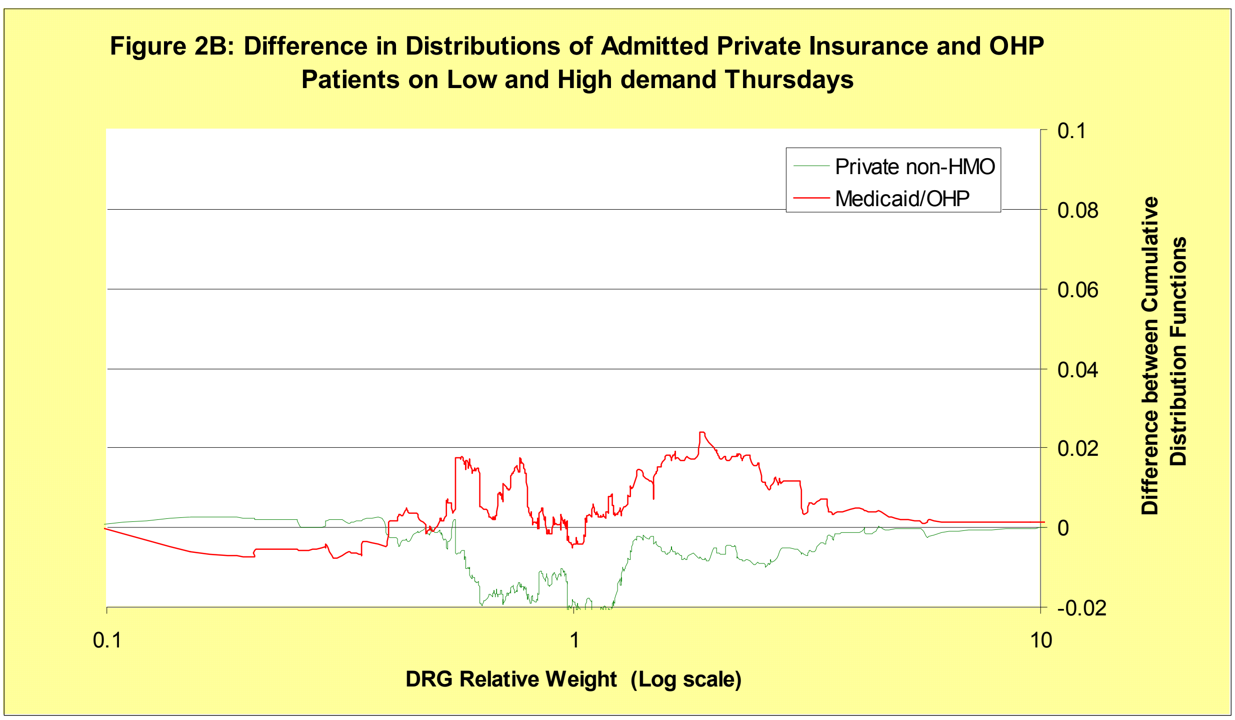


Figure 2B: Difference in Distributions of Admitted Private Insurance and OHP Patients on Low and High demand Thursdays
Private non-HMO
Medicaid/OHP
0.1
0.08
0.06
0.04
0.02
0
-0.02
Difference between Cumulative Distribution Functions
0.1
1
10
DRG Relative Weight (Log scale)

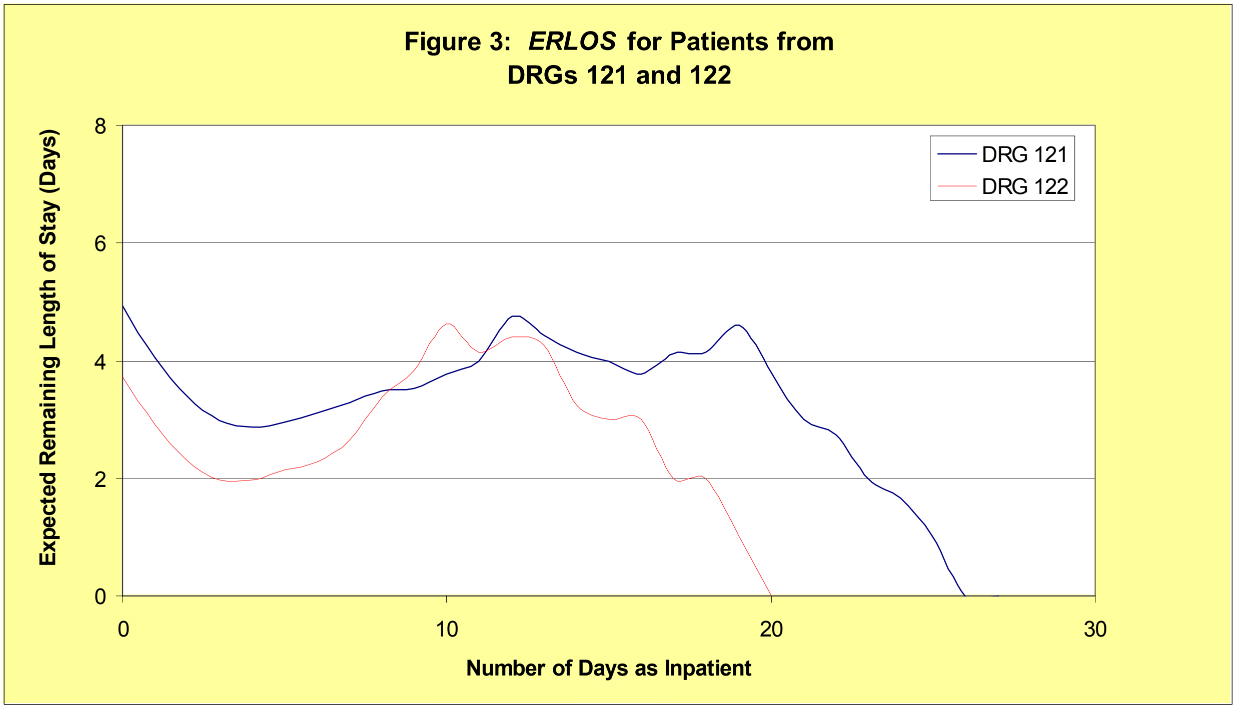


Figure 3: *ERLOS* for Patients from
DRGs 121 and 122
Expected Remaining Length of Stay (Days)
8
6
4
2
0
0
10
20
30
Number of Days as Inpatient
DRG 121
DRG 122

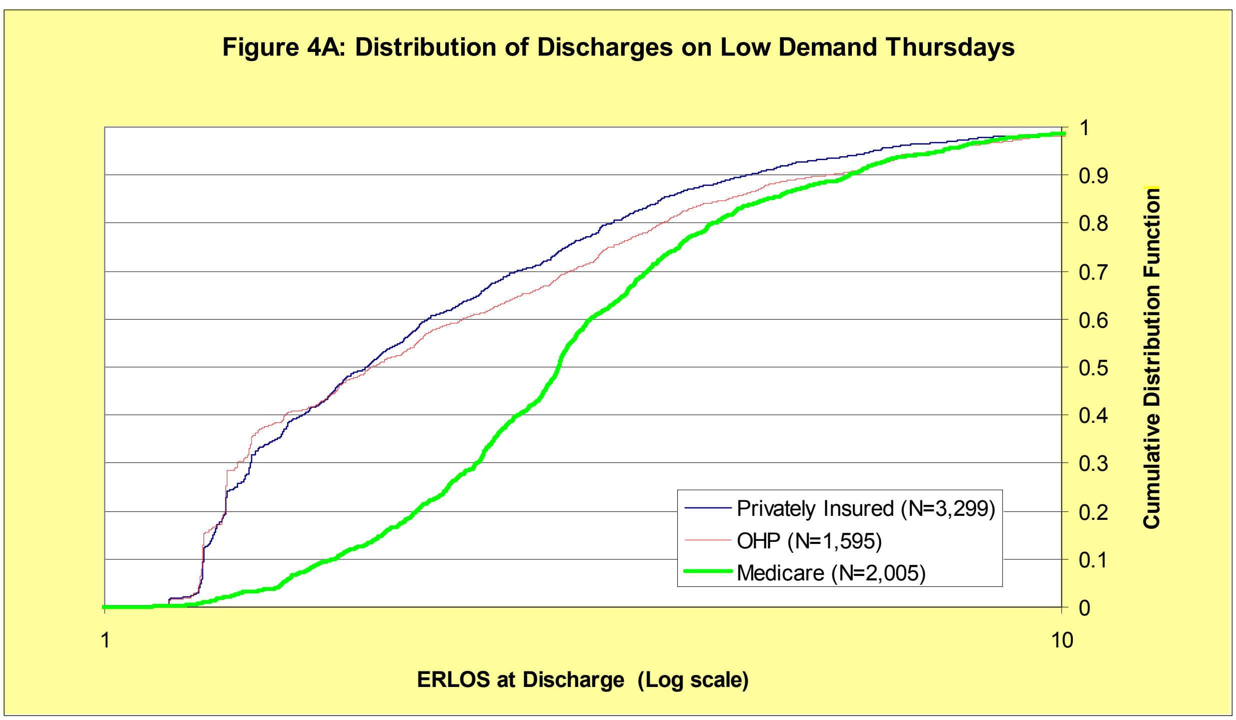


Figure 4A: Distribution of Discharges on Low Demand Thursdays
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Cumulative Distribution Function
Privately Insured (N=3,299)
OHP (N=1,595)
Medicare (N=2,005)
1
10
ERLOS at Discharge (Log scale)

Figure 4B: Distribution of Discharges on High Demand Thursdays

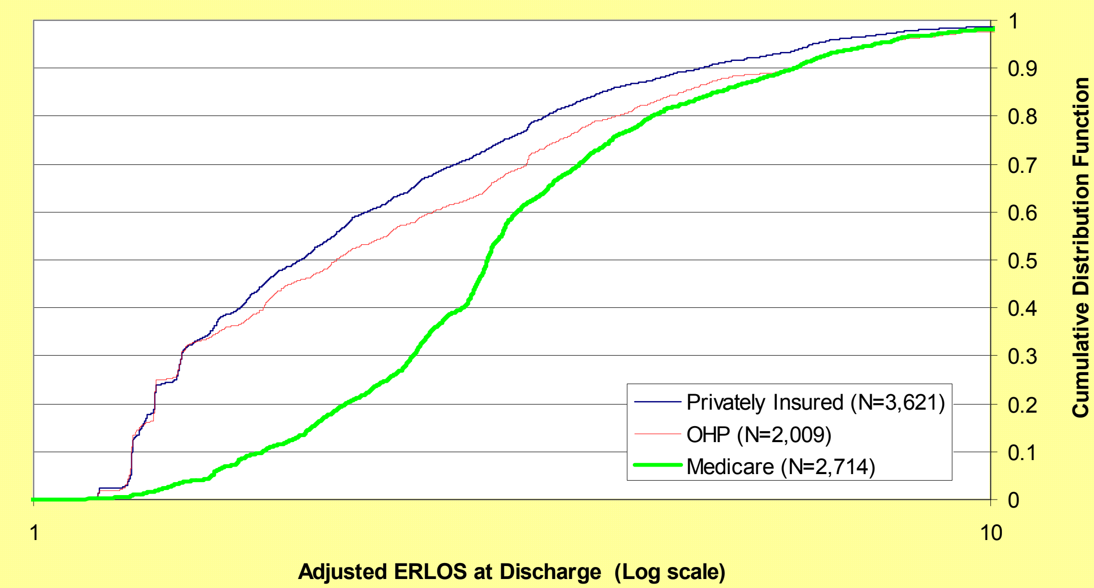
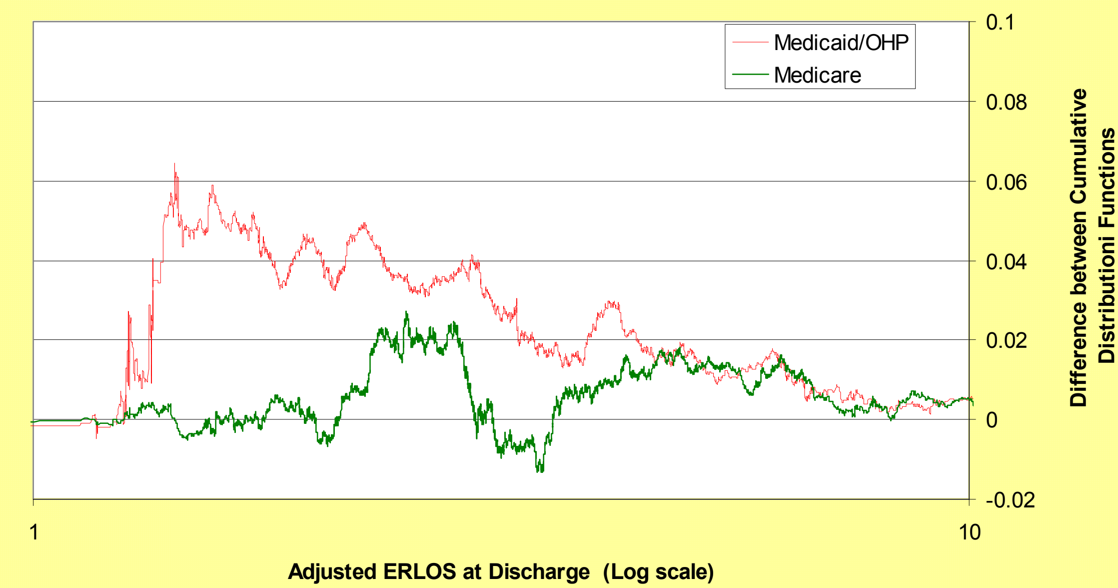


Figure 5A: Difference in Distributions of Medicare and OHP Patients Discharged on Low and High demand Thursdays



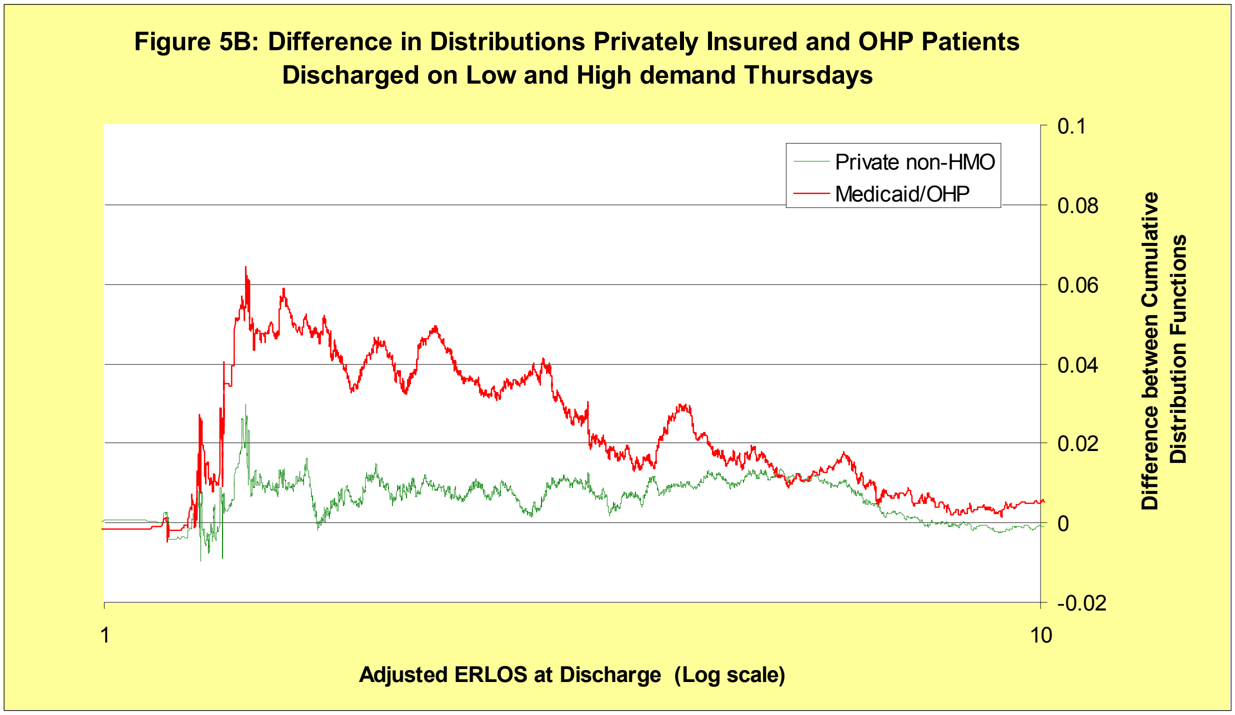


Figure 5B: Difference in Distributions Privately Insured and OHP Patients Discharged on Low and High demand Thursdays
Private non-HMO
Medicaid/OHP
0.1
0.08
0.06
0.04
0.02
0
-0.02
1
10
Adjusted ERLOS at Discharge (Log scale)
Difference between Cumulative Distribution Functions