Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks

Amitabh Chandra

Harvard University and National Bureau of Economic Research

Douglas O. Staiger

Dartmouth College and National Bureau of Economic Research

A large literature in medicine documents variation across areas in the use of surgical treatments that is unrelated to outcomes. Observers of this phenomenon have invoked "flat of the curve medicine" to explain it and have advocated for reductions in spending in high-use areas. In contrast, we develop a simple Roy model of patient treatment choice with productivity spillovers that can generate the empirical facts. Our model predicts that high-use areas will have higher returns to surgery, better outcomes among patients most appropriate for surgery, and worse outcomes among patients least appropriate for surgery, while displaying no relationship between treatment intensity and overall outcomes. Using data on treatments for heart attacks, we find strong empirical support for these and other predictions of our model and reject alternative explanations such as "flat of the curve medicine" or supplier-induced demand for geographic variation in medical care.

We are grateful to our editors, three astute reviewers, Jonathan Skinner, Elliott Fisher, Jack Wennberg, and many seminar participants for helpful comments. Jeffrey Geppert, Dan Gottlieb, and Lee Lucas provided expert assistance with the data. We received support from National Institute of Aging grant P01 AG19783-02. We obtained access to the proprietary data used in this paper through a data use agreement between the Centers for Medicare and Medicaid Services (CMS) and Dartmouth. Readers wishing to use these data must obtain them from CMS. The opinions in this paper are those of the authors and should not be attributed to the National Institute of Aging or the National Bureau of Economic Research.

[*Journal of Political Economy*, 2007, vol. 115, no. 1]

I. Introduction

Since Sir Allison Glover (1938) first documented significant variation in tonsillectomy rates across areas in the United Kingdom, an enormous body of literature in economics and medicine has documented variations in the use of intensive treatments (i.e., surgical or technologically intensive treatments) across comparable locales. Surprisingly, the use of more intensive procedures is not associated with improved satisfaction, outcomes, or survival but is associated with significantly higher costs. Traditional explanations such as sampling variation, differences in income and insurance, patient preferences, and underlying health status do not explain these variations.[1] These facts stand in sharp contrast to the results of randomized clinical trials that consistently find gains from the surgical management of acute conditions that are routinely interpreted as evidence in support of more intensive management of patients.[2]

These apparently conflicting findings can potentially be explained by a model of diminishing returns. Whereas randomized clinical trials are performed on a group of patients most likely to benefit from the intervention, the lack of a cross-sectional relationship between intensity and outcomes is explained by a "flat of the curve" argument, where physicians perform the intervention until the marginal return is zero (Enthoven 1980; Fuchs 2004). This is the interpretation that is used by McClellan, McNeil, and Newhouse (1994) and McClellan and Newhouse (1997) to explain the small returns from more intensive treatments for heart attacks. This explanation emphasizes inefficiency of medical spending, suggesting that welfare improvements may be realized by reducing spending in high-use regions. Indeed, Fisher et al. (2003*a*, 2003*b*) use this logic to argue that a 30 percent reduction in Medicare spending, such that spending in high-use regions is reduced to that of low-intensity areas, would not have any deleterious effects on patient outcomes or satisfaction.

While the diminishing returns model is intuitively appealing, it has a number of problems. First, there is no reason to expect wide variation in the use of treatments across areas that are similar, without making additional assumptions such as area norms or supplier-induced demand.

[1] Similar observations have been made by comparing treatments and outcomes in the United States with those in Canada and Europe. Phelps (2000) provides an economist's perspective on the immense literature on geographic variation in physician practice style. The medical literature on this topic is succinctly summarized in Fisher et al. (2003*a*, 2003*b*), Baicker and Chandra (2004), and the *Dartmouth Atlas of Health Care* (Wennberg and Cooper 1999; Wennberg and Birkmeyer 2000).

[2] In the context of treatments for heart attacks, see, e.g., Andersen et al. (2003) and Jacobs (2003). A review of over 23 trials by Keeley, Boura, and Grines (2003) noted the demonstrated superiority of the intensive intervention over alternative therapies for the management of heart attacks.

Second, such models still predict a positive relationship between medical spending in a region and patient outcomes unless all areas are in the range of zero or negative marginal benefits; this has never been found in the literature. A more fundamental problem with the diminishing returns model is that it predicts that the marginal benefit from more intensive patient treatment is lower in areas that are more aggressive. But U.S.-Canadian comparisons (Beck et al. 2003; Pilote et al. 2003) suggest the opposite: the marginal benefit from more technologically intensive treatment in heart attack patients is larger in the United States, where management of heart attacks is much more aggressive. These are facts that the flat of the curve model cannot reconcile. If it is the incorrect model, then embracing its policy implications of cutting spending could result in deleterious outcomes for many patients.

In this paper we develop a simple model of specialization in health care that uses a prototypical Roy model of treatment choice as its principal building block. In our model, patients are treated medically or intensively depending on their clinical appropriateness for these treatments, but we allow productivity spillovers to also affect the returns from each treatment choice. These spillovers may arise from knowledge spillovers, where physicians build experience by learning from one another, but may also arise from other sources, such as the selective migration of the best intensive physicians to certain areas and the corresponding migration of physicians who specialize in less intensive treatments to other areas. With these spillovers, as the proportion of patients in an area who are treated intensively increases, productivity spillovers increase the return to intensive treatment, while simultaneously reducing the return to the competing medical treatment. Thus this model naturally generates higher returns to receiving intensive treatment in intensive areas yet, because of the negative externality on patients receiving the competing treatment, does not necessarily generate any relationship between specialization and overall health outcomes. As we discuss in Section II, the model generates a rich set of sharp predictions that we test using detailed data on a sample of Medicare beneficiaries who had a heart attack (clinically referred to as an acute myocardial infarction [AMI]). Several of these predictions are consistent with other models such as the flat of the curve model, but others are unique to the Roy model with productivity spillovers. We find strong empirical support for our model and reject alternative explanations, such as "flat of the curve medicine" or supplier-induced demand, that are commonly proposed to explain geographic variation in medical care.

In Section II, we generalize a prototypical Roy model of specialization to allow for productivity spillovers and derive a wide range of testable predictions of this model. In Section III, we justify our focus on heart attacks to test this model and describe our data and estimation strategy.

Section IV tests the implications of our model. Section V presents conclusions and discusses the policy implications of our work, including its implications for the interpretation of randomized clinical trials. In the Data Appendix to this paper we provide details of our estimation and sample.

II. A Roy Model of Heart Attack Treatments with Productivity Spillovers

We consider a simple Roy model of patient treatment choice in the presence of productivity spillovers. In this model, an individual patient chooses between two alternative treatments: nonintensive management (emphasizing medical management, denoted with subscript 1) and intensive intervention (emphasizing surgery, denoted with subscript 2). The physician chooses the treatment option for each patient that maximizes utility on the basis of the expected survival rate (Survival_1, Survival_2) and cost (Cost_1, Cost_2) associated with each option. The survival rate from a given treatment option depends on patient characteristics (Z) but is also positively related to the fraction of all patients who receive the same treatment (P_1, $P_2 = 1 - P_1$).[3] Similarly, the cost from a given treatment option also depends on patient characteristics and may be negatively related to the fraction of all patients who receive the same treatment. Thus, a patient's choice of a specific treatment has a positive productivity externality, raising survival and reducing costs for all other patients receiving the same treatment.

The fundamental assumption of productivity spillovers in our model is quite plausible in health care. Much of physician learning about new techniques and procedures occurs from direct contact with other physicians ("see one, do one, teach one"), which leads to natural Marshallian knowledge spillovers from practicing in an area in which physicians have specialized in a particular style of practice. In their seminal paper on knowledge spillovers in medicine, Coleman, Katz, and Menzel (1957) found that doctors who were the most integrated in a social network were the first to adopt a new drug. More recently, a randomized control trial found that providing information to "opinion leaders" in a hospital resulted in large increases in the use of appropriate medications following heart attacks and decreases in the use of outdated

[3] This is similar to how network externalities are modeled by Katz and Shapiro (1985, 1986). The social network literature commonly uses the fraction of the population adopting a particular practice as a proxy for information flows since it captures the density of adopters in a network (see Rogers 2003). An alternative model would have the production externality depend on the number (rather than proportion) of patients in an area who receive the ith treatment. In this alternative model, larger areas are better at the use of both interventions, an implication that we test and reject (see table 8 below).

therapies (Soumerai et al. 1998). Strong evidence of knowledge spillovers comes from Meltzer (2006), in which hospitalists (physicians who specialize in hospital care as opposed to a specific disease) were randomly assigned to physicians groups. Physicians in the randomized groups exhibited social network effects by learning from the experience of the hospitalists to whom they were randomized.

An alternative mechanism generating productivity spillovers may operate through the availability of support services in area hospitals (cath labs, cardiac surgeons, cardiac care units, nursing staff), which will depend more on the overall practice style in the area than on any individual patient's needs. This second mechanism is analogous to the increased availability of software as more users adopt a given computer operating system, which has been emphasized in the literature on network externalities (e.g., Goolsbee and Klenow 2002). Finally, productivity spillovers may occur through the matching process of physicians to areas, since physicians who are more skilled at a particular treatment may self-select into areas in which use of this treatment is more common. Note that this last mechanism has different welfare implications, since selection of physicians into one area imposes externalities on other areas. We will return to this point in our discussion of welfare implications.

A. *The Model*

More formally, conditional on the fraction of patients receiving the treatment, let the survival rate and cost associated with each treatment option take the following simple form:

$$\text{Survival}_i = \beta_i^s Z + \alpha_i^s P_i + \epsilon_i^s \quad \text{for } i = 1, 2 \tag{1}$$

and

$$\text{Cost}_i = \beta_i^c Z + \alpha_i^c P_i + \epsilon_i^c \quad \text{for } i = 1, 2. \tag{2}$$

The subscript i indexes the treatment option, and we omit the patient subscripts for convenience. We assume that each patient's indirect utility function (U) depends on both the price and quality of care, as captured by the expected survival rate and costs associated with each treatment option:

$$U_i = \text{Survival}_i - \lambda \text{Cost}_i = \beta_i Z + \alpha_i P_i + \epsilon_i \quad \text{for } i = 1, 2, \tag{3}$$

where $\beta_i = \beta_i^s - \lambda\beta_i^c$, $\alpha_i = \alpha_i^s - \lambda\alpha_i^c$, and $\epsilon_i = \epsilon_i^s - \lambda\epsilon_i^c$. The parameter λ represents the value of life (survival per dollar) and captures the tradeoff being made by the patient and physician between improved survival and increased costs. A typical value of λ used in cost-effectiveness studies would place the societal value of a life year at around $100,000 (Cutler

2004), although physicians and patients may use $\lambda = 0$ if they do not pay directly for the cost of treatment (e.g., because of insurance). The term $\beta_i Z$ represents an index of how appropriate a given patient is for each treatment on the basis of medically relevant characteristics (Z). Patients who are more appropriate have greater utility from the treatment. The second term ($\alpha_i P_i$) captures the productivity spillover, which is positive if $\alpha_i > 0$: area specialization in a given treatment improves survival or reduces costs (or both) for all patients receiving that treatment. The final term (ϵ_i) represents factors that influence survival and costs and are known to the patient (or physician) at the time of choosing a treatment option but unobserved to the econometrician.

An individual patient receives the intensive treatment ($i = 2$) only if his or her utility is higher from the intensive than from the nonintensive treatment ($U_2 > U_1$). Therefore, the probability that an individual patient receives the intensive treatment is given by

$$\begin{aligned}\Pr\{\text{intensive treatment}\} &= \Pr\{i = 2\} = \Pr\{U_2 - U_1 > 0\} \\ &= \Pr\{(\alpha_1 + \alpha_2)P_2 - \alpha_1 + (\beta_2 - \beta_1)Z > \epsilon_1 - \epsilon_2\} \\ &= \Pr\{\alpha P_2 - \alpha_1 + \beta Z > \epsilon\}, \qquad (4)\end{aligned}$$

where $\alpha = \alpha_1 + \alpha_2$, $\beta = \beta_2 - \beta_1$, and $\epsilon = \epsilon_1 - \epsilon_2$. Integrating equation (4) over the distribution of patients (Z) in the population yields the market demand curve for the intensive treatment.

Among patients who choose the intensive treatment, the expected utility gain is given by

$$E[U_2 - U_1 | U_2 - U_1 > 0] = \beta Z + \alpha P_2 - \alpha_1 + E[\epsilon | U_2 - U_1 > 0]. \qquad (5)$$

Thus the expected utility gain among patients receiving the intensive treatment (the treatment on the treated) depends on a patient's relative appropriateness for the treatment, the proportion of all patients who receive the intensive treatment, and a term representing the selection effect. Because equation (5) has a simple Tobit structure, the truncated mean will increase with the mean of the underlying variable for any log concave distribution for ϵ (Heckman and Honore 1990). Thus patients who choose the intensive treatment will have a higher expected utility gain from the treatment (in terms of higher survival or lower costs) if they are more appropriate (higher βZ) or live in more intensive regions (higher αP_2).

B. Equilibrium

Equations (3)–(5) represent a standard Roy model except that, in equilibrium, P_2 must be equal to the demand for the intensive treatment

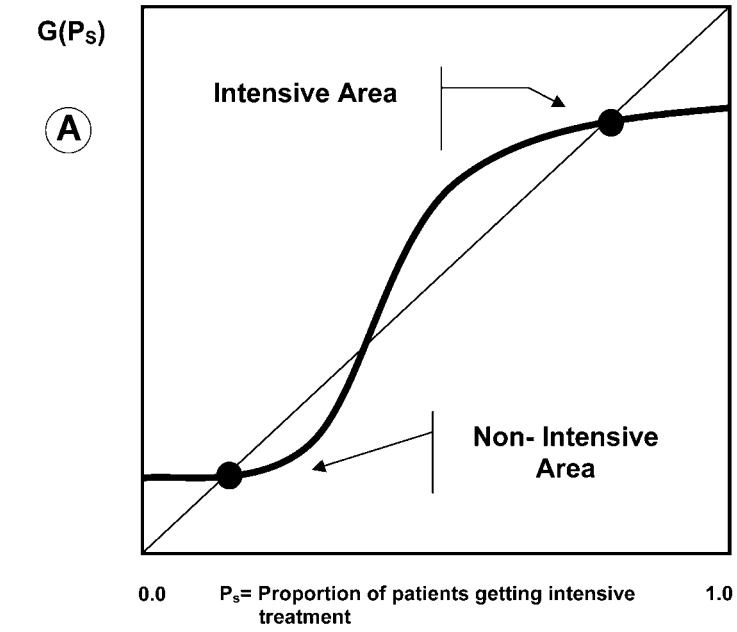
according to equation (4). If we let $f(Z)$ represent the distribution of Z in the population, this implies that the following equilibrium condition must hold:

$$P_2 = \int_Z \Pr(\alpha P_2 - \alpha_1 + \beta Z > \epsilon) f(Z) dZ$$
$$\equiv G(P_2). \quad (6)$$

Thus equilibrium in this model is defined as a solution to equation (6) (a fixed point). In other words, in equilibrium the proportion of patients choosing the intensive treatment must generate benefits to intensive treatment that are consistent with this proportion actually choosing the treatment. For example, if all patients choose the intensive treatment but this generates utility benefits such that some patients would prefer the nonintensive treatment, then rates of intensive treatment will decline. As fewer patients choose the intensive treatment, costs will rise and survival will fall (through the spillover mechanism), further lowering the benefits to intensive treatment. This process continues until we reach an equilibrium in which only the most appropriate patients are left receiving the intensive treatment. Thus cost and quality of care determine demand for each treatment option and adjust to equilibrate the market.

C. *Empirical Implications*

This simple model has a number of very strong empirical implications. First, as illustrated in figure 1, equation (6) can generate variation across areas in the proportion of patients receiving the intensive treatment for two different reasons. For a given distribution of patients (which fixes the function G), there can be multiple equilibria because of the productivity spillovers. For example, figure 1*A* illustrates a situation with two equilibria: an intensive equilibrium in which most patients receive intensive treatment and the returns to this treatment are high, and a nonintensive equilibrium in which few patients receive intensive treatment and the returns to this treatment are low (the middle crossing point represents an unstable equilibrium). Alternatively, differences across areas in the distribution of patients (which changes the function G) will also lead to different equilibria. For example, figure 1*B* illustrates a situation in which most of the patients in one area are more appropriate for intensive treatment leading to an intensive equilibrium, whereas most of the patients in the nonintensive area are not appropriate for intensive treatment. Even in the single-equilibrium scenario the productivity spillover increases the differences across areas: Having



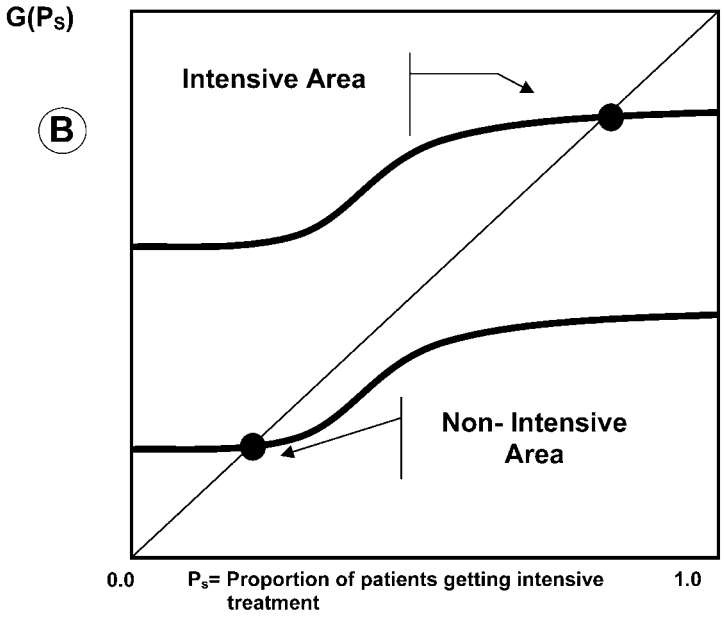


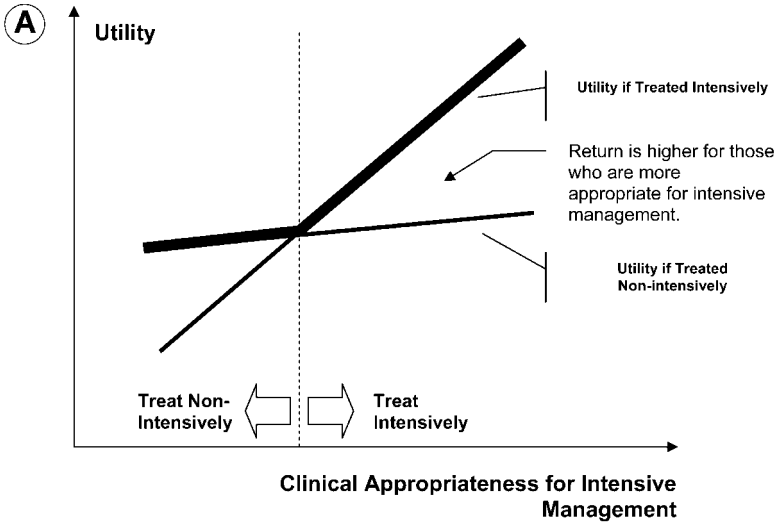
FIG. 1.—Characterizations of area variations: *A*, multiple equilibrium; *B*, single equilibrium.

more patients appropriate for intensive treatment in an area (a shift upward in G) increases the return to intensive treatment for other patients, which in turn leads to a further increase in intensive treatment (a move to the right along G). Thus small differences in the distribution of patients can potentially generate large equilibrium differences in specialization across areas.

A second empirical implication of the model that follows immediately is that identical patients will be more likely to be treated intensively in an area in which more patients are appropriate for intensive management of the heart attack (on the basis of Z). Any shift in the distribution of patients (Z) that results in a higher proportion being treated intensively (P_2) results in a higher probability of choosing intensive treatment for each patient as a direct consequence of equation (4). This implication is similar to that tested by Goolsbee and Klenow (2002) for individuals' purchase of home computers and is a unique implication of spillovers: Preferences of the population have spillovers on the choice of an individual. While this implication of the model is always true in the single-equilibrium case, it may not hold in the multiple-equilibrium case if a shift in the distribution of Z affects the choice among the equilibria. Unfortunately, our static model is silent on what determines the choice among multiple equilibria.[4]

A number of additional implications of the model are illustrated in figure 2. For simplicity, we ignore unobserved patient characteristics (ϵ) in this figure in order to focus on the intuition of the model; under standard regularity assumptions about ϵ (see Heckman and Honore 1990), this results in no loss of generality but significantly simplifies the exposition. Figure 2*A* plots patient utility (survival net of costs) as a function of a patient's appropriateness for intensive management (which depends on Z) for each of the two treatment options (intensive and nonintensive). Patients to the left of the intersection of these two curves are treated nonintensively and patients to the right are treated intensively. Patients further to the right are more likely to both receive intensive treatment and experience higher returns to the treatment if treated. Thus the model predicts that the return to intensive treatment (treatment on the treated) is highest among patients with the highest probability of treatment. This must occur through a combination of larger survival benefits or lower treatment costs among pa-

[4] Arthur (1989) emphasizes the importance of "lock-in" by historical events. In our application, two regions might have differed in terms of their distribution of initial patient types. This difference is what caused the choice of present specialization.



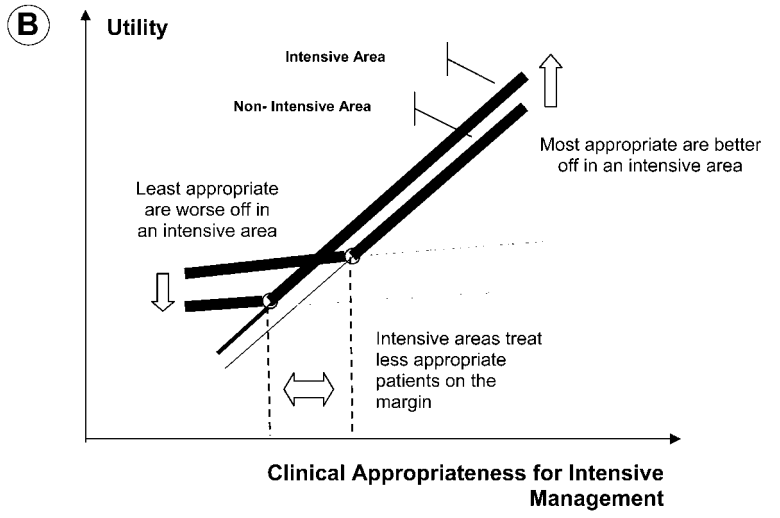


FIG. 2.—Graphical illustration of the Roy model with productivity spillovers. *A*, Relationship between two alternative ways to treat patients within an area. The production possibilities frontier describes the best treatment for a patient of given clinical appropriateness. The model predicts that the returns to intensive management are increasing in patients' appropriateness for such interventions. *B*, Contrasts the care across two areas that differ in their surgical intensity. As a result of the productivity spillover, patients appropriate for intensive management are better off in the surgically intensive areas, whereas patients appropriate for nonintensive management are worse off in such areas.

tients with the highest probability of treatment (the "more appropriate" patients).[5]

Figure 2*B* illustrates four additional implications of the model by comparing patient utility in intensive and nonintensive areas. First, the utility associated with nonintensive management is worse in areas that are intensive. This suggests that there should be a negative relationship between the proportion of patients who receive surgically intensive treatment and indicators of the quality of care for patients treated medically; that is, intensive surgical treatment crowds out good medical management. A second implication of the model follows directly: In intensive areas, patient utility will be higher among patients who are most appropriate for intensive management, but lower among patients who are more appropriate for medical management. In other words, the area's specialization in intensive management helps patients who are appropriate for this type of care through higher survival or lower costs, but patients who require less intensive management are harmed by the area's specialization in intensive management. For example, to the extent that older age is a crude proxy for requiring less intensive management, our model predicts that very old patients will have lower survival and higher costs in an intensive area, whereas those who are younger benefit through higher survival and lower costs from their areas' specialization in intensive management. On net, however, average benefits across all patients may be higher or lower in surgically intensive areas.

A third implication of our model is that the marginal patients receiving the intensive treatment in intensive areas will be less appropriate for the treatment than the average patient receiving the intensive treatment. As shown in figure 2*A*, patients are given the intensive surgical treatment if their appropriateness is above the point at which the nonintensive and intensive utility curves intersect. As seen in figure 2*B*, this intersection is at a lower appropriateness level in intensive areas. Therefore, the additional patients receiving the intensive treatment in these intensive areas are less appropriate than the patients receiving the intensive treatment in nonintensive areas.

A final implication of our model is that among those receiving intensive treatment, the benefit to receiving intensive treatment in the intensive area is larger than the benefit in the nonintensive area. In other words, the treatment effect on the treated will be larger in more intensive areas. As can be seen in figure 2*B*, this higher return is the

[5] More precisely, our model makes unambiguous predictions about the utility gains to a patient receiving the intensive treatment ($U_2 - U_1$). Utility can increase in three ways: (1) survival increases or cost declines or both, (2) survival increases that outweigh cost increases, or (3) cost declines that outweigh survival declines. Our discussion focuses on the first case as the most relevant, but the others are possible.

net result of a higher benefit if patients are treated intensively and a lower benefit if patients are treated nonintensively in the intensive areas. Thus the high return to intensive treatment in intensive areas is the result of both a positive productivity externality on intensive treatment and a negative productivity externality on medical treatment.

All the model's implications are consistent with the key empirical regularities summarized in the Introduction. In particular, our productivity spillovers model can generate (1) substantial differences across areas in the use of intensive procedures that are unrelated to average patient outcomes, (2) a negative correlation between surgical intensity and the quality of medical management of a condition, and (3) large returns to receiving the intensive intervention, particularly in high-intensity areas. Thus a very simple equilibrium model of productivity spillovers can rationalize all the main stylized facts in the literature.

Some of the empirical implications of our model are the result of productivity differences across areas rather than spillovers per se. In our model, productivity differences across areas arise solely through the productivity spillover term ($\alpha_i P_i$) in equation (3). But there may be other reasons for productivity differences across areas that are independent of the fraction of patients receiving each treatment. If this were the case, then areas that happened to be relatively good at the intensive treatment (with high patient utility relative to nonintensive treatment) would have high returns to intensive treatment and also be more likely to use the intensive treatment for any given patient—just as predicted by our productivity spillover model.

However, there are two empirical implications of our model that depend on productivity spillovers. First, spillovers imply that a patient is more likely to be treated intensively in an area in which other patients are more appropriate for intensive treatment (on the basis of Z). This implication would not arise without a direct link between productivity and the fraction of patients being treated intensively. Second, spillovers imply that specialization in the intensive treatment must help patients who are most appropriate for intensive treatment, but at the same time must harm patients who are least appropriate by reducing specialization in the nonintensive treatment. This implication would not necessarily arise without the negative relationship between the productivity of intensive and nonintensive treatments that is implied by spillovers. For example, suppose that productivity in both the intensive and nonintensive treatments was driven by a latent quality factor in the area, but that this latent quality had a larger impact on productivity for the more complex intensive treatment. Then areas with higher latent quality would treat more patients intensively, but all patients in these areas would have better outcomes.

Most of the empirical implications of our model would not be pre-

dicted by the standard flat of the curve explanation for area variation in treatment rates. In our modeling framework, the flat of the curve argument has no spillovers ($\alpha = 0$). Instead, areas vary in terms of the trade-off they make between survival and costs (λ varies): some areas are willing to incur greater costs per life saved (lower λ), perhaps because they have higher income, have excess hospital capacity, or are paid on a cost-plus (e.g., fee for service) basis. In this flat of the curve model, the areas that have the lowest λ place the least weight on costs in making treatment decisions and therefore are more likely to use the intensive (and more costly) treatment for patients who receive little survival benefit. But such a model predicts that the average survival benefit of intensive treatment will be *lower* for patients receiving the treatment in high-intensity areas, whereas the incremental costs will be *higher*. That is, the flat of the curve story fundamentally implies that high-intensity areas simply choose to treat additional patients who yield relatively few lives saved per dollar. Moreover, since externalities play no role in the flat of the curve argument, this model does not imply a lower quality of medical management or worse outcomes among patients least appropriate for intensive treatment in high-intensity areas. Finally, the appropriateness of other patients in the area plays no role in determining treatment of a given patient in the flat of the curve model, whereas this is a fundamental feature of models with productivity spillovers. Instead, the flat of the curve model suggests that area factors such as income and hospital reimbursement structure will determine the area intensity.

D. Welfare Implications

Any change in the fraction of patients receiving the intensive treatment has positive externalities on some patients and negative externalities on others. Therefore, Pareto improvements over any equilibrium are not possible in our model without side payments from winners to losers. We focus instead on whether any deviation from a given equilibrium will result in higher average utility in the population, implicitly assuming that either appropriate side payments could be made or the societal objective is to maximize average utility. Finally, for this to be an appropriate measure of welfare, we must assume that all spillovers are captured within the model and, in particular, that specialization in one market has no impact on other markets (e.g., the spillover mechanism is not working through attracting quality physicians from other areas).

Because each patient's choice of treatment affects the outcomes of all other patients, there is no reason to expect that equilibrium in our model is optimal. Let P^* be the proportion of patients receiving the intensive treatment in equilibrium (a solution to eq. [6]). Small in-

creases in P^* away from its equilibrium value, accomplished by changing the treatment for the marginal patient, will not affect the marginal patient who is indifferent between the two treatments in equilibrium. Thus the effect of such an increase in P^* on expected utility depends solely on the externality it has on other patients: the positive externality on patients receiving the intensive treatment balanced by the negative externality on patients receiving the nonintensive treatment. Let U be the average utility in the population given the proportion of patients receiving the intensive treatment (P_2), so that $U = \int_Z \max(U_1, U_2 | P_2 = P) f(Z) dZ$. Then it is straightforward to show that

$$\left.\frac{\partial U}{\partial P_2}\right|_{P_2 = P^*} = \alpha_2 P^* - \alpha_1(1 - P^*) = -\alpha_1 + \alpha P^*, \tag{7}$$

where $\alpha = \alpha_1 + \alpha_2$.

Equation (7) is quite intuitive. The impact on expected utility of a small increase in intensive treatment away from its equilibrium value is simply the weighted average of the positive externality (α_2) on the intensive treatment and the negative externality ($-\alpha_1$) on the nonintensive treatment, where the weight is the proportion of patients receiving the intensive treatment in equilibrium. For $P^* > \alpha_2/(\alpha_1 + \alpha_2)$, increasing the intensive treatment above its equilibrium value will have a positive effect on expected utility, whereas for $P^* < \alpha_2/(\alpha_1 + \alpha_2)$, the opposite is true. Thus expected utility of patients is not in general maximized in equilibrium.

Note that the right-hand side of equation (7) is exactly the same as the term representing the impact of spillovers on patient choice of treatment in equation (4). Thus, if an area does more intensive treatment in equilibrium than it would in the absence of spillovers (implying that the right-hand side of eq. [7] is positive), then there are additional welfare gains from doing even more intensive treatment. Similarly, if an area does less intensive treatment in equilibrium because of spillovers, then there are additional welfare gains from doing even less. From a welfare perspective, there will be too little area variation in treatment rates if the marginal person ignores the externality that their treatment decision has on others. Said a different way, if physicians maximize social welfare (rather than patient welfare) and take this externality into account in making treatment decisions, then they will optimally do some harm (in a utility sense) to the marginal patient because of the externality on other patients, and the result will be more variation across areas in specialization.

In the case of multiple equilibria, the equilibria can be ranked in terms of expected utility. Thus shifts from one equilibrium to another can improve expected utility. Unfortunately, the model has no general

implications regarding which equilibrium will have the highest expected utility (and recall that no equilibrium is Pareto dominant without side payments). However, if all areas had identical patient distributions (Z) and the variation in treatment rates across areas arose from multiple equilibria, then one could compare patient survival and costs across areas to determine the optimal equilibrium rate of intensive treatment. This was the approach taken by Fisher et al. (2003*b*), who argue that a 30 percent reduction in Medicare spending (such that spending in high-intensity regions is reduced to that of low-intensity areas) would not have any deleterious effects on patient outcomes or satisfaction. Our model suggests that this conclusion is justified only if the observed variation represents multiple equilibria for regions with identical patient distributions. In the single-equilibrium case, our model suggests the opposite: when equilibrium treatment rates differ across regions because of differences in the underlying patient distribution, then expected patient utility in more aggressive regions would be raised by increasing—not decreasing—intensive treatment rates above their equilibrium values.

III. Data and Estimation

A. Data

We focus our empirical work on the treatment of AMI for four reasons. First, cardiovascular disease, of which AMI is the primary manifestation, is the leading cause of death in the United States. Because mortality post-AMI is high (survival rates at one year are less than 70 percent), a well-defined endpoint is available to test the efficacy of alternative interventions. Second, markets for heart attack treatment are geographically distinct. The acute nature of this condition requires immediate treatment and generally precludes a patient from traveling long distances to seek care. Therefore, mobility is limited, and it is possible to observe production in many distinct local markets. Third, the treatment of heart attacks is characterized by the choice between two competing interventions, both of which restore blood flow to the coronary arteries: intensive management characterized by the use of angioplasty or bypass and nonintensive management characterized by the use of fibrinolytics therapy (also known as thrombolytics), which dissolve clots that may have formed in a blood vessel. We measure the use of intensive therapy by focusing on the use of cardiac catheterization since it is a well-understood marker for surgically intensive management of patients (our decision follows McClellan et al. [1994] and McClellan and Newhouse

[1997]).[6] Patients who received bypass or angioplasty are included in the set of persons receiving catheterization and, therefore, intensive treatment. Fourth, some geographic areas specialize in the intensive management of heart attacks, whereas others specialize in nonintensive management. But in contrast to other markets, neither production choice can completely dominate a market because some patients are always more appropriate for a particular intervention (e.g., 95-year-old patients do not benefit from surgery and must be treated noninvasively). Thus the productivity of both intensive and nonintensive management can be observed in all markets.

Regardless of whether a patient is treated intensively or using thrombolytics, patients should also be prescribed beta-blockers during their hospital stay, which reduce the uptake of adrenalin and thereby slow the heart. Beta-blockers have been shown to improve outcomes for the majority of patients, their use is substantially below what most experts believe is appropriate, and the rate of beta-blocker use among AMI patients is a widely used measure of the quality of medical care (Jencks et al. 2000; Baicker and Chandra 2004). Because beta-blockers are a form of medical management of patients, their use serves as a marker of the quality of nonintensive medical management in an area.

Because AMI is both common and serious, it has been the topic of intense scientific and clinical interest. One effort to incorporate evidence-based practice guidelines into the care of heart attack patients, begun in 1992, is the Health Care Financing Administration's Health Care Quality Improvement Initiative Cooperative Cardiovascular Project (CCP). The CCP developed quality indicators that were based heavily on clinical practice guidelines developed by the American College of Cardiology (ACC) and the American Heart Association (AHA). Information about more than 200,000 patients admitted to hospitals for treatment of heart attacks was obtained from clinical records. The CCP is considerably superior to administrative data (of the type used by McClellan et al. [1994]) since it collects chart data on the patients: detailed information is provided on laboratory tests, the location of the myocardial infarction, and the condition of the patient at the time of admission. In the Data Appendix we provide a detailed account of the estimation sample used in this paper.

[6] In performing cardiac catheterization, a cardiologist inserts a thin plastic tube (catheter) into an artery or vein in the arm or leg, from where it is advanced into the chambers of the heart and into the coronary arteries. The catheter measures the blood's oxygen saturation and blood pressure within the heart. It is also used to get information about the pumping ability of the heart muscle. Catheters are also used to inject dye into the coronary arteries, which can then be imaged to assess arterial stenosis using an x-ray. The procedure in which catheters with a balloon on the tip are inflated in order to compress the atherosclerotic plaque to improve blood flow is referred to as angioplasty.

B. Defining Geography

To construct local markets for health care, we exploit insights from the *Dartmouth Atlas of Health Care*. The atlas divides the United States into 306 hospital referral regions (HRR), with each region determined at the zip code level by the use of an algorithm reflecting commuting patterns and the location of major referral hospitals. The regions may cross state and county borders because they are determined solely by migration patterns of patients. For example, the Evansville, Indiana, HHR encompasses parts of three states because it draws patients so heavily from Illinois and Kentucky. HRRs are best viewed as the level at which tertiary services such as cardiac surgery are received (although they are not necessarily the appropriate geographical level for primary care services).

Analysis at the HRR level is preferable to analysis at the city or state level since it uses the empirical pattern of patient commuting to determine the geographic boundaries of each referral region rather than assuming that the arbitrary political boundaries of states and cities also define the level at which the health care is delivered. Furthermore, for the purpose of studying geographic productivity spillovers, an analysis at the HRR level is superior to one at the level of an individual hospital for two reasons. First, patients can be assigned to an HRR on the basis of their residence rather than the hospital at which they received treatment (which may be endogenous). In addition, geographic productivity spillovers are likely to operate at a broader level than that of a given hospital; for example, these spillovers are expected to reach beyond the boundary of the firm to affect productivity of all firms in a region. Physicians often have operating privileges in multiple hospitals and interact (socially and professionally) with other doctors who may or may not practice in their hospital, and patients are commonly referred to other hospitals within the HRR for treatment. Through such interactions, the entire area (as measured by HRRs) is more appropriate for analyzing equilibrium implications of productivity spillovers.[7]

[7] As an alternative to the use of HRRs we have also reestimated our analysis using the U.S. Census Bureau's metropolitan statistical area (MSA) designation of urban areas. In general, HRRs may be thought of as encompassing the relevant MSA with the addition of surrounding rural areas whose patients travel to the MSA to seek care. Using the MSA sample results in a loss of sample size and introduces some noise into the instrumental variable analysis since within an MSA (such as Manhattan or Boston) there is much less variation in differential distance. The results reported in this paper are robust to the use of MSAs, with the exception of the instrumental variable estimates, which, while not statistically different from what we report, are very imprecise.

C. Estimation

The key estimating equations in this paper take the following form:

$$\text{Outcome}_{ijk} = \beta_{0k} + \beta_{1k}\text{Intensive Treatment}_i + \boldsymbol{X}_i\Pi_k + u_{ijk}. \quad (8)$$

Here, Outcome_{ijk} refers to either survival or costs for patient i in HRR j; k is an indicator variable that indexes the different groups for whom the effect of intensive management is sought (e.g., patients appropriate for intensive management vs. not or, alternatively, high-intensity HHRs vs. low-intensity HRRs). Following McClellan et al. (1994) and McClellan and Newhouse (1997), Intensive Treatment is measured by the receipt of cardiac catheterization (including angioplasty or bypass) within 30 days of the AMI. Alternatively, in some specifications we use total spending (Medicare Part A and Part B charges) in the first year after the AMI as a proxy for intensive treatment. The vector $\boldsymbol{X}_i$ includes the entire set of CCP controls listed in the Data Appendix. These variables provide a relatively complete summary of the patient's condition at the time of admission, that is, include all the relevant clinical information available to the physician in the patient's chart at the time of the heart attack. Our tables report estimates of β_{1k} and their difference between different subsamples of the data; the latter is the central parameter of interest for the purpose of testing our model. Standard errors are clustered at the level of each area.

Both our model and common sense suggest that the choice of intensive treatment for a patient will be endogenous. Even though we have excellent information on the patient's clinical condition at admission, the attending physician or cardiologist is likely to make the treatment choice on the basis of information that is not observable in the CCP (e.g., using information observed in the weeks following the initial admission). In particular, the selection problem that confounds ordinary least squares (OLS) estimation of the above equation is that intensive treatment is recommended to patients who will benefit most, and these patients are typically in better health (e.g., did not die in the first 24 hours after the heart attack). This selection of healthy patients into treatment biases OLS estimates toward finding a large effect of intensive treatment. We follow the work of McClellan et al. (1994) and estimate equation (8) using instrumental variables. In particular, we use differential distance (measured as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest noncath hospital) as an instrument for intensive treatment, with a negative value of differential distance indicating that the nearest hospital is a cath hospital. We capped differential distance at ± 25 miles on the basis of preliminary analysis that suggested little effect

of differential distances beyond 25 miles on the probability that a patient receives catheterization.

In contrast to regressions in which we study an individual's receipt of cath (and in which an instrumental variable is clearly indicated), we use OLS for tests of our model that require the use of area cath rates as the key explanatory variable. Observed variation in area cath rates would have to be driven by large differences in unobservable health status across areas to bias our results. We doubt that this is true, given the prior literature documenting that population health status is not correlated with intensive treatment at the area level (e.g., Fisher et al. 2003*a*, 2003*b*). But even if it were, we would be biased toward observing a positive association between cath rates and survival, which we do not.

To define a patient's clinical appropriateness for intensive management, we estimate a logistic regression model for the probability of receiving cardiac catheterization within 30 days of the heart attack. Specifically, we estimate

$$\Pr(\text{Cardiac Cath}_{ij}) = G(\theta_0 + \theta_j + \boldsymbol{X}_i\Phi). \tag{9}$$

Here, θ_j is the risk-adjusted logistic index for the use of cardiac catheterization in HRR j ($j = 1, 2, \ldots, 306$). This equation is analogous to equation (4) in our model, with the HRR fixed effects capturing the externality that causes similar patients to be treated differently across areas. Fitted values from this regression that exclude the HRR fixed effect, $\Pr(\text{Cardiac Cath}_{ij}) = \hat{G}(\theta_0 + \boldsymbol{X}_i\Phi)$, are used as an empirical measure of clinical appropriateness for cardiac catheterization.[8] For simplicity, in much of the empirical work we split the fitted values at their median to yield two equally sized groups; those above the median are appropriate for intensive management and those below are not. In order to classify areas as being intensive versus not, we construct HRR-level (risk-adjusted) catheterization rates by recovering the θ_j from the equation above.[9]

[8] Our model suggests that omitting the HRR fixed effects will bias the coefficients on the comorbidities ($\boldsymbol{X}$) because places with more appropriate patients will tend to do more cath. However, in practice this choice is empirically irrelevant for our estimates. The correlation between our preferred measure of appropriateness (using all the CCP risk adjusters and HRR fixed effects) and one that excludes the HRR fixed effects is 0.997 (at the patient level). Similarly, the correlation between our preferred measure and simpler measures using only age-gender-race interactions is 0.75.

[9] To verify that more and less aggressive areas rank patients similarly, we estimated separate logistic regression models (by whether an area's risk-adjusted cath rate was less than or greater than the median area's rate) of a patient's probability of receiving cath on all the CCP risk adjusters. For every patient in the data, we obtained predicted values from the two models and noted that their correlation was 0.986. This high correlation suggests that areas rank people similarly.

TABLE 1
INSTRUMENTAL VARIABLE ESTIMATES OF INTENSIVE MANAGEMENT AND SPENDING ON ONE-YEAR SURVIVAL BY CLINICAL APPROPRIATENESS OF PATIENT

	INSTRUMENTAL VARIABLE ESTIMATES OF		
	Impact of Cath		
SAMPLE	On One-Year Survival (1)	On One-Year Cost ($1,000s) (2)	Impact of $1,000 on One-Year Survival (3)
A. All patients (N = 129,895)	.142	9.086	.016
	(.036)	(1.810)	(.005)
B. By cath propensity:			
Above the median (N = 64,799)	.184	4.793	.038
	(.034)	(1.997)	(.017)
Below the median (N = 65,096)	.035	17.183	.002
	(.083)	(3.204)	(.005)
Difference	.149	−12.39	.036
	(.090)	(3.775)	(.018)
C. By age:			
65–80 (N = 89,947)	.171	6.993	.024
	(.037)	(1.993)	(.009)
Over 80 (N = 39,948)	.016	16.026	.001
	(.108)	(2.967)	(.007)
Difference	.155	−9.033	.023
	(.114)	(3.574)	(.011)

NOTE.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. Differential distance (measured as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital) is the instrument. Each model includes all the CCP risk adjusters, and the standard errors are clustered at the level of each HRR.

IV. Empirical Results

Our empirical results are organized into two sections. We begin by testing the basic implications of the Roy model—that patients are sorted into surgery versus medical management on the basis of the returns to each type of treatment. We then turn to testing the model's implications that depend on the presence of productivity spillovers.

A. Testing Implications of the Basic Roy Model

Table 1 presents instrumental variable estimates of the impact of cath (as a marker for intensive treatment) on one-year survival and one-year costs (in thousands of 1996 dollars). In panel A of table 1 we present the analysis for all patients, and then the remaining panels estimate separate regressions by the clinical appropriateness of the patient. In column 3 we instrument costs with differential distance to estimate the cost-effectiveness of spending on survival. The results in table 1 suggest that the returns to more intensive management of patients are increasing in clinical appropriateness, as predicted by our model. Patients ap-

propriate for catheterization who received the intervention saw an 18.4-percentage-point survival gain relative to appropriate patients who did not receive intensive therapy. For the less appropriate group, the effect of receiving catheterization is not statistically different from zero.[10] Column 2 of the table demonstrates that it also costs more to aggressively treat a patient who is clinically inappropriate for intensive management. One reason for this finding is the risk of iatrogenic (physician-induced) complications that require additional rehabilitation days in the intensive care unit or cardiac care unit. Column 3 uses total dollars spent on a patient as a measure of intensity. By this measure, increasing intensity (by spending an additional $1,000 on a heart attack patient) raises survival by 3.8 percentage points for patients who are appropriate for intensive management, but by nothing for patients who are not good candidates for intensive management. Panel C of table 1 provides an alternative breakdown of patients into appropriateness groups based on age (65–80 vs. over 80). Clinical guidelines recommend that older patients not be treated surgically, suggesting that the returns to such treatment should be negligible for the over-80 group. Panel C of table 1 confirms that this is indeed the case. Together, the results in table 1 provide support for our Roy model characterization of specialization: patients most appropriate for intensive treatment benefit most from such therapy, whereas those least appropriate for it do not benefit from it.

Table 2 explores the validity of differential distance as an instrument in our sample. Following McClellan et al. (1994), we split the sample in half and compare average characteristics of the sample above and below the median differential distance (−2.0 miles). Columns 1 and 2 show that among all patients, there is a 6.1-percentage-point difference in the cath rate between the samples above and below the median, with higher differential distance to a cath hospital associated with lower rates of cath. A similar pattern is seen when the sample is restricted to patients who are most or least appropriate for cath on the basis of their propensity score or their age. Even among the patients with low propensity scores or over age 80, there is a three to four percentage-point decline associated with being further from a cath hospital. These differences are all highly significant, even after we control for the full set of patient

[10] Following suggestions made by our anonymous reviewers, we note that our results are robust to controlling for distance to the nearest hospital, as well as controlling for the distance to the nearest hospital and rural classification of the zip code of the residence. These results are available on request. Briefly, after we control for actual distance to the nearest hospital, the coefficient (and accompanying standard error) on survival for appropriate patients was 0.182 (0.033); for those less appropriate it was 0.051 (0.081). After we controlled for distance and rural location, these coefficients changed to 0.181 (0.046) and 0.084 (0.13), respectively.

TABLE 2

Relationship between Differential Distance (DD) and Probability of Catheterization and Survival, and Differential Distance and Observable Characteristics (%)

	30-Day Cath Rate		One-Year Survival		One-Year Predicted Survival		30-Day Predicted Cath Rate for Patients Getting Cath	
Sample	DD Below Median (1)	DD Above Median (2)	DD Below Median (3)	DD Above Median (4)	DD Below Median (5)	DD Above Median (6)	DD Below Median (7)	DD Above Median (8)
All patients (N = 129,997)	48.9	42.8	67.6	66.7	67.5	67.2	63.3	63.2
By cath propensity:								
Above the median (N = 64,733)	74.0	67.1	84.6	83.8	83.4	83.5	72.6	72.6
Below the median (N = 65,244)	22.9	19.5	50.1	50.4	51.1	51.6	32.3	32.5
By age:								
65–80 (N = 90,016)	61.1	54.9	74.3	73.5	73.9	73.9	67.4	67.3
Over 80 (N = 39,961)	20.3	16.5	52.1	52.1	52.6	52.7	34.6	34.1

Note.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. Differential distance is measured as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital.

controls from the CCP (the first-stage *F*-statistics on differential distance are over 50 for all specifications reported in table 1).

Columns 3 and 4 of table 2 report survival rates for patients above and below the median differential distance. If differential distance is unrelated to patient mortality risk, then these estimates can be combined with the difference in cath rates from columns 1 and 2 to form a simple Wald estimate of the effect of cath on survival.[11] Among all patients, there is a 0.9-percentage-point decline in survival associated with being further from a cath hospital, suggesting that the intensive treatment associated with cath leads to improved survival. The Wald estimate suggests that cath is associated with a 15-percentage-point improvement in survival (0.009 gain in survival/0.061 increase in cath), an estimate quite close to the 14.2 percent estimate from table 1 that included the full set of patient controls. The remaining rows show results for other patient samples that are similarly consistent with the table 1 estimates, with a larger association between differential distance and survival seen among patients who are more appropriate for cath and no relationship seen among patients who are less appropriate for cath. Given the similarity of the Wald estimates to instrumental variable estimates with a full set of patient controls, it appears that differential distance is uncorrelated with observable differences in mortality risk. Columns 5 and 6 show that, as expected, there is little difference in the average one-year predicted survival rate (from a logit of survival on the patient controls) for patients above and below the median differential distance.

Columns 7 and 8 of table 2 compare average 30-day predicted cath rates (the propensity to get cath) for only those patients getting cath in the areas above and below median differential distance. If the additional patients getting cath in the low–differential distance sample were less appropriate for cath, we would expect to see that the average patient getting cath in these areas would have a lower propensity. In contrast, we see little difference in the sample that is nearer to a cath hospital. Thus it appears that differential distance is an instrument that increases cath rates among a sample of patients who are very similar (at least on observable factors) to the average patient being treated. Therefore, it appears reasonable to interpret the instrumental variable coefficients as estimates of the treatment effect in the treated population.

Finally, the Roy model predicts that as areas increase their intensity, the additional patients receiving the intensive management should be

[11] Further evidence that the differential distance instrument is uncorrelated with other determinants of receiving intensive treatment is provided in McClellan et al. (1994), MeClellan and Newhouse (1997), and McClellan and Noguchi (1998).

TABLE 3
RELATIONSHIP BETWEEN THE AVERAGE AND MARGINAL PATIENT RECEIVING CARDIAC CATHETERIZATION ($N = 303$)

Patient Characteristic	Characteristic of Average Patient Getting Cath across All Areas (1)	Difference between Marginal Patient and Average Patient Getting Cath in Higher-Cath HRRs (2)
Cath propensity	.633	−.045
	(.002)	(.008)
Over age 80	.125	.063
	(.002)	(.012)
Not eligible for cath using ACC/AHA guidelines	.028	.010
	(.001)	(.003)

NOTE.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. The sample is restricted to patients receiving cardiac catheterization within 30 days of an AMI. ACC/AHA guidelines reflect a binary variable assigned to each patient in the CCP that measures whether the patient is ideal, appropriate, or not eligible for catheterization on the basis of a review of the patient's chart.

less clinically appropriate for the intervention. In contrast, a noneconomic model would predict that intensive areas simply perform more intensive treatments on all patients; there is no triage of patients into the treatment based on their appropriateness. To test this insight we estimate

$$\text{Appropriateness}_{ij} = \mu_0 + \mu_1 \ln (\text{Cath Rate})_j + e_i. \tag{10}$$

The dependent variable is a measure of the appropriateness of a patient for cardiac cath (e.g., the propensity score), and the equation is estimated at the patient level *only for individuals receiving cath.* The explanatory variable of interest is the natural logarithm of the risk-adjusted area cath rate. As in the logic of Gruber, Levine, and Staiger (1999), the coefficient μ_1 measures the difference in mean appropriateness between the marginal and the average patient receiving the invasive treatment. If average appropriateness among patients getting cath is declining as the cath rate rises ($\mu_1 < 0$), we infer that the marginal patient getting cath was less appropriate than the average patient (analogously to deriving marginal cost curves from average cost curves).

The results of estimating equation (10) are reported in table 3. As the area's intensity increases, the marginal patient is 4.5 percentage points less likely to be appropriate for intensive treatments.[12] As a robustness check, we included two alternative measures of clinical appropriateness: age and a measure of clinical *ineligibility* for cath as defined

[12] Baicker, Buckles, and Chandra (2006) use this technique to demonstrate an identical result for geographic variation in the use of cesarean delivery. In their analysis, counties with higher cesarean rates are shown to perform the procedure in births that are less appropriate for it.

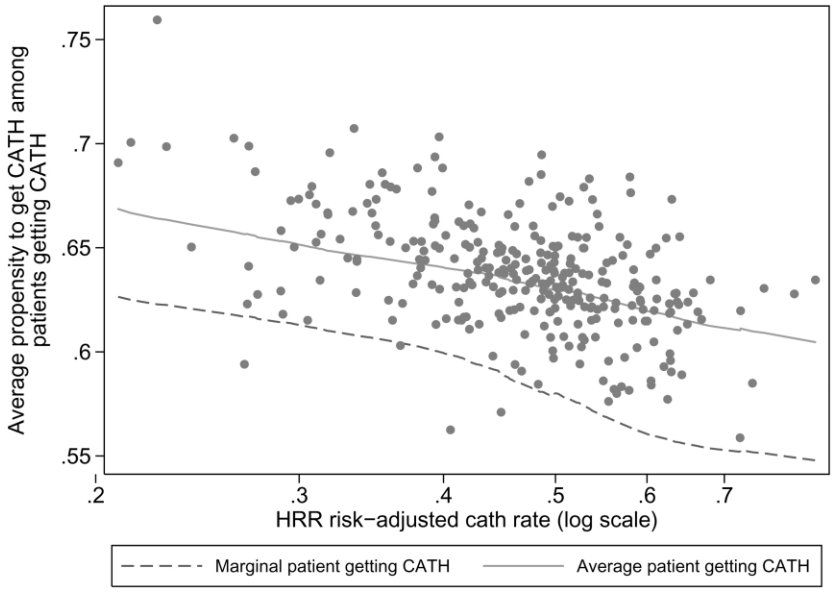


FIG. 3.—Relation between average patient and marginal patient receiving cardiac catheterization. For each of the 306 HRRs we graph the average propensity to receive cardiac catheterization (among patients who actually received it) against the log of the area risk-adjusted cath rate. Using local regression, we estimated the relationship between the average propensity and the risk-adjusted cath rate and the slope of this line at each point. These estimates were then used to plot the average (upper line) and marginal patient (lower line and estimated as the local difference in the average) receiving treatment.

by the ACC and the AHA. Table 3 indicates that by both measures, the marginal patient is significantly less likely to be appropriate for cardiac catheterization.[13]

Figure 3 provides a graphical illustration of this relationship. For each of the 306 HRRs, we graph the average propensity to receive cardiac catheterization (among patients who actually received it) against the log of the area risk-adjusted cath rate. Using local regression, we estimated the relationship between the average propensity and the risk-adjusted cath rate and the slope of this line at each point (which we also smoothed). These estimates were then used to plot average and

[13] Both measures are included as covariates in our estimation of the empirical appropriateness for cardiac catheterization. As such, these results are not three separate confirmations of the same prediction. The ACC/AHA measures are based on an evaluation of the patient's chart characteristics and classify each patient as being ideal, appropriate, or inappropriate for catheterization; see Scanlon et al. (1999) for details. This measure, however, is not our preferred measure because, as others have noted, expert panels have been shown to exhibit enormous variability, particularly for the use of procedures classified as inappropriate, and to be greatly influenced by the composition of the panel (Shekelle et al. 1998).

TABLE 4
HRR-Level Measures of Intensive Treatment, Medical Management, Support of Medical Treatment, and Demographic Characteristics

HRR Indicator	Mean	Standard Deviation	10th Percentile	90th Percentile	Correlation with HRR Cath Rate
Measures of intensive treatment:					
Risk-adjusted 30-day cath rate	46.3%	9.1%	34.5%	58.3%	1.00
Risk-adjusted 30-day PTCA rate	17.7%	5.1%	11.3%	23.6%	.81
Risk-adjusted 30-day CABG rate	13.4%	2.9%	10.2%	16.9%	.51
Risk-adjusted 12-hour PTCA rate	2.7%	2.6%	.6%	5.8%	.52
Measures of quality of medical management:					
Risk-adjusted beta-blocker rate	45.6%	9.5%	34.2%	58.3%	−.31
Support for intensive treatment:					
Cardiovascular surgeons per 100,000	1.06	.27	.70	1.40	.33
Cath labs per 10,000	2.40	.76	1.50	3.30	.39
Demographic characteristics:					
Log of resident population	13.96	.89	12.72	15.18	−.05
Log of per capita income	9.55	.20	9.31	9.85	.02
Percent college graduates	19.3%	5.5%	13.1%	26.6%	−.05

Note.—HRR surgical and medical intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath or beta-blockers on HRR fixed effects and CCP risk adjusters.

marginal patients receiving treatment. As seen in the figure, average appropriateness of patients getting cath declines in more intensive areas. The average appropriateness can decline only if the marginal patient is less appropriate: as the lower line in figure 3 indicates, the appropriateness of the marginal patient appears to be below that of the average patient and also declines as areas become more surgically intensive.

B. Testing Implications of Productivity Spillovers

The empirical results in tables 1–3 provide support for the basic assumptions of the Roy model: that patients are sorted into surgery versus medical management on the basis of the returns to each type of treatment. The remaining tables focus on testing for the presence of productivity spillovers. Table 4 tests the prediction that the quality of medical management is worse in areas that are surgically intensive. To measure the extent of intensive treatment in an area, we use the risk-adjusted 30-day cath rate in each HRR. The risk-adjusted cath rate reflects variation across areas in the probability that observationally identical patients will receive cath and therefore measures the productivity externality through its role of increasing the probability of receiving the

TABLE 5
OLS RESTIMATES OF THE RELATIONSHIP BETWEEN PROBABILITY OF RECEIVING CATHETERIZATION AND HRR PATIENT CHARACTERISTICS
($N = 138,873$)

HRR-LEVEL INDEPENDENT VARIABLE	PROBABILITY OF RECEIVING CATHETERIZATION	
	(1)	(2)
Average propensity to get cath	.529	.575
	(.172)	(.167)
Percent under age 65		.150
		(.135)
Log of resident population		−.003
		(.005)
Log of per capita income		.024
		(.024)

NOTE.—The table reports OLS estimates of the relationship between a patient receiving catheterization and the average appropriateness for catheterization in an HRR. Regressions control for patient risk adjusters, and standard errors are clustered at the level of HRRs.

intensive treatment. To measure the quality of medical care in an area, we use the risk-adjusted rate at which patients received beta-blockers in the HRR. Use of beta-blockers is a widely used marker for the quality of medical care. Not surprisingly, the risk-adjusted cath rate is positively correlated with other risk-adjusted rates of intensive surgical treatment such as bypass (CABG) and angioplasty (PTCA). More important, the negative correlation between the risk-adjusted cath and beta-blocker rates supports the view that the quality of medical management is worse in surgically intensive areas.

The remaining rows of table 4 report the correlation between risk-adjusted cath rates and other area-level characteristics of interest. Risk-adjusted cath rates are positively associated with cardiovascular surgeons per capita (physicians who perform cardiac surgery) and the number of cath labs per capita. These correlations are consistent with the view that a higher level of support services available in high-intensity areas may contribute to the externalities. The last three rows of this table demonstrate, perhaps surprisingly, that risk-adjusted cath rates are not strongly associated with demographic characteristics of the HRR such as population, income, or education.

A more fundamental implication of productivity spillovers is that the characteristics of other patients in the population will influence the treatment choice of an individual: all else equal, a patient will be more likely to receive cath if she lives in an area in which the *average* patient is more appropriate for cath. In column 1 of table 5 we regress a patient's receipt of cath on the average propensity to get cath in each HRR and patient risk adjusters. The coefficient is statistically significant and implies that for every one-percentage-point increase in the expected area

TABLE 6
INSTRUMENTAL VARIABLE ESTIMATES OF INTENSIVE MANAGEMENT AND SPENDING ON SURVIVAL, BY SURGICAL INTENSITY OF HOSPITAL REFERRAL REGION

	INSTRUMENTAL VARIABLE ESTIMATES OF		
	Impact of Cath		
SAMPLE	On One-Year Survival (1)	On One-Year Cost ($1,000s) (2)	Impact of $1,000 on One-Year Survival (3)
A. All patients:			
HRR risk-adjusted cath rate:			
Above the median (N = 63,771)	.256	6.691	.038
	(.061)	(3.510)	(.021)
Below the median (N = 66,124)	.09	9.835	.009
	(.059)	(3.155)	(.007)
Difference	.166	−3.144	.029
	(.085)	(4.720)	(.022)
B. Patients above the median cath propensity:			
HRR risk-adjusted cath rate:			
Above the median (N = 32,388)	.271	.347	.78
	(.064)	(4.370)	(9.820)
Below the median (N = 32,411)	.168	4.962	.034
	(.046)	(2.890)	(.021)
C. Patients below the median cath propensity:			
HRR risk-adjusted cath rate:			
Above the median (N = 31,383)	.206	16.21	.013
	(.129)	(5.130)	(.009)
Below the median (N = 33,713)	−.139	22.064	−.006
	(.165)	(6.870)	(.007)

NOTE.—HRR intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath on HRR fixed effects and CCP risk adjusters. Differential distance (measured as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital) is the instrument. Each model includes all the CCP risk adjusters, and the standard errors are clustered at the level of each HRR.

cath rate (based on patient characteristics), there is an additional 0.53-percentage-point increase in the probability that an individual receives cath because of spillovers. In column 2 we control for area demographics that could potentially confound this relationship, but they are insignificant and do not materially change the coefficient.

A key implication of our model (and any model in which productivity differences across areas generate specialization) is that the return to intensive management should be higher in high-intensity areas than in low-intensity ones. This prediction provides a sharp test of our model against the flat of the curve model, which predicts the opposite. To test this prediction, we use our estimates of HRR-level intensity from the estimation of equation (9) to classify patients as being treated in high- or low-intensity regions (as measured by whether the risk-adjusted cath rate is above or below the median rate). In table 6, we report instru-

mental variable estimates of the effect of receiving intensive management within each of these two groups. The results of this analysis are reported in a manner that mirrors the estimates in table 1. The survival return to intensive management in intensive areas is roughly three times the return observed in low-intensity areas; there is no statistically significant difference in the costs associated with the different areas.[14] In panels B and C of table 6, we further split the sample within high- and low-intensity regions according to whether patients were more or less appropriate for cath (on the basis of a cath propensity above or below the median). As predicted by our model, the most appropriate patients in high-intensity areas have the highest survival returns (and lowest costs) associated with intensive management, whereas the lowest survival returns (and highest costs) are seen among the less appropriate patients in low-intensity areas.

Table 7 tests a unique implication of productivity spillovers: Patients most appropriate for intensive treatments are better off being treated in high-intensity areas, whereas patients who are least appropriate for intensive treatments should be worse off in such areas. To test this prediction, we split patients into three equal-sized groups of appropriateness for intensive treatments. Within each group, we report the relationship between survival and the area cath rate in column 1. We also report the relationship between spending and the area rate in column 2. Panel A of table 7 replicates the central finding from the area variations literature: area intensity is associated with costs but not with improved outcomes among patients as a whole. However, as panel B demonstrates, this finding masks significant heterogeneity by patient appropriateness. Patients appropriate for intensive management clearly benefit from being treated in intensive areas. However, as the productivity externality predicts, patients least appropriate for intensive treatments are harmed as a result of being treated in intensive areas. The size of the negative externality is large: increasing area cath intensity by 10 percentage points (0.1) would reduce the survival of patients least appropriate for cath by 0.75 percentage points. This finding holds true (although is less significant) when we split the data by alternative indicators of appropriateness for intensive treatment, including age and ACC/AHA guidelines.[15]

[14] Analogous to n. 10, controlling for distance to the nearest hospital barely changed our coefficients. In these specifications, the instrumental variable returns to cath for more and less intensive areas were 0.264 (0.059) and 0.094 (0.058), respectively. Further controlling for rural designation of a patient's zip code yielded estimates of 0.301 (0.098) and 0.111 (0.077), respectively.

[15] In results available from the authors, we have verified that our results are similar (though less precise) in simpler specifications in which we replace the risk-adjusted HRR rate with the difference between actual and expected rates based on only age, race, and gender. For patients under age 80, the coefficient on the alternatively defined HRR rate is 0.021 (0.029), whereas for those over the age of 80, it is −0.041 (0.038).

TABLE 7
RELATIONSHIP BETWEEN HRR CATHETERIZATION RATE, SURVIVAL, AND COSTS, BY CLINICAL APPROPRIATENESS FOR INTENSIVE MANAGEMENT

	OLS ESTIMATES OF THE RELATIONSHIP BETWEEN HRR RISK-ADJUSTED CATH RATE AND			
SAMPLE	One-Year Survival (1)	One-Year Cost ($1,000s) (2)	Beta-Blocker in Hospital (3)	Catheterization within 30 Days (4)
A. All patients (N = 138,873)	.007	8.093	−.28	.702
	(.019)	(1.410)	(.073)	(.004)
B. By cath propensity:				
Top tercile (N = 46,287)	.052	10.012	−.366	.802
	(.019)	(1.439)	(.073)	(.032)
Middle tercile (N = 46,295)	.03	11.154	−.271	.906
	(.030)	(1.784)	(.082)	(.021)
Bottom tercile (N = 46,291)	−.075	2.763	−.209	.369
	(.028)	(1.612)	(.073)	(.021)
Difference (top − bottom)	.127	7.249	−.157	.433
	(.034)	(2.161)	(.103)	(.038)
C. By age:				
65–80 (N = 96,093)	.023	9.616	−.311	.775
	(.021)	(1.448)	(.072)	(.012)
Over 80 (N = 42,780)	−.031	4.738	−.215	.531
	(.028)	(1.603)	(.080)	(.022)
Difference (top − bottom)	.054	4.878	−.096	.244
	(.035)	(2.160)	(.108)	(.025)
D. By AHA/ACC criterion:				
Ideal (N = 89,569)	.027	9.845	−.302	.769
	(.023)	(1.599)	(.076)	(.010)
Appropriate (N = 31,800)	−.002	6.174	−.282	.752
	(.024)	(1.537)	(.080)	(.026)
Not appropriate (N = 17,504)	−.08	2.958	−.177	.264
	(.040)	(1.511)	(.065)	(.021)
Difference (top − bottom)	.107	6.887	−.125	.505
	(.046)	(2.200)	(.100)	(.023)

NOTE.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. HRR surgical and medical intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath or beta-blockers on HRR fixed effects and CCP risk adjusters.

Our model suggests that patients who are inappropriate for cath are worse off in intensive areas because (1) the quality of medical management in these areas is worse than in other areas, and (2) few of these patients receive intensive treatment, even in the more intensive areas. Columns 3 and 4 of table 7 explore these two dimensions of care directly by estimating the relationship between area cath intensity and the rate at which patients receive beta-blockers (as an indirect marker of quality of medical management) and cath (as a direct marker of intensive management). Use of beta-blockers is lower among all patient groups in the high-intensity areas, suggesting that quality of medical management is generally poor in these areas. At the same time, cath rates among those patients least appropriate for cath rise much less in high-intensity areas

than for other patients. Thus worse outcomes for this group appear to result from worse medical management in areas with high cath intensity, as our model predicts.

Finally, our analysis has assumed that productivity spillovers depend on the proportion of patients receiving a given treatment rather than on the absolute number. Alternatively, productivity spillovers could be more prevalent in larger areas; larger HRRs such as Los Angeles or Manhattan may excel at both the intensive and nonintensive delivery of care. To explore this hypothesis further, we estimate the relationship between alternative measures of HRR size and patient survival in table 8. We regress one-year survival on patient risk adjusters, the risk-adjusted cath rate, and the log of the resident population (cols. 1–4) and log of AMIs per hospital (cols. 5–8). Table 8 demonstrates that the size measures are largely insignificant: larger areas do not result in improved survival. On the other hand, the "externality" (as measured by the HRR cath rate) is always protective of patients who are appropriate for cath and harmful for inappropriate patients.

V. Conclusion

A very simple equilibrium model of patient treatment choice with geographic productivity spillovers appears to rationalize the main stylized facts concerning variation across areas in the use of technologically intensive medical care. The model yields a range of additional empirical implications, for which we found uniform support in our analysis of treatment for heart attacks. Alternative models, such as those based on "flat of the curve" medicine, have fundamentally different implications that are clearly rejected by the data. Thus there appears to be strong empirical support for the presence of productivity spillovers in medical care.

Our findings suggest that productivity spillovers play an important role in explaining geographic specialization in production, as first proposed by Marshall (1890). A large medical literature has documented the important role of social networks in physician adoption of new technologies, suggesting that knowledge externalities are the source of the productivity spillovers. While knowledge spillovers are the most natural interpretation of our model and empirical results, alternative mechanisms, such as the migration of specialized inputs into regions that favor one type of production, could generate the productivity spillovers we observe. These alternative mechanisms share the key feature that specialization in one sector improves productivity in that sector while reducing productivity elsewhere, thereby reinforcing the tendency to specialize.

The presence of productivity spillovers may justify a policy interven-

TABLE 8

Relationship between Survival and Alternative Measures of HRR Size, by Clinical Appropriateness of Patient

HRR-Level Independent Variable	All Patients (1)	By Cath Propensity: Top Tercile (2)	By Cath Propensity: Middle Tercile (3)	By Cath Propensity: Bottom Tercile (4)	All Patients (5)	By Cath Propensity: Top Tercile (6)	By Cath Propensity: Middle Tercile (7)	By Cath Propensity: Bottom Tercile (8)
Risk-adjusted cath rate	.006	.052	.031	−.077	.008	.053	.031	−.075
	(.019)	(.019)	(.030)	(.028)	(.019)	(.019)	(.030)	(.028)
Log of resident population	−.002	−.001	.002	−.007				
	(.002)	(.002)	(.003)	(.003)				
Log of AMIs per hospital					.005	.014	.006	−.007
					(.004)	(.005)	(.007)	(.006)
Observations	138,873	46,287	46,295	46,291	138,873	46,287	46,295	46,291

Note.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. HRR surgical and medical intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath on HRR fixed effects and CCP risk adjusters.

tion to ameliorate the deleterious effects of the externality. However, evaluating the welfare implications of any proposed intervention is not a trivial exercise. With multiple equilibria we may be able to identify areas stuck in a suboptimal equilibrium, and one-time interventions that "shock" the system to a different equilibrium would be called for (see, e.g., O'Connor et al. 1996). While other commentators (Phelps 2000; Fisher et al. 2003*a*, 2003*b*) have emphasized the welfare cost of excess geographic variation in practice style, our theoretical model (with a single equilibrium) suggests that there is too little variation from a welfare perspective: if aggressive areas became even more aggressive, average patient welfare would increase. Thus understanding the optimal policy response hinges importantly on whether the variations observed in the data are the consequence of single or multiple equilibria.

Our results also raise important questions about what can be learned from randomized controlled trials in medicine. While randomized trials are considered the gold standard for determining the effectiveness of a given medical treatment, they are designed to provide a partial equilibrium estimate of the treatment effect in a well-defined population. But with productivity spillovers, the general equilibrium effect of adopting a new treatment could be smaller or larger than the partial equilibrium estimate of treatment effectiveness because of the negative externality imposed on patients who are more appropriate for an alternative treatment and the positive externality on patients who are more appropriate for the new treatment. This general equilibrium effect is not identified in a typical randomized trial, but could potentially be estimated in a trial that randomized across areas rather than individuals. In addition, the effectiveness of the treatment depends on where the trial is conducted (just as our instrumental variable estimates depended on intensity in the area): Surgical interventions may perform poorly in an area that specializes in medical management of its patients and perform well in a more surgically intensive area. As such, the external validity of a randomized trial is compromised.

The implications of our findings extend beyond the treatment of heart attacks. To the extent that productivity spillovers are a common feature of many sectors, our results provide compelling evidence that such spillovers are an important feature of geographic specialization. Moreover, our results provide some of the first direct evidence of the negative externalities imposed on a subset of the population because of equilibrium pressures toward specialization. Such negative externalities are a central component of arguments for government intervention. Finally, our model has interesting empirical implications when applied to the more general question of human capital externalities. For example, our model would suggest that people living in areas with higher-ability populations would be more likely (with ability held constant) to go to col-

lege, the return to going to college in these areas would be higher, but wages of low-ability people in these areas would be lower. In principal, these are all testable implications.

Data Appendix

A. Construction of Hospital Referral Regions

Hospital referral regions (HRRs) are constructed using a two-part algorithm.

Step 1: All acute-care hospitals in the 50 states and the District of Columbia were identified from the American Hospital Association Annual Survey of Hospitals and the Medicare Provider of Services files and assigned to a location within a town or city. The list of towns or cities with at least one acute-care hospital ($N = 3{,}953$) defined the maximum number of possible hospital service areas (HSA). Second, all 1992 and 1993 acute-care hospitalizations of the Medicare population were analyzed according to zip code to determine the proportion of residents' hospital stays that occurred in each of the 3,953 candidate HSAs. Zip codes were initially assigned to the HSA in which the greatest proportion (plurality) of residents were hospitalized. Approximately 500 of the candidate HSAs did not qualify as independent HSAs because the plurality of patients resident in those HSAs were hospitalized in other HSAs. The third step required visual examination of the zip codes used to define each HSA. Maps of zip code boundaries were made using files obtained from Geographic Data Technologies, and each HSA's component zip codes were examined. In order to achieve contiguity of the component zip codes for each HSA, "island" zip codes were reassigned to the enclosing HSA or HSAs were grouped into larger HSAs. This process resulted in the identification of 3,436 HSAs, ranging in total 1996 population from 604 (Turtle Lake, ND) to 3,067,356 (Houston) in the 1999 edition of the *Dartmouth Atlas*. Intuitively, one may think of HSAs as representing the geographic level at which "front-end" services such as diagnoses are received.

Step 2: HSAs make clear the patterns of use of local hospitals. A significant proportion of care, however, is provided by referral hospitals that serve a larger region. HRRs were defined in the *Dartmouth Atlas* by documenting where patients were referred for major cardiovascular surgical procedures and for neurosurgery. Each HSA was examined to determine where most of its residents went for these services. The result was the aggregation of the 3,436 HSAs into 306 HRRs that were named for the HSA most often used by residents of the region. Thus if a Medicare enrollee living in Hartford, CT, was admitted to a hospital in Boston, the utilization would be attributed to Hartford, and not to Boston. This assignment avoids the serious shortcoming of unusually high utilization rates in large referral centers such as Boston or Rochester, MN.

B. Construction of CCP Estimation Sample

The CCP used bills submitted by acute-care hospitals (UB-92 claims form data) and contained in the Medicare National Claims History File to identify all Medicare discharges with an International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal diagnosis of 410 (myocardial infarction), excluding those with a fifth digit of 2, which designates a subsequent episode of care. The study randomly sampled all Medicare beneficiaries with

AMI in 50 states between February 1994 and July 1995 and in the remaining five states between August and November 1995 (Alabama, Connecticut, Iowa, and Wisconsin) or April and November 1995 (Minnesota); for details, see O'Connor et al. (1999). The Claims History File does not reliably include bills for all the approximately 12 percent of Medicare beneficiaries insured through managed-care risk contracts, but the sample was representative of the Medicare fee for service patient population in the United States in the mid-1990s. After sampling, the CCP collected hospital charts for each patient and sent these to a study center, where trained chart abstracters abstracted clinical data. Abstracted information included elements of the medical history, physical examination, and data from laboratory and diagnostic testing, in addition to documentation of administered treatments. The CCP monitored the reliability of the data by monthly random reabstractions. Details of data collection and quality control have been reported previously in Marciniak et al. (1998). Finally, the CCP supplemented the abstracted clinical data with diagnosis and procedure codes extracted from Medicare billing records and dates of death from the Medicare Enrollment Database.

For our analyses, we delete patients who were transferred from another hospital, nursing home, or emergency room since these patients may already have received care that would be unmeasured in the CCP. We transformed continuous physiologic variables into categorical variables (e.g., systolic blood pressure <100 mm Hg or ≥00 mm Hg, creatinine <1.5, 1.5–2.0, or >2.0 mg/dL) and included dummy variables for missing data. We used date of death to identify patients who did or did not survive through one year after the AMI. For all patients, we identified whether they received each of six treatments during the acute hospitalization: reperfusion (defined as either thrombolysis or PCI within 12 hours of arrival at the hospital), aspirin during hospitalization, aspirin at discharge, beta-blockers at discharge, angiotensin converting enzyme inhibitors at discharge, counseling for smoking cessation, and each of three treatments within 30 days of the AMI: cardiac catheterization, PCI, or CABG. We used the ACC/AHA guidelines for coronary angiography to identify patients who were ideal (class I), uncertain (class II), or inappropriate (class III) for angiography; these details are provided in Scanlon et al. (1999).

For each AMI patient we computed Medicare Part A and Part B costs within one year by weighting all diagnosis-related groups and relative value units nationally. This measure of costs abstracts from the geographic price adjustment in the Medicare program.

C. *Construction of Clinical Appropriateness Index, HRR Catheterization Rates, and HRR Beta-Blocker Rates*

To compute this index we estimate a logistic regression model for the probability of receiving cardiac catheterization within 30 days of the heart attack. We estimate

$$\Pr(\text{Cardiac Cath}_{ij}) = F(\theta_0 + \theta_j + \boldsymbol{X}_i\Phi)$$

and

$$\Pr(\text{Beta-Blockers}_{ij}) = F(\theta_0 + \Pi_j + \boldsymbol{X}_i\Phi).$$

Fitted values from this regression, $\Pr(\text{Cath} \mid \boldsymbol{X})$, are used as an empirical measure of clinical appropriateness for cardiac catheterization. The HRR fixed effects for the use of cardiac catheterization and beta-blockers are obtained as θ_j and

TABLE A1
RISK ADJUSTERS INCLUDED IN THE MODEL

Age, race, sex (full interactions)	*hx* angina missing (ref = no)	Peak CK >1,000
Previous revascularization (1 = *y*)	*hx* terminal illness (1 = *y*)	Nonambulatory (ref = independent)
hx old MI (1 = *y*)	Current smoker	Ambulatory with assistance
hx CHF (1 = *y*)	Atrial fibrillation on admission	Ambulatory status missing
History of dementia	CPR on presentation	Albumin low (ref ≥3.0)
hx diabetes (1 = *y*)	Indicator MI = anterior	Albumin missing (ref ≥3.0)
hx hypertension (1 = *y*)	Indicator MI = inferior	Bilirubin high (ref <1.2)
hx leukemia (1 = *y*)	Indicator MI = other	Bilirubin missing (ref <1.2)
hx EF ≤40 (1 = *y*)	Heart block on admission	Creat 1.5–<2.0 (ref = <1.5)
hx metastatic cancer (1 = *y*)	CHF on presentation	Creat ≥2.0 (ref = <1.5)
hx nonmetastatic cancer (1 = *y*)	Hypotensive on admission	Creat missing (ref = <1.5)
hx PVD (1 = *y*)	Hypotensive missing	Hematocrit low (ref = >30)
hx COPD (1 = *y*)	Shock on presentation	Hematocrit missing (ref = >30)
hx angina (ref = no)	Peak CK missing	Ideal for cath (ACC/AHA criteria)

Π_j, respectively. Note that these fixed effects are not used to obtain our empirical measure of appropriateness for cardiac catheterization. The risk adjusters included in this model are listed in table A1.

The choice of variables was based on those selected by Fisher et al. (2003*a*, 2003*b*) and Barnato et al. (2005). With the exception of two variables that are both measured by blood tests, albumin and bilirubin (where the rates of missing data were 24 percent), we do not have a lot of missing data (rates were less than 3 percent). Furthermore, there is no relationship between the presence of missing values for albumin and bilirubin and the intensity of an area. Nor does having missing values for these variables affect the probability of receiving catheterization.

References

Andersen, Henning R., et al. 2003. "A Comparison of Coronary Angioplasty with Fibrinolytic Therapy in Acute Myocardial Infarction." *New England J. Medicine* 349 (August 21): 733–42.

Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-in by Historical Events." *Econ. J.* 99 (March): 116–31.

Baicker, Katherine, Kasey S. Buckles, and Amitabh Chandra. 2006. "Geographic Variation in the Appropriate Use of Cesarean Delivery." *Health Affairs* 25 (September/October): 355–67.

Baicker, Katherine, and Amitabh Chandra. 2004. "Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care." *Health Affairs* (April 7): w184–w197 (Web exclusive). http://www.healthaffairs.org.

Barnato, Amber E., F. Lee Lucas, Douglas Staiger, David E. Wennberg, and Amitabh Chandra. 2005. "Hospital-Level Racial Disparities in Acute Myocardial Infarction Treatment and Outcomes." *Medical Care* 43 (April): 308–19.

Beck, Christine A., John Penrod, Theresa W. Gyorkos, Stan Shapiro, and Louise Pilote. 2003. "Does Aggressive Care Following Acute Myocardial Infarction Reduce Mortality? Analysis with Instrumental Variables to Compare Effectiveness in Canadian and United States Patient Populations." *Health Services Res.* 38, no. 6, pt. 1 (December): 1423–40.

Coleman, James S., Elihu Katz, and Herbert Menzel. 1957. "The Diffusion of an Innovation among Physicians." *Sociometry* 20 (December): 253–70.

Cutler, David M. 2004. *Your Money or Your Life: Strong Medicine for America's Health Care System.* Oxford: Oxford Univ. Press.

Enthoven, Alain C. 1980. *Health Plan: The Only Practical Solution to the Soaring Cost of Medical Care.* Reading, MA: Addison-Wesley.

Fisher, Elliott, et al. 2003*a*. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Ann. Internal Medicine* 138 (February 18): 273–87.

———. 2003*b*. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Ann. Internal Medicine* 138 (February 18): 288–98.

Fuchs, Victor R. 2004. "More Variation in Use of Care, More Flat-of-the-Curve Medicine." *Health Affairs* (October 7). (Web exclusive). http://www.healthaffairs.org.

Glover, Allison F. 1938. "The Incidence of Tonsillectomy in School Children." *Proc. Royal Soc. Medicine* 31: 1219–36.

Goolsbee, Austan, and Peter J. Klenow. 2002. "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *J. Law and Econ.* 45, no. 2, pt. 1 (October): 317–43.

Gruber, Jonathan, Phillip B. Levine, and Douglas Staiger. 1999. "Abortion Legalization and Child Living Circumstances: Who Is the 'Marginal Child'?" *Q.J.E.* 114 (February): 263–91.

Heckman, James J., and Bo E. Honore. 1990. "The Empirical Content of the Roy Model." *Econometrica* 58 (September): 1121–49.

Jacobs, Alice K. 2003. "Primary Angioplasty for Acute Myocardial Infarction—Is It Worth the Wait?" *New England J. Medicine* 349 (August 21): 798–800.

Jencks, Stephen F., et al. 2000. "Quality of Medical Care Delivered to Medicare Beneficiaries: A Profile of State and National Levels." *J. American Medical Assoc.* 284 (October 4): 1670–76.

Katz, Michael L., and Carl Shapiro. 1985. "Network Externalities, Competition, and Compatibility." *A.E.R.* 75 (June): 424–40.

———. 1986. "Technology Adoption in the Presence of Network Externalities." *J.P.E.* 94 (August): 822–41.

Keeley, Ellen C., Judith A. Boura, and Cindy L. Grines. 2003. "Primary Angioplasty versus Intravenous Thrombolytic Therapy for Acute Myocardial Infarction: A Quantitative Review of 23 Randomised Trials." *Lancet* 361 (January): 13–20.

Marciniak, Thomas A., et al. 1998. "Improving the Quality of Care for Medicare Patients with Acute Myocardial Infarction: Results from the Cooperative Cardiovascular Project." *J. American Medical Assoc.* 279 (May 6): 1351–57.

Marshall, Alfred. 1890. *Principles of Economics.* New York: Macmillan.

McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality? Analysis

Using Instrumental Variables." *J. American Medical Assoc.* 272 (September 21): 859–66.

McClellan, Mark, and Joseph P. Newhouse. 1997. "The Marginal Cost-Effectiveness of Medical Technology: A Panel Instrumental-Variables Approach." *J. Econometrics* 77 (March): 39–64.

McClellan, Mark, and H. Noguchi. 1998. "Validity and Interpretation of Treatment Effect Estimates: Using Observational Data." Manuscript, Stanford Univ., Dept. Econ.

Meltzer, David. 2006. "Diffusion of Innovation through Social Network Connections among Physicians." Manuscript, Univ. Chicago.

O'Connor, Gerald T., et al. 1996. "A Regional Intervention to Improve the Hospital Mortality Associated with Coronary Artery Bypass Graft Surgery." *J. American Medical Assoc.* 275 (March 20): 841–46.

———. 1999. "Geographic Variation in the Treatment of Acute Myocardial Infarction: The Cooperative Cardiovascular Project." *J. American Medical Assoc.* 281 (February 17): 627–33.

Phelps, Charles E. 2000. "Information Diffusion and Best Practice Adoption." In *Handbook of Health Economics*, vol. 1A, edited by Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: Elsevier Sci.

Pilote, Louise, Olga Saynina, Frederic Lavoie, and Mark McClellan. 2003. "Cardiac Procedure Use and Outcomes in Elderly Patients with Acute Myocardial Infarction in the United States and Quebec, Canada, 1988 to 1994." *Medical Care* 41 (July): 813–22.

Rogers, Everett M. 2003. *Diffusion of Innovations*. 5th ed. New York: Free Press.

Scanlon, Patrick J., et al. 1999. "ACC/AHA Guidelines for Coronary Angiography." *J. American Coll. Cardiology* 33 (May): 1756–1824.

Shekelle, Paul G., et al. 1998. "The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures." *New England J. Medicine* 338 (June 25): 1888–95.

Soumerai, Stephen B., et al. 1998. "Effect of Local Medical Opinion Leaders on Quality of Care for Acute Myocardial Infarction: A Randomized Controlled Trial." *J. American Medical Assoc.* 279 (May 6): 1358–63.

Wennberg, David E., and J. D. Birkmeyer, eds. 2000. *The Dartmouth Atlas of Cardiovascular Health Care*. Chicago: American Hospital Pub.

Wennberg, John, and M. M. Cooper, eds. 1999. *The Quality of Medical Care in the United States: A Report on the Medicare Program, the Dartmouth Atlas of Health Care*. Chicago: American Health Assoc.