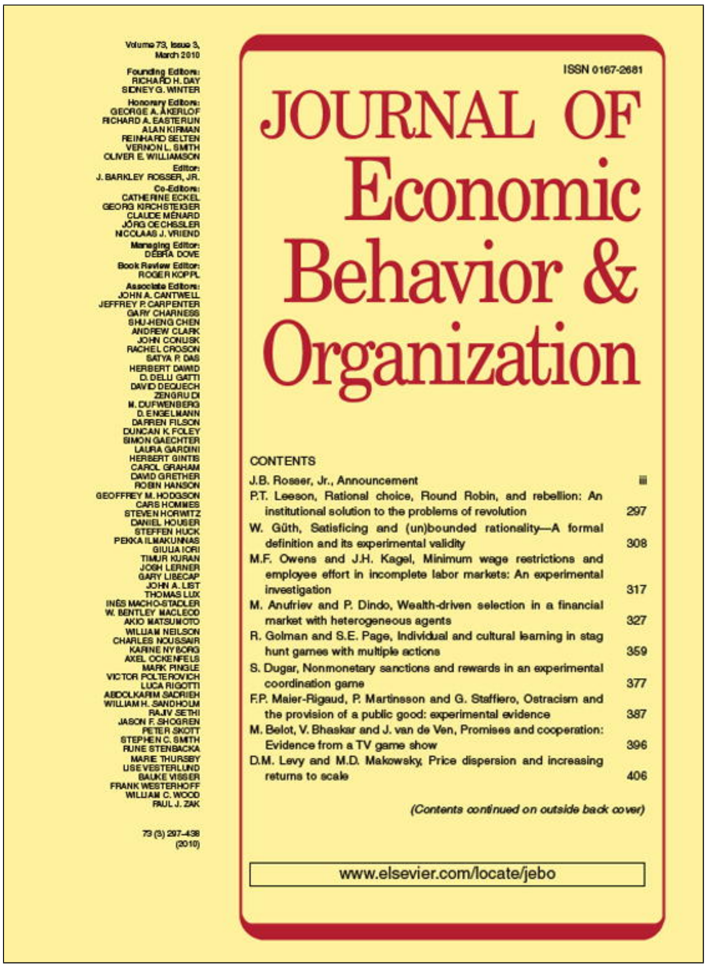

Volume 73, Issue 3,
March 2010

73 (3) 297–438
(2010)

ISSN 0167-2681

JOURNAL OF Economic Behavior & Organization

CONTENTS

(Contents continued on outside back cover)

www.elsevier.com/locate/jebo

Journal of Economic Behavior & Organization 73 (2010) 418–432

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Geographic variations in a model of physician treatment choice with social interactions

Mary A. Burke[a], Gary M. Fournier[b], Kislaya Prasad[c,*]

[a] Federal Reserve Bank of Boston, Boston, MA 02205, United States
[b] Department of Economics, Florida State University, Tallahassee, FL 32306, United States
[c] Robert H. Smith School of Business, University of Maryland, Van Munching Hall, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 2 March 2007
Received in revised form 12 October 2009
Accepted 20 November 2009
Available online 2 December 2009

JEL classification:
C73
Z13
I10
I11

Keywords:
Geographic variations in medicine
Evolutionary game theory
Interacting particle systems
Social influence
Local interactions
Medical practice guidelines

ABSTRACT

Region-specific norms of behavior are a widespread phenomenon. In the case of medical practice, numerous studies have found that geographic location exerts an inordinate influence on the choice of treatments and procedures. This paper shows how the presence of social influence on treatment decisions can help explain this phenomenon. We construct a theoretical model in which physicians' treatment choices depend on patients' characteristics and on the recent choices of nearby peers—either because there are local knowledge spillovers or because physicians want to conform to local practice patterns. In this setting, regional differences in the patient mix give rise to geographically divergent treatment patterns—the treatment a patient receives depends on where she lives. Investigation of Florida data reveals significant geographic variation in treatment rates consistent with the predictions of our model. Implications for patient welfare are explored.

1. Introduction

This paper develops a general theory of the emergence of region-specific norms of behavior and uses it to suggest a possible resolution of a famous public health puzzle—the well-documented effect of geography on the choice of medical procedures. Over the years, a number of studies have found that medical choices tend to be relatively uniform within regions and yet quite diverse across regions. Surprisingly, the phenomenon persists today despite vast improvements in communication technology. In this paper, we show that the observed geographic treatment patterns are likely to be a consequence of *local social interactions* among physicians—interactions that result in a positive correlation between the choices of a given physician and the choices of her local colleagues. We expect such a relationship to arise in the medical context for (at least) two reasons: (1) physicians may learn and acquire skills from one another, and (2) there may be social pressure to conform to local practice norms. We posit that both of these mechanisms are grounded in face-to-face encounters and therefore should operate most strongly at the local level. A model adapted from the theory of *interacting particle systems* represents

* Corresponding author. Tel.: +1 301 405 9637.
E-mail address: kprasad@rhsmith.umd.edu (K. Prasad).

0167-2681/$ – see front matter
doi:10.1016/j.jebo.2009.11.004

medical treatment choice in the presence of local social interactions and reproduces the stylized facts of the geographic variations puzzle. The theoretical model, although tailored to the requirements of the puzzle at hand, is quite general. The model predicts that, within a region, patients of all ages will tend to receive the treatment that is best suited to the patient of modal age for that region. As a result, a given patient will be treated differently in different regions, depending on the dominant age group in the region. We find strong support for these predictions among extensive data on coronary patients from Florida. We also use the model to show that patient welfare need not improve under a system of practice guidelines designed to eliminate regional treatment variations.

The extent of geographical variation in medical care in the United States is quite striking. Consider the case of two procedures used to treat heart conditions. In a comparison of hospital referral regions across the country, rates of Coronary Artery Bypass Grafting among Medicare enrollees varied by a factor of more than 3.5, while the rates of Coronary Angioplasty ranged from 2.5 to 16.9 per 1000 enrollees (see Wennberg and Cooper, 1999). Such high variation has been recorded for many procedures and treatments, spanning all areas of medicine, and often persisting over time. The systematic study of treatment variations appears to begin with Glover's (1938) presentation to the Royal Society of Medicine. In the United States, Wennberg and his colleagues have documented the phenomenon comprehensively over a number of years (e.g. Wennberg and Gittelsohn, 1973; Wennberg and Birkmeyer, 1999; O'Connor et al., 1999; Pilote et al., 1995).

The variations have proved remarkably robust to controls for incidence of illness and demographic and socioeconomic factors (Phelps and Mooney, 1993; Wennberg and Gittelsohn, 1982). Wennberg and Gittelsohn (1982) attributed geographic variations to geographic differences in physician "practice style," defined as a set of beliefs about the efficacy and appropriateness of alternative forms of care. However, inquiries into the role of practice style in treatment variations have yielded mixed results concerning its importance (e.g. contrast Grytten and Sorensen, 2003 with Folland and Stano, 1989). Phelps and Mooney (1993) present a model, based on Bayesian updating, of how regional norms might persist once they are in place, but the model does not explain how diverse norms might arise in the first place. Unlike other papers in the literature, we show how regional variations can emerge, starting from random initial conditions. More recently, Chandra and Staiger (2004, 2007) attribute medical treatment variations to local productivity spillovers. Although this explanation overlaps with our own, we show that conformity effects can just as readily explain treatment variations. Furthermore, the empirical evidence to date cannot neatly distinguish between these two sources of choice interactions.

We develop a formal model to explain how geographical variations in medical care arise, and why there may be resistance to efforts to standardize practice. In this model, two key features drive the emergence of regional treatment norms: local choice interactions among physicians and regional variation in patient characteristics. Concerning the former, we assume that the choices of a physician's nearby colleagues exert an influence on her own choices, either because of local increasing returns or because of pure conformity effects. As an example of local increasing returns, a surgeon's own expertise in a given treatment or procedure may improve as her peers gain experience in the same treatment and share their insights. If so, a given treatment will yield better outcomes, and so become increasingly favored within a group of interacting peers, the more frequently it has been used in the past. Alternatively, choice spillovers among neighboring physicians could reflect pure conformity effects. Conformity of behavior within a group may arise because individuals have an innate preference for social esteem, as in Bernheim (1994), or because of incentives in the institutional environment. For example, in the United States, malpractice claims are judged by comparing a doctor's actions to standard practices within the local medical community, thereby discouraging deviation from such practices.[1] Either form of social influence contributes to the emergence of geographical variations, but the welfare implications will be quite different between these two cases.[2]

The second key feature of the model is that patient characteristics differ, on average, across geographic regions and these characteristics influence the treatment choice. Specifically, we assume that the "ideal" treatment for a given patient—the treatment that maximizes the probability of a successful outcome independent of the local treatment history—depends on underlying traits such as age, in addition to (unmodelled) transitory symptoms. This assumption agrees with the observation, for example, that heart bypass surgery carries greater risks among the elderly than among the young and hence is less common among the former group.[3] To see how this assumption contributes to the emergence of regional treatment variations, consider what happens when a young patient arrives for treatment in a region with a high proportion of elderly patients. The young patient's physician will take the patient's age into account, but will also be drawn towards the treatment she observes her neighbors performing most often, which will likely be the treatment that is ideal for elderly patients. In the long-run, the dynamic feedback between physicians' choices and patients' types produces a stable treatment pattern in which the regional treatment norm represents the ideal treatment for the modal patient in the region. In this long-run equilibrium, patients of different types in the same region will receive the same treatment, and a patient of a given type will receive a different treatment depending on where she lives because the modal patient varies across regions.

[1] This is the dominant legal standard in the U.S. In Florida, "the prevailing professional standard of care for a given health care provider shall be that level of care, skill, and treatment which, in light of all relevant surrounding circumstances, is recognized as acceptable and appropriate by reasonably prudent similar health care providers." "Similar health care providers" are, inter alia, ones who practice in the same or similar medical community. (Florida Statutes 766.102).

[2] It is worth noting that other forms of social influence exist. In another study, of the adoption of new technologies in medicine, a significant influence of the choices of "star" physicians was found (Burke et al., 2007).

[3] See, for example, Hannan and Burke (1994) and Alexander and Peterson (1997).

We test the model's predictions using a census of Florida patient discharge records, focusing on patients over 25 with a primary diagnosis of either coronary atherosclerosis or acute myocardial infarction (AMI, or "heart attack"). The usable sample covers over 500,000 inpatient stays during the periods 1995–2001. Consistent with our predictions, we find that younger patients are less likely to get invasive heart treatments—either coronary angioplasty or heart bypass surgery—at hospitals with a larger proportion of older patients, and older patients are more likely to get an angioplasty at hospitals with a larger proportion of young patients. (In the case of all heart surgeries, the latter effect has the expected sign but is not statistically significant.) Based on the assumption that regional treatment variations imply net welfare losses, policies have been proposed that would seek more uniform compliance with guidelines for medical practice that are based solely on patient characteristics and symptoms (see, for example, Congressional Budget Office, 2008). However, our model implies that the welfare implications of treatment variations depend on the underlying motivation for the social influences on treatment choice. In the presence of increasing returns, treatment variations have benefits, in the form of gains from specialization that accrue to the dominant patient type in a region, as well as costs that fall on the minority-type patients in a region. If the benefits outweigh the costs, treatment variations will dominate a system of strict treatment guidelines. On the other hand, if conformity pressure drives the choice spillovers, patients would be better off if physicians were constrained to make treatment decisions based solely on patient characteristics.

2. A model of medical treatment choice

We construct a formal model of treatment choice that incorporates local social influence and demographic variation. We represent the social environment as a one-dimensional, stochastic interacting particle system, as in Liggett (1999).[4] Physicians reside at fixed locations along a line. At a given location, patients arrive at random intervals, one at a time. Patients can be one of two "types," and the physician must choose one treatment (between two options) for each patient.[5] The patient's "type" can refer to any characteristic that is relevant to the treatment decision, but we find it useful to let type refer to age, such that some patients are 'old' and some are 'young.' At a given treatment opportunity, the physician selects the treatment that maximizes her payoffs. The payoff functions (which are identical for all doctors) depend on the patient's type and also on the recent treatment choices of the physician's two adjacent neighbors, where the latter dependence can represent either source of social influence described above. The line consists of two (connected) regions, each with a distinct mix of old versus young patients. Because the neighborhoods of adjacent physicians overlap, social influence percolates across the line of physicians and results, in equilibrium, in a single treatment being applied to all patients in a given region regardless of age. At the same time, regional differences in the age mix mean that the dominant treatment—the one that is "ideal" for the majority-type patient in the region—will differ across regions. Furthermore, the emergence of region-specific norms does not require extreme differences in the age distribution across regions. If the share of old patients is greater than fifty percent in one region and less than fifty percent in the other, the treatment norm will differ between the regions.

2.1. Theoretical model

Physicians are indexed by their location on $\mathbb{Z}$, the set of integers. For each $x \in \mathbb{Z}$, $\{x-1, x+1\}$ denotes the set of neighbors of x. There are two types of patients, denoted α and β, and two treatments, A and B. We characterize the physician payoffs to a given treatment choice in terms of a single number, $U(z, h, L, R, \ldots)$, where $z \in \{A, B\}$ represents the chosen treatment, $h \in \{\alpha, \beta\}$ represents the patient's type, and L and R represent, respectively, the last treatments chosen by the physicians to the left and to the right of the given doctor (L and R each belong to $\{A, B\}$). To capture the case of social influence derived from knowledge spillovers, we identify the physician's utility with the objective quality of the treatment choice, where we assume that this quality is fully captured by the probability that the given treatment succeeds. That is, a given treatment can either "succeed" or "fail" and physicians choose treatments to maximize the (objective) probability of success. Symbolically, we let $\pi_z(\cdot)$ denote the probability function of treatment success and we assume the following identity[6]:

$$U(z, h, L, R) = \pi_z(h, L, R).$$

[4] Such systems have been used to study problems in evolutionary game theory. The modelling framework builds upon ideas in Ellison (1993), Morris (2000), Young (1998), Young and Burke (2001), and Burke et al. (2006). Our main theoretical contribution is to generalize Ellison's model to allow payoffs (and hence choices) to depend on a factor which varies across locations, where in the medical example this factor is the patient's type.

[5] For convenience, we assume that physicians control treatment choices. In practice, treatment choice is a complicated combination of physician and patient preferences, as well as of constraints imposed by institutions like insurance companies, hospital administrations, and governments. We show below that our key results hold when patient preferences regarding treatment are incorporated. Institutional constraints on choice will be evaluated in our discussion of welfare issues.

[6] The fact that patient and physician interests are completely aligned does not require altruism on the part of physicians. For example, physician's incentives (which we do not model explicitly) could be aligned so as to make their interests coincide with patient welfare. In general, physician payoffs may depend on a number of factors, including, for example, the money rewards to different treatment choices (which depend, in turn, on reimbursement systems such as Medicare) and the variability of treatment outcomes. We also do not allow outcomes to depend on unobserved factors such as the quality of complementary inputs and the physician's skill and effort levels.

For example, $\pi_A(\alpha, A, B)$ denotes the probability that treatment A succeeds on a patient of type α, assuming the neighboring physicians to the left and right applied treatments A and B, respectively, at their last treatment opportunities, and this probability captures the physician's utility of choosing treatment A in the given context. We assume that all physicians know these conditional probabilities and hence the function π_z is identical across physicians. This formulation captures knowledge spillovers because, as described in detail below, we assume that a treatment's success probability increases in the number of neighboring physicians that used it at their last treatment opportunity. To capture a preference for conformity rather than knowledge spillovers, we can use a mathematically identical payoff structure and simply shift the interpretation of the utility function. In this alternative interpretation, physicians care about treatment success and about conformity to local practice patterns. Treatment success now depends only on patient characteristics, but conformity benefits depend on the recent choices of neighbors, and physicians trade off the benefits of acting like their neighbors against the dictates of patient characteristics. Such a model can lead to identical descriptive results but, as described below, carries different welfare implications. In the exposition below we adhere to the knowledge spillovers interpretation except when specified otherwise.

The essential assumptions about physicians' payoffs are the following: (1) payoffs from using a treatment increase with the number of neighbors who use the same treatment; (2) neither treatment dominates the other (that is, neither is better regardless of what neighbors do); and (3) when neighbors are evenly split between the two actions, then A is optimal for type α and B is optimal for type β.[7] Property 3 is a stronger version of the reasonable requirement that, for any fixed neighborhood, A yields higher payoffs when used on an α type than when used on a β type (and B yields higher payoffs when used on a β type than on an α type). In our two action setting, the above properties are equivalent to:

Property P. Preferences satisfy the following two conditions:

(a) Treatment A is optimal for α-patients if one or more neighbors use A, but B is optimal if both neighbors use B.
(b) Treatment B is optimal for β-patients if one or more neighbors use B, but A is optimal if both neighbors use A.

This is not a strong condition. To illustrate its reasonableness we present, and graph, an example of such preferences. Suppose, as above, that $U(z, h, L, R, \ldots) = \pi_z(h, L, R)$. Let

$$\begin{aligned} &\pi_A(\alpha, B, B) = 0.3, \qquad &&\pi_A(\beta, B, B) = 0.2,\\ &\pi_A(\alpha, A, B) = 0.4, &&\pi_A(\beta, A, B) = 0.3,\\ &\pi_A(\alpha, A, A) = 0.5, &&\pi_A(\beta, A, A) = 0.4. \end{aligned}$$

Similarly, for treatment B the payoffs are

$$\begin{aligned} &\pi_B(\alpha, B, B) = 0.4, \qquad &&\pi_B(\beta, B, B) = 0.5,\\ &\pi_B(\alpha, A, B) = 0.3, &&\pi_B(\beta, A, B) = 0.4,\\ &\pi_B(\alpha, A, A) = 0.2, &&\pi_B(\beta, A, A) = 0.3. \end{aligned}$$

Fig. 1 illustrates physician payoffs from using each treatment on an α-patient. Observe that the preferences satisfy property **P** (a).[8]

Patients arrive randomly at each location, with inter-arrival times that are exponential with parameter λ. Without loss of generality we take $\lambda = 1$. The concentration of patient types varies by region. We partition $\mathbb{Z}$ into two regions, East and West. The negative integers constitute the West, while the non-negative integers constitute the East. The probability that a patient who arrives at any given location in the East (West) is of type α will be given by p_E (p_W). The *state* of the system is a function from integers to $\{A, B\}$ ($\omega : \mathbb{Z} \to \{A, B\}$). An '$A$' at location x (i.e. $\omega(x) = A$) indicates that the physician at x used treatment A on her most recent patient. A 'B' denotes the use of treatment B on the most recent patient. The set of states is denoted by Ω.

When a patient arrives at a specific location $x \in \mathbb{Z}$, the physician must make a choice between A and B. The choice depends on the type of patient, as well as the choices made (in the recent past) by neighboring physicians. Following the norm in the evolutionary game theory literature, we assume best-response dynamics — physicians maximize $\pi_z(h, L, R)$. The state of the system can be visualized as an infinite sequence, with values at each location indicating the most recent choice made by the

[7] The first two assumptions are required for the very possibility of coordination on two different standards. The third property is closely related to the concept of risk dominance in game theory. Recall that in two-person coordination games an action is risk-dominant if it is the case that the action is optimal whenever there is a better than 50 percent chance that the opponent also uses the same action. For its application within a local interaction system similar to ours, see Ellison (1993). We have modified the property to allow for multiple types.

[8] Conformity effects can be modeled formally using a standard formulation (as in Brock and Durlauf, 2001). Relabel the choices $z \in \{-1, 1\}$, and let $U(z, h, L, R) = \pi_z(h) - J(z - \bar{z})^2$. $\pi_z(\cdot)$ still represents the success probability of treatment z, but this now depends only on the patient's type. The term $J(z - \bar{z})^2$ represents the non-conformity penalty for deviating from the (most recent) average behavior ($\bar{z}$) in the physician's neighborhood. Let $\pi_A(\alpha) > \pi_B(\alpha)$ and $\pi_B(\beta) > \pi_A(\beta)$. If the neighboring choices differ from each other, the non-conformity cost is just J for either treatment choice, and treatment follows the patient's type—that is, A (B) is chosen if the patient is of type α (β). If both neighbors choose the same action and if J is sufficiently large, the physician will choose the same action as her neighbors regardless of the patient's type.

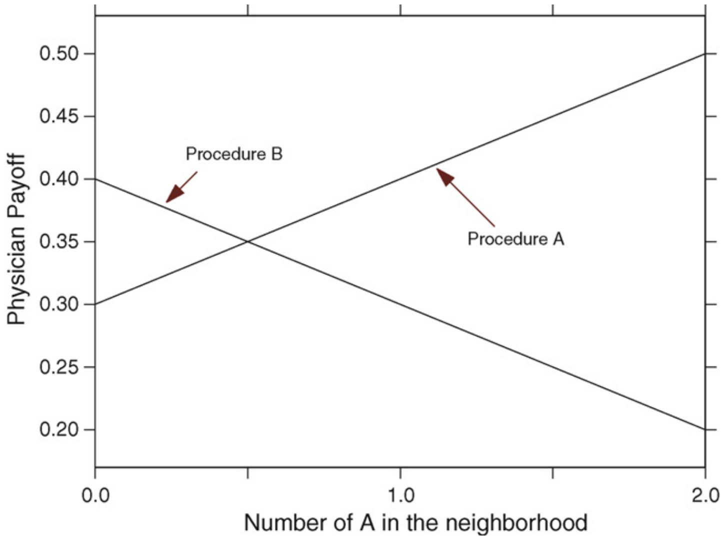

Fig. 1. Physician payoffs for procedures A and B and an α-patient.

physician there

$$\cdots AABBBABAAABA \cdots .$$

At random dates there is a transition: the value at one location changes from A to B or vice versa. The process is a continuous time Markov chain, X_t, and we are interested in the invariant (equivalently stationary, or equilibrium) distributions of this process.

Let $\mathbf{A} \in \Omega$ denote the state ω with $\omega(i) = A$ for all $i \in \mathbb{Z}$. In other words:

$$\mathbf{A} \equiv \cdots AAAAAAAAAAAA \cdots .$$

Similarly, $\mathbf{B} \in \Omega$ denotes the state ω with $\omega(i) = B$ for all $i \in \mathbb{Z}$:

$$\mathbf{B} \equiv \cdots BBBBBBBBBBBB \cdots .$$

The configuration at a particular date t will be identified by ω_t.

Let δ_ω be the probability that puts all of its mass on ω. Clearly, $\delta_{\mathbf{A}}$ and $\delta_{\mathbf{B}}$ are invariant measures. If we somehow reach the configuration $\mathbf{A}$ (or $\mathbf{B}$), the process can never escape from this state. Following Liggett (1999), we say the process *coexists* if there is an invariant measure that is not a mixture of $\delta_{\mathbf{A}}$ and $\delta_{\mathbf{B}}$. Alternatively, the process coexists if for i and j, $\lim_{t\to\infty}$ Prob$\{\omega_t(i) \neq \omega_t(j)\} > 0$. We show that the process X_t defined above coexists by identifying an invariant distribution in which both procedures are used with strictly positive probability at the same dates.

Define the set of states $S \subset \Omega$ as follows: $\omega \in S$ if there exists $m \in \mathbb{Z}$ such that $\omega(i) = A$ for all $i < m$ and $\omega(i) = B$ for all $i \geq m$. In other words, S consists of states such as

$$\cdots AAAAAABBBBBB \cdots .$$

S is *irreducible* —every state in S is reached with positive probability from any other state in S. It is *closed* —once in S, we can never escape. It is *recurrent* —we eventually return to every state in S—but not periodic.[9]

We prove the existence of an invariant distribution that has S as its support. For simplicity, the distribution is characterized by the location of the boundary point between the region in which treatment A is used and the region in which treatment B is used. In the proposition below, $\rho(\cdot)$ specifies the probability distribution of this boundary point. Proofs are in Appendix A.

Proposition 1. *Suppose preferences satisfy property* **P**. *Let* $p_W > 1/2$ *and* $p_E < 1/2$. *Then the physician choice process coexists. Specifically, there is an invariant measure* ρ, *with support* $\mathbb{Z}$, *such that*

$$\rho(m) = \frac{1}{K}\left(\frac{1-p_W}{p_W}\right)^{-m} \quad \text{if } m < 0,$$

$$\rho(m) = \frac{1}{K}\left(\frac{p_E}{1-p_E}\right)^{m} \quad \text{if } m \geq 0.$$

K is a real number constant which can be chosen to ensure that ρ is a probability.

[9] For formal definitions of these properties, see Norris (1998).

The proposition above tells us that the location of the East–West boundary is random. The probability $\rho(m)$ gives us the likelihood that the boundary will be m. Imagine the process as follows: each state consists of an infinite string of A's followed by infinitely many B's, but the boundary between the two regions can drift to the left or to the right along the line (by at most one unit at a time), according to the probabilities governed by $\rho(\cdot)$.[10] We refer to *the long-run outcome* ρ to describe the steady state in which the states from S appear according to probability ρ.

Remarks. (1) In case $p_W < 1/2$ and $p_E > 1/2$, we get a similar result, only the support now consists of a string of B's followed by A's. In case $p_W < 1/2$ and $p_E < 1/2$, the invariant distribution is $\delta_{\mathbf{B}}$. If $p_W > 1/2$ and $p_E > 1/2$, it is $\delta_{\mathbf{A}}$. (2) When $p_W = p_E = 1/2$ the state always remains in S and the boundary performs a symmetric random walk, as in the one-dimensional linear voter model (see Liggett, 1999). Despite the fact that the state always remains in S, the process does not coexist. This is because $\lim_{t\to\infty} \Pr\{\omega_t(i) \neq \omega_t(j)\} = 0$. (3) The proof of Proposition 1, as well as Proposition 2 below, requires infinitely many locations (i.e. $\mathbb{Z}$). In the finite case we would reach either **A** or **B** with positive probability, and then be trapped. However, we suspect that one can recover geographical variation by adding small noise to the model (see Burke et al., 2007) for suggestive simulation results). (4) If $p_E = p_W = 1$, then only α patients arrive at each location. This case corresponds to the model studied in Ellison (1993), and the risk dominant equilibrium **A** will be played.

In our model, choices depend upon recent decisions of neighbors but not on the recent decisions of the decision-maker herself. However, we can allow dependence on own past decisions with no alteration of results. To see this, assume that the neighborhood of x is defined by $N \equiv \{x-1, x, x+1\}$. Now suppose property **P** is modified so that (1) if all three players in N choose an action, then the physician at x chooses the same action. Otherwise, (2) she chooses A (B) if the patients is of type α (β). Now consider a state $\omega \in S$. We need only consider states at the boundary, which are states where the two adjacent neighbors make different choices. For such states, a choice of A is made for an α patient, and B is chosen for a β patient. This is exactly as before, and must lead to the same conclusion as Proposition 1.

So far we have ignored the role of patient preferences in the treatment decision. In practice, patient preferences are important and could be subject to social influences—for example, patients may request a treatment based on the recommendation of a friend. Accordingly, we can recast our treatment choice model in the case that patients influence treatment choices, and adopt the conformity model of social influence described in footnote 8. We can think of each location on our lattice as a "family" within which, say, cardiovascular events arise occasionally. The two regions of the lattice differ in the age composition of families, such that patients are younger, on average, in one region. Patients are randomly matched with physicians, who are not themselves subject to social influence. If choices in the neighborhood are diverse, then the physician–patient pair chooses treatment A for α-patients and treatment B for β-patients. If both neighbors make the same choice, however, and if social influence among patients is sufficiently strong (J is large), the physician–patient pair conforms to the neighbors' choice regardless of the patient's type, Property **P** is satisfied, and Proposition 1 holds.[11] As before, regional variation in medical practice will emerge and treatment will depend on where a patient lives.

2.2. Emergence of norms

Since the Markov process described above has several invariant distributions, we would like to identify the distribution which is most likely to be selected in the long run from randomly chosen initial conditions. It turns out that we can show (Proposition 2) that, starting from *almost any* initial state, the system eventually converges to an equilibrium that involves regional treatment norms (as opposed to globally uniform treatment choice). That is, the uniform states **A** and **B** are, in a well-defined sense, exceptional, and the system's long-run behavior will typically be described by the invariant distribution ρ from Proposition 1.[12]

Proposition 2. *Suppose the initial distribution is ν_θ, the Bernoulli product measure with density $\theta \in (0, 1)$, and let $p_W > 1/2$ and $p_E < 1/2$. Let π^t denote the distribution of the Markov chain at time t. Then π^t converges weakly to ρ as $t \to \infty$.*

2.3. Welfare

The fact that medical treatment depends on geographic location leads naturally to the concern that the quality of care might also vary geographically. This concern has been examined, from various perspectives, in the medical literature and in the health policy literature. For example, Krumholz et al. (2003) find that the treatment of myocardial infarction (MI) differs in New England compared to other regions along several dimensions (after controlling for patient, hospital, and physician characteristics), such as in the usage rate of beta-blockers, and that 30-day mortality rates are also lower in New England. A

---

[10] This drift occurs because, at the given boundary location at any time, the last treatment choices on either side differ from each other. Therefore, the next treatment decision at the boundary will be governed solely by the next patient's (random) type. Depending on this type, the next treatment could differ from the previous one, leading to a shift in the boundary position.

[11] Since physicians in this formulation do not override patient preferences, we can expect similar results if patients seek out doctors who offer the type of treatment they desire rather than being matched randomly with physicians.

[12] The proof is patterned after Durrett (1988) and Bramson and Griffeath (1981), who investigate the so-called "biased voter model". Our process is not identical to the biased voter model, but the differences are inconsequential for the main arguments. One difference is that we have regions with different "bias" or patient mix and another is the transition rate at a site where neighbors make opposing choices. While we deal with the much simpler one-dimensional case, in light of Bramson and Griffeath our results should generalize to $\mathbb{Z}^2$ and higher dimensions.

2008 report by the Congressional Budget Office (CBO 2008 "Geographic Variation in Health Care Spending"), based on a survey of studies, describes large variations in Medicare spending across regions of the U.S. The report finds that greater spending does not generally lead to higher-quality care and, in some cases, results in lower-quality care, where the quality of care was based either on patient outcomes or on clinical appropriateness. The report suggests that greater standardization of care, for example through the provision of incentives for universal compliance with accepted medical practice guidelines, could reduce inefficiencies in health care delivery. Official guidelines are already widely disseminated (for examples of guidelines in coronary care see Eagle et al., 1999; Gunnar et al., 1990), but enforcement is weak and potentially controversial. Consistent with the findings of Chandra and Staiger (2007), the CBO report also acknowledges that a push toward globally uniform practice may result in losses for some patients.

Our theoretical framework can shed light on the tradeoffs involved between a treatment pattern involving region-specific norms and a hypothetical pattern involving global uniformity (subject to patient characteristics). In the context of the model, we can define patient welfare in terms of the success probability of the selected treatment, and aggregate welfare as the average probability of treatment success across all locations.[13] We can then compare aggregate welfare between the long-run distribution described above and an alternative long-run distribution that represents a policy of enforced treatment guidelines. We represent such a policy as one that requires that patients of type α receive treatment A and patients of type β receive treatment B.

Recall that in the model, regional treatment variations can reflect either the presence of local increasing returns (as a result of knowledge spillovers, for example) or a preference for conformity (among physicians or patients). While the choice between these motivating mechanisms does not affect the long-run treatment patterns, it does affect the welfare analysis. Under increasing returns, the likelihood of success of a treatment increases with local experience in the treatment and hence some patients benefit from local uniformity of practice. Under conformity preferences, however, no benefits from local specialization accrue and some patients receive a suboptimal treatment. Accordingly, we find that the net benefits of a policy that imposes uniform practice guidelines (based solely on patient characteristics) will depend on which mechanism is driving the treatment variations. Even in the case of increasing returns, however, we can show that the desirability of strict treatment guidelines is not clear cut.

Suppose that physician preferences are not subject to pure conformity effects and, as in Section 2.1, utility equals the likelihood of success of the treatment used. We consider a policy which involves enforcement of treatment guidelines requiring the use of A on α-patients and B on β-patients.[14] We consider whether the policy improves the *expected* likelihood of success for the patient population as a whole. Note that, from our definition of physician payoffs, the long-run average utility of the physician at location x (in equilibrium ρ) is a measure of the average success rate of treatments for the patient population. Consequently, we can use this as a measure of welfare. We prove:

Proposition 3. *The long-run outcome ρ need not maximize the success rate of treatments for the patient population. In particular, it may be dominated, at every location $x \in \mathbb{Z}$, by the policy of enforced treatment guidelines.*

The proof of the proposition involves identifying a plausible technology for which the guidelines policy proves superior to the long-run outcome ρ at every location. In general, guidelines are more likely to dominate the long-run outcome (a) the more similar are the population profiles between the two regions (for example, if both p_W and p_E are close to 1/2), (b) the greater the payoff advantage to having neighbors that choose two different treatments rather than the same treatment, and (c) the smaller the losses from having to reverse the treatment choice at locations at which the two neighbors choose the same treatment.[15] By an analogous argument, one can also show that ρ may dominate the policy of enforced treatment guidelines.

Proposition 4. *A policy of enforced treatment guidelines can be dominated by the long-run outcome ρ.*

The proof is immediate from inspection of the proof of Proposition 3, and hence omitted. To summarize, in the case where they are driven by knowledge spillovers between physicians, regional variations in practice may not be undesirable. Propositions 3 and 4 help to delineate some of the circumstances under which treatment guidelines may be desirable.

3. Empirical analysis

Our model holds the testable implication that the treatment a patient receives will be the one that is medically best-suited to the dominant patient type in the region. While the model, with its stylized assumptions, predicts absolute uniformity of treatment choice within regions, we do not expect such a stark outcome in the real world. However, the model suggests that, other things constant, the characteristics of the local patient pool will skew treatment in a predictable direction for all patients in a given region or at a given treatment location, such as a hospital. In the case of coronary care, given that

[13] This aggregate welfare criterion is equivalent to life-years gained, a standard criterion adopted in the public health and medical literature. Unlike the Pareto criterion, this welfare measure permits interpersonal comparisons. Unlike the Kaldor-Hicks criterion, our approach does not measure the willingness to pay for life-years gained in order to determine whether those who are worse off could be compensated by those who are better off.

[14] The policy alternative of moving patients to regions based on their characteristics is considered infeasible.

[15] Note that following treatment guidelines when others do so is not a best response—the policy requires enforcement, the costs of which we do not consider.

Table 1
Summary statistics on the panel data set from Florida, 1995 to 2001.

Variable	Mean	S.D.
Dependent variables		
Angioplasty rate per thousand, younger patients	450.25	109.41
Angioplasty rate per thousand, older patients	330.27	124.42
Surgery rate per thousand, younger patients	671.48	134.39
Surgery rate per thousand, older patients	555.25	150.14
Explanatory variables		
Hospital volume in quarter		
Total number of heart patients in quarter	395.25	227.12
Total volume of angioplasties in quarter	161.10	113.16
Total volume of surgeries in quarter	258.90	175.38
Hospital demographics		
Proportion of heart patients in younger group	0.35	0.10
Proportion of heart patients in older group	0.46	0.10
Proportions of patients by patient race		
Black	0.07	0.06
Hispanic	0.09	0.19
Other	0.01	0.04
Income, 1999	38615	4423
Proportion of patients with managed care plan	0.36	0.16

Pearson correlation coefficients	Patient age
Angioplasty (patient receives angioplasty)	−0.11222
p-Value under $H0 : \rho = 0$	<0.0001
Surgery (patient receives surgery)	−0.11435
p-Value under $H0 : \rho = 0$	<0.0001

Note: There are 1488 hospital-quarter observations, based on 572,316 patient records. Younger patients are defined as the group including youngest one-third of the sample, patients with age <63, while older patients are the oldest one-third in the sample, patients with age >72. Proportions are calculated relative to the full sample of all patients admitted to the hospital for coronary care.

heart surgery rates decline on average with age, we expect that the probability that a younger patient receives heart surgery (as opposed to non-invasive treatment of the disease) will be lower at a hospital that receives a relatively high proportion of older coronary patients. Conversely, surgery rates for older patients will be higher at hospitals with a relatively young patient pool than at hospitals with an older patient population.[16]

We test these hypotheses using a census of patient discharge records from Florida, focusing on the treatment of heart disease. Heart disease (defined below within the context of our data) is an appropriate choice for our empirical analysis because its treatment mirrors a number of features of our theoretical model. First, the treatment options can be divided into two discrete categories, surgery and "medical management." The surgical options include (1) percutaneous transluminal coronary angioplasty (PTCA or, in common parlance, coronary angioplasty), and (2) coronary artery bypass grafting surgery (CABG, or heart bypass surgery); medical management involves the use of drugs (such as beta blockers, calcium channel blockers, and ACE inhibitors) and other non-invasive therapies (such as diet and exercise modification).[17] Consistent with the payoff structure in our model, there is evidence that the invasive treatments are, on average, less suitable for older patients than younger patients (see, for example, Smith et al., 2001). Our data agree with the presumption of age-dependence, in that the raw (Pearson) correlation between patient age and the probability of receiving surgical treatment for heart disease is negative and significant (see Table 1). In addition, the age distribution of patients admitted to hospitals for coronary care varies considerably across hospitals (and regions) in Florida, and elderly patients (defined as 73 years of age or older) are well-represented.

We have access to the entire sample of quarterly patient discharge records from Florida hospitals for the years 1995–2001, from a legally mandated and audited census of inpatient stays. Each record gives the patient's age, race, sex, principal diagnosis and (where applicable) secondary diagnoses, treatments received, the hospital name and county location, the length of stay, and related information.[18] From this census, we extract the records of patients 25 and older who were admitted

[16] While our empirical analysis is focused on age effects, there may be other characteristics that lead to treatment variations in a similar manner. Nichols (2006) reports differences in the treatment and outcome patterns of African-American patients with AMI admitted to hospitals with disproportionately black patient populations. Consistent with the predictions of our model, he finds that providers with greater numbers of black patients adopt treatment practices that are particularly effective in treating conditions common among African-Americans.

[17] While the surgical options differ from each other, the distinction between invasive (i.e., surgical) and non-invasive treatments is standard in the literature (see, for example, Smith et al., 2001), suggesting a bright line between these treatment modes.

[18] Each observation is a single hospital stay rather than a longitudinal patient record. Identifiers for individual patients are masked, such that repeat hospitalizations are censored. Long-term outcomes, such as 30-day survival, are not available.

with a diagnosis of coronary artery disease, in the form of either atherosclerosis or acute myocardial infarction (AMI).[19] Among patients with these diagnoses, we keep those that received either (a) angioplasty (PTCA), (b) bypass surgery (CABG), or (c) neither PTCA nor CABG, where this last classification intends to capture those who received only drug treatments and/or other non-invasive therapies.[20] We include records from only those hospitals that held the legal authority ("certificate of need" license) to perform heart surgery and that maintained the required surgical facilities, and for which the hospital's records span at least 5 quarters. The final sample contains 572,316 patient records, spanning 62 hospitals and 25 quarters. The sample accounts for a high proportion of all hospital-based coronary treatment events in the state during the period of observation, and represent all 11 of the state's "Administrative Health Care Districts." Summary statistics are presented in Table 1.

3.1. Hospital-level panel regression model

We choose the hospital as the geographic unit of analysis, for several reasons. First, doctors working at the same hospital are likely to interact with at least some subset of their colleagues and observe each others' treatment choices. Second, we can control for (fixed) unobserved heterogeneity at the hospital level that could create the appearance of local treatment norms for reasons other than local social influence. Third, we can also control for time-varying hospital factors, such as total volume in cardiac surgeries, that (if omitted) could also yield spurious results. Therefore, the identification exploits within-hospital variation over time in the age distribution of the treated patients that is orthogonal to other observed changes in hospital characteristics. We also include time dummies to control for statewide changes in treatment tendencies over time.

From the usable patient records, we aggregate the variables of interest to the hospital-quarter level to obtain a panel of 1448 observations. To test the model, we consider whether hospitals with a greater proportion of young (old) patients will be more likely (less likely) to apply invasive treatments to their old (young) patients. We define "young" and "old" based on the bottom and top terciles, respectively, of the age distribution of heart patients in the sample. As a result, young patients are those 62 or younger and old patients are those 73 or older. The dependent variables refer to rates of invasive treatment (at the hospital-quarter level) among either old patients or young patients. When the dependent variable refers to a treatment rate for old patients, the main explanatory variable of interest is the share of young patients treated at the same hospital in the same quarter, and vice versa when the dependent variable refers to young patients. We construct two different measures of the rate of invasive treatment to ensure robustness of results. In the first case we use the angioplasty rate, and in the second case we use the rate of surgery in general, including either angioplasty (PTCA) or bypass (CABG).[21] In the case of the treatment of old patients, the model is as follows:

$$\text{treatment}(73)_{ht} = \alpha_h + c_t + \gamma Age(62)_{ht} + \beta X_{ht} + \delta V_{ht} + \epsilon_{ht}. \tag{1}$$

The model in (1) is estimated using ordinary least squares. The dependent variable is the rate (per thousand patients) of angioplasty (alternatively, either of the two surgical options) performed in the hospital for coronary patients aged 73 or older in the given quarter. This treatment rate depends on the fixed tendencies of the hospital, α_h, the current calendar date c_t (a specific quarter), the average income, racial composition, and insurance status, respectively, of the hospital's current patients, X_{ht}, the hospital's overall volume of angioplasty (surgery) procedures in the quarter, V_{ht}, and the proportion of young patients treated at the hospital in the quarter, $Age(62)_{ht}$. We predict that the coefficient γ is positive—old patients are more likely to get invasive procedures, all else equal, the greater the proportion of young patients treated at the same hospital in the same period.

For the treatment of young patients, the model is analogous, as follows:

$$\text{treatment}(62)_{ht} = \alpha_h + c_t + \gamma\,\text{Age}(73)_{ht} + \beta X_{ht} + \delta V_{ht} + \epsilon_{ht}. \tag{2}$$

In the above, the dependent variable, $\text{treatment}(62)_{ht}$, measures the treatment rate of young patients and the explanatory variable, $\text{Age}(73)_{ht}$, is the proportion of old patients at the same hospital in the same quarter. We predict that γ takes a negative sign.

3.2. Results

Table 2 reports the results of our panel regressions. Recall that the unit of observation is the hospital-by-quarter. The top panel shows the results pertaining to the treatment of "old" patients (73 and older). The first (left-most) column of numbers consists of the regression coefficients when the dependent variable is the angioplasty rate, and in the second column the

[19] The CCS Diagnosis Categories were used to identify the 56 ICD-9CM categories relevant to these patients and to identify broad categories of comorbidities.

[20] The third group includes some patients who received coronary angiography, a somewhat invasive diagnostic procedure, but only if this diagnostic was not followed by either PTCA or CABG. Most patients (greater than 60 percent) in the sample who received angiography subsequently received surgery and hence will be counted as receiving invasive treatment. Also, we exclude patients who receive heart transplants, heart valve surgery, implanted defibrillator devices, and pacemakers.

[21] We might predict, for example, that results will be weaker when using the rate of a single type of surgery (such as PTCA), since some hospitals, including those with many young patients, may specialize in just one type of surgery, such as CABG.

Table 2
Treatment rates by patient age and the effect of hospital demographics; panel regressions.

	Angioplasty			Surgery		
	Coef.	S.E.	*p*-Value	Coef.	S.E.	*p*-Value
Dependent variables: treatment rates, older patients						
Hospital volume, number of angioplasties	0.775 **	0.036	0.000	0.721 **	0.038	0.000
Hospital demographics						
Proportion of all patients aged 62 or younger	179.587 **	69.469	0.010	61.318	59.429	0.302
Proportion of all patients aged 73 or older						
Proportion patient race						
Black	-226.590 **	111.540	0.042	-530.797 **	120.448	0.000
Hispanic	-88.335 *	48.548	0.069	-95.639 *	56.477	0.090
Other	31.918	64.534	0.621	25.764	64.371	0.689
Income, 1999	0.005 **	0.002	0.035	0.005 *	0.003	0.052
Proportion patients with HMO or PPO insurance	48.223	30.871	0.118	38.588	32.512	0.235
Intercept	85.983	99.763	0.389	345.795 **	104.882	0.001
Dependent variables: treatment rates, younger patients						
Hospital volume, number of surgeries	0.829 **	0.037	0.000	0.584 **	0.038	0.000
Hospital demographics						
Proportion of all patients aged 62 or younger						
Proportion of all patients aged 73 or older	-170.349 **	51.927	0.001	-216.112 **	55.422	0.000
Proportion patient race						
Black	-302.083 **	86.240	0.000	-500.674 **	97.595	0.000
Hispanic	-105.609 **	52.578	0.045	28.369	53.322	0.595
Other	45.419	38.978	0.244	47.005	36.091	0.193
Income, 1999	0.003 *	0.002	0.097	0.003	0.002	0.219
Proportion patients with HMO or PPO insurance	22.634	26.155	0.387	-3.901	28.838	0.892
Intercept	311.441 **	84.136	0.000	565.340 **	90.046	0.000

Note: Standard errors are robust. Dependent variables are expressed as treatment rates per thousand patients. The model includes fixed effects for hospitals and time, not reported here.
* Significant at $\alpha = 0.10$.
** Significant at $\alpha = 0.05$ level.

dependent variable is just the surgery rate (either angioplasty or bypass). Standard errors are in parentheses below each estimate and significance is indicated with asterisks as noted in the Table. The bottom panel shows the analogous results for the treatment of “young” patients (62 and under). In three out of four cases, we observe the predicted effects of local demographics on treatment patterns: old patients are more likely to receive angioplasty the greater is the share of young patients treated at the same hospital in the same period, and young patients are less likely to receive angioplasty, and less likely to receive surgery in general, the greater the share of old patients at the hospital. While the share of young patients does not have a significant impact on the combined surgery rate for older patients, the coefficient takes the predicted sign.[22]

The results can be interpreted quantitatively as follows: a one standard deviation change in the proportion of younger heart patients (from 0.35 to 0.45) would raise the expected number of angioplasties given to older patients by about 18 (from 330 per thousand to 348 per thousand.) For surgeries, a one standard deviation increase in the proportion of young patients is predicted to add 6 surgeries for older patients. Similarly, at the mean number of angioplasties for young patients (450 per thousand), a one standard deviation change in the proportion of older heart patients at the hospital (from 0.46 to 0.56) would reduce the expected number of angioplasties performed on younger patients by about 17 (from 450 to 433 per thousand.) For the surgery treatments given younger patients, a one standard deviation increase in the proportion of older patients decreases by 21 the expected number of surgeries for younger patients.

We also observe (Table 2) that the angioplasty rate (total surgery rate) increases, for either young or old patients, with the hospital's total volume of angioplasties (volume of surgeries) during the quarter. This result most likely reflects specialization—for example, it has been observed that hospitals that perform relatively few invasive procedures have lower success rates for them.[23] To the extent that this volume variable captures (time-varying) specialization at the hospital level, including this control helps to ensure that the effects of patient age composition are not merely proxying for the effects of

[22] Together with the significant result for angioplasty alone, this result suggests that old patients are not significantly more likely to receive bypass surgery when there is a large share of young patients in the region.
[23] See, e.g. Birkmeyer (2000).

Table 3
p-Value adjustments for multiple comparisons.

	Unadjusted (*p*-value)	Bonferroni (*p*-value)	Holm (*p*-value)
Dependent variable: angioplasty			
Treatment rates, older patients			
Hospital volume, number of angioplasties	0.000**	0.000**	0.000**
Hospital demographics			
Proportion of all patients aged 62 or younger	0.010**	0.020**	0.020**
Proportion of all patients aged 73 or older			
Proportion patient race			
Black	0.042**	0.084*	0.042**
Hispanic	0.069*	0.138	0.138
Other	0.621	1.242	1.242
Income, 1999	0.035**	0.070*	0.070*
Proportion patients with HMO or PPO insurance	0.118	0.236	0.236
Intercept	0.389	0.778	0.389
Treatment rates, younger patients			
Hospital volume, number of surgeries	0.000**	0.000**	0.000**
Hospital demographics			
Proportion of all patients aged 62 or younger			
Proportion of all patients aged 73 or older	0.001**	0.002**	0.001**
Proportion patient race			
Black	0.000**	0.000**	0.000**
Hispanic	0.045**	0.090*	0.090*
Other	0.244	0.488	0.386
Income, 1999	0.097*	0.194	0.194
Proportion patients with HMO or PPO insurance	0.387	0.774	0.774
Intercept	0.000**	0.000**	0.000**
Dependent variable: surgery			
Treatment rates, older patients			
Hospital volume, number of angioplasties	0.000**	0.000**	0.000**
Hospital demographics			
Proportion of all patients aged 62 or younger	0.302	0.604	0.302
Proportion of all patients aged 73 or older			
Proportion patient race			
Black	0.000**	0.000**	0.000**
Hispanic	0.090*	0.180	0.138
Other	0.689	1.378	1.242
Income, 1999	0.052*	0.104	0.070*
Proportion patients with HMO or PPO insurance	0.235	0.470	0.236
Intercept	0.001**	0.002**	0.002**
Treatment rates, younger patients			
Hospital volume, number of surgeries	0.000**	0.000**	0.000**
Hospital demographics			
Proportion of all patients aged 62 or younger			
Proportion of all patients aged 73 or older	0.000**	0.000**	0.000**
Proportion patient race			
Black	0.000**	0.000**	0.000**
Hispanic	0.595	1.190	0.595
Other	0.193	0.386	0.386
Income, 1999	0.219	0.438	0.219
Proportion patients with HMO or PPO insurance	0.892	1.784	0.892
Intercept	0.000**	0.000**	0.000**

*Significant at $a = 0.10$; **Significant at $a = 0.05$.

specialization.[24] We find that the racial composition of the patient base also affects treatment systematically: surgery (or angioplasty alone) is less likely to be selected, for either old or young patients, the greater the proportion of black patients treated at the hospital. Similar effects are observed for the share of hispanic patients, but the significance is weaker on average. Hospitals with higher patient income (as of 1999) have a higher angioplasty rate for old patients, but income effects are otherwise only weakly significant. The share of insured patients exerts no significant effects on treatment tendencies.

We test multiple hypotheses using overlapping data, and therefore run an elevated risk of spuriously producing statistically significant results. Several options exist for adjusting p-values to obtain robust inference under these conditions.[25] In Table 3, we consider two different corrections. First, the Bonferroni adjustment multiplies the regression p-values by the number of tests used. This correction is a conservative measure which may overcompensate for the multiple comparisons problem (Legendre and Legendre, 1998). The second adjustment, a more powerful test attributed to Holm, involves a modified correction to the p-values (Holm, 1979).[26] Under either method, the corrected p-values are compared to the chosen α significance level. Table 3 shows the three different p-values associated with each coefficient estimate for each model: the unadjusted value, the Bonferroni-corrected value, and the Holm-corrected value. While some control variables lose significance under the corrected p-values, the significance levels of the key variables of interest—for example, the proportion of young patients at the hospital—are robust to either correction method.

4. Conclusion

This paper addresses the puzzle of robust geographic variations in medical treatment. We provide a theoretical explanation for the puzzle using a model that involves local social influence among doctors and regional variation in patient characteristics. Starting from almost any initial condition, simple adaptive behavior leads to region-specific treatment norms. In equilibrium, the variation in treatment rates far exceeds the underlying variation in patient characteristics, consistent with previous empirical findings. The theory makes sharp predictions about the relationship between local patient characteristics and treatment use, predictions largely confirmed in our data. In particular, we find that old (young) patients tend to be treated more invasively (less invasively) for heart disease the greater the share of young (old) patients treated at the same hospital in the same period. In each case, treatment choices for all patients are skewed toward the choice that is medically optimal for the dominant demographic group in the area.

Our model makes significant contributions to the theoretical understanding of treatment variations. First, we provide an explanation for how regional norms emerge rather than just why norms might persist once in place. Second, the model shows that either productivity spillovers or conformity pressure can result in the emergence of treatment variations and demonstrates an important difference in welfare implications between the two mechanisms. In addition, the model implies that regional norms are grounded in relatively stable underlying fundamentals (in our case, patient characteristics), a result that contrasts with the multiple equilibria that often arise in settings involving conformity effects and other spillovers. Although there are multiple equilibria in our model, the emergent stable equilibrium admits regional variations. This makes it feasible to do systematic empirical work.

For simplicity, we assumed that patients do not respond to the emergence of treatment norms by choosing hospitals that best match their preferences and characteristics. In practice, patients (and their health care providers) exert some choice over the hospitals in which they are treated, and may even select residential locations based in part on proximity to a particular hospital. While such sorting would mitigate the welfare losses experienced by minority-type patients in a region, there are likely to be considerable constraints on optimal hospital–patient matching and hence our results should be largely robust.

Our welfare analysis holds important policy implications. Under conformity pressure, a tendency to over-utilize invasive treatments may arise in some regions (and to under-utilize them in other places), contributing to inefficient care and elevated costs. On the other hand, we show that legislating strict treatment guidelines may be harmful in the presence of knowledge spillovers, because some patients will be deprived of gains from specialization. Regional treatment variations are not necessarily welfare-reducing, but further research along these lines is clearly warranted before policy recommendations can be made.

Acknowledgments

We thank Peyton Young, Steve Durlauf, Susan Lampman and three anonymous referees for their comments. We are also grateful to seminar audiences at the Brookings Institution, the University of Birmingham, the Santa Fe Institute, and at Stanford University's Institute for Theoretical Economics. Any remaining errors are our own.

[24] With hospital fixed effects in the model, fixed specialization is already accounted for. However, specialization could vary over time in a manner that would be proxied by changes in procedure volume.

[25] For discussion and critique of these adjustments, see Legendre and Legendre (1998), Perreneger (1998) and Nakagawa (2004) and the citations therein.

[26] With the Holm correction for k multiple tests, individual p-values are ordered from smallest to largest, i.e., $p_1 \leq p_2 \cdots \leq p_i \leq p_3 \cdots \leq p_k$. Probability values are first adjusted according to $p_i' = (k - i + 1)p_i$. Next, proceeding from left to right, if an adjusted p-value is smaller than the one to its left, the smaller one is set equal to the larger one. Corrected p-values may be greater than one.

Appendix A.

A.1. Proof of Proposition 1

Proof. Since the process restricted to S is irreducible and aperiodic, it has a unique invariant distribution. Each state can be specified in terms of m, the location of the first zero. First we define the probabilities $b(m)$ and $d(m)$ of transition $m \to m+1$ and $m \to m-1$ respectively. Recalling that the rate of arrival of patients is one, these are given by

$$b(m) = \begin{cases} p_W & \text{if } m < 0, \\ p_E & \text{otherwise.} \end{cases}$$

In other words, m moves to the right if an α-patient arrives at m, which happens with probability p_W in the West and p_E in the East:

$$d(m) = \begin{cases} 1-p_W & \text{if } m \le 0, \\ 1-p_E & \text{otherwise.} \end{cases}$$

In other words, m moves to the left if a β-patient arrives at $m-1$, which happens with probability $1-p_W$ in the West and $1-p_E$ in the East. The process is reversible, so that invariant distributions can be obtained from the detailed balance conditions:

$$b(m-1)\rho(m-1) = d(m)\rho(m).$$

We can confirm that these are satisfied. The conditions $p_W > 1/2$ and $p_E < 1/2$ ensure that K is finite in the definition of ρ, and the balance equations are satisfied for non-zero $\rho(\cdot)$. In case $m \le 0$, we can substitute for ρ and confirm that

$$\frac{b(m-1)}{d(m)} = \frac{p_W}{1-p_W} = \frac{\rho(m)}{\rho(m-1)}.$$

When $m > 0$:

$$\frac{b(m-1)}{d(m)} = \frac{p_E}{1-p_E} = \frac{\rho(m)}{\rho(m-1)}.$$

So $\rho(\cdot)$ is an invariant distribution. It is not a mixture of δ_0 and δ_1, hence the process coexists. □

A.2. Proof of Proposition 2

Proof. Let ξ_t^x denote the process at time t when the initial configuration has A at site x, and B elsewhere. In this case the A-region will always constitute an interval, unless ξ_t^x has no A's at all. Let $L_t \equiv \min_i\{i|\xi_t^x(i) = A\}$ and $R_t \equiv \max_i\{i|\xi_t^x(i) = A\}$, so that $[L_t, R_t]$ denotes the A-region (initially, $L_0 = R_0 = x$). We first show that for $x \in$ West, and conditioning on the event:

$$\Omega = \{R_t \ge L_t \text{ for all } t > 0\},$$

ξ_t^x, grows linearly in time until R_t reaches the East/West boundary (specifically, until $R_t = -1$). Thereafter, only L_t extends westwards. Given $p_W > 1/2$, $p_E < 1/2$, and if $0 > R_t > L_t$, R_t and L_t perform independent random walks according to

$$R_t \to \begin{cases} R_t + 1 & \text{at rate } \lambda, \\ R_t - 1 & \text{at rate } 1, \end{cases} \tag{3}$$

$$L_t \to \begin{cases} L_t - 1 & \text{at rate } \lambda, \\ L_t + 1 & \text{at rate } 1, \end{cases} \tag{4}$$

where $\lambda = p_W/(1-p_W) > 1$. Then, following Durrett (p. 38), and conditioning on Ω:

$$\frac{R_t - x}{t} \to (\lambda - 1) \quad \text{and} \quad \frac{L_t - x}{t} \to -(\lambda - 1) \quad a.s.$$

Once $R_t = -1$, conditional on Ω, R_t evolves like the boundary in Proposition 1. An analogous statement holds for the evolution of B regions in the East.

Next we consider an arbitrary configuration ξ and index the A and B regions as follows. Let A^0 denote the easternmost A-region that still occupies sites in the West: i.e. A^0 is a set of contiguous sites with (1)$x \in A^0 \Rightarrow \xi(x) = A$, (2)$A^0 \cap (\text{West} \cup \{0\}) \ne \varnothing$ and (3)$[i > \max A^0 \ \& \ i < 0] \Rightarrow \xi(i) = B$. Similarly B^0 denotes the set of contiguous sites with (1)$x \in B^0 \Rightarrow \xi(x) = B$, (2)$B^0 \cap \text{East} \ne \varnothing$ and (3)$[i < \min B^0 \ \& \ i \ge 0] \Rightarrow \xi(i) = A$. If ξ is chosen according to ν_θ then, with probability one, both A^0 and B^0 will exist, and share a common boundary (defined as in Proposition 1, as the location of the first B in B^0). Label the A-region immediately to the west of A^0 by A^{-1} and the nearest eastern region by A^{+1}, and so on. We do the same for B-regions, with Eastern regions having positive indices and western regions having negative ones. Now A regions grow in the West, B-regions grow in the East, and the A^0/B^0 boundary evolves like the boundary of states in the sub-chain on S in Proposition 1, unless one of A^0 or B^0 becomes extinct (the right boundary becomes smaller than its left boundary). In case A^0 or B^0 becomes extinct, we relabel indices according to the scheme above and get a new A^0/B^0 boundary.

Since, with probability one, there are initially infinitely many A and B regions, there are always A and B regions available to be relabeled. As $t \to \infty$, $|A^0| \to \infty$ and $|B^0| \to \infty$ and their extinction probability becomes zero. B-regions in the West and A-regions in the East tend to become extinct. As $t \to \infty$ the probability, for some location $x \in$ West, that $\xi_t(x) = B$ approaches the probability that the A^0/B^0 boundary is at $y \leq x$, which converges to $\rho(x)$:

$$\text{Prob}\{\xi_t(x) = B\} = \sum_{i \leq x} \rho(i).$$

So, observing that $\Omega = \{A, B\}^{\mathbb{Z}}$ carries the product topology, all the finite dimensional distributions converge as well, implying weak convergence of π^t to ρ. □

A.3. Proof of Proposition 3

Proof. Since a physician's expected utility at a location has been defined as the likelihood of success for the population profile at that location, we can speak of patient welfare in terms of these same payoffs. First we describe the expected utility at location x in the long-run co-existent outcome ρ. Suppose $x < 0$ (x is in the West). Expected utility at x is a weighted sum of three terms:

$$U_1 \equiv p_W \pi_A(\alpha, A, A) + (1 - p_W)\pi_A(\beta, A, A),$$
$$U_2 \equiv p_W \pi_B(\alpha, B, B) + (1 - p_W)\pi_B(\beta, B, B),$$
$$U_3 \equiv p_W \pi_A(\alpha, A, B) + (1 - p_W)\pi_B(\beta, A, B)$$

with corresponding weights (1) the probability that x is in the interior of a region of A's, (2) the probability that x is in the interior of a region of B's, and (3) the probability that x is at a boundary. These probabilities can be explicitly computed from Proposition 1. The expected utility for a location in the East can be obtained in a similar manner. With enforced treatment guidelines the expected utility at $x < -1$ (interior West) is a weighted sum of

$$V_1 \equiv p_W \pi_A(\alpha, A, A) + (1 - p_W)\pi_B(\beta, A, A),$$
$$V_2 \equiv p_W \pi_A(\alpha, B, B) + (1 - p_W)\pi_B(\beta, B, B),$$
$$V_3 \equiv p_W \pi_A(\alpha, A, B) + (1 - p_W)\pi_B(\beta, A, B)$$

with weights (1)p_W^2, (2)$(1 - p_W)^2$, and (3)$2p_W(1 - p_W)$ respectively. In the interior East the weights are p_E^2, $(1 - p_E)^2$, and $2p_E(1 - p_E)$ respectively. At $x \in \{-1, 0\}$, one neighbor is in the East and one is in the West so that the weights are $p_E p_W$, $(1 - p_E)(1 - p_W)$, and $p_E(1 - p_W) + p_W(1 - p_E)$ respectively. In the interior West, from the returns to scale assumption, $U_1 > V_1$, $U_2 > V_2$, and $U_3 = V_3$. Treatment guidelines can do better if $U_1 - V_1$ and $U_2 - V_2$ are small, $U_3 = V_3$ is larger than both U_1 and U_2 and has much greater weight under treatment guidelines than at the long-run outcome. These conditions can be satisfied by non-pathological technologies, e.g.

$$\pi_A(\alpha, B, B) = 0.1, \qquad \pi_A(\beta, B, B) = 0,$$
$$\pi_A(\alpha, A, B) = 0.4, \qquad \pi_A(\beta, A, B) = 0.1,$$
$$\pi_A(\alpha, A, A) = 0.5, \qquad \pi_A(\beta, A, A) = 0.11,$$

and similarly, for B:

$$\pi_B(\alpha, B, B) = 0.11, \qquad \pi_B(\beta, B, B) = 0.5,$$
$$\pi_B(\alpha, A, B) = 0.1, \qquad \pi_B(\beta, A, B) = 0.4,$$
$$\pi_B(\alpha, A, A) = 0, \qquad \pi_B(\beta, A, A) = 0.1$$

when p_E and p_W are close to 1/2.[27] For expected utility at the long-run outcome, the weight of the term U_3 becomes small as p_E and p_W become close to 1/2 (specifically, $\rho(m) \to 0$ as $p_E, p_W \to 1/2$). For treatment guidelines the weight of V_3 becomes close to 1/2, and so guidelines do better. This argument applies to the interior East with appropriate change of notation. For the case of $x \in \{-1, 0\}$, assuming p_W and p_E are both close to 1/2, expected utility under the guidelines is approximately equal to the expected utility in either the interior East or the interior West under guidelines. Therefore the long-run outcome ρ is dominated by the policy of treatment guidelines. □

References

Alexander, K.P., Peterson, E.D., 1997. Coronary artery bypass grafting in the elderly. American Heart Journal 134, 856–864.

[27] In contrast, for the technology given in Section 2.1, the long-run outcome is always superior to enforced treatment guidelines.

Bernheim, D., 1994. A theory of conformity. Journal of Political Economy 102 (5), 841–877.

Birkmeyer, J.D., 2000. High-risk surgery—follow the crowd. Journal of the American Medical Association 283 (9), 1191–1193.

Bramson, M., Griffeath, D., 1981. On the Williams–Bjerknes tumour growth model I. The Annals of Probability 9, 173–185.

Brock, W., Durlauf, S., 2001. Discrete choice with social interactions. Review of Economic Studies 68 (2), 235–260.

Burke, M.A., Fournier, G.M., Prasad, K., 2007. The diffusion of a medical innovation: is success in the stars? Southern Economic Journal 73, 588–603.

Burke, M.A., Fournier, G.M., Prasad, K., 2006. The emergence of local norms in networks. Complexity 11, 65–83.

Chandra, A., Staiger, D., 2004. Testing a Roy model with productivity spillovers: evidence from the treatment of heart attacks. NBER Working Paper No. W10811.

Chandra, A., Staiger, D., 2007. Testing a Roy model with productivity spillovers: evidence from the treatment of heart attacks. Journal of Political Economy 115 (1), 103–140.

Congressional Budget Office, 2008. Geographic Variation in Health Care Spending. Publication No. 2978. http://www.cbo.gov/ftpdocs/89xx/doc8972/02-15-GeogHealth.pdf.

Durrett, R., 1988. Lecture Notes on Particle Systems and Percolation. Wadsworth, Pacific Grove, CA.

Eagle, K.A., Guyton, R.A., Davidoff, R., Ewy, G.A., Fonger, J., Gardner, T.J., Gott, J.P., Herrmann, H.C., Marlow, R.A., Nugent, W.C., O'Connor, G.T., Orszulak, T.A., Rieselbach, R.E., Winters, W.L., Yusuf, S., Gibbons, R.J., Alpert, J.S., Garson, A., Gregoratos, G., Russell, R.O., Smith Jr., S.C., 1999. ACC/AHA guidelines for coronary artery bypass graft surgery: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (committee to revise the 1991 guidelines for coronary artery bypass graft surgery). Journal of the American College of Cardiology 34 (4), 1262–1347.

Ellison, G., 1993. Learning, local interaction, and coordination. Econometrica 61, 1047–1071.

Folland, S., Stano, M., 1989. Sources of small area variations in the use of medical care. Journal of Health Economics 8, 85–107.

Glover, J.A., 1938. The incidence of tonsillectomy in school children. Proceedings of the Royal Society of Medicine 31, 95–112.

Grytten, J., Sorensen, R., 2003. Practice variation and physician-specific effects. Journal of Health Economics 22, 403–418.

Gunnar, R.M., Bourdillon, P.D., Dixon, D.W., Fuster, V., Karp, R.B., Kennedy, J.W., Klocke, F.J., Passamani, E.R., Pitt, B., Rapaportet, B., 1990. ACC/AHA guidelines for the early management of patients with acute myocardial infarction: a report for the American College of Cardiology/American Heart Association task force on assessment of diagnostic and therapeutic cardiovascular procedures (subcommittee to develop guidelines for the early management of patients with acute myocardial infarction). Circulation 82, 664–707.

Hannan, E.L., Burke, J., 1994. Effect of age on mortality in coronary artery bypass surgery in New York, 1991–1992. American Heart Journal 128, 1184–1191.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70.

Krumholz, H.M., Chen, J., Rathore, S.S., Wang, Y., Radford, M.J., 2003. Regional variation in the treatment and outcomes of myocardial infarction: investigating New England's advantage. American Heart Journal 146 (2), 242–249.

Legendre, P., Legendre, L., 1998. Numerical Ecology, 2nd English ed. Elsevier, Amsterdam.

Liggett, T.M., 1999. Stochastic Interacting Systems: Contact, Voter and Exclusion Processes. Springer-Verlag, Heidelberg.

Morris, S., 2000. Contagion. Review of Economic Studies 67 (1), 57–78.

Nakagawa, S., 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behavioral Ecology 15 (6), 1044–1045.

Nichols, D., 2006. The impact of hospital racialization on racial differences in medical treatment. Unpublished working paper. Department of Economics, Washington University, St. Louis.

Norris, J.R., 1998. Markov Chains. Cambridge University Press, Cambridge.

O'Connor, G., Quinton, H.B., Traven, N.D., Ramunno, L.D., Dodds, T.A., Marciniak, T.A., Wennberg, J.E., 1999. Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project. Journal of the American Medical Association 281 (7), 627–633.

Perreneger, 1998. What's wrong with Bonferroni adjustments. BMJ Medical Publication of the Year 316, 1236–1238.

Phelps, C.E., Mooney, C., 1993. Variations in medical practice use: causes and consequences. In: Arnould, R.J., Rich, R.F., White, W. (Eds.), Competitive Approaches to Health Care Reform. The Urban Institute Press, Washington, DC.

Pilote, L., Califf, R.M., Sapp, S., Miller, D.P., Mark, D.B., Weaver, W.D., Gore, J.M., Armstrong, P.W., Ohman, E.M., Topol, E.J., for the GUSTO-1 investigators, 1995. Regional variation across the United States in the management of acute myocardial infarction. The New England Journal of Medicine 333(9) 565–572.

Smith, S.C., Dove, J.T., Jacobs, A.K., Kennedy, J.W., Kereiakes, D., Kern, M.J., Kuntz, R.E., Popma, J.J., Schaff, H.V., Williams, D.O., 2001. ACC/AHA guidelines for percutaneous coronary intervention (revision of the 1993 PTCA guidelines. Journal of the American College of Cardiology 37 (8), 2239i-lxvi.

Wennberg, J., Birkmeyer, J.D. (Eds.), 1999. The Dartmouth Atlas of Cardiovascular Health Care. Dartmouth Medical School and American Hospital Association, Chicago.

Wennberg, J., Cooper, M. (Eds.), 1999. The Dartmouth Atlas of Health Care. Dartmouth Medical School and American Hospital Association, Chicago.

Wennberg, J., Gittelsohn, A., 1982. Variations in medical care among small areas. Scientific American 182, 1102–1108.

Wennberg, J., Gittelsohn, A., 1973. Small area variations in health care delivery. Science 182, 1102–1108.

Young, H.P., 1998. Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press, Princeton, NJ.

Young, H.P., Burke, M.A., 2001. Competition and custom in economic contracts: a case study of Illinois agriculture. American Economic Review 91 (3), 559–573.