

# Custom-made versus ready-to-wear treatments: Behavioral propensities in physicians' choices<sup>☆</sup>

Richard G. Frank<sup>\*</sup>, Richard J. Zeckhauser

*Harvard University Boston, MA, United States*

Received 24 May 2007; received in revised form 17 August 2007; accepted 21 August 2007

Available online 6 September 2007

---

## Abstract

To customize treatments to individual patients entails costs of coordination and cognition. Thus, providers sometimes choose treatments based on norms for broad classes of patients. We develop behavioral hypotheses explaining when and why doctors customize to the particular patient, and when instead they employ “ready-to-wear” treatments. Our empirical studies examining length of office visits and physician prescribing behavior find evidence of norm-following behavior. Some such behavior, from our studies and from the literature, proves sensible; but other behavior seems far from optimal.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Norms; Physician decision-making; Prescribing

---

“Cognitive capacity is a scarce resource like any other . . . Given scarce cognitive resources, to rely on habit of other non-rational decision rules is fully rational.” (Robert Frank, 1987, p. 23)

“. . .the cost of coordinating a group of complementary specialized workers grows as the number of specialists increases.” (Gary S. Becker and Kevin M. Murphy, 1992, p. 1138)

“I Did it My Way.” (Frank Sinatra and Paul Anka, 1968)

---

<sup>☆</sup> An earlier version of this paper was presented at 25th Anniversary Conference of the Journal of Health Economics.

<sup>\*</sup> Corresponding author. Tel.: +1 617 432 0178.

*E-mail address:* [frank@hcp.med.harvard.edu](mailto:frank@hcp.med.harvard.edu) (R.G. Frank).

## 1. Introduction

Health care providers might base choices on norms for broad classes of patients, rather than customize their treatments in response to the economic and clinical circumstances presented by an individual patient.<sup>1</sup> That is, doctors might fail to optimize, either for themselves or for their patients, on a patient-by-patient basis. The idea that physician decisions are governed by norms conflicts with basic views of physician behavior in the field of health economics. Arrow (1963) launched the conception of the physician as an agent for the patient, implying that the physician would attend to each patient's situation, and act in her best interests. Subsequent research has discussed physician practice as a "workshop" (Pauly, 1980), or a "job shop" (Phelps, 2002). Either formulation suggests that physicians customize care at the individual level. Finally, it has long been observed that physicians price discriminate, behavior that represents another form of customized treatment (Kessel, 1958). This paper revisits the general question of whether doctors optimize at the level of the individual patient, and explores if not, why not. The medical care system has been transformed in the last 30 years, and new forces and new complexities surely influence physician decision-making. We review some recent studies, and present some sample empirical analyses on a set of physician choices.

We employ the evocative terms "custom-made" and "ready-to-wear" treatments, where the former would indicate selecting a treatment on a patient-by-patient basis, and the latter would imply following a norm. Though we discuss the literature on whether physicians respond to *financial incentives* at the individual patient level, our focus will be on the effect of clinical circumstances on *therapeutic decisions*. Thus, as two of our studies, we shall investigate prescribing behavior for antidepressants, and whether doctors respond to the complexity of a patient's condition when allocating their time in office visits.

One form of what came to be known as the norms hypothesis was that patients were treated according to the *incentives* in the modal or mean insurance plan in their geographic area, rather than the incentives in their particular plan (Ginsburg, 1976). Thus in areas where most patients had generous insurance coverage, a patient with a given level of coverage would receive more care than a patient with the same coverage in an area where typical coverage was stingier. We refer to this as the *incentives norms hypothesis*.

Newhouse and Marquis (1978) examined the proposition that medical care responds to area insurance arrangements, either in addition to or instead of individual insurance features. They did this by estimating demand models for individuals using data from the Ohio site of the RAND Health Insurance Experiment and national survey data. In their empirical analysis they showed first that individual insurance plan provisions do affect service utilization. They also showed, using a national survey, that average community insurance features do not significantly affect demand behavior. Their analysis supported the conclusion that when it came to responding to incentives, patterns of medical care were not driven by norms.

The Newhouse–Marquis findings emerged from an era dominated by fee-for-service medicine, where insurance contracts were simpler. Solo medical practice was the dominant organizational form in the 1970s. For example, in 1976, 50% of physicians were in solo office based practice, and about 30% in partnerships (NCHS, 1978). Today, only 30% of physicians are in solo practice (NCHS, 2006). Most doctors today practice in groups, or within some form of organized practice. Group practice can help individual physicians manage the care of individual patients

---

<sup>1</sup> Ginsburg (1976) offers this speculation.

through the use of information systems, and by managing highly differentiated reimbursement arrangements.

There are also strong forces in the modern era that make it much harder for doctors to tailor care at the individual patient level. Most importantly, the medical care system has become vastly more complex in recent years, as have its reimbursement practices. The typical American physician today holds 11 managed care contracts and serves patients from 15 health plans (NCHS, 2006). This makes it challenging to customize treatments on the basis of each particular patient's reimbursement arrangements.

Primary care physicians are also expected to do more today. Professional guidelines direct primary care physicians to screen for a variety of conditions (e.g., breast cancer, substance abuse, and depression), counsel patients on disease prevention measures, and manage a host of chronic conditions (e.g., diabetes, asthma, and congestive heart failure). In addition, payers are increasingly encouraging physicians to adhere to a variety of treatment guidelines, for common medical conditions and meet health-plan-specific criteria when admitting patients to hospitals or ordering high-cost tests. The cognitive challenge of providing customized treatments is enormous, and health plan rules magnify the challenge.

In today's world, the physician often has to work with numerous other parties, both within and outside her practice. For example, the increased role of organized medical groups means that physicians will interact with more fellow doctors and other professionals in 2007 than they did in 1978. The use of allied health professionals has expanded over time, so that today 48% of physicians use such staff compared to 30% in 1980 (CHSC, 2006). We argue below that when groups of disparate individuals must collaborate, it is likely to be more difficult to customize treatment.

In examining the choice between customized and ready-to-wear treatments, we focus on an area where developments over the past 30 years make it more likely that physicians would rely on norms, namely in selecting therapies for patients, leaving aside the issue of financial incentives. With the advance of science and medical knowledge, selecting the appropriate treatment for a patient has become a much more difficult task for a broad variety of conditions. Thirty years ago there might have been a single class of drugs (with a single mechanism of action) with a few primary drugs for treating a common condition, such as depression. Today there are multiple classes (that use different biochemical processes) of antidepressants, classes that might include half a dozen or more individual drugs. Similarly, today there are at least seven or eight primary treatments for prostate cancer, leaving aside combination treatments, in contrast to three treatments in decades past.<sup>2</sup> The advance of science has expanded the set of potential treatments for many conditions, and has increased the cognitive burden of selecting a treatment that best matches an individual patient's situation.

To be sure, there are also forces in modern health care that create pressure for more individualized care. The rise of consumerism and the expanded information available to patients through the Internet, health plans and popular publications leaves patients better able and more likely to demand care that accords with their preferences and clinical status. In addition, medical knowledge is increasingly providing information and tests for determining which treatments work best for particular types of patients, although many of these developments remain in early stages.

Our goal in this paper is to assess the contemporary resolution of the forces that push toward standard treatments, i.e., norms behavior, as opposed to those that encourage customized treat-

---

<sup>2</sup> Conversation with Mark Schoenberg, M.D., Professor of Urology and Oncology, Johns Hopkins Medical School.

ments for individuals in differing situations. The paper is organized into seven sections. Section 2 offers some background on the economic basis for norms and presents three general hypotheses about the use of norms in physician decision-making. Section 3 presents a conceptual approach to investigate the choice of treatments in the context of dynamic decision-making for treating a single patient. We posit that such decisions represent an example of the classic multi-armed bandit problem. That problem requires a choice among multiple actions whose distributions of payoffs are unknown, but can be learned through trials.

Section 4 examines a set of recent papers in health economics and health services research that suggest the possibility that physicians in the health care system of the late 1990s and early 2000s commonly use ready-to-wear treatments. Section 5 undertakes some illustrative empirical analyses of three specific medical decision-making situations. Section 6 presents our empirical results, examining each one in light of the evidence it provides on our three behavioral hypotheses. The final section offers a summary and conclusions.

## 2. Background and hypotheses

Ordinary consumers often use norms in making decisions, be they minor or monumental, sometimes to their detriment. Thus, they may adhere to budgets, even when special circumstances might call for their violation. They may drive the same route to work everyday, even when investigation of an alternative route might show that it is superior, and yield significant long-run savings. And they may trade in their cars periodically, even when they prove unusually reliable. But for the most part consumers' use of norms is sensible. To follow our sartorial metaphor, they buy ready-to-wear clothes, except possibly for a grand occasion, such as a wedding. And they buy "ready-to-wear" furniture except when they need custom-made built furniture for a particularly prominent location, or a hard-to-meet furniture challenge.

### 2.1. *The costs of customization*

Doctors, we posit, employ ready-to-wear treatments for the same reason individuals purchase ready-to-wear clothes: they are cheaper. If one tries to customize a treatment to each patient, a variety of costs are incurred. We identify them as the four Cs. Communication and cognition costs will be our focus. Two other relevant costs are coordination and capability.

*Communication costs:* Communication with patients to determine their symptoms, histories, and preferences can be both challenging and time-consuming. Groopman (2006) writes eloquently of the challenges that physicians encounter when they elicit information from patients to determine an optimal therapy. Physicians also must regularly attend to ideas and misconceptions that consumers bring to their illnesses and ideas about treatment. Consumers obtain information about diseases and their treatment from the Internet, direct to consumer advertising of prescription drugs, and from the popular press. Thus nested in the cost of communication are the costs of persuasion and negotiation over treatment.<sup>3</sup> All of these communication costs appear to be growing over time with the advent of expanded consumerism in health care (IOM, 2001), and expansion in the general complexity of medical care.

Sommers et al. (in press), examining treatment for prostate cancer, determines that patients vary significantly in their valuation of different side effects relative to life expectancy. Given

---

<sup>3</sup> The negotiation process has long been recognized in medicine. See for example Lazare et al. (1975). Kleinman (1987) details how patients come to visits with models of illness, agendas and expectations about how their physician will respond.

this, ideal customizing would select different treatments for patients with similar demographic characteristics and clinical symptoms. The authors observe from the literature, however, that such customizing is rare, presumably due to problems of communication and cognition. Drawing on a survey of 156 prostate cancer patients, and using a decision–theoretic model, they conclude: “. . . there is potentially significant benefit to be gained through using an impartial, evidence-based decision model that explicitly accounts for the preferences of each individual patient.” (Sommers et al., *in press*, p. 16)

*Cognition costs:* The challenges to human cognition have been elucidated by the research pioneered by such greats as Herbert Simon, and Amos Tversky and Daniel Kahneman. Given the difficulties of rational calculation, individuals often resort to satisficing behavior (Simon, 1958), and to the use of heuristics or rules of thumb (Tversky and Kahneman, 1974; Frank, 1987). Such heuristics usually are selected because they do a reasonable job for a broad array of cases.<sup>4</sup> However, there is systematic evidence that such heuristics can frequently lead decision makers astray in a range of situations, particularly when probabilistic outcomes are involved, as is almost always the case with medicine. The power of the field of behavioral economics has developed from the broad insight that heuristics can lead to significantly suboptimal behavior.

*Coordination costs:* As hard as it is for a single decision maker to act fully rationally, it is much harder for a group. Often information is not fully or effectively shared. For example, many individuals may have modest slivers of information pointing to B rather than A, but the group as a whole may persist with A. This is particularly true if those with information favorable to B would incur a cost in revealing that, say because they would have to be the “odd man out,” or take an action that would be personally costly.<sup>5</sup> This is known as herding behavior. (Banerjee, 1992)

Adam Smith’s famed pin factory showed the advantages to the division of labor. A major portion of the gains come because pins are a commodity with standard dimensions and composition. Posit counterfactually that pins were sold on a one-by-one basis, and that slightly stronger ones would be preferred for pants over shirts, say, and slightly longer ones would be better when the shirts were flannel, and slightly smoother ones for silk. The manufacturer might choose to produce a standard pin, rather than attempt to optimize as each customer came along. He would be even more likely to do so if his same production team also had to turn out other products to produce, such as staples and paper clips. By standardizing the pins, significant coordination costs would be avoided.

Medicine is a field bearing the three key features of our counterfactual pin factory: considerable specialization and division of labor, a greatly varied product mix, and “customers” with quite different needs being served in succession rather than batches. Medicine thus adds two elements to Becker and Murphy’s (1992) pioneering study of the degree of specialization as the prime source of coordination costs. We employ the label *coordination* to refer the distinctive challenges to flexible treatment when teams of practitioners face patients who, though superficially similar, may differ in some important respects.

For communication with patients, complexity of choice, and coordination with partners, it is convenient to think of a fixed cost per patient to determine and cater to her particular needs.<sup>6</sup> These

<sup>4</sup> On the use of heuristics see also Gigerenzer and Selten (2002) and Gigerenzer and Goldstein (1996).

<sup>5</sup> When there is a possibly unattractive value tradeoff implicit in a decision, that decision too will be inhibited. For example, a doctor may not wish to tell his receptionist: “Don’t schedule a follow up visit for this patient; her reimbursement is likely to be insufficient.”

<sup>6</sup> Glied and Zivin (2002) characterize such fixed costs as a reason underlying the use of incentive norms by physicians dealing with multiple payers. That work is discussed below.

costs can be avoided if a single therapy is selected for a broad class of patients. Hence, there is an incentive to select ready-to-wear treatments for patients who differ in important respects though they are superficially similar.

*Capability costs:* Doctors may also face a fixed cost in having a capability available, the equivalent of an up-front investment cost. For example, a doctor who is well acquainted with therapy C may have to spend some time to learn how to provide therapy D. That may lead her to specialize in C even if it is dominated by D. Thus, doctors, or some group practices, and even some locales might specialize in particular classes of treatments. If doctors do have particular expertise or experience with particular treatments, the patient may benefit if they receive that doctor's preferred treatment even if they might be better off with another doctor providing another treatment. That is precisely the conclusion that [Chandra and Staiger \(2007\)](#) reach (see discussion below) in considering treatments for heart attacks. Such treatments vary substantially across physicians and also across geographic areas. Even modest capability costs may influence choices, particularly for relatively simple therapies. Thus, a doctor may tend to prescribe a particular pharmaceutical when she has little experience with its competitors. Where there are capability costs, whether strong or mild, we would expect wide variation across individual physician practices.

The majority of this paper documents the tendencies of doctors to rely on ready-to-wear treatments for various conditions. We rarely judge whether taking all cost factors into account such practices are net undesirable. However, we do note in some instances that convenient heuristics can lead to substantially suboptimal treatments that are presumably harmful to patients relative to reasonable alternatives. For example, we shall observe that primary care physician treating depression tend to ignore changes in symptoms when determining whether to change the patient's dosage or medicine. Given that dosage and drug choice are so important in this context, and that primary care physicians meet frequently with their patients, implying that adjustment costs would be small, this situation reflects a troubling bias.

Chemotherapy treatment for a variety of cancers offers an intriguing example where standard treatment ignores the potential for customization. Most commonly, treatment is provided at fixed intervals, say monthly for four treatments. The periodic dose is determined by multiple factors; weight is usually the most significant. Dosage represents a trade-off between increased therapeutic benefit and increased side effects. The patient's white blood count is closely monitored. The second dose will be delayed if it has not returned to an acceptable range by the end of the month. By contrast, if the count falls less than expected and rebounds quickly, the dose and schedule are maintained.

Viewed from the standpoint of a decision theorist, this approach is odd. The original (at least implicit) cost-benefit analysis determining the dose was for a representative individual with certain characteristics, but there is a great deal of individual variation. Indeed, some drugs, such as the widely employed Cytosan (cyclophosphamide), are first processed by the liver and then set out to do their work, and patients' livers certainly differ in their effectiveness.<sup>7</sup> It would seem that the slowly recuperating patient has received, on balance, too high a dose, and that the patient whose blood count fell little received one that was too low. Yet except when patients have extreme side effects, their dosages are not adjusted on the bases of how they respond.<sup>8</sup> Rather, oncologists operate on the basis of the results from randomized controlled trials. For such trials, these types

<sup>7</sup> Colvin (1981), p. 25.

<sup>8</sup> One experienced oncologist observed: "sticking to the protocol [makes sense] if the goal is cure. Cure was demonstrated in the given regimen in the trial, and the goal is not titrating to some desired level of symptoms. [Better to] control the side effects." If the cure were sure, or if the side effects could assuredly be controlled, we would agree. However, neither

of adjustments in dosage, as suggested by logic, have not been tried. Experience with periodic radiation dosage is similar: the protocol is not adjusted as experience is gained with an individual patient.<sup>9</sup>

One thing is clear. Chemotherapy doses for standard treatment are determined on the basis of factors that have been tested on some fairly uniform basis in Phase III trials. The doses are not adjusted given experience with individual patients, neither in how they respond to doses in earlier rounds, nor in other factors that doctors feel may be significant, such as alcohol consumption. Cancer treatment of the future will be strongly oriented to specifics of the clinical condition of the particular patient, including individual genetic makeup.<sup>10</sup> Unfortunately, the future will be somewhat delayed because of the propensity, and sometimes the imperative, not to deviate from the norms that have emerged from clinical trials. Excess clustering of physician behavior may be reinforced by the proliferation of practice guidelines accompanied by increased public reporting on performance relative to recommended care patterns.

## 2.2. The rationality-bias spectrum for norms-based behavior

Physician decisions must all deal with the set of four costs we have outlined. Our focus in this paper is to provide a framework for understanding the conditions under which physician decision-making may move away from patient-by-patient optimality or custom-made treatments. We attempt to classify the empirical findings of this study, and of other studies we review, on a spectrum ranging from hyper-rationality at the top to bias at the bottom. *Hyper-rationality* implies that the physician effectively optimizes for each patient, given the information available at the time. Hyper-rationality assumes that costs of coordination and cognitive decision are *de minimus*, in some sense following the extreme version of rationality sometimes posited in economics.

Just down from hyper-rationality is a behavior that we label *decision-cost rationality*, or *DC-rationality*. DC-rationality recognizes that it is expensive to process information, coordinate actions, communicate with patients, etc. Thus, doctors take reasonable decisions in light of the costs they face. DC-rationality suggests a certain level of purposeful optimization, almost calculation, in choosing one's decision approach.

One of the central findings from behavioral economics is that individuals often employ a decision *heuristic* that works reasonably well over a broad array of circumstances, but can be far off the mark in others.<sup>11</sup> Following a norm or a ready-to-wear treatment would be one such heuristic. Uncertainty situations, particularly those involving learning and low probabilities, appear to be particularly vulnerable to off-the-mark decisions. Alas, many medical decisions have these characteristics. The choice of heuristics implies less attention to purposeful optimization than does our concept of DC-rationality. Simon's satisficing behavior (1958) would seem to represent an amalgam of DC-rationality and heuristic behavior.<sup>12</sup>

---

condition holds. It has also been suggested that concerns over malpractice lawsuits make following protocols a way to protect against such litigation.

<sup>9</sup> See Berkarda et al. (1984), p. 115 and Freireich and Bodey (1984).

<sup>10</sup> Conversation with David Rosenthal, M.D., head of Harvard University Health Services, and former President of the American Cancer Society.

<sup>11</sup> The classic works in this line of research include: Tversky and Kahneman (1974) and Kahneman and Tversky (1979).

<sup>12</sup> Bias arises when decision-making falls below the standard on invoking a heuristic that is a plausible decision approach. They fall prey to systematic biases, to follow the terminology of the Tversky and Kahneman (1974) article. For example, individuals tend to be far too confident of their own ability to assess uncertain quantities (Alpert and Raiffa, 1982). Here too, we shall see examples of biased decision-making, the lowest level of performance on our spectrum.

### 2.3. *Three behavioral hypotheses on the use of norms*

Our characterization of the modern physician's decision-making environment and its costs lead us to three general hypotheses. We consider medical encounters in three contexts to determine to what extent physicians' customize their treatments to their patients' clinical symptoms, and possibly preferences, and when they select treatments for a "representative" patient. Where they do the latter, we say that the Therapeutic-Norms Hypothesis applies.

**Therapeutic-Norms Hypothesis.** Doctors select their treatments for a representative patient within a category, rather than customize their treatments for each individual.

We do not indict doctors who employ therapeutic norms. That may be the only reasonable or feasible response to complicated decisions in the face of our four sets of decision costs. Consistent with the idea that a therapeutic norm may represent a sensible response to complexity and uncertainty, we pose a second broad hypothesis, namely that doctors will employ therapeutic norms where the returns to customizing will be the lowest. We do not suggest that doctors, make or act as if they make refined judgments on when to use rules-of-thumb in selecting therapies and when to specifically customize. Rather, we posit that their choices between customizing and following norms will show a broad logic. Customizing is more expensive and will be employed on average where it yields greater returns. By contrast, norms will be employed where the savings in cognitive or coordination costs are great relative to the loss in not selecting the optimum treatment for each patient at each point in time. This leads to

**The Sensible-Use-of-Norms Hypothesis.** Physicians will make use of norms where they make the most sense.

Much of our empirical work and analysis of the literature will be directed to testing the strength of this hypothesis, to which we apply the acronym SUN Hypothesis. To illustrate, the SUN Hypothesis would suggest that doctors are more likely to customize treatments for chronic conditions, where they will encounter the patient on successive occasions and learn how she responds, than for acute conditions, where the episode of care is short lived. There are two reasons: (1) doctors will have opportunities for learning, and (2) any fixed costs of customizing for the specific patient will be spread over more treatments.

Even if the SUN Hypothesis is supported for some classes of treatments and conditions it may not describe all or even a majority of clinical choices. In some instances, we conjecture, doctors will make treatment choices based on habit or familiarity with a procedure, sometimes with little consideration of the match between an individual patient and the treatment. This leads to our third behavioral hypothesis:

**My Way Hypothesis.** For many important conditions, physicians will regularly prescribe a therapy that is quite different from the choice that would be made by other physicians.

It makes sense for a physician to select a therapy for which she has differential expertise. But the My Way Hypothesis may also apply when a physician has had personal "good luck" with it, a plausible heuristic, but one that falls prey to the Availability Heuristic, namely overweighting evidence that one can bring easily to mind. The doctor's personal experience counts more than a possibly quite extensive literature. My Way choices may also reflect serendipitous factors, such as which drug for a condition a doctor encountered first, or which drug company promotion she first encountered. Worse than serendipity, we expect some occasions where doctors' idiosyncratic choices reflect significant biases. The My Way Hypothesis emphasizes variation in treatment



patterns across physicians. A related literature targets variation across areas and is grounded in geographic patterns of treatment and possibly local norms. (See Wennberg and Cooper, 1997; Phelps, 2000; and an extensive additional literature due to Wennberg and colleagues.) If individual choices by physicians are interrelated and related to local conditions, then My Way behavior may be connected to geographic variation.<sup>13</sup>

Note that our three behavioral hypotheses are strongly related. If there were no use of therapeutic norms, neither the SUN Hypothesis nor the My Way Hypothesis would likely be supported. More important, in many contexts the SUN Hypothesis works against the other two. If either the Therapeutic Norms or My Way Hypothesis is sufficiently powerful, some nonsensible norms are sure to be followed.

### 3. Dynamic decisions framework

To practice modern medicine effectively is like taking an examination in advanced decision analyses with wholly inadequate time. Beyond that, there is the expectation that the examinee – the doctor – has in hand a vast array of conditional probabilities, namely how medical conditions respond to alternative therapies. In addition, almost invariably choices must be made swiftly in the context of incomplete information about critical aspects of a person's medical condition. Far from ideal decisions are inevitable. Gawande (2002, p. 4) describes medical decision-making in the following terms:

“What you find when you get in close, however—close enough to see the furrowed brows, the doubts and missteps, the failures as well as the successes—is how messy, uncertain, and also surprising medicine turns out to be.”

The complexity of practicing medicine has increased markedly on all fronts. The United States is a more diverse country with a richer mix of cultures, races and ethnicities than in the past. Attitudes towards illness, responses to treatment, and genetic patterns are different from what they were just 20 years ago. Insurance arrangements have become more fragmented. As noted earlier, a typical primary care physician serves patients from numerous different insurance plans. Each health plan has different prices, coverage and constraints on utilization of care and prescribing. Finally, medical science has dramatically expanded the capacity to treat disease. There is a continuous flow of new drugs, diagnostic tests, surgical procedures and medical devices. New options for treating familiar conditions emerge regularly.

Given these uncertainties, and the complexity of the stream of decision tasks that must be undertaken in limited time, it is inevitable that physicians face the types of decision costs we outline above.<sup>14</sup>

To understand what doctors actually do, and how their actions compare to optimal choices, we start with the rational decision model (Gittins and Jones, 1974; Whittle, 1980). We characterize physician treatment choices as the classic multi-armed bandit problem, hereafter called

<sup>13</sup> There is an extensive literature on geographic variation dating to the 1970s. For a recent paper showing how optimizing behavior can result in norms see Chandra and Staiger (2007) that is discussed at some length later in this paper.

<sup>14</sup> Conlisk (1996) refers to such costs as deliberation costs. Frank (1987) suggests that in such situations adopting rules of thumb or norms may be second best efficient. Gigerenzer's ideas about fast and frugal heuristics see norms as efficient responses to deliberation costs (Gigerenzer and Selten, 2002). Gigerenzer and Goldstein (1996) argued that many heuristics are “ecologically rational”. Tversky and Kahneman (1974) shared this general insight, but they are somewhat less optimistic.

the *bandit problem*. The bandit problem represents a situation where an individual must choose between alternative lotteries whose payoff distributions are unknown, but can be learned through experience. In the original formulation, the decision maker plays  $N$  slot machines – i.e., one-armed bandits – in a sequence of trials, seeking to maximize her discounted expected reward. The challenge is that the slot machines are known to carry different and unknown pay-off distributions.

This model sets out an important trade-off that well characterizes physician decisions. Dealing with a single patient, there is a trade-off between costly exploration (trying out different alternatives in an effort to learn and increase future returns) and immediate returns (making the choice that yields the best payoff in the current period). Consider a patient who is likely to be on an anti-hypertensive medicine for the rest of his life. Medicine A, the first tried, proves quite satisfactory in managing his blood pressure and incurs only minimal side effects. Nevertheless, it may be worth trying medicine B, even though it is expected to be inferior, because it could turn out to offer even better blood pressure control or lesser side effects, and the patient will be using it for many years. The logic supporting such experimentation is clear, though for many it is far from intuitive. Note also, consistent with the bandit model, even when a patient has considerable experience with a medicine, there is a chance element about its efficacy and side effects in any brief period.

The precise nature of when to switch among options challenged statisticians and decision theorists for years. Fortunately, a simple conceptual solution was found. Each lottery – i.e., a medicine in this instance – can be characterized by its Gittins index, a number that depends only on the history of its pay-off to date (Gittins and Jones, 1974). The Gittins index characterizes the attractiveness of a lottery relative to some standard lottery offering a constant certainty payout. The lottery with the highest Gittins index is played at each point in time. As success improves, a lottery's Gittins index rises, and vice versa. With two highly uncertain lotteries, there can be lots of back and forth before one is finally selected. The lower the discount rate, and the greater the number of trials, the more switches there will be on average.

We describe the bandit model to illustrate a prime implication of our SUN Hypothesis, namely the way a physician should treat two kinds of patients, those with acute conditions and those with chronic conditions. Each condition can be treated with one of  $N$  drugs. Each treatment offers benefits defined by random variables  $b_{mjt}$  where  $m$  is the drug chosen,  $j$  indexes the patient, and  $t$  indexes time. The doctor will have prior distributions on the effectiveness of the various drugs, from experience and from the literature. In some cases, those priors may depend on the patient's characteristics, e.g., gender, age, comorbidities, not merely his condition. Given that patients respond differently to drugs, even the best informed doctor will be uncertain about the distribution of at least some of the  $b_{mjt}$  values. Thus, the bandit model applies, and there are potential gains to exploration.

The hyper-rational physician will, implicitly, compute the Gittins index for each drug, and administer the drug with the greatest value. She will continue with that drug unless and until its Gittins index falls below that of some competitor. In practice, of course, the doctor's implicit Gittins index values for the different drugs will only be rough approximations.

We focus on prescription choice from an office visit, since that is our empirical application below. This multi-armed bandit problem collapses to a simple solution for acute illness. Acute illnesses, such as an infection, are characterized by an immediate need for treatment and a condition that if properly treated will be of short duration with the individual returned to their usual level of activity. Thus, the physician may be most inclined to choose a drug that is most likely in their experience to alleviate the problem and does not threaten safety (e.g., as would allergic reactions to antibiotics). Side effects may be a secondary consideration because of the relatively short duration of care. Most important, there is likely to be a single period of use of the drug in most

instances; there will be little learning. This suggests that the doctor is likely to have a one-time only decision, unless the patient does not respond. Barring foreseeable differences in responses across patients, all patients suffering from a particular acute condition should be given the same drug. The solution to the bandit problem in this case supports the SUN Hypothesis that predicts considerable concentration of prescriptions for acute illnesses.

A patient with a chronic condition presents a full-blown bandit problem, with attention to uncertainties that get resolved over time. A physician would understandably select the same drug for the initial treatment of a chronic condition for different patients. However, over time experience will suggest that a drug is not working well, or is imposing troublesome side effects. Indeed, even though it may be working reasonably well, another may prove more satisfactory. We would not expect the doctor to approximate the optimal bandit strategy, but even use of a relatively crude heuristic would lead her to switch medicines as she learned. In effect, her implicit internal Gittins index for her current prescription may fall below that for a competitor drug. If doctors do behave roughly in this manner, as the SUN Hypothesis would imply, patients with a particular chronic illness would be treated with a larger variety of drugs than those with an acute illness.

We shall examine this proposition below. Recognizing behavioral propensities, such as status quo bias (Samuelson and Zeckhauser, 1988) suggests also that physician choices may fall short of DC-rational behavior or even sensible-heuristic behavior; we simply expect to observe a rough pattern of sensible behavior.

In prescribing effectively for the treatment of chronic disease, there is not only the question as to whether doctors learn from experience with a patient, but whether they make effective use of the information available. Consider a straightforward example where a patient's condition is indicated by an index. The doctor is trying to select drugs and dosages to improve the patient's score. To do this effectively, the rational or DC-rational, or even sensible-heuristic doctor will assess the patient's present condition and also how the patient responded to medicines and dosages in the past. In theory, she will work out an implicit algorithm that looks at such responses in determining future prescriptions. For example, she will likely keep the patient on medicines that have generated positive responses, and switch medicines for those that have had no response or incurred significant side effects. Similarly dosages may be increased when there was little response, and decreased where side effects were incurred.

To set out the decision problem more formally, at each point in time the doctor must select a medicine,  $m_t$ , and dosage,  $d_t$ . The information she has is the vector of medicines and dosages given to date, and the patient's condition values to date, label them  $C = (c_1, c_2, \dots, c_t)$ . The SUN Hypothesis would predict that the likelihood that  $m_t$  differs from  $m_{t-1}$ , and that  $d_t$  differs from  $d_{t-1}$  will depend on other elements of the  $C$  vector beyond  $c_t$ . A reasonable heuristic might only look at the past few values, but it surely would attend to differences between  $c_t$  and  $c_{t-1}$  when considering a change in medicine or dosage at time  $t$ . This is the prediction from the SUN Hypothesis that is studied in our analysis of prescribing for depression.

#### 4. The literature: recent findings on norms behavior

A number of recent papers show how norm following affects both clinical choices and the allocation of resources by doctors. Their topics include decision-making about prescription drugs, the use of physician time, and the application of surgical procedures. As a group, these papers uncover empirical regularities suggesting norm following behavior, the use of ready-to-wear treatments. We review some prominent contributions.

Hellerstein (1998) examines physicians' choices in prescribing generic versus brand-name pharmaceutical products, where the physician serves as agent to the patient and, though less recognized, also as agent for the health plan. Brand name and generic drugs are very close to being perfect substitutes in production. In fact the U.S. Food and Drug Administration refers to generic drugs as "chemical carbon copies". Hence, the generic-versus-brand name choice represents an ideal test of whether the doctor customizes her choices as agent to save dollars for patients and plans. If she does, she would be more likely to choose generics when the patient had less generous coverage. Similarly, we might expect some patients, e.g., older patients, to be more nervous about generic drugs. Therefore one might expect insurance arrangements and patient factors to play a relatively strong role in decisions about whether a branded or generic product is prescribed.

Hellerstein uses the National Ambulatory Medical Care Survey (NAMCS) to analyze this question. She finds that payment arrangements, as measured by a dummy variable measuring whether or not an individual had health insurance through an HMO – HMOs tend to provide strong incentives to choose generics – in fact increases the likelihood that a generic would be prescribed. No other characteristics of the patient had any significant impact of generic prescribing rates.

By contrast, observed and unobserved characteristics of the physician were very important in explaining the propensity to prescribe generics after controlling for patient factors, insurance variables and regulations. In fact, unobserved physician-specific factors explained about 30% of the residual variance in generic prescribing. While the NAMCS measures patient demographic, diagnostics and insurance arrangements, these measures clearly do not capture all features of the patient's clinical conditions. Thus clustering of patients across physicians by severity of illness may lead to attributing patient choices to physician-specific variation. Nevertheless, Hellerstein (1998) interprets the portion of the variation in generic prescribing associated with the individual physician in a fashion akin to our My Way Hypothesis. We see such doctor-specific propensities on a task that is encountered everyday. This suggests strong adherence to standard procedures, to norms behavior, which is more remarkable given that it is unrelated to such factors as age or medical specialty. Other explanations for this observed variation cannot be dismissed. For example, since the characteristics of local areas are only crudely accounted for, there may be local market factors, such as those identified by Wennberg and Cooper (1997) and Chandra and Staiger (2007) that would generate such findings. On balance, Hellerstein's (1998) basic findings are consistent with the My Way Hypothesis.

The Hellerstein analysis considered physicians as agents for patients. Glied and Zivin (2002) explore physician choices that affect their own financial well-being. They look at visit durations, tests ordered, medications prescribed and the scheduling of follow-up visits. If the Newhouse–Marquis results extrapolate to different circumstances, and somewhat different questions, a quarter century later, the authors should find that physicians choose the quantity and intensity of their treatments in response to the incentives generated by the payment arrangements of their individual patients.

Glied and Zivin also use the NAMCS data in their study. They observe physicians that serve patients from multiple payers, some under managed care others not. They assess whether patients treated in the context of managed care arrangements receive shorter visits, fewer tests and drugs, and are less likely to be advised to return for a follow-up visit as might be expected under capitated managed care arrangements.

Glied and Zivin find that an individual's payment arrangements matter in the expected direction, which suggests some customization. Perhaps more surprising, they find that the patient's actual arrangements matter less than the payment arrangements of other patients in the same practice.

That is doctors utilize modes of practice that respond to their overall patient population. Using our terminology, though doctors customize a bit, they have a significant tendency toward ready-to-wear practices.

Glied and Zivin posit that such patterns may be due to “fixed” per-patient costs, which is one strand of the decision-cost rationality story we outlined above.<sup>15</sup> In discussing these “fixed” costs they suggest that such costs may be cognitive, which is consistent with the idea of a “norm” in the sense used by Frank (1987). It is also consistent with coordination costs within a practice of switching from one approach to treatment to another. The Hellerstein and Glied and Zivin studies together suggest a form of the incentives norms hypothesis, once rejected by Newhouse and Marquis, may have been given new life by the more complex medical circumstances of the modern era.

Tai-Seale et al. (2007) use videotapes to probe the content and time allocations of 392 common primary care physician visits. The median visit duration across practice types – i.e., academic medical center, inner city clinic, or managed care group practice – differed significantly. Our primary interest is in variation within practice types. Optimization of care would lead visit lengths to vary according to personal characteristics, the types of problems in need of attention and the number of problems needing care. They find that within an organizational type there was little variation in duration of visits, even when the number and complexity of topics presented by a patient varied. The observed regularity of visit durations within a setting is consistent with the use of a “rule of thumb” about visit length. If so, a type of visit length norm is being employed. One consequence of the visit length norm is that more time was devoted to issues raised early in a visit at the expense of topics raised later in the encounter. The net result is that all topics were generally fit into the established visit length norm, which may have led some important topics to be short changed (Tai-Seale et al., 2007).

Treatments of chronic illnesses are thought to involve sophisticated trial and error to find the “right” combinations of care elements that will allow a condition to be effectively managed. Henke et al. (2007) examined treatment adjustments by primary care physicians for their patients with major depression. They analyzed detailed information on treatment adjustments, patient symptoms, personal characteristics of patients, and physician characteristics. The clinical data were obtained through independent clinical assessments by researchers. Thus the clinician may or may not have been aware of the symptoms and their changes. The statistical analysis of about 1820 depressed patients showed that while numerous adjustments are made to treatment over time, physicians appear to respond to the level of symptoms and not changes in symptom patterns. Thus, after controlling for severity of symptoms, patients showing improvement were as likely to have their treatment adjusted as were patients showing no improvement or even a worsening in symptoms. These results suggest that physicians are failing to either observe or use information about a patient’s progress. This may be due to the absence of information in clinical records, the failure to ask the right questions, or the inability of patients to communicate changes. Nevertheless the focus on a single readily observable and critical indicator to make decisions, namely level of symptom severity, represents a troubling practice pattern. Consistent with the idea of high cognitive and communication costs, Henke et al.’s results reflect a focus by doctors on the most available of clinical information. Such a focus is likely to miss vital aspects of an individual patient’s clinical circumstances.

---

<sup>15</sup> This observed pattern of physician choices is related to a literature that views quality of care as a local public good. In this case however, physicians appear to tailor their treatments to the modal payment arrangements holding constant patient characteristics, a practice likely to conflict with many notions of quality at the individual patient level.

These studies offer views of physician decision-making that conflict with simple notions of optimization. Some suggest the use of rules of thumb or norms as a response to the high cognitive costs of optimization at the individual patient levels (Glied and Zivin). Others imply that some existing norms may be unproductive compared to alternative rules that might yield sufficient reductions in cognitive loads to be manageable. This is perhaps best illustrated by Hellerstein where extant norms about generic prescribing may be raising costs for essentially no clinical gains in contrast to a “always prescribe generics if they exist” rule. Henke’s results highlight a troubling inefficiency stemming from relying on the most available clinical indicators. In what follows we place internal norms, such as those discussed above, in a decision-making framework that allows for some simple tests of the role of norms to be formulated.

We should not conclude that just because doctor behavior is consistent with our behavioral propensities that it is far from optimal. A deeper probe may reveal a reassuring explanation. Recent research by Chandra and Staiger (2007) examines the substantial geographic variation in the use of procedures to care for heart attacks. This reflects the possibility that there are “unproductive norms,” such as My Way behavior, at work. Chandra and Staiger (2007) propose instead a model that focuses on productivity spillovers related to local specialization. That is excellence in one clinical approach in a market raises the average skill of the practitioner of that approach operating in the same market. This in turn results in greater specialization in the market and reduces both the absolute and relative productivity of practitioners using alternative approaches. Using data on treatment for heart attacks from Medicare claims, they test this productivity spillover explanation for geographic variation in treatment patterns. Their econometric analysis produces evidence that is strongly consistent with the spillover theory. Thus homogeneity in clinical approach within a geographic area, and substantial heterogeneity across areas, reflect “norms” that may be productive because they stem from positive spillover effects from local specialization. Though some patients would be better off were they treated differently by doctors in other areas, Chandra and Staiger propose that such “norms” may be efficient given current capabilities within their area.

## 5. Empirical analyses

The Armed Bandit model and the SUN Hypothesis suggest that in the context of physician decision-making norms or rules of thumb may be most useful where: (1) little is known about a patient and their responsiveness to various treatments, (2) treatment is expected to be short lived, and or (3) there is little difference in the impact of different treatments on patients. These features lead us to propose a set of empirical analyses to examine whether patterns of treatment are consistent with our ideas about decision costs and the trade-off between ready-to-wear and custom-made treatments. If so, we then inquire whether they are consistent with productive use of norms. In constructing our empirical analyses we focus on physician prescribing and time allocation decisions. We do so because these choices reflect more general therapeutic decisions and are relatively simple to observe and measure. Obviously, a vast range of studies of norms-based behavior in treatment could be undertaken. Ours were selected on the basis of relevance and data availability, and their findings should be thought of as illustrative rather than reflecting some immutable pattern.

The first case follows directly from the discussion of the Bandit model and the SUN Hypothesis. In particular, we identify a set of acute and chronic medical conditions that each can be treated by any of a number of different drugs. The conditions are suffered by groups of heterogeneous patients. The drugs used to treat the chronic conditions offer different combinations of side effects. In addition, the different drugs may vary in their effectiveness across different types of patients. We observe the choices of primary care physicians who are commonly responsible for

treating and managing both sets of conditions. The acute conditions we observe include: upper respiratory illnesses, ear infections, pneumonia, and strep throat. The chronic illnesses include: high cholesterol, diabetes, asthma, hypertension, and depression. Our ideas imply that a norm or rule of thumb may be most productive if deployed for acute conditions (short periods of treatment) where there are few *expected* differences in performance across patients, and multiple relatively equivalent drugs are available. Given such conditions, we expect some doctors to mostly prescribe drug A, others to mostly prescribe drug B, etc.

Therefore we examine the concentration by physician of prescribing in acute versus chronic conditions with multiple treatment options. Because the period of treatment is different and the long-term consequences of differences in side effects and effectiveness are different we expect to observe the following pattern of prescribing:

Prescribing concentration by physician for the treatment of acute conditions will be significantly higher than for chronic conditions.

If we observe such patterns, that would be consistent with SUN Hypothesis, namely that doctors make sensible use of norms. Obviously, it would also be consistent with the My Way and Therapeutic-Norms Hypotheses.

Our discussion of coordination and persuasion costs (a component of communication costs) suggests that there may be substantial difficulties in customizing visit times so that they fit the economic and clinical circumstances of individual patients. [Glied and Zivin \(2002\)](#) view these as fixed costs of customizing, and [Hellerstein \(1998\)](#) observes physician-specific idiosyncrasies in choices among perfect substitutes that carry substantially different prices (often to their patients). Thus following [Tai-Seale et al. \(2007\)](#), we examine the importance of physician-specific characteristics in determining the time spent in patient visits. Coordination costs would suggest that physician-specific factors would be most important in explaining variation in visit times. Persuasion costs, including justifying short visits to patients, may also help to explain uniformity in visit times. Thus we will examine data on patterns in visit times. We would expect to observe patterns consistent with the following statement:

The inherent complexity of decisions will have limited impact on the duration of physician visits. Physician-specific factors will account for disproportionate variation in the duration of physician visits.

If the analysis of our data shows substantial uniformity in visit times across patients and significant physician-specific variation, that would be consistent with both the Therapeutic-Norms Hypothesis and the My Way Hypothesis. It is important to note that if “sicker” patients visit their physicians more frequently, that may serve as a mechanism to grant them more time. Thus we are able check the degree to which different physician and patient-related factors explain which patients have more frequent visits. However, this analysis will be unable to separate patient demand factors from physician responses to illness features. Inflexibility in visit times not compensated by differences in visit rates would also be disheartening, since in an optimal world more complex cases would receive much more time than simple cases.

Finally, we examine the basis upon which adjustments in the management of medication treatments are made.<sup>16</sup> These include changes in drugs used to treat a particular condition (depression)

---

<sup>16</sup> Note that our work differs from [Henke et al. \(2007\)](#) in two respects. First, we focus on two specific clinical changes: changes in drugs prescribed and dosage adjustments. Second, we examine the source of unobserved variance and decompose it between MD and patient.

and adjustment to dosages. We examine data from a clinical study on quality improvement for treatment of depression. We study the factors that explain the likelihood that a medication switch takes place and whether an adjustment is made to the dose prescribed. The Bandit model would suggest that a physician would observe a patient's response to care and the tolerability of the treatment in assessing the benefits to exploration. Patients that show improvement would lead a physician to continue the existing treatment, whereas reports of either disturbing side effects or no progress in symptom relief (or indeed a worsening of the illness) would lead a physician to explore other treatments. We therefore examine the factors that explain variation in medication changes and dosage changes by primary care physicians treating depression, a common chronic medical condition.

The Henke study leads us not to be optimistic about what we will find in this investigation. We thus posit finding support for the Therapeutic Norms and My Way Hypotheses, but securing results that are strongly in conflict with the SUN Hypothesis. (We outlined the competition between the SUN Hypothesis and the other two at the conclusion of Part II.) That is, we are positing finding doctor behaviors that appear to make little sense.

When multiple drugs are available for a chronic condition, and where dosages vary substantially across patients, drug switching and dosage adjustments will be driven by demographic, immediate clinical factors, or physician-specific factors. Patient responses to treatment will play little role.

If clinical indicators such as severity of symptoms and changes in symptoms are important in explaining changes in antidepressant treatments, that would be consistent with both rational behavior, and indeed customization of treatment. However, if these factors are not found to be important, the SUN Hypothesis would have to be rejected.

### 5.1. Data

We make use of two data sets to examine the three hypotheses set out above, the National Ambulatory Care Survey (NAMCS) and Quality Improvement in Depression study (QID).

*NAMCS:* The 2004 NAMCS is a national probability sample of physician offices conducted by the National Center on Health Statistics (NCHS). The unit of observation in the data is the patient visit to an office-based physician. The NAMCS uses a multi-stage probability sample design. This involves nesting of visits within physicians, and then physicians within a primary sampling unit (a geographic area made up of a county or group of counties). The data are based on the responses of 1372 physicians who agreed to participate in the survey. (The response rate was 70%.) Each physician reported on about 30 randomly selected visits that occurred over a 1-week period. This produced a sample of 25,286 visits.

The information about each visit includes: the reason for the visit, physician diagnosis (using ICD-9-CM), medications prescribed (identified via NDC codes), tests ordered, referrals made, duration of visit, demographic characteristics of the physician and the patient, insurance coverage of the patient, type of practice (e.g., solo, group), whether the patient was a new or a continuing patient, specialty of the physician, location by the four census regions, and location with respect to being in an metropolitan statistical area (MSA). In the analyses presented below, we rely primarily on the NAMCS prescribing information to construct the outcome variables for testing prescribing patterns for chronic versus acute illnesses and for examining visit times to primary care physicians.



In the analysis of the NAMCS data we use two specific subsets of the data. The first subset uses the physician as the unit of observation. That is, we examine the concentration of prescribing for acute and chronic conditions listed above for individual physicians. We required that a physician treat at least four cases of a particular condition to be included in the data set.<sup>17</sup> Concentration is measured as the percentage of prescriptions for a given condition accounted by the most prescribed medication used by the physician. Thus the subset of data used to study patterns of prescribing concentration is prescribing behavior by disease for each physician. We regress physician-prescribing concentration for each condition on physician demographic and practice characteristics, patient demographic and insurance variables, geographic location, and whether a condition was an acute or chronic illness.

The second subset of data focuses on the individual visit to primary care physicians (PCPs) as the unit of observation. In this analysis, we examine the visit duration. We include a set of individual patient characteristics (demographics, insurance and diagnosis), physician practice characteristics (solo practice, whether the practice is closed), and controls for location. We then analyze variation in visit duration for patient visits nested within practices.

*QID*: The Quality Improvement for Depression study is a set of four randomized effectiveness trials that were conducted using comparable measurement and study design methods to assess a quality improvement intervention for the treatment of depression in primary care settings. Primary care practices were randomized to either an intervention or the usual care condition. The intervention site received training and support to adopt quality improvement models for treating depression (Wells et al., 2003).

Patients who screened positive for depression during a specified enrollment period were invited to participate in the study if they expected to be treated in the practice where they were screened. All patients enrolled in the combined study met DSM IV criteria for major depression. Detailed clinical, treatment and demographic information was collected from patients at baseline and at three additional points in time spaced 6 months apart. Thus the duration of observation was 2 years. Physicians filled out questionnaires at baseline, which included information on their education and training, background in depression care, and their views of barriers to effectively treating depression.

In the sample analyzed here, we use the data set constructed by Henke et al. (2007). This is a sample of people who were in treatment with an antidepressant medication for two consecutive waves of data collection. We focus on this sub-sample because it allows us to observe cases where a physician could have been able to see changes in symptoms over time and alter treatment in response to observed patient responses. The resulting sample consisted of 1528 patients per wave over two waves of data collection.

Depressive symptoms are measured using the modified Center for Epidemiological Studies Depression scale (CES-D). For a complete description of how this scale was used, see Rost et al. (2001). The CES-D score, which ranges from 0 to 100, is among the most widely used measures of symptoms of depression. It is used to create a measure of symptom levels. We also use the CES-D scores taken at two waves to measure changes in symptoms and to classify those changes according to how changes are described in the clinical literature. Patients responding with a greater than 50% reduction in the CES-D symptom score were classified as “full responders;” those with

---

<sup>17</sup> We also examine the sensitivity of our results when we require that five or six prescriptions be made. As will be discussed below the qualitative finding persists, but the precision of estimates is greatest when using the four-prescription standard. Note further that our ideas do not involve the absolute level of prescribing concentration but rather the relative concentration between acute and chronic illnesses.

reductions of 25–50% were considered “partial responders,” and those with reductions in the CES-D of less than 25% were classified as “non-responders.”

In addition, variables describing a patient’s age, gender, race/ethnicity, educational attainment, number of comorbid conditions, health insurance coverage (Medicaid, private insurance, etc.), and a set of site fixed effects were included in the regression models. We examine the degree to which whether changes in medication are related to changes in clinical status.

## 5.2. *Specification and estimation*

**NAMCS:** The model of physician prescribing concentration pools disease-specific prescribing rates for chronic and acute care conditions that can be treated with a large numbers of possible pharmaceutical agents within physicians. We include covariates that describe: the demographics of patients getting treatment for each condition (age, sex, and race), the insurance arrangements of patients (private insurance, Medicare, Medicaid, managed care, number of managed care contracts a physician holds), physician practice characteristics (solo, specialty within primary care), and region of the country. We also include covariates that describe patient demographics and insurance arrangements. They are intended to account for patient heterogeneity in preferences, and possibly clinical circumstances within a diagnostic category. The count of managed care contracts is included to control for institutions like formularies that might affect prescribing concentration independent of other considerations. We experiment with a few specifications of the demographics, insurance and clinical indicators. One possibility that we investigate is that patients with more contact with a physician will tend to have less concentration in prescribing. Because chronic medical conditions typically require a patient to see a doctor more frequently, they allowing for more learning and negotiation of treatment by the patient. We conduct two sets of analyses. First, we examine whether the level of contact by a patient, as measured by having six or more visits in the past year affects the impact of illness type on concentration in prescribing. Our regression model uses a random effects model, where a physician random effect is included to account for the repeated measures of conditions treated within a physician’s practice. Finally, we re-estimate the model using prescription counts for physicians with five and six prescriptions per condition, which reduces the number of conditions observed within physicians.

For studying visit time variation, the unit of observation is the individual visit within a physician primary care practice. We therefore include characteristics of the patient (diagnoses and demographics), characteristics of the patient’s insurance coverage, and descriptors of the physician’s practice and location. We characterize the error as having two components: a physician-specific component and a patient visit error. We focus on the physician-specific error and practice characteristics to examine the potential role of norms. We experiment with different covariate combinations to examine the sensitivity of the estimate for the physician-specific error to changes in covariates. We include a variable that describes whether a patient has had a high level of contact with a physician, six or more visits in the previous 12 months. This allows for a trade-off between the time spent per visit and the number of visits. In addition, to allow for more clinical complexity, we include some interaction terms between groups of diagnoses (e.g., diabetes and hypertension). The specific right-hand side variables used in the various NAMCS specification are described on [Tables 1a and 1b](#).

The second analysis we conduct estimates a logit of whether a patient had six or more visits during the past year. This allows us to estimate whether there are visit differentials which might result in total time differences according to patient illnesses. The same set of covariates is used in the logit models as in the regression model on visit duration.

Table 1a  
NAMCS variables for physician-disease observations

Variable	Mean	S.D.
Medicaid (%)	0.21	0.17
Private insured (%)	0.59	0.23
Medicaid (%)	0.08	0.12
New patients (%)	0.07	0.09
With 6+ visits (%)	0.30	0.36
Solo practice	0.37	
Female patients (%)	0.59	0.13
Black (%)	0.14	0.23
Age 50–65 (%)	0.26	0.12
Metropolitan statistical area	0.84	
Internal medicine	0.32	
MC contracts = 0	0.08	
MC contracts 1–3	0.10	
MC contracts 4–10	0.51	

### 5.3. QID

In the analysis of the QID data we estimate regressions using generalized estimating equations (GEE) with a logistic link function. We estimate two specifications. The first has as the dependent variable whether or not a medication switch was made. A switch is defined if an antidepressant drug is added or removed from the treatment regimen. The second model has a dependent variable that takes on a value of one if an increase in antidepressant medication dosage was made and

Table 1b  
NAMCS primary care physician visits

Variable	Mean	S.D.
Medicare	0.06	
Medicaid	0.24	
Private insurance + other	0.07	
Mental health	0.63	
Hypertension	0.04	
Diabetes	0.17	
Arthritis	0.03	
Cholesterol	0.09	
Asthma	0.04	
Respiratory infection	0.04	
Other infection	0.03	
Age	49	22
Black	0.11	
Hispanic	0.10	
Other non-white	0.05	
Female	0.59	
Solo practice	0.35	
Metropolitan statistical area	0.82	
Accepts new patients	0.95	
Internal medicine	0.28	
Six plus prior year	0.27	

Table 2  
QID variables

Variable	Definition	Mean
Male	1 if male	0.27
Schooling	Years of schooling	11
Married	1 if married	0.45
Age	In years	34
Hispanic	1 if Hispanic	0.12
Black	1 if black	0.10
Asian	1 if Asian	0.03
Native American	1 if native American	0.05
# comorbid	# comorbid conditions	2
Medicare	1 if Medicare	0.11
Medicaid	1 if Medicaid	0.11
Uninsured	1 if uninsured	0.05
CESD	CES-D score	42.7

zero otherwise. The right-hand side of these two specifications is the same. The right-hand side variables fall into four classes: (1) two indicators for partial and non-response to treatment; (2) three indicators of the level of symptoms; (3) patient demographic and health insurance measures; and (4) treatment site indicators. The specification of the level of symptoms is of particular significance. We use the CES-D score transformed into a spline with “knots” placed at the terciles of the distribution of CES-D scores in the sample. These are defined in Table 2. In the analysis of the QID data, we estimate the variance components. We identify patient, clinical and physician-specific components of the unexplained variance.

## 6. Results

The results of the tests of our three Treatment Hypotheses are reported in Tables 3–5. The results in Table 3 show results for the regressions on the fraction of prescriptions accounted for by a physician’s most prescribed drug in a therapeutic area. We begin by noting that prescribing is generally quite concentrated across all illnesses. The mean level of prescribing is 60%; this indicates a high degree of concentration across conditions. This level of concentration is far higher than the market share of any leading drug in any of the clinical areas examined. The pattern of results reported in Table 3 shows that few of the patient, insurance or practice characteristics explain much of the variation in the concentration of prescribing. The percentage of black patients in a therapeutic category, the age structure measured by the percentage of patients aged 50–64 in a therapeutic area, and having a high share of Medicare patients (a measure of both age and insurance), all have effects on prescribing concentration that are significant at the 0.10 level. Only the indicator for the share of patients in the age group between 50 and 64 has more than a modest estimated impact on prescribing concentration.

The key parameter in this model is the coefficient indicating whether a condition is chronic or acute. The variable indicating a chronic condition has an estimated coefficient of  $-0.133$ , which is significantly different from zero at the 1% level. The estimate suggests that concentration in prescribing is about 13 percentage points lower for patients with a chronic condition. This result is consistent with the SUN Hypothesis, which would posit that the gains to exploration are relatively greater for chronic illnesses. The estimated main effect for the frequent visit

Table 3  
regression results NAMCS data: prescribing share for top drug

Variable	Coefficient estimate (S.E.)	Coefficient estimate (S.E.)
Chronic	−0.133** (0.012)	−0.133** (0.014)
New patients (%)	−0.073 (0.073)	−0.068 (0.073)
Medicaid (%)	−0.041 (0.063)	−0.044 (0.063)
Medicare (%)	−0.032 (0.046)	−0.031 (0.046)
Private (%)	−0.002 (0.032)	−0.002 (0.032)
Female (%)	0.020 (0.094)	−0.027 (0.094)
Black (%)	0.096 (0.064)	0.095 (0.063)
50–65 (%)	−0.191+ (0.106)	−0.195+ (0.106)
Internal medicine	−0.021 (0.028)	−0.020 (0.028)
Region	Yes	Yes
MSA	−0.027 (0.033)	−0.029 (0.028)
No MC contracts	0.049 (0.048)	0.046 (0.047)
1–3 MC contracts	0.006 (0.042)	0.004 (0.042)
4–10 MC contracts	−0.024 (0.028)	−0.026 (0.028)
6+ visits		−0.026 (0.028)
6+ visits × chronic		0.003 (0.029)
Constant	0.697** (0.081)	0.705** (0.081)
$\rho$	0.603	0.602
Wald Chi <sup>2</sup>	138**	142**
N	1723	1723
Range of estimates chronic		
From inclusion of other demographic and insurance variables <sup>a</sup>	0.122 to 0.171	

+ $p < 0.10$ , \* $p < 0.05$ , and \*\* $p < 0.01$ .

<sup>a</sup> For example, we estimate this model by replacing dummies for 80% insurance shares with simple % of patients for each type of payer. In that case chronic is 0.123 (0.006) and  $\rho$  is 0.78.

indicator (six plus) was negative and not significant. The interaction of the frequent visit and chronic disease indicators was estimated to be  $-0.003$ , which is not significantly different from zero at conventional levels of significance.<sup>18</sup> This suggests that physicians with more patients making more visits for chronic diseases do not have different prescribing concentration rates than those with chronic patients making fewer visits. Other sensitivity tests that increase and decrease the demographic and insurance information yield a set of parameter estimates for the chronic illness indicator that range from  $-0.122$  to  $-0.171$ . All of these estimates are significant at the 1% level.

Table 4 reports results for our analysis of visit times to primary care physicians. It reports estimates from the random effects regression models on the time spent by physicians in individual visits. Physician visits vary around a mean of about 17 minutes. Several results are notable. First, the most important factor affecting the duration of a visit is whether a patient is a new patient. New patients have visits that are roughly 4.6 minutes longer than continuing patients, a modest difference of about 25% of the mean. This is due to the need to obtain more basic information about a patient and their health status. In addition, patients' diagnosed with an upper respiratory problem typically have visits that are 2.2 (column 1) to 2.3 (column 2) minutes longer than otherwise similar patients. Startlingly, no other characteristic of an illness (diagnosis), insurance or patient

<sup>18</sup> We conducted this investigation in response to the observation by Ben Sommers that patients may play a role in affecting a range of physician decisions.

Table 4  
Regression results

Variable	Duration of primary care visit (min) <sup>a</sup>	Duration of primary care visit (min) <sup>b</sup>	Probability of 6+ visits in past year 1, 2
Hypertension	0.337 (0.331)	0.332 (0.331)	−0.019 (0.015)
Diabetes	0.830+ (0.451)	0.830+ (0.454)	0.080** (0.211)
Arthritis	1.010 (0.717)	1.016 (0.712)	0.023 (0.033)
Cholesterol	1.499** (0.424)	1.439** (0.424)	−0.041* (0.020)
Mental health	1.827** (0.597)	1.824** (0.597)	−0.126 (0.027)
Infection	1.581 (1.048)	1.578+ (1.051)	0.013 (0.049)
Upper respiratory	2.238** (0.638)	2.250** (0.638)	−0.041 (0.080)
Other	−0.812 (0.927)	−0.821 (0.926)	0.019 (0.043)
Self-pay	−1.878** (0.601)	−1.892** (0.600)	−0.058* (0.028)
Medicare	0.947** (0.354)	0.923** (0.651)	0.076** (0.016)
Medicaid	0.481 (0.499)	0.509 (0.504)	0.109** (0.023)
New patient	4.622 (0.452)	4.563** (0.458)	
Age	0.031** (0.007)	0.031** (0.010)	0.0023** (0.0003)
Black	−0.616 (0.439)	−0.609 (0.439)	0.022 (0.021)
Hispanic	0.033 (0.451)	−0.033 (0.452)	−0.002 (0.002)
Other non-White	0.478 (0.654)	0.465 (0.652)	−0.047 (0.030)
Female	−0.179 (0.235)	−0.167 (0.234)	0.029** (0.010)
Solo	1.707* (0.771)	1.702* (0.77)	−0.009 (0.030)
Internal medicine	0.493 (0.812)	0.497 (0.814)	0.017 (0.032)
Metropolitan statistical area	0.893 (1.019)	0.883 (1.018)	0.041 (0.041)
Accept new patients	0.475 (1.731)	0.487 (1.733)	0.052 (0.069)
6+ visits		−0.253 (0.597)	
Managed care contracts <sup>c</sup>	Yes	Yes	Yes
Region <sup>c</sup>	Yes	Yes	Yes
Constant	15.940** (2.028)	15.966** (2.026)	0.097 (0.081)
$\rho$	0.28	0.283	0.220
Wald Chi <sup>b</sup>	211	192	420
<i>N</i>	5425	5425	5434
Range of estimates			
For $\rho$ under alternate specifications of diagnosis insurance and patient demographics	0.27 to 0.31		

+ $p < 0.10$ , \* $p < 0.05$ , and \*\* $p < 0.01$ .

<sup>a</sup> Coefficient (standard error).

<sup>b</sup> GEE estimates.

<sup>c</sup> Note that yes meant that vector of variables were included in the regressions.

demographics systematically affects visit length by more than 2 minutes. The complexity of the situation to be considered thus had virtually no effect on the length of visits. This finding strongly supports the use of Therapeutic Norms, but strongly rejects the SUN Hypothesis for this context. Setting equal times for medical visits regardless of difficulty is strongly at variance with what optimal or even sensible treatment would require. One other factor, unrelated to medical condition, affected treatment time: patients who are supposed to pay the entire bill out-of-pocket (presumably because they are uninsured), have visits that are about 1.9 minutes shorter than similar patients with private insurance.

The estimates reported on Table 4 also reveal some differences in visit times by region of the country and mode of practice. Physicians in the south have visits that are significantly ( $p < 0.01$ ) shorter than those in the west, by about 2 minutes. Also patients seen by physicians in solo

Table 5  
QID data results—GEE estimates/logistic link

Variable	Medication switch	Dosage increase
(1) Clinical indicators		
Response to treatment	NS	NS
Level of symptoms	NS	NS
Comorbid conditions	NS	$p < 0.05$
(2) Demographic indicators		
Race/ethnicity	$p < 0.10$	NS
Gender	NS	NS
Schooling	$p < 0.01$	$p < 0.05$
Age	$p < 0.05$	$p < 0.05$
(3) Insurance		
$\rho$ MD	0.149	0.027
$\rho$ patient	0.851	0.97

practice have somewhat longer visits than those seen in a group practice (1.7 minutes). Together these results suggest a high degree of clustering around the mean visit time of 17 minutes. This is consistent with the data from physician visit videotapes reported by [Tai-Seale et al. \(2007\)](#). Column 2 of [Table 4](#) includes the indicator for whether a patient has had six or more visits in the past 12 months. Our conjecture is that patients making more frequent visits might be given shorter visits because there have been more opportunities to exchange information and negotiate treatment plans. The indicator variable for patients who made six or more visits is negative and small ( $-0.25$ ), and very imprecisely estimated. Moreover, including that variable had little effect on other estimates.

[Table 4](#) also reports estimates of  $\rho$ , the physician-specific variance component. The estimate of  $\rho$  was 0.28 in both specifications reported on [Table 4](#). In this model 28% of the unobserved variance in visit times is physician-specific. Again, this means that some physicians will tend to see patients for longer than other physicians, and measured patient characteristics are not the explanation. It is also over and above the impact of observed physician characteristics, such as whether they practice solo or in a group, the nature of their specialty, or whether their practices are full (as measured by an indicator that they are willing to accept new patients). Thus, some physicians will supply longer visits to all their patients. This is consistent with a coordination-communication cost driven norm, and also with the My Way Hypothesis.

The third column of [Table 4](#) reports results for a quite different dependent variable. It is the logit of the probability of having six or more visits during the past year. The results are notable in that they display a pattern of results that is somewhat different from those obtained in the visit duration analysis. Specifically, only patients with diabetes and those with high cholesterol have significant differences in the likelihood of having six or more visits during the past year. Other illnesses showing small differences in visit duration showed no significant differences in the likelihood of having six or more visits. In addition, a substantial portion of the unexplained variance in number of visits is explained by unobserved physician-specific factors, 22%. Thus, there appears to be some compensation in visit volume, suggesting a modest bit of tailoring in time allocation by illness. This offers a modicum of support for the SUN Hypothesis. Nevertheless, evidence remains supportive of on the use of therapeutic norms for visit times, after correcting for the possibility of multiple visits.

Table 5 reports the analysis of the QID data in relation to Treatment Hypothesis 3. There is considerable amount of alteration of treatment both by changing medications and by increasing dosages. The first column of Table 5 reports the significance levels for classes of explanatory variables that explain changes in medications for treatment of depression by primary care providers. The results show that there is some customizing of treatment to individual characteristics. For example, demographic factors (such as Hispanic ethnicity) and educational attainment significantly affect the likelihood of medication changes. It is striking, however, that the clinical factors included in the model provide no significant explanation for medication switches. The variables measuring the degree of a patient's response to care (partial responder; non-responder), that are based on measured changes in symptoms, over a 6-month period have no significant impact on the likelihood of having a medication switch. Neither did the complexity of a patient's health conditions as indicated by the number of their comorbid conditions. Thus, the decision to switch medications appears to be unrelated to usual measures of clinical progress or complexity. This result would be quite surprising were we not already familiar with the findings of the study by Henke, which they reinforce. The variance components estimates indicate that about 15% of the unobserved variation is attributable to physician factors. Thus, most variability is due to idiosyncratic patient factors unrelated to medical condition. The results suggest a form of tailoring but not one focused on measurable symptoms or changes in measurable symptoms. Conceivably this pattern is due to the difficulty in eliciting symptom and monitoring changes.

The second column of Table 5 reports results for the logit regressions for changes in dosing of SSRI antidepressants. The pattern of results is quite similar to those related to drug switches. In particular, demographic factors (patient age, educational attainment) significantly affect the likelihood of a change in dosage. However, none of the clinical measures of symptoms or response to treatment are found to have any significant impact on dosage adjustments. In this case it appears that key changes are not made according to well-established measures of symptoms, or accepted medical practice (AHCPR, 1993) but are made on the basis of other factors. Because symptoms used in the regression models are those independently measured by research staff and not by treating clinicians, we cannot differentiate the impact of a physician not responding to levels or changes in symptoms from simply not recognizing the levels and changes in symptoms. Either explanation would be disturbing, and at variance with the SUN Hypothesis. The variance components estimates show that for dosage changes nearly all changes stem from patient specific factors not captured by clinical and demographic factors in the model.

This pattern of results is consistent with high cognitive and communication costs for eliciting information and recognizing changes. The results suggest that characteristics of the patient other than clinical progress may play a role in guiding physician perceptions about the best course of treatment.

### 6.1. Concluding remarks

This paper has revisited ideas about norms in medical care. We have considered two classes of norms in medical practice, incentives norms and therapeutic norms. These norms are analytically distinct but are intertwined in their impacts on clinical choices. For example, if a strong therapeutic norm exists, then medical practice may be unresponsive to financial incentives. In such a case, the therapeutic norm undermines the observed incentives norm. In other cases an incentives norm may drive the adoption of a particular therapeutic practice (Glied and Zivin, 2002). Newhouse and Marquis (1978) carefully assessed the evidence on incentives norms in the 1970s, and rejected the notion that such norms importantly influenced service utilization responses to



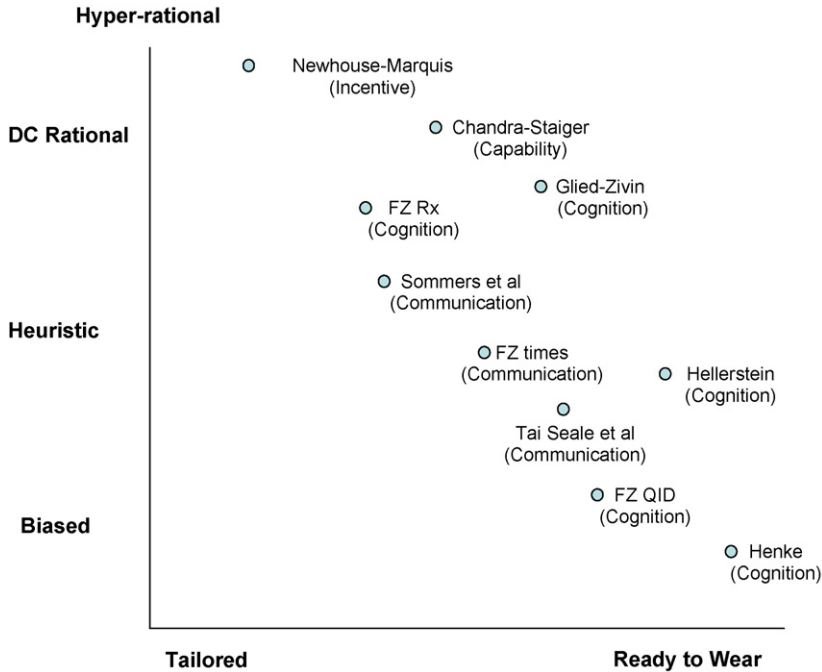


Fig. 1. Evidence on the use of norms and decision-making efficacy.

insurance provisions. New evidence from the literature conducted in the context of a far more complex medical care environment just beyond the turn of the 21st century has raised the possibility that both therapeutic and incentives norms may exert important influences on clinical choices.

We have described some factors that may make the development of norms a sensible strategy for physicians dealing with rapidly expanding clinical science, complex payment arrangements, and diverse patient populations. These are described as four types of costs (4Cs): communication, cognition, coordination and capability. The levels of these costs will influence choices made by physicians between *custom-made* and *ready to wear* treatments. The literature points to the presence of both incentives and therapeutic norms in medical practice.

The new empirical work presented here finds evidence of treatment patterns consistent with the use of ready-to-wear treatments. In some cases those treatments appear to be sensible responses to the four Cs, thus supporting our Sensible-Use-of-Norms (SUN) Hypothesis. In other cases, we find evidence of idiosyncratic individual doctor behavior, what we label My Way behavior, unrelated to key clinical patterns in patients. What is disturbing is that our investigations of physician visit times and prescribing patterns for depression show some norms-based behavior that are in severe violation of the SUN Hypothesis. Physicians appear to be engaging in some highly suboptimal therapeutic practices.

Fig. 1 summarizes the evidence we have gathered on the evidence about decision-making and the use of norms. The horizontal axis represents the continuum between strictly custom-made and ready to wear treatments. The vertical axis presents four degrees of “rationality” in decision-making described in the paper: hyper-rational, decision-cost rational, heuristic, and biased. We have assigned evidence according to how we interpret what the evidence suggests about these

two dimensions of behavior and decision-making. We also indicate the source of decision-making costs involved in each study.

The Newhouse–Marquis study clearly supports the concept of custom-made care with respect to incentives, and represents support for hyper-rational decision-making. At the other end of the spectrum are the studies by Henke et al., by Tai-Seale et al., and our analyses of the QID data on prescription behavior for depression, and of the NAMCS data on visit length, which offer support for physician behavior that is both ready-to-wear and biased. In between are the results produced by Chandra and Staiger, Sommers et al., Glied and Zivin, and our results on prescribing concentration. Our assessment is that there is much ready-to-wear treatment supplied by physicians. In many cases this appears to be a sensible response to a complex decision-making environment. Nevertheless, there is also evidence of the supply of ready-to-wear treatment that is based on idiosyncratic physician-specific preferences or severe biases in the application of heuristics.

A theme of this analysis is that effective therapies must be directed to the condition at hand. Poor decision-making is the malady raised by this work. Decision support tools, whether provided in training, through more refined national standards, through a desk-top computer, or by surrounding personnel, would seem a promising approach to curb biased norms behavior.

## Acknowledgements

We thank Dan Ford, Lisa Meredith, Lisa Rubenstein and Katherine Rost for use of the QID data. We appreciate comments received from Victor Fuchs, Tom McGuire, Joe Newhouse, Ben Sommers, Cass Sunstein and two anonymous referees. Rachel Henke provided invaluable help in using the QID data. Max Kates and Christina Fu provided expert programming assistance. We are grateful to Samuel Osher, M.D., David Rosenthal, M.D., and Mark Schoenberg, M.D. for helpful discussions. Financial support from the NIMH is acknowledged.

## References

- Agency for Health Care Policy and Research, 1993. Depression in Primary Care. AHCPR Publication#93-0551. USDHHS, Rockville MD.
- Alpert, M., Raiffa, H., 1982. A progress report on the training of probability assessors. In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, NY, pp. 294–305.
- Arrow, K.J., 1963. Uncertainty and the welfare economics of medical care. *American Economic Review* 53, 941–969.
- Banerjee, A.V., 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* 107 (3), 797–817.
- Becker, G., Murphy, K., 1992. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics* 107 (4), 1137–1160.
- Berkarda, B., Kaurer, K., Mathe, G. (Eds.), 1984. *Antineoplastic Chemotherapy*. Thieme-Stratton, New York, p. 115.
- Center for Health Systems Changes (CHSC), 2006. Losing Ground: Physician Income 1995–2003. [www.hschange.com/content/851](http://www.hschange.com/content/851).
- Chandra, A., Staiger, D.O., 2007. Productivity spillovers in health care: evidence from treatments of heart attacks. *Journal of Political Economy* 115 (1), 103–140.
- Colvin, M., 1981. Cyclophosphamide and analogues. In: Croke, S.T., Prestayko, A.W. (Eds.), *Cancer and Chemotherapy*. Academic Press, New York, p. 25.
- Conlisk, J., 1996. Why bounded rationality. *Journal of Economic Literature* 34 (2), 669–700.
- Frank, R.H., 1987. Shrewdly rational. *Sociological Forum* 2 (1), 21–41.
- Freireich, E.J., Bodey, G.P., 1984. Evaluation of new agents: planning, execution and evolution of clinical studies. In: Kuernerle, H.P. (Ed.), *Clinical Chemotherapy*.
- Gawande, A., 2002. *Complications: A Surgeon's Notes on an Imperfect Science*. Picador Press, New York.

- Gigerenzer, G., Selten, R., 2002. *Bounded Rationality: The Adaptive Toolbox*. MIT Press, Cambridge.
- Gigerenzer, G., Goldstein, D.G., 1996. Reasoning the fast and frugal way: model of bounded rationality. *Psychological Review* 103 (4), 650–669.
- Ginsburg, P., 1976. Comment. In: Rosett, R. (Ed.), *The Roles of Health Insurance in the Health Services Sector*. NBER, New York.
- Gittins, J., Jones, D., 1974. A Dynamic allocation index for sequential allocation experiments. In: Gani, J. (Ed.), *Progress in Statistics*. Amsterdam, North-Holland, pp. 241–266.
- Glied, S., Zivin, J.G., 2002. How do doctors behave when some (but not all) of their patients are in managed care? *Journal of Health Economics* 21 (2), 337–354.
- Groopman, J., 2006. *How Doctors Think*. Houghton Mifflin, New York.
- Hellerstein, J.K., 1998. The importance of the physician in the generic versus trade name prescription decision. *RAND Journal of Economics* 29 (1), 108–136.
- Henke, R., et al., 2007. *Factors Associated with Treatment Adjustment for Depression in Primary Care Department of Health Care Policy*. Harvard University.
- Institute of Medicine (IOM), 2001. *Crossing the Quality Chasm*. NAS Press, Washington.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291.
- Kessel, R.A., 1958. Price discrimination in medicine. *Journal of Law and Economics* 1, 20–53.
- Kleinman, A., 1987. Anthropology and psychiatry: the role of cross cultural research on illness. *British Journal of Psychiatry* 151, 447–454.
- Lazare, A.S., Eisenthal, L., Wasserman, J., 1975. The customer approach to patienthood: attending to patient requests in a walk in clinic. *Archives of General Psychiatry* 32, 553–558.
- National Center for Health Statistics (NCHS), 2006. *National Ambulatory Care Survey: 2004 Summary*, Advance Data #374.
- Newhouse, J.P., Marquis, M.S., 1978. The norms hypothesis and the demand for medical care. *Journal of Human Resources* 13 (Suppl.), 159–182.
- Pauly, M.V., 1980. *Doctors and Their Workshops*. University Press, Chicago, Chicago.
- Phelps, C.E., 2002. *Health Economics*. Addison, Wesley, Boston.
- Phelps, C.E., 2000. Information diffusion and best practice adoption. In: Culyer, A., Newhouse, J.P. (Eds.), *Handbook of Health Economics*. North-Holland Press, Amsterdam.
- Rost, K.M., Duan, N., Rubinstein, L.V., et al., 2001. The quality improvement for depression collaboration: general analytic strategies for a coordinated study of quality improvement in depression care. *General Hospital Psychiatry* 23 (5), 239–253.
- Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision-making William Samuelson. *Journal of Risk and Uncertainty* 1, 7–59.
- Simon, H., 1958. Theories of decision-making in economics and behavioral science. *American Economic Review* 49, 253–283.
- Sommers, B., Beard, C. D'Amico, A. Dahl, D., Kaplan, I., Richie, J., Zeckhauser, R., in press. A decision analysis using individual patient preferences to determine optimal treatment for localized prostate cancer. *Cancer*.
- Tai-Seale, M., McGuire, T.G., Zhang, W., 2007. Time allocation in primary care office visits. *Health Services Research*.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124–1131.
- Wells, K., Sherbourne, C., Schoenbaum, M., et al., 2003. Five-year impact of quality improvement for depression: results of a group level randomized controlled trial. *Journal of the American Medical Association*.
- Wennberg, J., Cooper, M.M., 1997. *Dartmouth Atlas of Health Care*. American Hospital Publishing, Chicago.
- Whittle, P., 1980. Multi-armed Bandits and the Gittins Index. *Journal of the Royal Statistical Society* 42 (2), 143–149.