

NOTE:

The responsibility for the content of this article rests with the author and does not necessarily represent the views of the Institute of Medicine (IOM) or its committees and convening bodies.

Revised IOM Report

**EXPLAINING GEOGRAPHIC VARIATION
IN HEALTH CARE SPENDING, USE AND QUALITY,
AND ASSOCIATED METHODOLOGICAL CHALLENGES**

Willard G. Manning
University of Chicago

Edward C. Norton
University of Michigan

Adam S. Wilk
University of Michigan

May 18, 2012

The authors acknowledge the help of Jenefer Jedele on earlier drafts of this document.

I. CONCEPTUAL FRAMEWORK

Our conceptual framework for what causes geographic variation in health care use, expenditures, and outcomes is based on variation in demand, supply, the interaction between demand and supply, prices, and quality of care, as well as differences in public policies across areas. We lay out a conceptual framework because it helps understand what is important to control for—and what is not—when comparing health care use, expenditures, and outcomes across geographic areas. The conceptual framework also helps define the econometric challenges in estimating geographic variation. Instead of describing these challenges repeatedly, we explain how study design and threats to validity affect empirical studies for all of these issues. Other common methodological issues are missing data, measurement problems, endogeneity, and using panel data. We briefly discuss dynamic issues in geographic variation. After the conceptual framework (including the empirical issues), we move to the main topic for this paper, which is a critical review of the empirical literature.

1. FIRST NEED TO CHOOSE THREE THINGS

Before describing what causes geographic variation, we need to choose three essential elements: the geographic areas, the patient populations, and the kinds of health care. We take these three things as given, presumably determined by the research question, and do not reflect further on those choices.

The geographic area can be defined by economic markets (by patient flows, e.g., Health Service Areas, and Hospital Referral Regions), or by a number of alternatives such as political boundaries (e.g., county, state), administrative areas (e.g., zip codes, Census tracts), or by where people live (e.g., metropolitan statistical area). Patient populations can be defined by demographics (e.g., age, gender), by disease or medical condition (e.g., breast cancer, multiple sclerosis, hip fracture), by insurance (e.g., Medicare, uninsured), or by the geographic area (e.g., all residents of a city). The health care delivered can be defined by type of provider (e.g., inpatient, ER), by type of insurer (e.g., anything that Medicare pays for), by type of disease or medical condition (e.g., all treatment following admission for AMI), or by type of treatment (e.g., angioplasty vs. stent).

We assume that the researcher chooses the geographic areas, patient populations, and the kinds of health care based on the research question and data availability.

2. CATEGORIZE EXPLANATIONS INTO SEVERAL TYPES

Conditional on these three things, we can measure geographic variation in use, expenditures, and outcomes, and try to explain their causes. In the simple version of our conceptual framework, health care use depends on demand-side and supply-side factors. Health care expenditures depend not only on use, but also on variation in price, which also depends on both demand-side and supply-side factors. In simple notation, this can be summarized as

$$Use = U(D_i, S_j) \quad (1)$$

$$Expenditure = U(D_i, S_j) \times P(D_i, S_j) \quad (2)$$

where i refers to an individual patient, j refers to a provider, U denotes use, D indicates demand-side factors, S indicates supply-side factors, and P stands for price.

Because many of the policy questions surrounding geographic variation concern differences in provider behavior and practice patterns, we want to control for differences in demand-side factors. For example, an area with more elderly persons, who typically have high demand for health care because of their higher prevalence and severity of chronic disease, is expected to have higher health care expenditures than a similar region that has fewer elderly. When comparing to another area we would want to control for age and health status to focus on supply-side differences. Therefore, we really want to compare use and expenditures across geographic areas after controlling for all demand-side factors. In our notation, we want to compare use and expenditures

$$Use_j = U(S_j|D) \quad (3)$$

$$Expenditure_j = U\left(S_j|D\left(P(S_j|D)\right)\right) \times P(S_j|D) \quad (4)$$

where the conditioning implies controlling for all demand-side factors such that the areas would be comparable.

Demand is derived from constrained utility maximization. Consumers demand health care based on their health status, preferences for types of care, and a budget constraint. Therefore in empirical work, it is important to control for all those factors that determine demand, so as to focus on geographic variation that is due solely to supply-side factors. For now, we ignore the role of market equilibrium and public policies that jointly determine some of these variables that are labeled as either demand-side or supply-side. We return to these issues below. As one example in the context of insurance premiums, copayments are determined in part by market equilibrium where lower out-of-pocket costs shift demand out against a possibly upward-sloping supply curve.

Given this basic conceptual framework, next we need to explain each part in more detail. We start with the demand-side variation in the factors that determine the constrained utility maximization, including health status (including case mix and severity), demographics, income and assets, and insurance coverage. This is what we want to control for in principle in empirical work.

Then we describe supply-side variation due to differences in provider choices and decisions that affect treatment patterns. This variation is of most interest to policy makers, and where potentially the most gains can be made in terms of improving health care costs and quality. But there are other important sources of variation on the supply side. Providers vary in their input costs, include land prices and wage rates. There are also differences across regions in the number of providers and their type. Competition and density varies widely between urban and rural areas, for example.

To some degree the local, state, and federal governments all affect supply and demand through regulations, price setting, certificate of need, tax policy, zoning, building state schools of medicine and nursing, state public health initiatives, and legislation. The interplay between demand and supply playing within rules set up by the government affect the geographic equilibrium for health and labor markets.

Beyond the usual neo-classical model of demand and supply jointly determining prices and quantities, we also believe there is a gray area, where demand- and supply-side factors interact in such a way that it is less clear whether these should be controlled for or not. These interactions are not often studied, and it is not clear to what extent researchers should control for them. Physician agency on behalf of patients affects the reliability of empirical results due to the extent that they may vary systematically across areas.

After explaining demand- and supply-side variation, we move on to price variation, which affects expenditures directly through prices and indirectly through the demand function (such as the impact on health insurance premiums via the budget constraint). Some price variation is due to local market structure and the effects of competition or monopoly power. Because a large fraction of health care is paid by the government through administered prices, geographic variation in Medicare and Medicaid prices are important.

Quality of care (health outcomes) is sufficiently different from health care use and expenditures in terms of conceptual framework and empirical issues that we devote a separate section just to quality. Quality, for example, is multidimensional and hard to measure. Patients and policy makers care about quality of care. The goal should never be simply to cut expenditures, which can be achieved easily but with calamitous results for quality of care. Geographic variation in quality of care may be both due to variation in health care use, and also may cause variation in health care use.

When assessing the validity and importance of studies of demand-side or supply-side factors contributing to observed variation in health care use and expenditures, it is important to recall there will always be a stochastic element to medical care processes.

Skinner's chapter on regional variation (Handbook of Health Economics, 2012) lays out a conceptual framework that is based explicitly on simple demand and supply models. Following Chandra and Skinner (2011), he uses a two-period model in which consumers maximize current and future consumption; health care increases the probability of surviving until the second period. On the supply side, providers maximize a function of both patient health and income, subject to resource and ethical constraints. In the end, both his conceptual framework and ours abstract from some of the complexities of actual health care markets and interactions, while still being flexible enough to illustrate how demand, supply, their interaction, prices, and quality of care may all affect geographic variation in health care use, expenditures, and outcomes.

3. DEMAND-SIDE VARIATION IN USE

Individual demand for health care depends on that person's health status, preferences for care, and their budget constraint, including type and generosity of health insurance coverage. Health is an important part of utility, and one way (certainly not the only way) to improve health is to purchase health care. Therefore utility maximization involves tradeoffs related to the purchase of health care—the purchase of health care or other consumption or savings, the time needed to receive care or the time doing other things, and expected utility from future health status with or without treatment (including possible side effects).

Health care demand depends predominantly on health status of the patient. Sicker patients require more care and cost more to treat. Therefore it is crucial to control for health status when comparing health care across patients and regions because there is evidence that health varies geographically. Ideally, researchers would have the full chart history of all patients on all dimensions of health. Empirical models would have full controls on current and past health status, on all co-morbidities, allergies, genetics, and biomarkers. This is rarely feasible. Administrative data generally has limited information on current health status and even less information on the history of health status. Researchers will control for demographics as a way to control for health status because demographics are often correlated with health status, so. Age, gender, race and ethnicity, education, income and wealth, and family status have all been shown to be correlated with health, which justifies their use as proxies for health status. Moreover, the latter two of these are elements of the budget constraint. Health behaviors, such as smoking, drinking, drug use, and risky sex, are also demand-side factors that affect demand for health care. To the extent that these health behaviors affect health status or affect treatment choice, they should be controlled for. These are sometimes recorded in administrative data, more often in electronic medical records, and are also correlated with demographics.

Patients with identical health status may demand different amounts of health care if their preferences for consuming health care differ. For example, some patients (or their families) embrace hospice and palliative care while others embrace intensive treatment at the end of life. Some patients have greater tolerance for complicated drug regimens or willingness to deal with side effects, and so are willing to accept complex pharmaceutical solutions. Christian Scientists typically shun most modern health care. Preferences are rarely, if ever, recorded in any data set, but may be weakly correlated with demographics.

Patients also vary in their budget constraint and price responsiveness. Demand depends on the out-of-pocket price and on income (net of taxes and premiums), which vary by insurance. Health insurance lowers the out-of-pocket cost of care and also increases access to care directly because some providers only see patients with certain types of health insurance. Demand-side cost sharing can also vary throughout the year if there are deductibles, limits on benefits, and stop-losses. More generous insurance with its higher premiums reduces disposable income, which can lower demand. A further complication is that health insurance is endogenous with respect to health status. Price elasticity may also depend on the patient's income and wealth. Health care is a normal good, so conditional on the out-of-pocket cost, those with higher income and wealth have higher demand. For those on Medicaid, there is wide state-level variation in eligibility limits and benefits, and these differences also have changed over time and within

states. In general, over the last several decades eligibility has expanded, complicating regional variation assessments for the Medicaid population.

A patient's demand can also vary due to non-price responsiveness. The classic example is distance to the provider, with longer travel distance lowering demand. Another example is waiting time, either waiting in the office or ER, or waiting for a scheduled appointment. Geographic variation in the density of providers and competition among those providers leads to variation in the travel distance and time waiting. In addition, family members can affect one's response to out-of-pocket prices. For example, if a family member has an ER visit, then they will likely pay the family deductible thereby lowering the out-of-pocket price faced by other family members and increasing demand.

4. SUPPLY-SIDE VARIATION IN USE

The geographic variation that policymakers worry about is observed when identical patients treated by different providers receive different treatment. In other words, holding all demand-side factors constant, what supply-side factors influence variation in use? Let us begin with decisions by the physician (or nurse or other individual provider). Extensive research has been dedicated to examining the determinants of physicians' decisions regarding the most appropriate diagnosis to make, whether subsequent care beyond the initial office visit would be appropriate, and, if necessary, what subsequent care would best meet their patients' needs. Less research, it appears, has been dedicated to an exploration of physician determinations with respect to the extent, timeliness, and appropriateness of the office visit that would initiate a medical care episode. Physicians may recommend different treatments because of their differing abilities to diagnose. This may be driven, in part, by differences in medical school training, which may also affect how physicians treat particular medical conditions. It is well known that physicians who train in the same place (medical school, internship, and residency) have more similar treatment patterns than those who train apart (Woodward et al. 1990; Tamblyn et al. 2003). As such, it is understandable that different treatment patterns would emerge based on that training. Perhaps unfortunately, evidence suggests that these patterns are slow to change (Epstein and Nicholson 2009).

Furthermore, there is variation on the supply side from the hospital level up through the state that affects the level of use that patients experience. Much of this variation may be attributable to variation at the individual physician and physician group levels, to the extent their decisions drive health care use at higher levels. Relman (1985) and Eisenberg (2002) are among those who have each estimated that that physician decisions determine how the vast majority of health care resources are allocated, either directly or indirectly. Yet variation in inpatient use and expenditures merits study in its own right. Variation in inpatient use and expenditures regionally may stem from variation in admission and readmission rates per capita, the mix of patients' diagnosis-related groups (DRGs), average length of stay, in-hospital efficiency and care use (e.g., of hospitalist and specialist visits, nurse staffing, tests, procedures, drugs), and other components of inpatient resource use. The factors that researchers have suggested drive much of the observed variation within and across these components are often intertwined and difficult to untangle. This was the conclusion of a recent review by Pyenson and colleagues (2010). While

their particular identification strategy may be subject to bias owing to their sample's potential for regressing to the mean, their work is illustrative of the difficulties of identifying and isolating the primary causes of regional variation in inpatient hospital use.

One reason for this variation is that hospitals vary tremendously in the amount of specialty care offered. Some hospitals have cardiac units, cancer centers, MRIs, emergency rooms, and emergency helicopters, while others do not—similar variation exists among physician practices in their levels of investment in medical technology, though perhaps on a smaller scale. The reasons for this variation are complex, but include competition among other hospitals for patients and for physicians (the latter often referred to as the “medical arms race”). Variation in the ability to provide specialty care at the hospital level may explain part of observed variation in use.

In considering the determinants of supply of inpatient care facilities, it is important to note that construction and relocation costs can be very high, and so relocation is rarer and slower for hospitals than for individual physicians and small physician groups. When capital is difficult to obtain, changes in hospital supply are determined simultaneously with changes in demand through the interaction of market forces. Ownership status—whether the hospital is for-profit (FP), non-for-profit (NFP), or government-operated (GO)—is another critical factor to consider in the context of hospital supply. In his seminal 1963 monograph, Arrow proposed to explain the fact that more than half of American hospitals are NFP (which persists today) under a theoretical framework based on information asymmetries. In particular, he suggested that because patients were largely unable to differentiate hospitals on the basis of the quality of care they offered, patients may interpret a hospital's FP status (which is observable) as a signal of a lesser willingness to expend additional resources in order to improve quality of care, eyeing bottom line costs.

Newhouse (1970) later proposed that what distinguished FP and NFP hospitals was that they aspire to maximize different outputs. While FP hospitals seek to maximize profits, NFP hospitals seek to maximize the quality of their care and the quantity of patients they serve. The implications of this distinction include that FP hospitals would be more likely to cut operational costs where possible than their NFP peers. The means by which they might do so would likely be limited to those that did not observably reduce quality of care, but less directly observable measures of quality might be affected by these cuts. As a result, one would expect regions served more by FP hospitals to have lower-than-average operational costs incurred for similar patients. FP hospitals might also be more selective about the types of services they offer and the types of patients they serve, preferring to provide more profitable services to more profitable patients. Thus regions served more by FP hospitals may experience higher overall inpatient use and expenditures, at least for those select services and patient types perceived to be more profitable (e.g., privately insured patients). Use and expenditures may be higher in regions served more by NFP hospitals in terms of those services and patient types perceived to be less profitable (e.g., uninsured or Medicaid patients). This work by Newhouse as well as that by Arrow (1963) precipitated a large collection of research distinguishing hospitals and markets on the basis of ownership.

In addition, given hospitals' incentives to reduce their production costs because of the DRG payment system by which inpatient hospital services are bundled and paid for at a single fixed rate based on diagnosis, procedures, and comorbidities, it is not surprising that wider variation in the use of these post-acute services would also emerge. While overuse may be of concern with some of these services, there are concerns about the underuse of others, such as hospice care.

There are two important sources of variation at a geographic level larger than a hospital and smaller than a state. These are medical malpractice risk and local emergency response. The risk of medical malpractice is partly due to the provider's skill and partly due to the judicial system, which typically operates at the county level. Some judicial courts are reputed to be friendly to plaintiffs' malpractice cases and thus attract more lawsuits. To the extent that physicians respond to the threat of malpractice by ordering more tests, a practice called defensive medicine (the empirical extent of this is debated in the literature), some geographic variation may be due to differences in malpractice threat. A second supply side factor at the county level (or geographic area smaller than a state) is emergency response time. Some areas have 911, while others do not. People in rural areas may live far from emergency response teams such as firefighters and EMTs. Funding for such response teams is subject to political pressures and so varies over time and across localities.

State laws affect clinical practice and therefore affect state-level variation in use. Cost-effective public health programs (some of which are led at the county-level) may reduce the use of acute care or other costly health care services. Certificate-of-need laws affect the purchase and building of expensive high-technology equipment, such as advanced imaging. Construction of medical schools, which are often in state universities, often depends on state legislature decisions. State laws affect minimum staffing rules in nursing homes and minimum length of stays for newborns (to prevent "drive-through deliveries"), among many other aspects of health care delivery. These state laws may affect the level of use, and thus provide state-level variation, although most empirical work has found that CON laws have little net effect on construction (Salkever and Bice 1976; Lanning et al. 1991; Harrington and Curtis 1996; Conover and Sloan 1998; Salkever 2000; Grabowski et al. 2003; Lang et al. 2004; Rothberg et al. 2005; Kane et al. 2007; Almond and Doyle, Jr. 2008; Sochalski et al. 2008; Rivers et al. 2010; Serratt et al. 2011).

5. PATIENT-PROVIDER INTERACTIONS

The simple dichotomy of demand and supply, with factors neatly falling into one side or the other, misses some important causes of geographic variation. Interactions between demand and supply—or more accurately between patient and provider—affect use. In particular, imprecise communication may dramatically affect treatment decisions and short-term patient outcomes. Thus the interactions of supply-side and demand-side factors, rather than either side independently, may bring about increased variance in medical care use and expenditures. Other examples include adherence to treatment regimens, which depends not only on the patient's ability to understand the treatment and willingness and ability to stick to it, but also the physician's ability to explain the importance of adherence and choosing the most appropriate drug or therapy for that patient. Both sides affect adherence. A related consideration is whether

the patient might willfully resist adherence to a physician's recommended treatment because of poor communication, distrust (e.g., if undue supplier-induced demand is perceived), or for other reasons. Similarly, recommendations for changing health behaviors such as smoking, diet, and exercise depend on patient behavior but also on how the physician presents and recommends behavior change. Another related issue is whether a patient seeks a second opinion. Some patients are more accepting of a single opinion; some physicians are more encouraging of patients asking for a second opinion. Patients also differ in their willingness to push back against managed care limits (or more generally against limits imposed by their insurance company). A final example is statistical discrimination (Balsa and McGuire 2001; 2003), which depends on characteristics of both patient and provider.

Unlike the pure demand or supply-side factors, it is far less clear to what extent researchers should control for these interactions. Equations (3) and (4) are hard to implement when there are interactions. There is less research on these topics, so the literature is less well developed.

Eisenberg (2002) offered a convenient framework for categorizing many of the factors that may affect a physician's decisions when seeing a patient: factors focused on the patient's benefit (i.e., with the physician acting as a dutiful agent), factors focused instead on the physician's own benefit, and factors representing the acknowledgement of, or accounting for, social and individual resource constraints. We rely on this framework in much of our literature review discussing supply-side determinants of geographic variation in use and expenditures.

Uncertainty can elevate the variance of observed medical care use and expenditures. Information asymmetry can differentially interfere with the physician's ability to serve her patients best. Arrow highlighted in his seminal essay (1963) the many uncertainties of the care process—in identifying when the patient is in need of care, in communicating the patient's symptoms, in specifying the nature and severity of the patient's condition, in selecting an appropriate course of treatment, in anticipating the patient's response to the intervention, in measuring the effects of the intervention, and in determining when best to terminate the intervention, to name a few—that may affect the physician's decision-making process. Such uncertainty can elevate the variance in observed service use and expenditures. We also expect the effects of the information asymmetries on regional health care expenditures may vary with each region's health status and disease profile as well as its providers' specialty mix. Because the effects of some of these uncertainties (e.g., regarding imprecise communication between patient and physician) are not solely supply-focused, we discuss these factors elsewhere.

Several theoretical constructs predicting how physicians behave in response to uncertainty have been presented and tested by their proponents and other scholars. For one, the concept of uncertainty in this context suggests the relevance of principal-agent models, as summarized by McGuire (2000). To the extent a physician's actions may be considered unobservable to the patient (e.g., because of insufficient health literacy or poor communication), this model suggests physicians are presented an opportunity to re-optimize behavior depending on inconsistent preferences and information. Developers of evidence-based guidelines seek to minimize variation of this kind, though the effective dissemination and application of such

guidelines in practice is inhibited for several reasons discussed by Cabana and colleagues (1999) among others.

Other theories predicting physician practice patterns and responses to uncertainty include that posited by organizational theorists Goodrick and Salancik (1996), who suggested that health care providers would look to the performance of local peers and leaders in determining the most appropriate course of action. Burke and colleagues (2010) similarly identified social influences on individual providers as a potentially key determining factor. Many have argued that practice pattern variation may be in part attributable to the influences of nearby medical schools, academic medical centers, and research centers, which themselves do not always demonstrate similar practice patterns (Fisher et al. 1994; Wennberg et al. 2004). Each of these theories is, in some measure, consistent with observed geographic variation in practice patterns and, thus, use and expenditures.

Other organizational theorists have suggested that in the absence of an established practice standard, physicians (and like-minded organizations) will be more likely to decide on the basis of their resource-sharing relationships and dependencies so as to mitigate financial risks or build toward the greatest possible financial gains (Goodrick and Salancik 1996; Lemak and Alexander 2001). If treatment decisions are based in part on financial incentives, then there is some concern that providers are not acting solely in the best interest of the patient.

Supply-side effects extend beyond the most elemental patient-provider interaction. Provider responses can be aggregated up to a market level, as in a more traditional analysis where agency is not considered an issue. Even at this market level, patient populations can mediate supply-side processes. For example, markets are often defined spatially because of the reluctance of patients to travel far or to suffer the possible consequences of traveling a greater distance to a bigger or more complete facility (as in a heart attack). In large part because of variation in market size, the availability of different types of specialty care varies between rural, smaller cities and MSAs, the suburbs, and larger urban areas.

Moreover, there may be meaningful variability across markets in terms of the number of insurers for the not-public, not-ERISA small group market and individual markets. Some markets may have relatively few insurers serving these populations, while others may have ten or more. We note health insurer market concentrations have also increased in recent years (Robinson 2004; Robinson 2006; Melnick et al. 2011). The effects of this market-level variation may be muted by the market dominance of Medicare, Medicaid, and SCHIP as well as their administratively set prices. The roles of the Medicaid and SCHIP programs in local health insurance markets are driven principally by state policies.

It has long been thought that physicians respond to financial and other incentives. This is consistent with the model of physician utility maximization proposed by McGuire and Pauly (1991), in which the utility-maximizing physician modifies her care practice so as to maximize the personal benefit she receives from income, leisure activities, and a preferred style of practice. In their study, McGuire and Pauly assessed a range of models incorporating varying strengths of income effects that therefore offered wide-ranging hypotheses as to the behavior of physicians in response to changes in fees for their services. Their first result was the rejection of the “literal

target income” model, which proposed that physicians seek to maintain a target income with some consideration of disutility from inducing service use (e.g., due to positively valued leisure) and incorporated significant income effects to account for this theory. They also rejected the simpler model that dropped the income effects and proposed that physicians act solely to maximize profit with increasing incentives to induce use as fees increased. Their proposed middle-ground alternative included some limited income effects and consideration of the disutility of inducing service use due to lost leisure and other individual physician-level factors. This model has underlain the work of many scholars studying the physician decision-making process, including Eisenberg’s. The formal approach of McGuire and Pauly yields a general economic model of physician behavior and decision-making.

One illustrative example of the importance of physician incentives is the difference in incentives between fee-for-service and managed care. Compared to physicians under a managed care system, physicians under a fee-for-service system tend to provide more visits and less prevention (Greenfield et al. 1992; Krieger et al. 1992; Phillips et al. 2000; Emanuel et al. 2002; Landon et al. 2004). There is also strong anecdotal evidence that financial incentives matter, and use is positively correlated with reimbursement. This is one of the most important areas to explore in any empirical investigation of geographic variation.

To some extent, the influence of the physician in determining use and expenditures for a given patient is moderated by behavioral responses of patients and policies of the patient’s health insurer. Economists have represented the physician’s “market power” in this context using a variety of models, such as models of monopolistic competition (McGuire 2000).

Chandra and colleagues (2011) recently produced a review of theoretical and empirical literature investigating many of these and related factors at the levels of the patient, the physician, and the clinical interaction. Their discussion is presented in the context of regional variation in practice patterns and thus overlaps with our discussion at intervals. We refer the reader to their work for this immediately relevant theoretical discussion as well as their collection of evidence largely gathered from the clinical literature.

Research on supply-side variation in use also extends to the effects of overall and regional physician supply and mix on variation in use and expenditures. One of the most significant reports highlighting this factor was Wennberg and Cooper’s *The Quality of Medical Care in the United States: A Report on The Medicare Program*, otherwise known as *The Dartmouth Atlas of Health Care* (1999). The large and important report’s eye-catching maps and figures and accessible analytic frameworks drew the attention of policymakers and researchers and solidified the report’s place as a seminal work in the study of geographic variation.

Physician decisions in terms of practice location are germane to this discussion. It has been argued this decision may be modeled through the neoclassical economic construct of the utility maximizing physician (Chou and Lo Sasso 2009). A physician’s choice of location may well depend on his or her prospective income in a market as well as the historical density and mix of patients most likely to seek care, anticipated changes in the area’s demographics and care-seeking preferences, the type of setting (e.g., solo practice, large group practice, integrated delivery system) in which she would prefer to practice, similarities in the population’s culture

and language preferences to the physician's own, and anticipated malpractice insurance premiums (Chou and Lo Sasso 2009). The location theory-based approach of Newhouse and colleagues (1982b) is a contribution of primary importance in this area.

Conditional on the establishment of variation in physician practice location, it remains to show that physician volume and mix are correlated with use and expenditures to conclude that drivers of practice location are also drivers of geographic variation in these outcomes. In a classical economic model of a market for a given undifferentiated, normal good, an increase in the supply of the good corresponds with a reduction in its price due to competitive pressures. Moreover, within this classical economic framework, if alternatives exist that may be considered substitutes for the good, then the relative prices between the goods may affect the volume of each good purchased (Varian 2009). Yet while numerous studies have noted that the volume of health care services has expanded over time along with the volume of physicians, the prices of their services have not fallen as the simple model would predict. Numerous other predictions of classical economic theory regarding the supply of health care services have proved inaccurate. We refer the reader to, for example, Newhouse (1992) for further discussion of the growth in health care use and prices both over time or Fisher et al. (2009) for a brief synopsis of geographic variation in health care expenditure growth, both relevant in this context.

In the discussion section of their primary article on applications of location theory to physicians, Newhouse and colleagues (1982b) noted that others had expressed skepticism that their relatively simple location theory-based approach could adequately reflect the physician's capacity to induce demand for their services, which would affect the location theory model's implicit assumption that the demand within a given region is fixed. The monopolistic competition model described by McGuire (2000) synthesized and formalized this argument, specifying the nature of use inducement in this perspective as a market failure whereby physicians could leverage their market power to set the quantity of services provided (since the services they provided are non-retradable) when prices are fixed (as in the case of Medicare's prospective payment systems). Promoters of the idea that demand inducement materially affected the location decisions of physicians have highlighted the persistence of regional inequities in access to physician services as support for their preferred theory (Newhouse et al. 1982b). To address these failures, policymakers have introduced several policies in an effort to incentivize location in areas designated as "healthcare professional shortage areas" or other areas with limited access to care. Still, the measures of access they have used in developing these incentives may not have adequately accounted for provider specialty and the rates at which patients cross market boundaries in the process of seeking care.

Another factor that may generate variation in use and expenditures is the concept of discrimination, which may manifest as another characteristic of physicians and physician practices and determine use and expenditure patterns in a region. Specifically, by discrimination we mean that two nearly identical patients treated by the same physician might receive different treatment based on their demographics (where demographics are unrelated directly to their clinical diagnosis or prognosis), as discussed by Balsa and McGuire (2003). In their study, Balsa and McGuire described multiple models of discrimination based on the mechanisms of prejudice, clinical uncertainty, and stereotyping. These models illustrated that, as select characteristics of physicians, physician groups, and markets—such as the racial composition of patient and

physician populations, awareness of the local prevalence and incidence of diseases by patient group and research regarding these rates at a national level, physician bias, and others—vary, so might medical practice patterns vary by region. For example, in their discussion of a model of the role of clinical uncertainty in racial inequities in care based on miscommunication and misunderstanding, Balsa and McGuire noted that inter-cultural miscommunication regarding symptoms may result in overtreatment if treatment is likely to benefit the average patient. As such, we may conclude that regions with greater diversity and likelihood of inter-cultural patient-physician pairings, which may be more likely than intra-cultural patient-physician pairings to experience miscommunication, may be susceptible to overtreatment for select conditions. While Balsa and McGuire did not test their models empirically, they provided several substantive frameworks that can support the further exploration of discrimination in this context. Subsequent work by Balsa and colleagues (2005) has begun these empirical explorations, focusing on the effects of clinical uncertainty on observed racial differences in the diagnosis of three select conditions. Still, further study will be required to compare and size the effects of these mechanisms across a wider range of clinical conditions, a larger proportion of physicians' practices, more specifically identified minority groups, and non-urban settings.

6. MARKET-LEVEL BEHAVIOR

In most of the discussion so far in this conceptual framework for the report, we have focused on issues of a very micro-level behavior of consumers (patients), and suppliers (physicians and other health providers) or allowed for their interplay in the context of asymmetric information and the agency problem. But both sets of individuals and organizations are embedded in markets, some of which are local or regional, while others are national (e.g., the pharmaceutical market). The two sides of the market jointly determine the outcomes of price, quantity, quality, and their variants.

Because each actor in the market is typically too small to affect the outcome individually, unless there is oligopoly or monopoly, the individual actors (patients and providers from the demand and supply sides) will act as if they are price takers. But the collective effect of all involved works to determine these outcomes jointly. Thus, some things may appear fixed at the individual or firm level, once we look at the market level, the interplay of demand (sums of individual demands) and supply (the sums of individual firm supply functions) will jointly determine the price, delay to appointment, and waiting times in the market that the individual consumers and firms take as given.

Further, what happens in the output market (for medical visits or inpatient stays) can also affect the input markets as well. Because the demand in the input markets are derived demands based on demands in the output markets (in perfectly competitive markets, the input demand is the sum of marginal revenue products at a given wage or price for that input in the output market), the input prices are jointly determined with output price if the specific output market is large in that input market. But if the input market has a perfectly elastic supply or if the specific output market is a small demander in the input market then the output market will treat that input price as given.

The number of suppliers is itself determined in market-level equilibrium. Entry and exit in the output market will occur until the marginal firm just breaks even (including a return on investment). A similar process can occur in the input markets.

If we move beyond the standard perfectly competitive model, there are some other matters to consider. If there is evidence of indivisibilities or increasing returns to scale, then the number of firms may be limited by how large these economies are relative to the size of the local market. The limiting case is natural monopoly. But situations may not be that extreme; there may be room for a small number of firms, which leads to duopoly or small group oligopoly in that geographic market. In the health context, this is most likely to be a problem for either hospitals or certain types of specialists in rural or smaller metropolitan areas. Absent regulation, a small number of sellers will tend to have higher prices than if the market were large enough to sustain multiple providers and see either monopolistically competitive or perfectly competitive behavior.

The implication of this line of argument is that demand and various economies of scale and scope lead to the number of providers who operate in specific geographic or other markets. This is one of the neoclassical, econometric counter-arguments against the market level version of supplier-induced demand in the longer run; it is counter in this case because demand and supply are jointly determined and the system may not have enough structure to test one model against the other.

If there are multiple output markets that are geographically separated, then some of the economies will lead to regionalization. Maybe there is sufficient demand in the aggregate to justify one tertiary care facility or one very specialized specialist over a much larger area than for community hospitals or primary care providers. At that point, the competition could be regional or with highly urbanized areas—consider Eastern Colorado/Western Kansas as smaller regional markets and Denver and Kansas City as larger urban markets that serve a large regional market. In the rural health literature, a population's location within two overlapping hospital markets is an important consideration. For related theoretical discussions, please see Lösch (1939; 1943) on location theory, Hotelling (1929) on the Hotelling model, and Steiner (1952; 1961), Rothenberg (1962), and Spence (1976a; 1976b) on excess duplication.

The geographic “mosaic” of markets also provides another level of competition. Firms and newly trained physicians can choose among markets to work in. Competition within a specific geographic market may mean that new entrants in some markets cannot earn as much as they would if they had located or relocated elsewhere. If there is an increase in the number of new providers (net of retirements), then we would expect that some of them will enter markets that they would not have if the overall number of physicians were smaller. Examples of this phenomenon can be seen in the specialty location decisions analyzed by Newhouse and his colleagues in a series of papers (various dates, 1982-2005) which found evidence of a surge (caused by shifts in the amount of funding of medical education by the federal government) in newly trained physician specialists moving into smaller markets.

As a practical matter, none of these equilibrate instantaneously. Physicians are often reluctant to relocate given the time it takes to build a practice and referral networks. Hospitals

may be slow to enter and exit in part because of regulatory issues, side-payments by local communities or governments, or of a market dominated by non-profit or public firms (Keeler et al. 1999; Ricketts 2000; Fonkych and Melnick 2010; Shen et al. 2010).

Nevertheless, the prices in both input and output markets and the number of providers on either the inpatient or the outpatient side are partially determined by the market to the extent that there are not administered prices (as with Medicaid and Medicare). Provider preferences, agency, public policies on prices, and location may alter the equilibrium. However, they do not negate the fact that the number of suppliers is not exogenous in the longer run.

So far we have focused on the behavior of patients and providers, their interactions at the finest level and how this leads to market-level behavior and market equilibrium at a point in time, or in how decisions are made over time (e.g., providers moving into areas). One implication is that resources will be drawn into areas of potential profitability relative to other markets. This reduces the geographic variability of prices but not eliminate them. Given the nature of medical markets, we would expect entry and exit to be slower than in other markets. In the next few sections, we will focus on three issues of particular concern and how they could vary geographically: price, quantity, and quality.

7. PRICE VARIATION

Prices vary geographically, and this variation contributes directly to variation in health care expenditures. It also contributes indirectly to variation in health care use through the demand and supply functions. First, the prices per unit of goods and services vary across insurers but also within insurer across markets. The prices of things like a routine office visit, lab services, durable medical equipment, and a day in the hospital are all determined through a combination of government price regulations, negotiated fees, and the private market (Melnick et al. 1992; Zwanziger and Mooney 2005). Berndt and colleagues (2000) described many of these and other related issues in some detail, though their focus was restricted, at times, to higher-level medical care price indices less relevant to our purposes. Medicare sets prices explicitly allowing for some geographic variation (for a comprehensive review, see *Geographic Adjustment in Medicare Payment: Phase I: Improving Accuracy*, Second Edition, edited by Edmunds and Sloan (2011), produced for the IOM). Insurance companies generally negotiate prices with large providers. Competition affects the ability of insurers and providers to negotiate better prices (OAG 2011).

Second, prices for episodes of care also vary geographically for similar reasons. For health care that is set prospectively, such as Medicare inpatient care (Medicare PPS), the price per episode for a specific DRG reflects differences in input costs and market competition. To the extent that prices depend on the usual and customary price in an area, prices will reflect not only standard competition but also the extent of managed care in the area and other factors.

Third, these prices not only have a direct effect on total health care expenditures, but may indirectly affect the quantity. This could be reflected in equation (4), where the amount of demand depends in part on the out-of-pocket price. An increase in out-of-pocket price leads to

lower demand. So higher prices raise health care expenditures directly but indirectly lower health care expenditures through use. The supply could also be considered a function of the final price received by the provider.

The underlying reasons for geographic variation in prices are input markets and competition among insurers and providers. Non-physician labor is a large fraction of health care costs. There is a market for nurses, lab staff, and other administrative support. The cost of labor varies with the unemployment rate, opportunity cost of similar work, and the unionization of nurses and hospital staff. Another input cost is medical malpractice premiums. Medical malpractice rates vary across states (and across specialties), and these rates are partially passed along to consumers. The many principles applied when Medicare adjusts its rates are discussed in a recent report edited by Edmunds and Sloan (2011). The medical arms race literature has documented that prices are higher in urban areas with greater competition.

8. QUALITY OF CARE VARIATION

Although the geographic variation literature has focused largely on health care use and expenditure, we want to emphasize that there is a third outcome to consider. Quality of care is closely related to use and expenditures, but is distinct. Quality of care is determined simultaneously with other health care decisions, subject to a budget constraint. The goal of health policy should never simply be to minimize health spending, or make health care use equal across all populations, but instead to spend wisely and get the greatest value for the constrained resources. For that reason, it is essential to consider quality of care along with how much is spent.

One way that quality of care differs from use and expenditure is the ability to measure it. Finding operational definitions of quality is challenging. Mortality, which is relatively easy to measure, is rare and irrelevant for many health conditions. Measures of process do not necessarily measure health outcomes well. The best measures of quality of care would take into account morbidity, mortality, and prognosis. Such measures are hard to quantify, and as a result the study of geographic variation in the quality of care is sparse.

If one could satisfactorily measure quality of care, we believe there are three levels at which quality variation matters. At the treatment level, there is variation in the quality of care received for a given treatment. This is the simple static story conditional on treatment and provider. At the patient level, variation in quality of care affects what type of treatment a patient receives, and who provides that treatment. Patients with better access, insurance, and information seek out the highest quality providers. At a larger level and over a longer time span, variation in quality of care affects hiring and training of health professionals. Certain areas become known as being more advanced in some treatments (Chandra and Staiger, 2007). Variation in quality, through a variety of mechanisms, may also be considered a determinant (if an indirect one) of variation in use and expenditures on some level.

9. RESEARCH DESIGN AND EMPIRICAL METHODS

Behavioral Framework. Much of the evidence on geographic variation is based on either cross-sectional or panel analysis of short panel data that are strongly correlated over time at whatever level of geographic detail is examined. As the preceding discussion indicates, we would expect that geographic variation will in large part be due to a combination of market forces reflecting differences among areas in demand-side and supply-side factors, heavily affected by differences in variations in state policy (e.g., Medicaid) and federal payment rules (Medicare, principally) that partially dictate differential payment for different areas. What is observed is an equilibrium of economic and policy factors in contrast to a more traditional neoclassical equilibrium of geographic markets in the absence of regulation or administered prices.

Because the results depend on this interplay of health care demand and supply, payment and provider policies, addressing geographic variation in health care use, expenditures, and outcomes is analytically more difficult than in other situations because of the simultaneity with which this is all determined. But in what follows, we will make some assumptions to make this more tractable. First, we will assume that the population in each area is fixed so that we do not need to address the issue of what happens when potential patients move (although we know that some people move because of health; e.g., those with respiratory problems to the southwest; the frailer elderly back to their families from retirement settings) or geographic shifts occur in the population.

The second is that the mix of inpatient and outpatient providers is fixed. In the short run, this is probably a reasonable assumption that allows us to comment on the literature on links between provider supply and demand. But in the longer run, this is an untenable assumption for a number of reasons. The number and mix of providers has changed over time in response to a number of forces, including geographical shifts in the population at large, payment differences embedded in Medicare and Medicaid (e.g., the closing of many rural hospitals post the introduction of DRG/PPS), and expansions in the number of doctors (Newhouse et al. 1982b; McGuire 2000; Rosenthal et al. 2005; Chou and Lo Sasso 2009) as a result of a federal infusion of funds into physician training.

We will return to this last one because the number and type of local providers is not immutable. Assuming that it is has led to some confusion in the literature on what can be controlled for and what should be treated as fixed or what can be learned from short term associations.

With these two working assumptions, the differences among areas can be reformulated into a “reduced form” function of the characteristics listed under the conceptual framework that can be treated as fixed: population structure (in terms of demographics), health status, income, assets, and insurance coverage from the demand side; access (delay to appointment, travel and waiting time in the office), mix of providers, agency, and the structure of local and national labor and other input markets from the supply side; and public policies concerning health such as Medicare and Medicaid payment policies, certificate of need, and other regulatory features that either vary geographically or treat different areas differently. If we further ignore the role of monopsony or oligopsony in input markets (hospitals vs. registered nurses), then the various

outcomes (price, quantity, quality, treatment patterns) would vary as a function of these fixed input prices at the area level (at least in the short run).¹

Econometric Implications. Although there may be behavioral links among price, quantity, quality, treatment patterns, simple correlations among them are strictly associations or partial correlations if there is some adjustment. Those outcomes are jointly determined and the analyst typically does not have observable data to deal with residual confounding. Without further restrictions or measures, one cannot place any causal interpretation on the links between price, quantity, expenditures per capita, and treatment patterns.

There is some structure to the relationships. Demand depends on demographics, health status (case-mix and severity), income, assets, out-of-pocket price (based on insurance) and time costs (e.g. travel + waiting and delay to appointment). Supply of health care depends on input prices, (in the shorter run) the fixed stock of providers and skills/facilities given state and local regulatory policies. The supply by providers depends on the mix of public insurers with fixed administered prices, and private and ERISA insurance. The provision of insurance depends on the national rules for ERISA-based insurance and state regulation of insurance for non-ERISA group and individual insurance. We can conclude that suppliers' behavior responds to demographic mix only via demand. But we cannot infer that demand responds only to supplier mix and location via price and geographic access and appointment delays. If good data on the latter are missing, then one interpretation of supply-side effects on individual consumers is that they are picking up aspects of geographical access such as delays to appointment and the cost of travel and waiting time.

Without further structure on this model for the health care market, the full system is fundamentally under-identified. Thus, one cannot estimate a causal structure for the full blown model, only the reduced form that we have tried to focus on in this review. (This assumes that there is no set of instruments that can be used as a *deus ex machina*, or a set of local natural experiments that facilitate the estimation of certain key parameters.)

This view suggests that different types of health care use and expenditures should be studied as a function of these fixed factors and not other variables in the structure because the full structure cannot be estimated consistently. There is not sufficient information to identify the whole structure here. Thus the modeling needs to be done in reduced form, rather than in a structural modeling sense – endogenous variables as dependent variables with fixed or exogenous variables as independent variables.

The analytical issue then becomes one of choosing among data and estimators that are appropriate for the research question, while allowing the observations to be treated as clustered by geographic market in the cross-sectional case, and market by time in the panel context. That many of the key covariates (e.g., time costs in the short run, input prices) are area level variables suggests certain types of estimators may not be appropriate—pooled models without cluster or

¹ Labor in health is small relative to the larger labor market, but may not be in some skilled worker groups. In the very short run, an area market may be large relative to these labor categories. But in the longer run, workers can move from one market to another if the return on the difference is sufficient. See the discussion on the number of physicians in a geographic market.

panel corrections, simple random effects models (in the econometric sense as opposed to the broader statistical use of that term to include random coefficients), and simple fixed effects models (because they sweep out market level variables that need to be corrected for, and only rely on within market variation, not cross market variation). Given that the goal is often to examine area differences net of a basic core set of individual, household, and market level variables, a more appropriate approach would be to use some variant of multi-level or hierarchical modeling, and then to test if the shifts across markets are more than simple additive or multiplicative shifts, and also to examine variation across characteristics. Such models would allow one to say something about the variability in incremental effects of health, being insured, or other demand-side factors.

Natural experiments in policy, especially at the state level may provide insights into how the relevant local markets shift to obtain new equilibria for prices, services per capita, quality of care and other aspects of health care. Whether the specific natural experiment allows some outcomes or other aspects to be studied depends on the specific intervention. But the literature in health economics, health services research, and health policy generally is full of shifts in Medicaid and Medicare policy, including major shifts in coverage (Medicare and Medicaid's introduction, the addition of Medicare Part D), market structure (managed care and carve outs in Medicaid, the impact of Hill-Burton on poor Blacks), and supply of medical professions (expansion in funding of medical training by the federal government). These allow particularly interesting parts of the geographic variation to be explored.

One concern in all of these contexts is about policy endogeneity with respect to why the state or federal government changed certain policies and not others. Some authors have advocated the use of fixed effects approaches or difference-in-differences models to sweep out time invariant unobservables from panel studies of the impact of policies on health outcomes, such as shifts in health policies. But this does not sweep out why the state or federal government may have chosen that policy change. In general, when assessing the effect of any change in state or federal policy one has to consider the issue of policy endogeneity; for example see the discussion in Poterba (1994). For example, if we were to consider the effect of cigarette prices on smoking and treated state or local excise tax changes as a natural experiment, can the excise tax shift be treated as exogenous when smoking sentiment and behavior varies geographically?

When assessing the importance of studies of demand-side or supply-side factors contributing to observed variation in health care use and expenditures, it is important to recall there will always be a stochastic element to medical care processes. Newhouse and colleagues (1993), in a study seeking to bound the explainable variation in predicted health expenditures at the person level for a pediatric population, concluded that only 35 percent of those future expenditures could be explained, as measured using a random effects model. This study's data were limited to outpatient health care expenditures among children aged five to thirteen, for whom outpatient health care expenditures comprised a larger proportion of total use and expenditures than the population-wide average. (Therefore, the proportion of explainable variation may be considerably greater in an older population who more commonly suffers from chronic diseases, which are more predictable than acute care costs.) In a previous paper, Newhouse and colleagues (1989) measured variation in health care use at the person-year level, using controls for demographic characteristics and individual fixed effects. They focused on

adolescent and adult RAND Health Insurance Experiment participants. The authors concluded no more than 14.5 percent of total expenditures (though as much as 50 percent of outpatient expenditures), could be explained by their control variables. Inpatient expenditures, given their relative infrequency, high costs, and often unpredictable nature, dominated the authors' measure of variance in health care expenditures. Relaxing the authors' assumptions as to what would be the most reasonable specifications, the full range of the authors' estimates included a minimum of 0.25 percent (using only the individual's health plan identifier) and a maximum of 62 percent (using all available covariates) of variation in total expenditures explained. These two studies focused on the maximum explainable variance due to time-invariant but not fully observed factors. In a related study, Manning and colleagues (1982) examined the effects of measures of health status on demand for health care services as well as their ability to "post-dict" a previous year's health expenditures using current measures of health status. They demonstrated that even if the general scope of the measures applied in such regression analyses are similar, there can be significant differences in the amount of variation explained based on how the measures are structured and timed relative to the outcome of interest.

A broader literature has examined theoretic and methodological alternatives for risk-adjusting measures of use or expenditures and the extent of the variation that can be explained with these approaches (Van de Ven and Ellis 2000). Much of this risk-adjustment literature pertains to analyses conducted at the individual level, while geographic variation analyses are conducted at the area level (though the levels of analysis vary considerably across these studies, leading to exaggerated perceptions or misperceptions of compared R^2 statistics). Still, it is an empirical regularity among these studies that the amount of sample variation explainable can vary significantly with the application of different risk-adjustment methods and models (Cumming and Cameron 2002).

Across studies exploring alternative methods for explaining a population's variation in use and expenditures, differences in authors' choices of explanatory variables and in their methods for cleaning their data, trimming outliers, and making other adjustments make their estimates difficult to compare. Still, there are several lessons learned. Among these is that considerably less than all of the observed variation in health care use in a population can be explained using available data and methods while retaining meaningful explanatory power. Fixed effects, random effects, and other model specifications may be introduced in order to increase a study's R^2 statistic, but few additional insights about the nature or impact of variation explained in this manner can be gained.

Two other issues are important when looking at R^2 measures. The first is they depend on what variables are included in the equation estimated. The second is that the numbers at the aggregated or state level are inherently different from those at the facility or state level because they rely on cross state variation in the x 's included (e.g., age, race/ethnicity mix). For variables that vary at the individual level but not at the state level, these variables are washed out of the analysis. State-level analysis mostly captures what differs across states. For further discussion of this aggregation effect, see this report's methodological appendix.

We examined many papers from this literature as a part of the review. Different analyses have relied on different types of data and econometric methods. Some analyses have employed

only cross-sectional evidence, while others have employed panels of varying length. The purely cross-sectional approaches in the absence of further structure or suitable instruments only provide evidence that the variables move together, not that one is causally related to some other. We ought to be concerned about either joint determination by a process similar to the one described above or the presence of some omitted (unobservable) third factor.

Some analyses have relied on panel data, with varying degrees of treatment of the panel or cluster aspect of the data. In some, there has been no formal treatment of the panel aspects in the analysis. Many have made corrections for cluster effects (a less parametric version of the intraclass correlation model that includes the random effects as a special case). This has been true of studies with varying duration (years) of data. We have not found instances of fixed effects estimators. Given the concerns with the existence of geographic variation and its sources, there is a natural reluctance to use fixed effects estimators to sweep out the time invariant characteristics of the market or state that are not observed by the analyst. After all these are the heart of this study.

We found a few papers that allow for some convergence in health care expenditures over time. We expect that markets and individual actors (patients or providers) are involved in dynamic processes including entry, exit, and responding to policy shifts. These processes take time if the market is out of equilibrium. This recognition raises a number of issues. First, although the geographic patterns have been relatively stable over time, it is not clear that time-invariant factors (implied by the random effects and fixed effects formulations) are the only determinant of geographical variations, especially over longer time frames. Second, the inferences drawn in such settings are often biased because the intertemporal correlation is not intracluster but slowly decaying over time. If not adequately addressed, this misspecification of the error structure could impart a false sense of confidence in the results (see, for example, Bertrand and colleagues (2004) on trusting difference-in-differences models) if the correlations are not oscillating and the time trend in area variables is of the same sign as the correlation structure. If of the opposite pattern, then one could reach a false sense of imprecision for the results. Moreover, any researcher that examines differences over time has to address the question of why the covariates evolve over time. Third, some variants of the reduced form interpretation may no longer be valid.

Note that such dynamics reflect more than temporal correlations. One must also consider the implications of decisions made in response to perceived trends. One example can be found in the planning activities of health system executives. Such officials in areas like Nevada and Arizona, which have experienced dramatic population growth over the past several decades, may have invested in their systems' expansion in anticipation of continued population growth, only to find their projections were too high. In this example, growing demand sparked a significant increase in supply that could potentially resort to inducing demand later on.

Wang (2009) used information on health care per capita expenditures by state for the period from 1980-2004 to test for convergence across regions. Using time series and panel methods, he found evidence of convergence that differs across type of health care. But the magnitude of the shifts suggests that market forces alone are not sufficient to guarantee full convergence. Panopoulou and Pantelidis (2012) used the same data base to find evidence that the

health markets not only differ, but markets are converging to different equilibria. States in the South and West tend to be in a low spending class, while states in the North and East tend to be in a high spending group.

Focusing on Extremes. One of the recurring issues in the geographic variation literature is the focus on extreme cases in the cost per patient or cost per capita. The early work in this area by the Dartmouth group and other small area variation analysts was criticized for failing to allow for the fact that sampling variability issues alone can lead to differences between the maximum and minimum values (Kazandjian et al. 1989; Diehr et al. 1990). Any formal, statistical inferences need to be corrected for the fact that they are not randomly drawn but are selected based on the data.

This concern about over interpretation of extremes has largely been lost in the recent literature on geographic variation, especially in the work that is based on small to moderate sample sizes or cell sizes, which have occurred in many of the papers that we reviewed. (Note: although one rarely sees the extremal quotient anymore as a formal tool, the same issue applies to analyses that contrast areas based in part on that they are more extreme in their observed rates net of adjustments.)

Alternatively, one could ask the question that if the underlying data was a model of individual costs and use (overall or for a specific disease), how much of the variance across areas would exist if there were neither area-specific random nor area-specific fixed effects, after adjusting for the core set of covariates that we have argued should be a part of the covariate adjustment?

Expanding the specification to allow for additional area effects, such as input prices and cost of living adjustments (COLAs), how much of the overall variance in individual cost or use is attributable to area effects? If we think of these as either random effects models or intraclass correlation models, what fraction of the overall residual variance is attributable to these unspecified area effects, compared to unobserved person/household variables?²

Consider a simple random effects model at the person level where the dependent variable of outcome is y , the observed independent variables are x , μ are area effects that are homogeneous within cell s , and ε is an error term independent of x and μ

$$y_{is} = x'_{is}\beta + \mu_s + \varepsilon_{is}$$

Then the area (s for state) level is

$$\bar{y}_{\bar{L}s} = \bar{x}'_{\bar{L}s}\beta + \mu_s + \bar{\varepsilon}_{\bar{L}s}$$

² Note that a quick and dirty analysis of the MEPS 2004 adults suggests that the intraclass area correlation is on the order of 0.0087 for total expenses without any adjustment for covariates or design effects. The cluster effects as a set are significant ($p < 0.0001$) if treated as fixed effects. Of course, this raises the questions of “how big is big” for such a correlation, and do we temper our results based on the variances in the outcomes.

Where the notation of an “over bar” and a “sub dot” indicates an average over the observations i but retaining the distinction by s . The last term for the individual independent and identically distributed (*iid*) error rolled up to the area average has mean zero and variance

$$\sigma_{\varepsilon}^2 / N_s$$

If the underlying *iid* error ε is normally distributed, then so will be $\bar{\varepsilon}_{\bar{s}}$.

The most extreme cases in terms of area effects observed are likely to be areas with smaller populations at risk if the data are analyzed at a rolled-up or aggregate level.

There appears to be an absence of formal hierarchical modeling or of correction for intra-cluster correlations and area level covariates. Although the correlation may be small, what matters is the product—which could be large—of the ρ with N_s in area s . Such an approach would allow for the use of individual or household level covariates and data. An additional benefit is that the model could be extended to allow for random slopes on key covariates.

One of the other issues that arise when analyzing rates of spending or visit or admission rates is that places or institutions with smaller populations will be more variable than the corresponding per capita rate for larger states or institutions. The most widely accepted approach to account for this is to do a weighted analysis with a weight proportional to the population at risk in the state or institution. This is the correct approach if the data are independent within state or institutions. That may not be the case if there are state or institution level effects such as μ capturing state or institution effects. If there are such concerns in the model, the most efficient correction is to weight each state by an appropriate correction for weighting for rates. For efficiency in GLS the area rate should be weighted by

$\left[\left(\frac{\sigma_{\varepsilon}^2}{N_s} \right) + \sigma_{\mu}^2 \right]^{-1}$ rather than by the area’s number of patients at risk N_s . If the underlying sample size N_s is small, then the result is that large states get larger rates than they would under a population weight N_s . If the underlying sample sizes are quite large, then each state will have roughly equal weight.

Area-Level Covariates for Individual or Household Variables. In this literature, it appears that several of the papers have used area averages of individual or household covariates when doing analysis of area-level expenditures and use per capita; many other studies have used area-level measures of person/household variables to impute such information about individuals when the administrative data (e.g., Medicare claims) do not have them. Each of these two practices raises issues in its own right, including aggregation bias and Berkson measurement error and potential transmitted bias issues. (For further discussion of these issues, see Berkson 1950, and Buzas et al. 2005.)

Aggregation bias is often used to refer to situations where the behavior of the aggregate or average (as in area mean) is different from what would be observed in the underlying micro

data (Theil 1971). In some cases, this may reflect underlying confounding that has not been corrected for in the aggregate analysis. In others, it may reflect the tendency in observational studies to attribute the behavior of the average to the underlying cases, as in Freedman's (2002) comment on Durkheim's analysis of suicides that inferred that Protestants were more likely to commit suicide because countries with more Protestants had higher suicide rates; this is also referred to as ecological fallacy. Finally, there is the narrow sense of the term as used in econometrics that the macro- (aggregate) level parameters reflect the micro- (individual) level variables whether one is aggregating over individuals, commodities, or time. In general, simple averages are not enough, but instead there needs to be some weighted averages that depend on the underlying micro variables and their distribution (Theil 1971).

An additional issue is that if the underlying micro model is nonlinear in the covariates, then an aggregate-level model of the average micro-level variable is a form of misspecification that can lead to bias if the underlying variables are not symmetric around some mean. This is the same problem that says that estimating a simple linear function of y on x , rather than on x , x squared, etc., provides a consistent estimate of the average marginal effect. This is true if the covariate is symmetrically distributed or if the underlying curvature in the range of interest is minimal.

II. CRITICAL REVIEW OF THE EMPIRICAL LITERATURE

1. INTRODUCTION TO DEMAND-SIDE VARIATION IN USE

As we discussed in the conceptual framework, demand-side factors that influence the demand for health care include an individual's health status, sociodemographic variables, health behaviors, preferences for treatment, and constraints on income and insurance. There is a long-standing recognition of the association between sociodemographic characteristics and both economic factors and health status based on many studies in the literature. The literature on area variation, while including sociodemographic and economic factors as covariates, has tended to focus on explaining the effect of supply-side factors on variation in health care use and expenditures and has paid less attention to the contributions of these demand-side variables.

The empirical evidence on the relationship among health expenditures and sociodemographic and economic factors, as well as theory, supports their inclusion in models attempting to explain variation in spending. The question is: at what level should these variables be included? Demand for care has often focused on characteristics and factors for a specific individual or family. However, there is evidence that these individual and household characteristics averaged over larger population characteristics and factors may also influence the demand for care. The research question can direct the inclusion of individual characteristics, population characteristics, or both.

The empirical literature on geographic variation in health care use, expenditures, and outcomes can generally be divided into two types of outcomes. One body of work examines health care spending by geographic area for an entire population. For these aggregated outcomes, information about individual-level demand-side characteristics is aggregated to predict percent use or per capita spending at the population level. In most of these studies, the aggregated individual-level demand-side factors were included in analyses as covariates in predicting unadjusted population-level outcomes. In other studies, these factors were used to adjust the outcome measures of spending and health care use.

The second body of work examines use and expenditures by individuals for a specific illness, medical procedure, or hospitalization. In these micro analyses, demand-side variables are measured at the individual level. Population-level measures are included predominantly when the data are not available at the individual level. For example, income is often not available in administrative or claims databases. So Census data is often used to extract income data for the area and applied to the individuals in that area.

2. HEALTH STATUS MEASURES

Health status at the individual level is clearly a significant determinant of the demand for health care. In addition, at the population level health status is not evenly distributed geographically in the United States. The prevalence of many chronic illnesses, such as coronary heart disease, cancer, and arthritis varies geographically (CDC 2011). Some regions of the

United States have recently been labeled as the “diabetes belt” (Barker et al. 2011), the “stroke belt,” and the “heart failure belt” (Mujib et al. 2011).

Empirical evidence links health status, measured in various ways, with health care expenditures as well as with the receipt of medical care (e.g., Wolfe 1986). Analyses attempting to describe or explain geographic variation in health care spending have frequently included at least one measure of health status at the individual level or the population level, occasionally both. Measures of health status encountered in the literature on geographic variation can be grouped into two broad categories: clinical conditions and self-rated health.

Importantly, “health status,” “case mix,” and “severity” are often used interchangeably as reflections of health status, perhaps depending on the author’s background in epidemiology or health services research. The literature is somewhat agnostic as to the appropriateness of using one or another of these in different contexts, despite the fact that they are distinct, if highly related, concepts (Patrick and Erickson 1993; McDowell and Newell 1996). In this section, we discuss the uses of the components that, in various combinations, comprise these concepts, including the identification of patients with certain clinical conditions or levels of severity with a condition, counts of conditions, and measures of overall health.

Clinical Conditions. A common measure of clinical health status is the presence (or number) of a set of co-morbid conditions experienced by the individual. Co-morbid conditions are frequently identified by ICD-9 diagnosis codes in the claims record. Health status is often indicated by the presence or absence of a particular condition, rather than the severity of a particular condition (e.g., stage of cancer). Conditions most commonly controlled for include: acute myocardial infarction, coronary heart disease, diabetes, stroke, cancers, and hypertension.

To create a more comprehensive picture of an individual’s health status, rather than include an indicator for each co-morbidity, a classification system is sometimes used as a single overall measure of health status. One diagnostic classification system used is the Hierarchical Condition Category (HCC) classification system. The HCC system categorizes ICD-9 diagnoses codes. The conditions included in the HCCs are selected based on a set of ten criteria including clinical meaningfulness, ability to predict expenditures, and higher disease prevalence (Pope et al. 2004). Rather than an indication of the total number of co-morbidities, the HCC system imposes a hierarchy of severity so an individual’s classification is weighted for the most severe among related diagnoses (Pope et al. 2004). The HCC system does not indicate the total number of diagnoses but does reflect the range of disease categories with which an individual has been diagnosed as well as the severity within a disease category.

The Charlson index of co-morbid illness is another frequently used index (Charlson et al. 1987) in analyses of variation in use and expenditures. The Charlson index is a weighted index of the number and severity of co-morbid conditions. The index includes 19 diseases selected based on their ability to predict mortality. It has been validated in a variety of populations; however, given the criteria used in identifying the index’s core conditions, it could argued the index is more appropriately used in studies of variation in mortality rather than in use or expenditures (de Groot et al. 2003).

Clinical health status also can be measured through patient-level surveys. The use of such measures—based on responses to questions like, “Has a doctor ever told you that you have diabetes?” as indicators of health status—can be problematic. Because they require access to health care, such measures are potentially subject to variation in physician diagnostic practices and patient care-seeking behavior or whether the potential patient has visited a provider or filled a prescription in the recent past. Song and colleagues (2010), for example, demonstrated substantial regional variation in diagnostic practices, controlling for Medicare beneficiary characteristics. Access to care has been shown to vary with sociodemographic characteristics (see, for example, Morales et al. 2004; Bryce et al. 2009) and place of residence (see, for example, Stearns et al. 2000; Paquette et al. 2011). Self-reports of sensitive behaviors or of health issues are also often reported with bias. Butler and colleagues (1987) report measurement errors associated with a self-reported measure of arthritis finding that the measurement error was systematically different across socioeconomic groups.

Other health status variables are measured at the population level. These health status variables included general mortality or mortality rates for specific conditions such as cancer, liver failure, or infectious diseases. Population-level mortality rates are calculated from state or federal vital statistics. These data are typically measured at the state or metropolitan statistical area (MSA) level and are applied at the hospital referral region (HRR) level.

Another population-level measure of health status is based on the rate of hospitalization for five conditions (i.e., hip fracture, heart attack, stroke, gastrointestinal bleeding, and surgery for lung or colon cancer) used by the Dartmouth Atlas of Health Care and those using Dartmouth data. The argument for using these conditions to reflect a population’s health status overall is that their incidences are reasonable proxies for the incidence of disease generally (CECS 1998).

The data sources for health status that researchers have had access to include administrative data such as Medicare claims or hospital discharge records. Measures of health status have also been extracted from survey data such as the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS), and the Medicare Current Beneficiary Survey (MCBS). Few studies examining the association between health status and geographic variation have fielded their own surveys to collect self-reported data on health status. Although these surveys are typically nationally-representative sample surveys, analyses using these data are limited to measures collected through survey methodology and were not conducted with the secondary study’s research question in mind. Individual-level health status variables from these various sources are sometimes included as proxies for population-level health status; however, data taken from administrative databases are not necessarily representative of the entire geographic area of interest (e.g., HRR) but only of the beneficiaries or sample members assigned to that geography.

Population-level health status variables that might be more representative of the geographic area are often extracted from Census data, American Hospital Association (AHA) data, or the Area Resource File (ARF). Problems could potentially arise with including these measures at the population level because the data are often only available at a different level than the level at which the outcome variable is measured. For example, data on mortality or disease rates, rates of disability or limitations have not been available at the HRR or HSA level but rather

only at the MSA or state level. Because of spending data limitations, spending is typically defined by HRR or HSA. Demographic information is not often available at the same level so analyses include demographic data measured at the level of the MSA. Studies have argued that there is strong overlap between HRRs and MSAs; therefore, the error that is added by the different levels of measurement will be minimal.

It is not always clear from the literature if studies included population-level variables in order to evaluate the contribution of the larger population's health on area aggregated individual health care expenditures or if these population measures were included when individual-level measures were unavailable.

Self-rated Health. Another frequently used measure of health status is the individual's self-reported response to the question, "In general, would you say your health is excellent, very good, good, fair, or poor?" This single item measure and its multi-item (and often multi-dimensional) measure(s) have been used in a series of studies and general population data sets.

Self-rated health has been widely used in large national surveys such as the NHIS and NHANES. This also sometimes referred to as self-assessed health (SAH). Self-rated health is attractive because it is considered an easy item to implement as well as a succinct way to summarize the underlying array of complex health dimensions (Krause and Jay 1994). Self-rated health has been used extensively in the economics literature to measure health status and its popularity is in part due to its ability to account for large portions of variation in demand as well as its ability to affect the magnitude of other coefficients (Manning et al. 1982), although multi-item and multi-dimensional measures of health are clearly better. As with more clinical measures of health, self-rated health has also been shown to vary at the state and local levels (CDC 2011).

Despite the predictive strength of a single item self-rated health, as an indicator of health status it may be unclear as to what exactly this single item is measuring as respondents may use different frames of reference when answering. In a series of in-depth interviews, Kraus and Jay (1994) report a range of frames respondents were using when responding. Self-rated health by itself should not be considered a comprehensive measure of health status. Including measures of other aspects of health to create a more comprehensive picture significantly improves the performance of models predicting demand. For further discussion of these issues, please see the citations in Manning and colleagues (1982) and relevant subsequent literature.

Another issue with the traditional use of self-rated health is that present health status is used to explain past expenditures, especially if they are collected in the same interview. The later health status variables may be endogenous; this leads to a post-diction bias, but finding good instrumental variables is difficult. Manning and colleagues (1982) found substantial and significant differences in a comparison of "prediction" versus "post diction" models (using baseline health to predict subsequent health care expenditure). The post diction model resulted in larger R^2 values as well as larger absolute values of the coefficient for the health status variables. In addition to the pre- versus post-diction limitation, no distinction is typically made regarding stable, declining, or even improving health over the course of the spending period.

One study by Zuckerman and colleagues (2010) was able to evaluate a change in health status by measuring conditions at baseline and newly diagnosed conditions during the year of study. The amount of variation explained by changes in health status was not reported; however, Zuckerman and colleagues did find a significant difference between spending quartile in the proportion of some newly diagnosed conditions. Continuously collected datasets like the Medicare Current Beneficiary Survey, the MEPS, and the HRS provide opportunities for using predictive modeling to examine the relationship between previously collected health status and future spending.

Effect of Health Status on Variation. Despite the inclusion of measures of health status in many of the studies reviewed, few evaluated the amount of geographic variation in health care spending explained by health status alone, net of other characteristics. Those that reported the amount of variation explained by health status reported a wide range of estimates, most likely due to variation in the measures of health status employed. The Congressional Budget Office (CBO) reported in their 2008 report that previous work by MedPAC had concluded 16 percent of the variation in spending could be explained by measures of health status.³ Previously Cutler and Sheiner (1999) had reported that as much as 66 percent of variation in spending could be explained by their measures of health status. Much of this discrepancy can likely be explained by their different levels of analysis and methodologies: MedPAC's (2003) analysis had been conducted at the state level weighting by population, and Cutler and Sheiner (1999) had conducted their analyses at the HRR level (drawing on information about some of their covariates from MSA- and state-level data sets) with estimates left unweighted.⁴ Newhouse and colleagues (1989; 1993) reported explainable variance estimates from stable health measures (time-invariant in the short run) for more explicit health status measures at the individual level. They reported estimates of approximately 10 percent for both dichotomous and continuous clinical measures (e.g. hemoglobin levels); and 5 percent for self-rated measures of general and mental health and physical limits while also adjusting for demographics. Combined with various other health status measures, the amount of variance explained was between 23 percent and 25 percent. Zuckerman and colleagues (2010) reported a similar level (29 percent) of the explained variance by combined health status variables at the individual-level. The explained variance is typically reported for a group of health indicators, or health indicators in addition to a set of sociodemographic characteristics.

One difficulty in drawing conclusions about the explanatory power of health status results from studies using different indicators of health status and possibly measured at different levels. Not all studies include both self-rated health and measures of clinical conditions, nor did all studies include the same conditions. As demonstrated by Manning and colleagues (1982), a

³ This figure could not be verified in the version of the June 2003 MedPAC report cited by CBO that was publicly available as of the preparation of this report.

⁴ We were unable to verify for either study whether the health status measures derived for analysis were based on evidence of patients' clinical diagnoses during the year for which their per capita measure of expenditures was analyzed (retrospective) or on based on evidence of patients' clinical diagnoses in the year prior (prospective). If either analysis were conducted retrospectively or prospectively, we should expect higher or lower R^2 statistics, respectively. Some of the range of these estimates could be partially explained if Cutler and Sheiner (1999) estimated the effects of health status on expenditures retrospectively, while MedPAC generated their estimates prospectively.

greater proportion of variation in spending that is explained depends on more comprehensive measures of health status, among other things.

3. DEMOGRAPHIC PROXIES FOR UNMEASURED HEALTH STATUS

Substantial evidence points to differences in health (see, for example, Dunnell et al. 1999; Taylor et al. 2005), access to care (see, for example, Morales et al. 2004; Bryce et al. 2009), and spending (see, for example, Taylor et al. 2006; Shang and Goldman 2008) associated with demographic variables. Demographic groups are not uniformly distributed across the United States—such variation is routinely used in studies such as that by Groenveld and colleagues (2005). As with health status, analyses attempting to explain any variation in health care spending should include as possible explanatory variables demographic variables. Age, sex, and race and ethnicity are demographic variables that are typically considered important confounders of health and health care spending and use and should be controlled for in analyses examining variation in health care spending and use. While demographic variables are not health status *per se*, they are related to health through such things as health insurance, income and employment, and health behaviors (Farley 1985; Lantz et al. 1998; Adler and Newman 2002; Case and Deaton 2005; DeNavas-Walt et al. 2006).

Age and Gender. Age and gender are extracted from administrative databases and are available in survey data. Although age or aging does not necessarily directly affect health, its clear association with health (Yang et al. 2003; Case and Deaton 2005) has led to its inclusion to address missing or inadequate adjustment for health status, case mix and severity. Age is often grouped into 5-year age groups with the highest age group as a catchall for those over a certain age. Two direct mechanisms explaining the link between gender and differential health care utilization and expenditures are, first, pregnancy and family formation—though this mechanism’s effects on disparities in health care use have declined in recent decades in the U.S. due to increasing women’s workforce participation among other factors (Sorensen and Verbrugge 1987; Blau 1998; Haas et al. 2004)—and, second, the linkage between gender and select behavioral and psychosocial patterns tied to health (Green and Pope 1999; Kawachi et al. 1999; Koopmans and Lamers 2007). Both age and gender may be introduced into regression analyses as proxies for unmeasured behavioral and other health status predictors. However, since age and gender distributions tend, by and large, not to vary significantly across regions, they may not contribute as much to studies of geographic variation as other demand-side variables.

Race and Ethnicity. The race or ethnicity at the individual-level is often extracted from administrative databases. For population-level estimates of racial or ethnic composition, Census or other state or federal data are used. Issues for this category of variables can arise due to missing information or because of small sample sizes. For example, race is often categorized into black or non-black, either because finer race and ethnicity information was not collected or the number of persons in a given racial or ethnic category is too small for subpopulation analyses. For population-level data, questions have arisen regarding the appropriateness of using Census data as a substitute for individually-obtained race and ethnicity information (Chen et al. 2004).

Age, sex, and race and ethnicity are most frequently used as case-mix adjustments to the expenditure or health care use outcomes of interest. Such adjustments have been made to the outcome data prior to examining area effects. In analyses using the Dartmouth Atlas data, Medicare expenditures are adjusted first to reflect regional price differences, and then adjusted to reflect differences in the age, sex, and race of the beneficiaries between the HRRs (Gottlieb et al. 2010b). Cells of individuals are created by 5-year age groups, sex, and race (dichotomized as Black or Non-Black). Adjusting the expenditures in this way provides estimates of what expenditures should be given the demographic make-up of the Medicare beneficiaries in each HRR. Moreover, this process is often used as a way of indirectly adjusting for case mix, the demographic variables serving as proxies for more precise health status measures. However, as we have discussed previously, researchers risk introducing aggregation bias or Berkson measurement error when using area-level averages or imputing individual covariates from area-level data, even if those values are based on data collected at levels as fine as the Census tract (Berkson 1950; Buzas et al. 2005).

Moreover, to the extent other variables (such as racial and ethnic indicators) are used as proxies for unobserved socioeconomic variables such as education or income, researchers risk introducing other forms of bias. One issue to consider when using proxy variables or census values for other covariates is that we are imputing new values to replace missing data (missing from the data source). In the case of a single covariate, we know that classical measurement error will bias the estimate of the coefficient for that covariate toward zero; this is often referred to as attenuation. But if there are multiple unobserved covariates that maybe correlated with each other, then the story is more complex; the resulting bias may be indeterminate because it depends on the correlations among the covariates and the different extents of measurement error. If the two or more covariates are not uncorrelated either in truth or in measurement error, then the measurement error in one will be transmitted to the others (see Greene 2012 for discussion of the single-covariate and more general case). For example if an analysis were to include both income and education (which are positively correlated), and if there is more measurement error in income than in years of education, then it is easy to show that the coefficient of income will be biased toward zero while that of education is biased away from zero if the measurement error is classical.

A related issue comes up when using means or Census tract values to impute missing data. If two covariates are correlated, then simple mean replacement will lead to bias because of the correlation between covariates that are not imputed and those that are. This is why in general one wants to use conditional mean (conditional on other covariates) rather than unconditional means to impute missing data (Little 1992).⁵

Education. A significant body of evidence has been accumulated pointing to the relationship between education level and health status with higher levels of education associated with better health outcomes (for example, see Winkleby et al. 1992; moreover, for more comprehensive discussion, see Ross and Wu 1995) as well as health care use and spending (for example, see Bombardier et al. 1977); these concepts are often linked by the mechanism of health literacy (Baker et al. 1997; AMA 1999; Scott et al. 2002). When included with income in analyses of health outcomes, education is more predictive of chronic conditions and the onset of

⁵ Also, as Wickens (1972) has shown, if the focus is on the reducing bias, a poor proxy is better than none.

functional limitations; income is related more to the progression of functional limitations and chronic conditions (Herd et al. 2007). Part of this reflects measurement error issues in self-reported income compared to education. Income in most surveys has nontrivial reporting issues (including bias and reliability). The economically correct income measure (permanent income) is rarely available. In addition, these measures are strongly correlated with each other, making it hard to separately identify the effects of each. Education level has also been shown to be correlated with insurance status (Hall et al. 1999), which subsequently affects patterns of health care utilization.

Despite its relationship to a variety of income and health-related variables, a measure of education is not often included. If included, coding of education is not consistent across studies and measurement is at the population level.

Implications. The difficulty in drawing conclusions about the amount of variation explained by demographic factors is that often estimates of explained variance are not reported for each measure with or without other often correlated measures. Further, demographic factors were combined with health status variables, or only an overall estimate of the variance explained is reported accounting for all variables in the model. The range of estimates that have been reported for the possible variance explained with demographic factors is wide, and estimates may vary significantly with choices of the level of analysis (Krakauer et al. 1996; CBO 2008; Anthony et al. 2009). As an example, Dartmouth Atlas researchers in their 1998 report concluded that age, sex, race and illness did not explain much of the variation in health care utilization, but this could be an artifact that most of the variation in these variables is within area, and less between areas. While individual health needs are predictive of demand for health care at the individual level, the authors conclude, they do not explain variation in utilization among hospital referral regions (CECS 1998).

Several difficulties also arise in drawing conclusions about the amount of variation among geographic areas explained by demographic proxies for health status. First, estimating the amount of variation explained by individual-level demographics is difficult given that most studies used these variables as adjustments to the outcome variable (such as age and gender) rather than including them as independent variables in the model. A second difficulty is that the studies reviewed do not use a consistent set of demographics or use different coding for the same construct. Demographic variables are frequently included in modeling geographic variation and probably explain some proportion of the variation in health care spending. However, measurement issues, inconsistency in coding, and a lack of reporting of the amount of variation accounted for specifically by demographic factors do not allow for conclusions about the importance of these factors in explaining variation in spending.

4. HEALTH BEHAVIORS

Health behaviors, such as smoking or drinking, are also demand-side factors that may affect demand for health care. Behaviors such as smoking, alcohol and substance use, and other risky behaviors have been shown not only to vary by sociodemographic characteristics, but also to vary geographically and by SES.

Smoking. Smoking can have a significant effect on one's health and has been identified as a risk factor for numerous health problems. Smoking rates not only differ by gender, race, education, and income, they also differ geographically (CDC 2001; CDC 2009b; CDC 2011b) which further contributes to variation in smoking related diseases such as lung cancer and coronary heart disease. Plus, it is known that smokers receive less preventive care themselves and for their children (Vogt 1984).

However, the contribution of smoking in explaining geographic variation in spending or health care use has not been adequately assessed. Only a handful of studies include an indicator of smoking status. Fisher and colleagues (2003a) included an indicator for whether the individual was a current smoker in assessing trends in HRR end-of-life spending, and found no association for a general population cohort and spending quintile. For an acute myocardial infarction cohort however the authors report a significant trend with the proportion of smokers decreasing with increasing end-of-life spending. However, the relationship they report is not linear; only the proportion of smokers in the highest quartile is lower than the proportion in the lowest quartile. Zuckerman and colleagues (2010) expanded the individual-level indicator to include never, current, and former smokers in a model of HRR spending quintiles (Medicare spending per beneficiary at the HRR level). No estimate of the geographic variance explained by smoking status was reported; however, the authors report no association between smoking status and quintile of spending.

At a population-level, Cutler and Sheiner (1999) assessed the impact of smoking rates in each MSA on Medicare spending in HRRs. The authors included smoking status as one of several illness variables and did not assess the impact of nor do they report the associated explained geographic variation in spending with smoking status. Skinner and colleagues (2001) also include a measure of the proportion of the population that was current smokers. Using Centers for Disease Control and Prevention data, they derived HRR-level estimates of current smokers by weighting based on the relative state population in each HRR. Estimates of the proportion of explained variance were not reported for smoking; however, the proportion of current smokers was not associated with HRR Medicare expenditures.

Binge Drinking. Binge drinking has been identified as a risk factor for health issues such as hypertension and acute myocardial infarction as well as with other health risk behaviors such as smoking and illicit drug use (Miller et al. 2007; CDC 2011a). Binge drinking also varies with sociodemographic characteristics (CDC 2009a; Naimi et al. 2003) and geography (CDC 2011b). Binge drinking has only been included in one reviewed study of geographic variation (Skinner et al. 2001). The proportion of the HRR population that report binge drinking was included in a model of Medicare expenditures per HRR. Despite a positive coefficient indicating greater spending with a higher proportion of binge drinkers, the relationship between this proportion and HRR Medicare expenditures was not significant.

Obesity. In addition to smoking and binge drinking, obesity is a significant risk factor for many health problems such as cancer, diabetes, and asthma (DHHS 2001). As with smoking, obesity rates not only vary by sociodemographic characteristics such as race and education they also have also been shown to vary dramatically across States and MMSAs (Ford et al. 2005;

CDC 2011b). A wide range of physical activity levels have been observed across States as well as MMSAs (Ford et al. 2005; CDC 2011b). When controlled for, obesity was included as a population-level indicator based on data extracted from CDC, state, or MSA level data. Cutler and Sheiner (1999) included the proportion of persons who are obese in each MSA; however, the authors report no results for obesity specifically, only that “illness,” represented using a broad collection of illness indicators, explained 66 percent of the variation in Medicare spending at the regional level (i.e., HRR). Skinner and colleagues (2001) included the estimated proportion of obese per HRR, reporting no significant association between proportion obese and HRR Medicare expenditures.

Seat Belt Use. Seat belt use, despite all states having laws, varies across the United States. Strine and her colleagues (2010) found that self-reported seat belt use (i.e. always wearing a seat belt) varied from as high as 89.6 percent of those living in the West to 80.4 percent in the Midwest. Seat belt use also varied by metropolitan designation with those in rural areas less likely than those in the most densely populated areas to report always wearing a seat belt. The only study including seat belt use found no association with Medicare expenditures (Skinner et al. 2001).

Bad Health Index. Rettenmaier and Saving (2010) combine the percent current smokers and percent obese into what they called a “bad health index.” The data used to calculate the index were taken from the CDC. The authors did not find a consistent relationship between worse health and higher per capita personal health care spending or per capita Medicare spending.

Implications of Health Behaviors for Geographic Variation. Health behaviors and other measures of risky behaviors are endogenous with health status. For the former, excise taxes can be sometimes used as an instrument. This is somewhat controversial because of its correlation with sentiment at the state level about smoking and other behaviors (Evans and Ringel 1999, DeCicca et al. 2008).

The impact that health behaviors have on geographic variation has not been sufficiently assessed in the literature. A minimal number of studies have included measures of various health behaviors; however, none has reported estimates of the amount of variation in health care spending that these measures explain and few have reported any significant associations between the behavior and geographic variation in spending. More work is needed to begin to understand what if any impact health behaviors have on geographic variation. In principle it would not be hard to make substantial progress on these issues because several of the larger survey and administrative datasets contain measures of many of these variables, as well as geographic identifiers.

5. TREATMENT AND CARE PREFERENCES

As previously mentioned, health status is a significant determinant of the demand for health care. But not all individuals with the same level of health will demand the same type or even the same intensity of care. Preferences for care may be based on cultural or religious

beliefs or simply individual desires for and comfort with particular treatment methods. Differences between racial and ethnic groups in their preferences for care have been for end-of-life care (Barnato et al. 2007b) and having a do-not-resuscitate or a do-not-hospitalize directive (Levy et al. 2005). Preferences have also been associated with spending. In a study of Medicare beneficiaries, limited care advance directives were associated with lower spending (Nicholas et al. 2011). To the extent that preferences are associated with health care expenditures and preferences vary by region and other demand-side factors, it is important to include measures of such preferences when attempting to explain geographic variation in health care use and spending.

Only a limited number of studies have examined the contribution of patient preferences to the explanation of area variation in spending or use. One study focused on individual preferences for seeing a doctor right away, for getting a test even if the doctor said one was not needed, and for seeing a specialist (Anthony et al. 2009). These items are coded as “yes” or “no” responses to a hypothetical medical situation. Anthony and colleagues (2009) found significant differences between quintiles and preferences with those in the lowest quintile expressing preferences for less intense treatment; however, the differences across quintiles were small. Another study by Barnato and colleagues (2007b) evaluated the effect of individual preferences for end-of-life care including the level of concern for the amount of medical treatment potentially received, the location for spending one’s remaining days, the desire for life-prolonging drugs that would make one feel worse, the desire for drugs that make one feel better but shorten one’s life, and the desire for life-prolonging mechanical ventilation.

Effect of Preferences on Geographic Variation. Associations between preferences and expenditures and utilization have been reported by Anthony and colleagues (2009). However, the amount of geographic variation explained specifically by patient preferences is unknown. Anthony and colleagues (2009) reported R^2 values ranging from 0.087 to 0.095 for the amount of geographic variation explained by each preference but these values are based on models that also included various demographic factors. Barnato and colleagues (2007b) found little evidence of a relationship between regional expenditures on end-of-life care and patient preferences, but these preferences were obtained by survey from patients not yet in need of such care. It is difficult to interpret whether these preferences as stated reflect well the preferences of patients (or their families) at the end of life in the same region.

One of the limitations of the limited empirical work on treatment preferences and geographic variation in spending and use is that end-of-life preferences have been determined by the treatment a patient received and not the actual articulated preferences of the patient. Caution should be taken when interpreting any relationship between preferences and medical care. Assessing the contribution of patients’ preferences for treatment is also hindered by a lack of data on patient preferences. Preferences are rarely, if ever, recorded in administrative data and very few studies have collected such information. A significant limitation in drawing conclusions from the few studies that have included preferences has been the lack of consistent definitions of treatment and care preferences. A second limitation is that articulated preferences for treatment or care may not translate into actual choices or care received. It may also not be clear whose preferences are actually being measured. With end-of-life treatment Hirschman and colleagues (2010) show that it may not always be the patient’s preferences that are the

determining factor but those of the patient's family. Finally, treatment preferences are, as with many of the other demand-side variables considered in this section, correlated with demographics also of interest in explaining geographic variation.

To the extent that patient preferences do account for some of the geographic variation in spending, whether or not such factors are a concern for policy makers would depend on the financing mechanism for the care provided. Care financed by the private sector may not be a public policy concern (CBO 2008).

6. BUDGET CONSTRAINTS

The level of expenditures and health care use are also dependent on the budget constraints that individuals face. Constraints such as income and assets, having insurance coverage, the terms and cost sharing one's health insurance can affect decisions to use and how much to use of health care services. Such constraints not only vary demographically, they also vary by socioeconomic status (SES) and geographically.

Insurance Status and Level of Benefits. Whether or not an individual has health insurance can significantly affect the use of health care services. For those who do have insurance, the specific benefits associated with the specific plan can play a major role in determining how much health care is consumed as well as what health care is provided (e.g., Manning et al. 1987; Hadley and Holahan 2003).

Insurance coverage and the generosity of those insurance plans are not uniformly distributed across the United States. The structure of deductibles, copayment, stop losses, limits on benefits, and other internal limits (such as doughnut holes) and how they vary by service and family structure will affect the use of health care. Rates of uninsurance vary across metropolitan statistical areas (e.g., Ahluwalia et al. 2009) and rates of under-insurance (due to limited scope of benefits, or high cost-sharing) have been shown to vary between rural and urban locations (Ziller et al. 2006). Those without health insurance tend to use fewer health care services, delay seeking care, are less likely to control chronic diseases well, and are also less likely to receive costly interventions or to use preventive services (Hadley 2003).

In addition to variation in having any coverage, variation in the type of insurance and in the generosity of health insurance coverage also exists. Some of the variation is due to differences in state regulations of private insurance and in eligibility for public programs such as Medicaid. Although eligibility for Medicare does not differ across states, differences in owning supplemental insurance coverage have been observed between rural and urban Medicare beneficiaries (Laschober et al. 2002) as well as health maintenance organization (HMO) and Medicare Advantage alternatives to traditional Medicare.

Medicaid programs vary across states and by subgroups. This also affects the dually eligible population as well, a group with much higher use and expenditures than the general population enrolled in either Medicare or Medicaid but not both simultaneously.

For many of the studies of variation in spending that were reviewed, individual-level insurance is well-controlled given that the most frequently utilized data source is Medicare and often limited to those covered under traditional fee-for-service. Other individual-level insurance differences, such as supplemental insurance coverage, have largely been ignored. Any differences in insurance status were mostly accounted for with measures at the population level. The proportion of the population uninsured (Kaplan 2011; Chen et al. 2009; Cutler and Sheiner 1999), the proportion of non-elderly with HMO coverage (Cutler and Sheiner 1999; Fisher et al. 2003a), and the proportion of the entire population eligible for Medicaid (Lee et al. 1997; Zhang et al. 2010) were frequent measures of insurance status.

Given the emphasis on public insurers in the geographic variation literature, we wish to note that the role of the government in affecting the health care system and geographic variation in use and expenditures within it should not be understated. On the demand side, the government serves as the largest “single” payer for health care services and selectively improves access to care for millions of Americans. However, the government also plays important roles on the supply-side by affecting provider supply and location, manipulating the incentives faced by providers across settings, and, for some populations, serves as the primary provider of care. We discuss the effects of the government on use and expenditures via supply-side mechanisms in greater detail later in this report.

Significant associations between population insurance coverage and expenditures were reported (e.g., Rettenmaier and Saving 2010); however, the amount of geographic variation specifically explained by the level of population insurance coverage was not reported. Cutler and Sheiner (1999), who analyzed variation in expenditures at the HRR level (with some covariate information drawn from MSA- and state-level data sets), reported an increase in the R^2 from 0.71 with demographic and health status variables to 0.75 with the addition of the proportion of the non-elderly population in an HMO and supply-side variables such as number of beds; however, it is not possible to determine what share of that increased explained variance was due to the insurance variable. Even though it is not possible to discern the specific contribution of insurance status of the population in this case, the additional variance explained by both variables is small and thus not likely to be a major factor in explaining regional variation.

In an interesting comparison of HMO and unmanaged coverage using the Community Tracking Survey (CTS), Baker and colleagues (2010) found significantly more variation in hospitalizations across metropolitan areas for those with unmanaged care compared with those in an HMO. A similar study conducted by Herring and Adams (2010) focused on the effects of HMOs among Medicaid beneficiaries, finding little or no effect of HMO enrollment on total expenditures at the MSA level. Both studies’ results were limited by imprecision among select CTS data elements, and Herring and Adams’ results may have been affected by their limited sample.

Income and Employment. Income is associated with a variety of health-related variables including health status, health insurance coverage, and health care use with higher levels of income associated with more positive outcomes (Adler et al. 1993; Feinstein 1993; Adler and Ostrove 1999; O’Connor et al. 2003; Krause et al. 2011). Increases in income, all other things

equal, make it possible to buy more medical care or to afford out-of-pocket premiums for insurance if one is not eligible for coverage through a public program. Thus, income should improve access to health insurance and thus access to medical care. Variation in spending is explained further when both income and insurance status are included in models. This suggests that the process by which income affects health is significantly more complex than can be explained in the context of insurance coverage (Hadley 2003).

A variety of measures of income were used in the literature reviewed with most being measured at the population level. Studies with individuals as the unit of analysis also used population level measures, applying the income data measured at the MSA, county, or zip code level to the individual. Income data were extracted from state and federal statistics, such as the Census. Measures of income were often dichotomous indicators of either low or high income, with “low” and “high” inconsistently defined or measured as the median household income within a defined area. Other income-related variables included SSI income (Skinner et al. 2001), the proportion of the population in poverty (Skinner et al. 2001), and receipt of or eligibility for low-income subsidies (Zhang et al. 2010; Newhouse et al. 1993).

The level of employment (or conversely, unemployment) was a second budget constraint considered in explaining area variation. As with income, these variables were measured at the population level. Employment status was extracted from state or federal statistics on the proportion of unemployed or employed adults for each beneficiary’s zip code, MSA, or county. Population-level data appeared to be used because individual-level data were not available. Rettenmaier and Saving (2010) appear to be an exception, explicitly controlling for state level characteristics.

Despite a significant association between income and average or expenditure quartiles being reported (Fisher et al. 2003a; Sutherland et al. 2009; Hadley and Reschovsky 2006), no estimates of the amount of variation explained specifically by income measures was provided.

Income is difficult to measure and is often not measured well. Much debate has occurred over whether wealth, rather than income, should be used a measure of SES. In a systematic review, Pollack and colleagues (2007) conclude that wealth is significantly associated with health. Yet they and others (Allin et al. 2009) determined that wealth is not significantly associated with health care use. However, these studies do conclude that wealth may still be a more appropriate measure of SES for particular subgroups such as the elderly and minorities.

The endogeneity of individual income, wealth, and employment with health raises an important limitation. A good deal of debate has occurred over whether income predicts health, or health predicts income, or both. However, Ettner (1996) conducted an instrumental variables analysis and provided evidence that income has a causal effect on health. The relationship may not be an either/or choice but rather reciprocal.

Financial Strain. Most insured individuals are faced with some level of cost-sharing. If the burden of the cost-sharing is high, the financial strain that individuals feel may induce them to delay seeking care or limit the intensity of the care they do receive. There is a substantial body of research on the effect of insurance and cost sharing on health care use and expenditures.

These include the RAND Health Insurance Experiment, a randomized trial of health insurance conducted between 1974 and 1992 (Newhouse et. al 1993). For a review of the literature, see Zweifel and Manning (2000).

A few of the studies reviewed included a self-reported measure of financial strain in the decision to obtain medical care. Individuals were coded as having financial strain if they reported that financial issues were important in deciding whether to obtain medical care. Anthony and colleagues (2009) report no association between an individual's reporting financial strain and that individual's outpatient visits; nor did she and her colleagues find any variation across HRR end-of-life spending quintiles in the proportion of individuals reporting financial strain. In a similar analysis of end-of-life spending among Medicare beneficiaries, Barnato and colleagues (2007b) found no significant differences across spending quintiles in the proportion of individuals reporting financial strain. Given the lack of variation across HRRs, the amount of variation in HRR spending explained by financial strain is most likely insignificant.

7. NON-BUDGET CONSTRAINT

Use of medical care and the level of expenditures can also be influenced by non-budget constraints such as distance to provider, wait times (both in the office and in delays to obtain a visit), and dependent health care use. Such variables have been used frequently as instruments. Examples of the use of differential distance to providers of varying characteristics as an instrument include studies by McClellan and colleagues (1994), Geppert and colleagues (2000), and Chandra and Staiger (2007).

Urban vs. Rural. The distance to a health care facility or health care provider can affect use and spending. Health care facilities are not evenly distributed throughout the United States nor are the distances to the nearest facility the same across regions. Klein and colleagues (2009) for example examined the distribution of burn center hospitals and found that regional differences existed in the distance one lives from the nearest burn center hospital. Geographic proximity to health care facilities has been also investigated as a factor in determining health care use and spending. LaVela and colleagues (2004) examined proximity to VA health care facilities and found that those further away utilized services less frequently.

Urban versus rural designation of the geographic area or the proportion of urban and rural residents within a geographic area can serve as proxies for distance to provider. A measure of the urban or rural status of the area was included in some of the analyses. Zip code of the beneficiary or patient was linked to Census data at the MSA level to assign the urban or rural status to the geographic area in which he or she lives. Kelley and colleagues (2011) reported a significant association between urban residence and expenditures. Actual distance to provider, however, has not been explored in explaining geographic variation in health care spending.

We note that, while we discuss the urban-rural dichotomy here principally in the context of its effects on expenditures through demand-side mechanisms of the patient's time, travel, and delay costs in obtaining care, there are supply-side aspects to consider as well. We discuss

important related issues, such as physician practice and hospital location hypotheses, among other supply-side issues later in this report.

8. PATIENT-PROVIDER INTERACTIONS

Much of the extant literature on geographic variation identifies and adjusts for demand side factors such as health status and demographic characteristics. Typically these factors are not the researchers' interest, though as we illustrate earlier in this report, there are important subtleties to the understanding of such factors that may be underappreciated.

Some supply side factors are often addressed as well. The classic example is the level of input prices. As input prices rise, the market and payment rules under Medicare, Medicaid, and some other plans, allow administered prices to increase. For private patients, whether uninsured or not, local market equilibrium prices for medical care will rise as long as input prices rise as well.

There are other mechanisms at work as well. For example, it is generally considered interesting and problematic that patients in two different localities but with the same complaints, backgrounds, characteristics, preferences, and access to medical care resources should receive markedly different health care services based on different physician recommendations. Observational studies and more comprehensive descriptive analyses attesting to such differences in care have driven considerable interest and research development in geographic variation in health care use and expenditures and the supply-side determinants in particular. Still, many concepts in this area (e.g., the relationship between provider supply and use patterns) are not yet well understood and continue to develop rapidly as the subject of new work. Furthermore, virtually all of these studies are unable to explain significant variation.

In this section we discuss the research dedicated to identifying and assessing sources of variation in use and expenditures. Studies reviewed for this report are organized by provider type. We first explore issues central to the understanding of variation in the use of physician services, both in volume and mix (intensity). Our second focus is on issues associated with variation in the use of inpatient hospital services, many of which are similar to those discussed in the context of physician services. A brief discussion is provided with respect to other service types including nursing home care and home health care. We then expand our discussion to include relevant issues at a broader system level including those related to the public health system and other local governmental functions. Finally we discuss variation attributable to policy and legal differences at the state level.

9. SUPPLY-SIDE FACTORS: INDIVIDUAL PHYSICIANS

At the heart of most health care use and expenditures is the interaction between the patient and the clinical provider (physician, nurse practitioner, or other individual provider making patient care decisions). As such these interactions are naturally the focus of much of the geographic variation literature. Estimates of the proportion of medical care costs attributable to the physician's role in this decision-making process vary, but it is widely agreed that physicians

are accountable for the majority. In one discussion of the ethical issues confronting physicians during these interactions, Arnold Relman wrote that, “Although physicians receive only about 20 percent of all the money spent on personal health care in this country, their decisions determine most of the remaining expense” (Relman 1985). Similarly, Eisenberg (2002) posited that as much as 90 percent of health care resource use could be attributed to these outcomes of patient-physician interactions and ultimately the physician’s recommendations.

These estimates highlight the importance of the physician in determining health care expenditures primarily during and following the initial patient-physician interaction. A more detailed discussion of efforts to examine the importance of this decision in influencing geographic variation in health care expenditures is provided below.

Care Decision Points. The two key decisions made during the visit by the patient to the physician are whether any additional care will be provided and, if necessary, which particular services will be provided downstream (i.e., the intensity of the care). Both have significant cost implications. These decisions also affect the number of subsequent patient-physician interactions, thereby either cutting off the cascade of use or iterating the decision-making process and extending the episode of care. Given the importance of these individual decisions to overall health care use and expenditures and the particular importance of the physician in making them, many researchers of geographic variation have sought to identify the key factors that affect physicians’ decisions. In this section we discuss how many of these factors have been understood and studied to date.

Extensive research has been conducted into the determinants of physicians’ patient care decisions: whether to see a patient, whether any subsequent care will be necessary, and what the nature of such care ought to be. Less research, it appears has been dedicated to studying the factors affecting the decision to seek and provide initial care that would initiate an episode of care. In a 2002 article, John Eisenberg offered a framework for understanding many of the factors that influence physicians and the patient care decisions they make. To Eisenberg, the factors most influential on the physician’s decisions could be divided into those related to the patient’s benefit and those related to the physician’s own self-interest. In the following sections we discuss the key factors within each category, if and how they vary geographically, and how they contribute to observed variation in health care use and expenditures both individually and through interactions with other like factors.

10. SUPPLY-SIDE FACTORS: INDIVIDUAL PHYSICIANS AND THE PATIENT’S BENEFIT

As discussed in the conceptual framework, information asymmetry between patient and physician can contribute to regional variation in expenditures. While research has been conducted regarding the role of patient education and health literacy in determining health care use and expenditures, we are not aware of empirical work assessing the role of the physician’s information level in affecting geographic variation.

Physician Opinions and Training. Two given physicians often do not recommend the same course of treatment. That physicians exhibit differential abilities to take the complete

information available and make appropriate use of it, that is, to diagnose and treat patients, is widely recognized as one source of variability in care decision-making. One relatively recent paper providing additional evidence of this variability was written by Stern and Trajtenberg (1998). To a degree, this is the result of varying curricula and quality of education and training in physicians' medical schools, internships, residencies, continuing medical education, and informal continuing education among peers. It is well known that physicians who train through different programs are less likely to have similar treatment patterns than those who trained in the same programs (Woodward et al. 1990; Tamblyn et al. 2003). Yet these differences may also be attributed to inherent physician abilities and interactions between the incoming medical student's characteristics and those of the training program.

In recent years researchers have attempted to examine the effects of the physician's innate abilities through the consideration of their licensure examination test scores and other standardized tests – other data for evaluating these physician characteristics being limited – and have found material differences in subsequent practice patterns (Tamblyn et al. 2002). While the data to support conclusions about physician education and inherent diagnostic and decision-making abilities are less than ideal, underlying regional differences in the quality of medical schools, pre-graduate education systems, and other local factors influencing physician skill sets outside of the training sphere strongly suggest a relationship between physicians' decision-making abilities and observed geographic variation in use and expenditures.

Evidence-Based Guidelines. Developers of evidence-based guidelines seek to minimize variation of this kind, but this ideal case breaks down for several reasons, many of which have been discussed by Cabana and colleagues (1999) among others. For example, in their systematic review, Cabana and colleagues identified 77 studies that had sought to measure the impact on guideline adherence of variation among physicians in their awareness of or familiarity with select clinical practice guidelines. Overall, the authors concluded that there was considerable variability in these studies' findings as to the apparent significance of these factors as barriers to guideline adherence. Median studies reported 54.5 percent and 56.5 percent of their respondents considered lack of awareness and lack of familiarity, respectively, to be significant barriers, and there were wide ranges among those studies. More specifically, many of the reviewed qualitative studies had asserted many physicians may fail to adhere to clinical practice guidelines because of the volume of information the physicians must absorb, the limited time in which to consume that information, and the variable accessibility of guideline resources. Some of the variation in adherence to clinical practice guidelines may then be attributable to known variation among physicians with respect to their age (Choudhry et al. 2005; McKinlay et al. 2007), sex (Ely et al. 1998), and other demographic characteristics, their proximity to academic medical centers, the range of conditions with which their patients present, and their levels of educational achievement, each of which can be associated with the physician's capacity to absorb and retain information from clinical practice guideline resources (Cabana et al. 1999). Because many of these factors vary geographically, it is reasonable to assume they may contribute to geographic variation in health care accordingly.

Adherence to clinical practice guidelines may also vary for reasons that cannot be attributed to individual physicians but rather to the limitations of the guidelines themselves. Guidelines released over time and by different biomedical research and medical specialty

organizations may be in conflict with each other. Not all guidelines are easily interpreted. Physicians know that the evidence used by guideline developers to formulate their recommendations is never entirely consistent or reliable. Furthermore, many clinical decisions must be made without clinical practice guidelines specifying the most appropriate choice due to gaps in the spectrum of clinical practice guidelines—less common clinical decisions and those about which published evidence is sparse may not have been prioritized by guideline development groups. As Sirovich and colleagues (2008) observed, of the over 3,000 treatments reviewed by *BMJ Clinical Evidence*, more than half have unknown effectiveness due to inconsistent evidence and gaps in clinical guidelines. This idea is also consistent with the internationally-focused findings of McPherson and colleagues (1982) who found that variation in the use of select surgical procedures was correlated more with the degrees of controversy and uncertainty about the appropriate use of the individual procedures themselves than with patterns of organization or financing of health services internationally. For all of these reasons, we may expect to observe additional variation in adherence to clinical practice guidelines. As Cabana and colleagues (1999) note, in different regions physicians may choose to follow one set of guidelines and another region may follow another set, which may drive some variation in medical care use geographically. There is evidence to support this (O'Connor et al. 1999).

11. SUPPLY-SIDE FACTORS: INDIVIDUAL PHYSICIANS AND SELF-INTEREST

The second category of factors Eisenberg considered in the context of physician decision-making pertained to the physician's own self-interest (2002). In this discussion, Eisenberg made reference to McGuire and Pauly's neo-classical construction of the utility-maximizing physician (1991) that alters his or her care practice so as to maximize the personal benefit they receive from income, leisure activities, and their preferred style of practice. The specific mechanisms by which physician self-interest may drive this latter variation are the subject of a great deal of interest and research.

Among the reasons for the heightened interest in this subject was the recent New Yorker article "The Cost Conundrum" by Atul Gawande (2009). Gawande did not reach firm conclusions in this article, but he did regularly highlight that McAllen, Texas providers seemed to be increasingly concerned about "leaving money on the table." He also quoted a local surgeon who was critical of local medical practice, saying of his peers, "We took a wrong turn when doctors stopped being doctors and became businessmen" (p.8).

Discretionary and "Supply-Sensitive" Services. Much of the econometric research in this area has naturally focused on the set of clinical decisions in which guideline-based recommendations are less certain and the choice to provide or not to provide a service is more discretionary, given the increased likelihood of variation for such services. Sirovich and colleagues (2005; 2008) have conducted multiple studies of variation in discretionary clinical practice using survey tools. In these studies, a geographically diverse sample of physicians is given these surveys in which they are asked to respond to clinical vignettes highlighting a variety of patient conditions and specify whether or not they would provide particular services or sets of services for the described patients. The authors' models indicate that increased local per-capita Medicare spending rates are strongly associated with increased probabilities of local physicians recommending that the discretionary services (and, in their 2008 article, cancer screening

services for populations for which the services' value is not well established) be provided in response to the vignettes. While the authors claim that the regional spending rates predict the increased inclination to recommend services be provided in the survey, the direction of causality is not made convincingly. Furthermore, the emphasis on clinical scenarios known to be more discretionary suggests that the authors' conclusions may overstate how strongly variation in the intensity of physician practice patterns is correlated with local Medicare spending. Still, even with relatively few and potentially biased vignettes chosen for evaluation in the surveys, the use of standardized narratives describing the vignettes (thereby eliminating variation in the surveyed physicians' case mix) significantly increase the credibility of the authors' argument that increased use and expenditures in a region—at least for the Medicare population—is correlated with increased likelihoods among physicians to recommend the provision of additional services in discretionary cases. Sirovich and colleagues do not devote material discussion to the possibility that the variation they observe may be an outgrowth of varying manifestations of physician self-interest geographically, nor do they address the idea that variation in information asymmetries between patients and physicians may help to explain some of the observed variation.

Fisher and colleagues (2003a) presented a similar investigative strategy using the simplifying assumption that, assuming equal case mix across regions or that the effects of case mix on variation are precisely estimated and adjusted for, the remaining differences in use (i.e., the number of different types of services provided to the population) would largely reflect differences in physician practice patterns. This assumption also reflects the prior work of Dartmouth Atlas researchers and their own work in an accompanying paper (2003b), which indicated that increased use is not associated with quality of care and patient outcomes. However, in these 2003 papers, Fisher and colleagues performed case-mix adjustment of their measure of regional expenditures—to which local physicians were “exposed” (thereby serving as the study's intervention)—by limiting the expenditures to those incurred on behalf of Medicare beneficiaries during the last six months of life. The authors defended this methodology by suggesting that end-of-life spending had been shown to be “unrelated to underlying illness levels” by region while still highly correlated with local per-capita Medicare expenditures. However, one may be concerned that the correlation statistics as cited reflecting the relationship between end-of-life spending and health status may be biased by the choice of unit of analysis, the definitions of the measures used to reflect illness levels, and the methods of unknown rigor applied in standardizing the output prices paid for the end-of-life care. It is also unlikely that this measure of end-of-life spending fully reflects all of the demand-side factors—such as severity and patient preferences—that one would control for under ideal circumstances. Moreover, in the study's methods discussion, the authors claimed that they had “confirmed” the effectiveness of their study's “natural randomization” of assignment to areas with different exposure to levels of end-of-life spending by observing that the groups' baseline health status measures were similar. However, this claim too is questionable given that the measures used for this comparison did not account for all relevant aspects of the groups' health (e.g., condition severity, levels of physical fitness, diet) or other demand-related factors (e.g., supplemental insurance coverage status, access to needed medical services, local medical transport capacity) that would also vary by region and be correlated with both regional end-of-life spending and these patients' individual use patterns. Because of concerns about the construct validity and face validity of the method of using an end-of-life expenditure measure to avoid concerns about underlying variation in health

status, it is uncertain to what extent the authors' assertions about the sufficiency of their case-mix adjustment methods should be considered reliable.

Fisher and colleagues constructed several models using local Medicare spending to predict use of inpatient visits, inpatient consultations, tests, minor procedures, specialist care, and hospital admissions. Using these models, the authors found that higher-spending regions received approximately 60 percent more care on average across these services and that these differences could be isolated to variation in physician practice patterns with respect to discretionary, supply-sensitive services. The credibility of these authors' findings may suffer from their limited efforts to adjust for differences in case mix and other potentially relevant demand-side factors, relying solely on their use of end-of-life spending to obviate controls for such factors. Still, the magnitude of the findings leaves little doubt that physicians' use of many, primarily discretionary services is higher in high-spending regions than in low-spending regions, since, as the authors write, "it is highly unlikely that the 60% differences...could be due to residual confounding by unmeasured illness levels," though their inferred conclusion may have outstripped the data and analysis they provided.

The authors ascribed their observed elevated intensity of use primarily to the local supply of physicians and other medical care resources. More specifically, they noted that most of the variation in use by region is associated with "discretionary" or "supply-sensitive" services and highlighted previous research that identified a "causal relationship" between provider supply and use. While the authors did not explicitly identify physician self-interest as one of the primary drivers of their observed findings, the implication of physician self-interest as an important contributing factor is clear in Fisher and colleague's arguments because of the importance of discretionary services to their findings. The variation in observed use of discretionary services is particularly striking when considering that Fisher and colleagues limited their sample to Medicare beneficiaries with unbroken coverage who were not enrolled in an HMO and whose patterns of illness were relatively consistent across quintiles of use. Other researchers have been more explicit about attributing such variation to physician self-interest (see, for example, Tussing 1983).

Initial Care-Seeking Behavior. The studies of Fisher and colleagues and Sirovich and colleagues, like many others on the subject of geographic variation in physician practice patterns, focus on the intensity of services provided after the decisions to seek and provide initial care have already been made. As we alluded to early in our discussion of individual physicians, this initial agreement to initiate health care use is often overlooked, in part because a thorough investigation of the subject may involve levels of complexity beyond the scope of most studies (e.g., decisions to seek care, adhere to prescribed treatment regimens, or return for follow-up care may be determined in part by a patient's past experience with the health care system). Instead the focus is on average levels of expenditures across a population or, often, among subpopulations with established health care needs and use. Among the studies that have considered the initial care-seeking process, the identified determinants of patient care-seeking behavior have included demographic, health status, insurance status, personal and family health and health care history, and cultural factors, each of which we discuss in early sections of this report. Similarly, the identified determinants of the physician's decision to see the patient (or initiate the care delivery process) have included the patient's insurance status and care history

and the physician's specialty, productivity, and values (e.g., whether the value of the additional income from the patient's visit exceeds the opportunity costs), many of which we discuss below in this section. Wilensky and Rossiter (1983) presented one such analysis including a variety of patient, market, and physician characteristics and concluded that higher physician-to-population ratios and, on an individual level, increased physician age predicted increases in physician-initiated care. Both demand and supply-side factors considered include elements that vary by region.

Few studies have explicitly examined geographic variation in the initial and subsequent care-initiation decisions. Among these is a study by Tsang and Resneck, who observed significant geographic variation in wait-times before initial appointments with dermatologists, even among (simulated) patients with urgent care needs (2006). Unfortunately it is not clear to what extent the physicians themselves were involved in the decision to see these patients on a timely basis either directly (through consulting with the visit scheduler) or indirectly (through establishing policies about the timeliness of visits for patients with different complaints), and, more broadly, it is doubtful that these experiences and observations are representative of the care-seeking process for other specialists and in primary care settings. A related question of intervals of time between visits following the initial contact has been addressed by Tobacman and colleagues (1992), who found variation within and across geographic regions with respect to the proposed intervals of time between visits for preventive care or the management of chronic ailments, and DeSalvo and colleagues (2003), who found variation in patient re-visit intervals associated with select physician characteristics that vary geographically as well as clinical factors, among other researchers. These studies do not make explicit connections to physician self-interest or utility maximization more broadly in their discussion of factors determining physician recommendations and accommodations in the initial care-seeking process.

Fraud and Abuse. In the extreme, manifestations of provider self-interest may include fraud and abuse. Researchers and policymakers alike have long suspected that some measure of observed variation in health care use and expenditures should be attributed to fraud and abuse of Medicare payment policies (Brown 1997), but empirical evidence of systematic variation in fraud and abuse geographically is generally lacking. Welch et al. (1996) proposed that there might be a connection between geographic variation in expenditures and local business practices, since such practices often differ between for-profit from non-for-profit home health agencies and the mix of agency ownership status varies regionally. While the incentive to increase profitability through practices that push the limits of Medicare payment policies may vary across for-profit and non-for-profit providers (and thus regionally)—we discuss in greater detail the potential connection between providers' for-profit or non-for-profit status and regional variation in health care expenditures below—the extent to which variation in these practices effect meaningful regional variation in expenditures also depends on the permissiveness of local claims processing staff (e.g., at Medicare's regional fiscal intermediaries or state Medicaid agencies). Because assiduous claims-review processes are highly resource-intensive, local budget constraints and priorities can determine the effectiveness of these processes in a state or region; one report offered empirical evidence that changes in budget allocations drove some variation in the robustness of claims review processes at Medicare fiscal intermediaries (GAO 1996). Others have suggested that the structure of Medicare's outlier payment policies make them particularly likely to be gamed by providers with patients for whom expenditures approached but did not

exceed program thresholds (Rawlings and Aaron 2005), and so regions with sicker patients or patients more likely to approach outlier payment thresholds would be likely to experience increased submission of fraudulent claims on their behalf.

Several of the key factors driving variation in physician practice patterns, including those at work at the system level factors in this context, are discussed later in this section. We address the influences and incentives underlying price and reimbursement structures in our discussion of geographic variation in prices later in this report.

12. SUPPLY-SIDE FACTORS: INDIVIDUAL PHYSICIANS AND RESOURCE CONSTRAINTS

As Relman noted, physicians have significant influence over the use of society's medical care resources (1985), including staff time and skills, facilities, and technologies as well as financial resources. This implies that collectively physician decisions have a substantial impact on societal welfare. Yet individual physicians do not regularly weigh societal constraints with respect to these resources in their treatment decisions, absent provider-level policies or capacity constraints that more directly affect the individual physician (e.g., triage of large patient volumes following a local disaster, stepping down patients from intensive care units to other hospital departments with greater capacity) (Strauss et al. 1986; Jain 2009). As such, physicians do not weigh their resource stewardship responsibilities in their decision-making as much as would be socially optimal or even optimal for other providers within the system.

Manipulating Physician Incentives. Many purchaser organizations—health plans and employer groups—have acknowledged this mismatch between individual physician decisions and the effects of their resource use, moving to implement policies and programs that constrain or encourage physicians to act as if they more consciously consider the purchasers' and the regional health care system's resource constraints in their decision-making processes. These efforts often appeal to the physician's self-interest by introducing new incentives with financial ties through quality and efficiency measurement and reporting programs, administrative oversight and intercession (e.g., use review, prior authorization requirements, nurse care management, physician profiling, tiered physician networks), and risk-sharing arrangements. Such practices are often referred to as managed care. The effectiveness of these incentives in bringing about material changes in practice patterns and resource use remains an open question, studies of these programs having more often focused on effects on quality and access (Petersen et al. 2006; Rosenthal 2007; Peckham and Wallace 2010). Among the few efforts to assess the effects of HMO enrollment on use and expenditures are two recent studies by Baker and colleagues (2010) and Herring and Adams (2010). These studies produce conflicting results as to the effectiveness of HMO enrollment in reducing use and expenditures, though it is not strictly appropriate to compare their results directly—Baker and colleagues assessed the effect of HMO enrollment on privately insured beneficiaries' inpatient hospital care use, and Herring and Adams assessed the effect of HMO enrollment on Medicaid beneficiaries' overall use and expenditures. Both were also potentially subject to bias due to imprecise CTS data elements.

Some national efforts such as the inclusion in the Physician Quality Reporting System (PQRS) administered through the Medicare program of measures of overuse and efficiency (e.g.,

Overuse of Imaging Studies in Stage 0-IA Melanoma, Avoidance of Overuse of Bone Scan for Staging Low-Risk Prostate Cancer Patients) do not vary nationally and so would contribute to geographic variation in use insofar as organizational capacity to support participation in and full implementation of the program varied. Also, because a relatively small proportion of this and similar efforts' measures pertain to overuse and efficiency (less than 10 percent of the 439 quality indicators in the RAND Quality Tools measurement set) (Sirovich et al. 2008), whether these measures in particular materially affect physician decision-making is less certain.

Payer-Driven Incentives. It is known that the extent to which measurement, profiling, and reporting efforts have been deployed by private insurers, state Medicaid programs, and other stakeholders as well as the magnitude of the incentives attached to those efforts do vary significantly geographically (Rosenthal et al. 2006). Key factors contributing to this geographic variation in program implementation include organizational resource constraints in terms of capital financing, funding for infrastructure, and staffing (Barr et al. 2006), the capacities of other stakeholder organizations to support these efforts' fruition (Birkmeyer and Birkmeyer 2006), the compatibility of local delivery system structures with established programs (e.g., pay-for-performance programs requiring the identification of a primary care gatekeeper may not be compatible with established medical practice in a market where primary care responsibilities are dispersed across a group practice) (Rosenthal et al. 2006), the scope of state-level measurement and reporting requirements (which, in turn, vary with health department budgets and priorities, state budget deficits, and other factors) (Kuhmerker and Hartman 2007), and the intensity of local market competition (Weiner et al. 2005).

Moreover, the inclusion in many studies of proportion of HMO coverage as an explanatory variable in many studies of geographic variation in health care expenditures (see, for example, MedPAC 2003) is justified in part because of the increased likelihood that HMOs and like organizations would deploy such programs than health plans with Preferred Provider Organization (PPO) or Point-of-Service (POS) structures; moreover there is little evidence that like programs, when implemented by PPO and POS plans, are equally effective (Hellinger 1996; Mays et al. 2004). However, given that the incentives associated with such programs have varied in magnitude and effectiveness (Rosenthal 2007; Baker et al. 2010), it is not known to what extent physician practice patterns and health care use and expenditures will vary accordingly. Given the regional differences in the market penetration of some of these organizational arrangements, there could be an implication for geographical variation in use and expenditures as a result. Importantly, however, such analyses ignore the potentially important questions as to why the geographic variation in these payer organizational forms has existed.

13. SUPPLY-SIDE FACTORS: DETERMINANTS OF PHYSICIAN SUPPLY AND MIX

In 1999, Wennberg and Cooper released *The Quality of Medical Care in the United States: A Report on The Medicare Program*, otherwise known as *The Dartmouth Atlas of Health Care*, which suggested the volume of providers in an area was associated with increasing use and expenditures on the services they provided there. Because of the volume of analyses they presented with similar findings and because of the simplicity of their approaches, their findings are not considered controversial as a descriptive characterization.

There have been concerns about the nature of the causal linkages. Numerous studies have since been dedicated to exploring the causal linkages behind the supply-expenditures association with varying levels of success, acknowledging the difficulties of contending with concerns of simultaneity bias in most any analytical approach to the key question of how to connect expenditures and provider supply would be threatened. Before discussing the literature linking expenditures and provider supply in some detail, in the following section we explore the determinants of physician supply in an area. In particular, we discuss the extent to which these factors align with or are distinct from those we have discussed previously in the context of the neoclassical model of physician utility maximization and whether a broader, market-level model with competitive aspects helps to explain some of the observed trends in provider location and supply that underlie part of the geographic variation. This is not an alternative to the physician-level analysis, but it reflects that individual physicians are embedded in markets.

Physician Practice Location. The choice of physician practice location following graduation from medical school is an archetypical decision that can be explained through the neoclassical economic construct of the utility maximizing physician (Chou and Lo Sasso 2009).

Newhouse and colleagues (1982b) elaborated on the role of the utility maximizing physician in a market competition model of physician location based on location theory. The historical context for their papers was a rapidly expanding pool of specialist physicians and concerns about whether this pool seeking too infrequently to establish their practices in underserved areas. Their work had implications for contemporaneous national distributions of physicians. In this study, the authors discussed the idea that, all other things equal, one should expect a physician to locate his or her practice in a market area where the average projected patient load in that area is at least as great as it would be in another market area—otherwise the physician would relocate where the demand for their services was greater. Thus as an overall population of specialists grows, the increase would be dispersed unevenly as determined by patient population volume and market boundaries: in particular, the specialists would not all locate in urban areas already served by a number of specialists. Some would locate there, but some would begin to locate in markets that were less well served before the specialist volume increase. Still, they would not disproportionately locate in communities identified as underserved unless their projected patient load would be larger there than in other “better served” areas. Given that physicians make and then reassess their choice of location at intervals, McGuire (2000) described a similar dynamic.

The authors tested these hypotheses by observing the natural experiment offered by the rapid growth in volume of various physician specialties between 1970 and 1979. Their data bore out several of their predictions, including that larger volume specialties would be more widely distributed across towns of all sizes, that specialties with more rapidly increasing volume in percentage terms (e.g., internal medicine, psychiatry, radiology) would have located in markets of smaller size more often than more slowly increasing specialties of similar physician volume, and that despite the rapid growth of many specialties, the number of towns with at least one such specialist would not grow as dramatically. A related study observed that, over the same study period, on average patient travel distance was most significantly reduced for several of the same medical and surgical specialties that Newhouse and colleagues had identified as most rapidly growing (Williams et al. 1983). Newhouse et al. (1982b) also found evidence that, although the

overall population of physicians grew substantially during the study period, the decline in general and family practitioner volume—who previously had been established disproportionately in smaller-population towns—offset the increases in the number of specialists entering smaller-population towns, thereby illustrating why many smaller-population towns were left without physicians of any specialty despite the overall growth in physician numbers nationally. This finding supported the idea that the market factors influencing physician location decisions are largely specialty-specific, though it is important to note the boundaries between specialties and specialty markets are not necessarily fixed (especially in the longer term), since physicians may provide new and different services in response to competitive pressures.

Assuming physicians are not solely motivated by the compulsion to aid in meeting the health care needs of a population to the detriment of their own personal well-being, it is reasonable to expect that a physician's location decision will be affected by other factors influencing a community's demand for health care services (e.g., income, insurance status)—and thereby the physician's income—as well as the other amenities in which the physician may be interested when comparing communities. A similar framework was discussed by Fuchs (1978) among others. Yet even when they relax their assumptions of no differences among towns in terms of patient population needs for care, nearby population density, other population health care demand factors, and amenities, Newhouse and colleagues found their hypotheses theoretically consistent, though the existence of specific cut-points for each specialty, determining the projected case volume a region must offer in order for the physician to locate there, would not be observed; such cut-points would necessarily vary by region and community. These findings are also consistent with the ideas of the expanded model of the utility-maximizing physician, as they affirm that the physician's choice of location depends not only on prospective income but also many other factors that might belong in the physician's implicit utility function.

A more recent study by Rosenthal and colleagues (2005) reviewed trends in the geographic distribution of physicians between 1979 and 1999. The authors observed a similarly dramatic increase in the volume of physicians and likewise assessed the distribution of physicians by town size while also evaluating a variety of measures of geographic access to physicians and using a variety of market definitions (e.g., county, county type by assigned rural-urban continuum code, metropolitan statistical areas (MSAs), areas adjacent and not adjacent to MSAs). Their findings evaluating the distribution of physicians by town size were very similar to those Newhouse and colleagues had derived: “Although most specialties experienced greater diffusion everywhere, smaller specialties had not yet diffused to the smallest towns.” Furthermore, Rosenthal and colleagues evaluated additional measures of access and through that analysis, they demonstrated the boundaries of the medical care markets used in prior research were highly permeable (i.e., significant numbers of patients traversed them to seek care). These market definitions had included more recent efforts to specify provider service areas based on billing data (Goodman et al. 2003; Guagliardo 2004) and other sophisticated modeling efforts (Wing and Reynolds 1988) intended to predict the location choices and distributions of physicians. Newhouse and colleagues (1982a) had previously hypothesized that the geographic distribution of physicians would be perceived less as a market failure once researchers effectively measured and accounted for the extent to which patients sought care outside their assigned market areas. Importantly, these results also led Rosenthal and colleagues (2005) to conclude that conventional measures of access and market definitions may often lead

policymakers to overstate the magnitude of observed urban-rural geographic disparities in access.

Newhouse and colleagues (1982b) also evaluated the extent to which differential patterns of physician location could be observed by region (the 23 states included in the study were each allocated to one of four regions) in 1970 and 1979. While they did observe differences by region that were statistically significant for individual specialties, the margins of difference were sufficiently small so as not to suggest the need for regionally variant policy solutions. These findings must be considered conservative in the context of most studies of geographic variation in use and expenditures conducted at finer levels of analysis, variation since the regions evaluated by Newhouse and colleagues were sufficiently broad as to wash out much of the variation that might otherwise have been observed. As an important side note, however, it is not clear that these finer levels of analysis, particularly in the use of hospital referral regions (HRRs), are appropriate for all of these analyses, since they may be too small to capture variation in patterns of referrals and resource use for some conditions (e.g., less common forms of cancer) and too large to observe all meaningful variation for others (e.g., low back pain).

Newhouse and colleagues asserted that, when considering the aggregate volume of physicians across specialties, it can appear as though physicians have disproportionately located in areas where needs are not unmet. If instead one took a closer look at physician location choices *by specialty*, their simpler model sufficed to explain those choices without accounting for induced use in any significant way, since physicians had been less likely to become generalists upon graduation from medical school during the period they studied, and the service areas newly without any physicians of any specialty were considerably more likely to have been served by a generalist who had relocated or, more often, left the workforce.

Other Determinants of Provider Supply. Many other studies have explored additional factors influencing a physician's choice of practice location. As we alluded to previously, medical school graduates disproportionately reside in the same state as their medical school (though they may disperse more widely than individuals graduating from professional schools in other disciplines) (Seifer et al. 1995), and so they are more likely to reside in regions (the Northeast in particular) where there are greater numbers of medical schools; nurses are even more likely to accept positions in facilities near their site of training (Kovner et al. 2011). Other medical school characteristics such as public ownership and levels of funding from the National Institutes of Health may also play a role (Rosenblatt et al. 1992, Chen et al. 2010). Furthermore, many physicians may prefer practicing medicine near medical schools, academic medical centers or large multi-specialty physician groups because of the access their patients might have to more technologically robust hospitals and other facilities, the prestige of affiliation with such organizations, and, increasingly, the attractiveness of becoming employed in an acute care setting rather than bearing the costs and risk of maintaining an independent practice with inferior economies of scale (though the importance of this factor on the physician's location choice may be determined by other local market factors). Nurse practitioners may also be interested in seeking employment at certain facility types including hospitals, outpatient facilities, and clinics (including the growing retail clinic sector), and so the robustness and trends in local labor markets may also affect this decision. Taking all of these factors into consideration, it comes as little surprise that physicians are more likely to locate their practices areas that offer a richer

array of appealing amenities, that are more ethnically and culturally like themselves, and that offer opportunities to work in a preferred setting type (Guzick and Jahiel 1976; Langwell 1980; Harrison, Stinson, and Thurston 2002).

14. SUPPLY-SIDE FACTORS: EFFECTS OF VARIATION IN PHYSICIAN SUPPLY ON USE AND EXPENDITURES

Having established the existence of variation in physician supply geographically, it remains to discuss the extent to which this variation drives geographic variation in use and expenditures.

It is commonly agreed that the associations between physician volume or mix and Medicare expenditures, each measured in various ways, exist and merit exploration. As discussed earlier, Wennberg and Cooper (1999) demonstrated through a variety of simple correlation studies that higher volumes of providers in a hospital referral region (HRR) were associated with higher expenditures on their services there. These findings are typically considered robust largely because of the variety of provider types, specialties, and services for which they conducted their analyses, only for their results to be similar in nature and magnitude. Similarly, in their discussion, Fisher and colleagues (2003a) attribute much of the variation in use and that they observed across HRRs to differences in physician practice patterns associated with changes in provider supply and mix. It should be noted, however, that Fisher and colleagues derived these associations somewhat unscientifically through direct comparisons of average supply and expenditure measure values for groups of HRRs organized by quintile of end-of-life expenditures (the validity of which as a measure we discuss above in this section). Also, since they make no adjustments or controls aside from those made to their expenditure outcome measures for age, sex, and race, it is not clear to what extent the similarities in the trends of their supply and expenditure measures may be attributable to unspecified factors.

It is important to note health services researchers' conclusions about the established associations between physician supply and expenditures or between physician specialty mix and expenditures are far from uniform. In a study focused on patients with ambulatory care sensitive conditions, Laditka and colleagues (2005) found that in urban areas there was a strong negative relationship between physician volume and expensive hospitalizations for those conditions (with the magnitude of the effects varying by patient age), though there was no evidence of such a relationships in rural areas. Welch and colleagues (1993) conducted a methodologically straightforward study with results indicating that there was no significant relationship between overall physician volume and several measures of expenditures and use among Medicare beneficiaries at the MSA level, though they did find an association between these outcomes and percentage of physicians who were in primary care (general practitioners, family practitioners, and general internists). Also, Carlisle and colleagues (1995) found mixed results when assessing the relationships between the volume of physicians capable of providing various services and the volume of those services provided in-hospital for residents of Los Angeles County at the zip code level, controlling for illness status (condition prevalence), hospital distance, median family income, and racial and ethnic mix measures: for many services the association was not

significant, while there was a significant direct association for pacemaker implantation and transurethral resection of prostate and a significant inverse association for hysterectomy.

Primary Care Provider Supply. In the case of primary care services, Fortney and colleagues (2005a) found some evidence that increased access to primary care (through reduced travel times) was associated with decreased use of specialty care and physical health inpatient admissions among a Veteran's Administration population. The quasi-experimental, matched sample design employed by Fortney and colleagues appears to be sufficiently robust as to imply a causal link between increased primary care access and decreased use of other services, and so a substitution effect does appear to be playing a role in reducing the population's use of specialty and inpatient care, given the market effects of the change in access for primary care services. However, the many unique qualities of the Veteran's Administration population and the level of coordination of services provided at the newly established Community-Based Outpatient Clinics, which serve as Fortney and colleagues' primary care intervention (though no further discussion of the exact types of services provided or specialties of the service providers is given), lead us to question the results' external validity. If these results may be extrapolated to other settings, they may help explain some geographic variation in use across service types. For example, it may be that areas with relatively low volumes of physicians providing primary care services may have relatively high specialty care costs or post-acute care service volume, such as has been observed in McAllen, Texas (Franzini et al. 2010), among other areas. McAllen had also been identified by Wennberg and Cooper (1999) as the hospital referral region with the fewest active physicians per 100,000 population—88.2 physicians.

On the other hand, the findings of Weinberger and colleagues (1996), among others, would indicate to the contrary that increasing access to primary care can increase demand for specialty care or inpatient hospitalizations, and so the service types are complements rather than substitutes. Still, many of the plausible mechanisms one might propose by which increased access to primary care services would reduce use of inpatient care services may conceivably require an extended period of time to take effect—the study reported on by Weinberger and colleagues only monitored patients for 180 days after the intervention. Moreover, given the chronically ill population observed, it is reasonable to suggest that increased access to primary care for this population would lead, in the short term, to the identification of additional cause for hospitalization rather than the prevention of those causes. We interpret these latter results to be reflective more of the choice of study period and intervention definition than the absence of a substitution effect between primary care and specialty care services.

An important recurrent point in this literature is that the consistent specification of primary care and which specialists may be included among those considered primary care specialists is critical to ensure results can be replicated and compared across settings. Typically studies have included materially different subsets of family medicine, general practice, internal medicine, pediatrics, geriatrics, and other non-traditional specialists among their definitions of primary care. Nurse practitioners and other non-physician clinicians—as well as other specialists—are also included variably among these definitions. Such diversity in study specifications inhibits meta-analysis and systematic review of these findings.

Reduction of Nonmonetary Costs. A 2008 Congressional Budget Office report on Geographic Variation in Health Care Spending included discussions of the monopolistic competition model previously laid out by McGuire (2000) and its predictions for market entry as one framework that might explain the connections between physician supply and health care use (CBO 2008). In this report the CBO observed that reduced nonmonetary costs (e.g., shorter travel times to nearer providers) associated with increased supply have been cited in studies of other industries as an effective mechanism for increasing the consumption of many goods. The CBO noted that this connection had been supported by the findings of Escarce (1992) in a study of the effects of surgeon supply on the volume of consultative visits with surgeons as well as those of Siciliani and Hurst (2004), who found a significant relationship between surgeon volume and wait times for elective surgery. This same mechanism could be used to explain increased intensity (and costs) of care choices, conditional on the decision to seek or provide care, in areas with increased physician supply.

The CBO also discussed the arguments of Fisher and colleagues (2003a) that, in areas with heightened competition, physicians may be more likely to increase their use of select services, which Fisher and colleagues called “supply-sensitive services.” This and several other articles draw connections between physician supply and use or expenditures (Fuchs 1978; Cromwell and Mitchell 1986). We discuss the concept of supplier-induced demand below.

15.SUPPLY-SIDE FACTORS: SUPPLIER-INDUCED DEMAND

As McGuire discussed (2000), given the market power ascribed to the physician in most patient interactions, it is not surprising that economists have suggested physicians could set the quantities of their services in a monopolistically competitive way. Related findings of other researchers include that supplier-induced demand may also be associated with select payment arrangements—fee-for-service in particular, and that a substantial proportion of services commonly provided are of no clinical benefit—a finding likewise often presented by Dartmouth Atlas Project researchers. A variety of theoretical frameworks have been proposed to explain and predict supplier-induced demand in health care.

Those researchers supportive of the concept of supplier-induced demand often ascribed it to an implicit desire among physicians to maintain a target income level, as Fuchs discussed. Sweeney (1982) specified an economic model formalizing these assumptions—arguing, if unconvincingly, that the physician’s target income could be estimated as a linear summation of exogenous factors—and tested them empirically. Based on his findings, and perhaps not surprisingly (given his unconvincing model assumptions), Sweeney concluded that his model yielded “few predictions” and that the model of target-income seeking behavior might only be feasible in areas with low population density and reduced market competition. Relatively few studies have been conducted more recently that depend on a target income approach, though one article by Rizzo and Blumenthal (1996) did find evidence of physician pricing behavior associated with individual physicians’ target income defined as relative to average income within the region—clearly their model’s assumptions differed markedly from those of Sweeney. This study was not extended to an analysis of use outcomes, though the authors’ model did retain the flexibility such that this would be possible. Still, work by McGuire and Pauly (1991), in the

process of formulating what would be known as their neoclassical physician utility maximization model, effectively demonstrated that the literal income hypothesis imposed unnecessary assumptions and restrictions on models of physician behavior.

The conclusion that supplier-induced demand is a meaningful consideration in economic models of physician decision-making has been drawn in numerous studies, of which we will mention a few. Hemenway and Fallon (1985) found that in regions with increased physician supply and therefore heightened market competition (measured across the internal medicine and general practice specialties that were the focus of their study), physicians were more likely to recommend higher-intensity medical treatment in response to standardized hypothetical clinical scenarios—i.e., in the authors' interpretation, to induce demand. The approach the authors used is reminiscent of the work of Sirovich and colleagues (2005; 2008) given their effective use of standardized clinical scenarios and physician surveys to address concerns about unmeasured confounding in models relying on actual use and practice information. Hemenway and Fallon (1985) made no comment on their certainty with respect to the directionality and significance of their causal inferences across markets. The articles prepared by Christensen (1992) and Nguyen and Derrick (1997) each estimated that select physicians' responses to Medicare fee cuts would include increases in use that offset between 40 and 50 percent of their revenue lost to the cuts. These articles do not discuss in extensive detail the relevance of their findings to the discussion of supplier-induced demand, though both note that their methods do not permit distinguishing increases in use due to supplier inducement from those due to increased demand owing to reduced out-of-pocket costs to patients. It should be noted that across these and other studies of supplier-induced demand, the estimated magnitude of the effects of supplier-induced demand on use and expenditures have varied markedly.

Hadley and Reschovsky (2006) examined the effects of changes in Medicare prices on service quantities under this construct. Their findings were connected to a broader conceptualization of supplier-induced demand that included changes in physician service mix and changes in patient volume or mix in response to financial services rather than simply adjusting the quantity of a service among existing patients. We noted that Newhouse and colleagues (1982b) had referred to this concept of physicians adjusting the mix of services they would provide in response to market dynamics in our discussion of their work earlier in this section. Hadley and Reschovsky found, for example, that physicians who owned their practices had higher Medicare price per relative value unit (RVU), a unit used to compare different mixes of physician services, than salaried physicians—this implied that physician practice owners more effectively managed the types of services they provided so that those garnering higher revenues were more commonly provided. Even employee physicians with variable incomes had higher Medicare price per RVU than employed fixed income physicians. Other factors one would expect to function as incentives for inducing demand for more or a more expensive mix of medical care services, such as survey responses indicating that potential income is very important or practicing in particularly competitive markets, were also associated with higher Medicare price per RVU measure values. Assessments of the association between these incentive-like factors and the number of Medicare beneficiaries treated yielded mixed results. Hadley and Reschovsky found no significant associations between these factors and measures of the average number of RVUs per patient. Each of these analyses was modeled alongside a variety of other market-level and patient characteristics other than Medicare price that were

significantly associated with the outcomes of interest, and the sensitivity to endogeneity bias in the determination of Medicare prices was of concern throughout. Most valuable among the authors' contributions despite these difficulties was the enumeration of several measures, the mix of services provided in particular, that should be considered in discussions of the merits of the supplier-induced demand concept and of the magnitude of the phenomenon's effects on use and expenditures. Thus, even in studies that did not find significant evidence of supplier-induced demand as measured through the quantity of a service provided, it is quite possible that other, subtler forms of supplier-induced demand have been overlooked.

Newhouse and colleagues (1982b) presented one such argument, noting that his location theory-based argument, with a few simplifying assumptions dropped, would account for the predictions of models allowing for supplier-induced demand. Chandra and Staiger (2007) presented another such argument in their specification and testing of a model of regional specialization with productivity spillovers. More specifically they proposed an alternative explanation for the observation made by many supporters of supplier-induced demand models that increasing competition for the provision of a given service in a market is associated with increased use of that specialty's service there. Chandra and Staiger's article provided evidence that, in fact, in a market with a higher percentage of its physicians in certain, more intensive specialties, the provision of more intensive services may be the result of knowledge sharing among those physicians. Moreover, they found that this increase in the provision of more intensive services might be to the benefit of their patients as well, even those for whom such services might not be recommended in another region where fewer specialists of this type have located. See Chandra and Staiger (2007) for further discussion of the potential implications of these findings.

16. SUPPLY-SIDE FACTORS: SPECIALTY MIX AND OTHER PHYSICIAN CHARACTERISTICS

Physician specialty mix and a handful of other physician characteristics have been discussed as possible determinants of geographic variation in health care use and expenditures. We discuss a select few of these characteristics below.

Specialty Mix. Numerous other studies have observed more specifically that regions with high volumes of medical specialists tend to have high expenditures, whereas high volumes of primary care providers are associated with lower expenditures (see, for example, Fisher and colleagues 2003a; Baicker and Chandra 2004a 2004b). The magnitude of the effects observed within these studies has varied significantly with the authors' choices of included physician specialties (e.g., "medical specialists" versus "primary care providers"), the level of analysis, and the types of expenditures measured. To some extent the magnitude of the results vary with estimates of the geographic variation in specialist mix and volume, though it is widely agreed that such variation exist and are significant (see Wennberg and Cooper 1999 for one illustration). A related study by Welch and colleagues (1993) found no significant relationship between overall physician volume and several measures of expenditures and use among Medicare beneficiaries in 1989 at the MSA level. Still, they did find an association between these outcomes and percentage of physicians who were in primary care.

While researchers have focused primarily on the effects of variation in the volume of medical specialists relative to other physicians, others have focused on the effects of variation in the volume of primary care providers relative to other physicians. In this context researchers often refer to ambulatory or primary care-sensitive complications, hospitalizations, readmissions, etc. Many researchers have asserted that increasing volumes of accessible primary care providers, particularly for more vulnerable patients or patients in poorer health, are associated with falling overall use and expenditures. Multiple studies have found evidence supporting this claim (Welch et al. 1993; Parchman et al. 1999; Basu et al. 2002). Given our earlier discussion of mixed evidence regarding whether primary care services are substitutes or complements for specialty and hospital care, it is no surprise that evidence regarding whether patients in areas with increased primary care provider volume have lower health care costs has been somewhat mixed (Carlisle et al. 1995; Schreiber and Zielinski 1997; Fortney et al. 2005b). A study conducted by Wennberg and colleagues (1997), for example, found no relationship between the volume of interventional cardiac procedures and the volume of cardiologists in the region, controlling for the number of cardiac catheterization facilities.

Recent evidence on the subject comes from two studies by Baicker and Chandra. In one of these studies (2004a), they analyzed the effects of changes in provider volume and mix on expenditures for Medicare beneficiaries at the state level, holding constant the total number of physicians. Baicker and Chandra found that increases in the volume of general practitioners were significantly associated with decreases in per-capita Medicare expenditures, increases in the volume of physician specialists were significantly associated with large increases in per-capita Medicare expenditures, and increases in the volume of nurses were not significantly associated with changes in per-capita Medicare expenditures. In the other study (2004b), they analyzed the effects of specialist physician workforce differences on per-capita Medicare expenditures at the HRR level. Holding other the volume of physicians of other types constant, Baicker and Chandra found that areas with greater numbers of family practitioners had lower average expenditures and areas with greater numbers of medical specialists (e.g., cardiologists) had considerably higher average expenditures. However, “the density of surgical specialists and internists has little effect on costs.” These results are consistent with the observations made by other researchers in the context of studying the determinants of physician supply, supplier-induced demand, and the role of substitution effects. Given this consistency, Baicker and Chandra’s findings may be given further credence as reflective of an important factor in determining geographic variation in expenditures and use.

Physician Ownership and Entrepreneurial Practices. Among the theories that have been proposed for these correlations is that regions with a greater number of specialists per capita may have higher expenditures if those specialists, who are more likely than their generalist peers to have an ownership stake in a specialty hospital, ambulatory surgical center, or other facility or technological capability, are then more likely to order or provide additional medical care services because of those services’ increased financial returns. This idea is consistent with the findings of Hollingsworth (2010), who observed that physicians with such ownership stakes provided different and more financially beneficial mixes of services and provided more services overall. Mitchell (2008) made similar observations considering trends in service volume and mix before and after physicians invested in hospital facilities. Other supportive evidence has

been obtained through multiple studies of the relationship between a physician's investment in an imaging facility (or other specialty facility) and increased self-referral for and use of such services (Hillman et al. 1990; Mitchell and Sass 1995; Casalino 2008; Hillman and Goldsmith 2010; Baker 2010). A review of this literature by Kouri and colleagues (2002) identified many of these same conclusions across other studies. Also, Stensland and Winter (2006) recently found that for higher-intensity procedures such as coronary artery bypass graft (CABG) surgeries, however, the effects of increased specialist ownership of cardiac hospital facilities on use rates are, for the most part, not significant. They also found no significant difference between high- and low-profit surgical procedures. The authors reasonably reconciled their apparently divergent findings by suggesting that a physician's use patterns for procedures such as surgeries that are "less discretionary and [carry] more risk for the patient" are less likely to be influenced by the increased financial incentives associated with facility investment and ownership.

Since information on the ownership of medical facilities is typically not available for large samples, there are few studies of geographic variation in physician ownership of specialty hospitals, ambulatory surgical centers (ASCs), and other like facilities. One relevant study was produced by the GAO (2003), who conducted a survey of executives at specialty hospitals (as well as general hospitals) in six states. Among the hospitals surveyed, the GAO observed significant variation in the levels of physician investment, which suggests that local standards and laws may have affected the extent to which physicians have an ownership stake in specialty hospitals. The GAO also observed that, as of 2003, specialty hospitals were a small but rapidly growing minority of hospitals, and so the importance of these hospitals to geographic variation and other far-reaching analyses of use and expenditures will have increased in subsequent years. Rapid growth of ASCs, which are commonly physician-owned, has also been observed (Hollingsworth 2010), likely with similar consequences for use and expenditures. Neither Hollingsworth nor the GAO dedicated material discussion to how their results varied geographically.

An important demonstration of geographic variation in this phenomenon was produced by Casalino and colleagues (2003). Across the twelve MSAs assessed as part of the Community Tracking Study, the authors found evidence of rapid growth of specialty hospitals and ASCs overall. More importantly for our purposes, there was also evidence of correlations between the growth rates of such facilities in these markets and other key market factors, such as the presence of a single dominant hospital with large market share (negative correlation) or a large single-specialty physician group (positive correlation). Other competitive factors and incentives could also influence patterns of physician investment. Variation in all of these factors nationally implies meaningful geographic variation in the presence of such facilities and physician ownership in them. Pham and colleagues (2004) affirmed these findings as part of their own interview-based study of market-level patterns of physician "entrepreneurialism." Under this heading, Pham and colleagues included investment in technologies with which to provide ancillary services and investment in or ownership of ambulatory surgical centers and specialty hospitals. Through their interviews of hospital executives, they likewise found evidence of geographically varying market forces leading to geographic variation in physician ownership of specialty hospitals and ambulatory surgical centers. Similar conclusions could be drawn from

this evidence about variation in physician investment in other entrepreneurial, business-oriented practices, such as ancillary services.

Groeneveld and colleagues (2005) investigated this question from a slightly different perspective, extending the discussion to include a focus on racial and ethnic disparities by including regional racial composition among their key variables of interest. They examined hospitals' investments in new procedural technologies and found different rates of adoption and use of these technologies in regions with different racial compositions, potentially as a result of different resource levels, different physician skill levels, or "a 'racial segregation effect,' whereby hospitals with large minority populations persistently lag behind predominantly white hospitals in innovation and quality of care." Because of various study limitations and considerations, some of which ought to be discussed, the authors were not able to isolate the pathway or pathways most likely to drive their results. The numerous inclusion criteria the authors applied in selecting procedures for their review might have included one or more that was significantly (though not by the intention of the authors) correlated with technology uptake in some facilities. One such criterion is the proportion of the procedure of interest performed on an outpatient basis. Also, since the pool of procedures analyzed was pared down to only five, the authors' results based on this analysis might not have been transferable to other procedures or facility types. Finally, it is not clear what influence local private health insurers may have had on procedure technology uptake and how the presence of insurers with certain characteristics (e.g., use of managed care principles and processes) may have been correlated with the geographically varying population characteristics Groeneveld and colleagues observed.

17. SUPPLY-SIDE FACTORS: INPATIENT HOSPITAL CARE

Considerable geographic variation research has been conducted with a focus on hospitals. Variation in inpatient use and expenditures regionally may stem from variation in admission and readmission rates per capita (Fisher et al. 1994; Wennberg and Cooper 1998), the mix of patients' diagnosis-related groups (DRGs) (Frick et al. 1985, Steinwald and Dummit 1989), average length of stay (Yuan et al. 2000), in-hospital efficiency and care use (e.g., of hospitalist and specialist visits, nurse staffing, tests, procedures, drugs), and other components of inpatient resource use (Mitchell 1985; Pilote et al. 1995).

We note that a large proportion of the observed variation in hospital costs may be attributable to variation at the individual physician or physician group level. Brownlee and colleagues (2011) recently illustrated the direct influence of physicians on geographically varying hospital admission rates by comparing admission rates for elective procedures for similar patients across the state of Minnesota. They noted that patient preferences regarding the choice of treatment for breast cancer may vary regionally, and, therefore, use and expenditures of mastectomy procedures may vary likewise. Importantly, however, previous work by Wennberg and Cooper (1998) had argued the recommendations of the physician are more influential in making final care decisions than patient preferences. (We note they did not consider effects of patients' selection of physicians in their analysis.)

A fuller understanding of the relationship between the physician's decision-making and the costs and use of inpatient care may be complicated by the physician's relationship to the hospital. It should be noted that many studies of care decisions made in a hospital setting do not distinguish between those directed by community physicians unaffiliated with the hospital (except as through the award of privileges) and those directed by affiliated hospitalists. Furthermore, it is not always clear whether affiliated hospitalist physicians are reimburse on a fee-for-service basis or are salaried. The internal validity of future studies of variation in the use of inpatient care will benefit when they account for such differences in addition to hospital-level characteristics and market factors.

18. SUPPLY-SIDE FACTORS: HOSPITAL SUPPLY AND OWNERSHIP

Hospitals face high costs of construction and relocation, which distinguishes them from individual physicians and physician groups. Ciliberto and Lindrooth (2007) recently studied determinants of market exit by hospitals during the late 1980s and 1990s, and concluded that the key factors were relative reimbursement levels and payer mix, operational efficiency, size, service mix (with the ability to offer technological advanced services associated with a lower likelihood of market exit), and for-profit versus non-for-profit ownership status, although the effects of these factors varied over time in terms of magnitude and statistical significance. Other factors affecting changes in hospital supply (including new constructions, additions, mergers, consolidations, relocations, and closures) may include the intensity of market competition and the restrictiveness of local regulations, including certificate of need (CON) regulations (we discuss CON regulations in greater detail later in this section) and other public incentive programs, such as supplemental payments for critical access hospitals and other facilities in designated shortage areas. Many of these were included in the theoretical framework developed by Joskow (1980) and in those of scholars building on his work.

Hospital Ownership Status. One of the critical factors to account for when modeling hospital supply is hospital ownership status—whether the hospital is for-profit (FP), non-for-profit (NFP), or government-owned. Sloan and colleagues (2001) were among the first to perform robust analyses to explore the hypotheses laid out previously by Arrow (1963) and Newhouse (1970), adjusting more effectively for case mix than preceding research using detailed clinical information from the National Long-Term Care Survey. In particular they compared FP and NFP hospitals through a cross-sectional analysis of expenditures during the six months following Medicare patients' health shocks (e.g., acute myocardial infarction events, strokes) as well as outcome-based quality metrics including measures of mortality, community living status, changes in the number of activities of daily living (ADLs), and cognitive status for the same Medicare patients. Their principal findings were that “for-profit hospitals were more expensive to Medicare, especially in terms of payments other than for index hospitalizations” than their NFP and government peer institutions, and that quality of care did not appear to differ materially between FP and NFP hospitals. With respect to expenditures, these findings suggested that regions with larger proportions of FP hospitals would observe higher inpatient and post-acute costs than regions with larger proportions of NFP hospitals, all else equal. This is the finding of Silverman and colleagues (1999) who conducted their analyses at the hospital service area level. They also observed that per-capita expenditures for beneficiaries in hospital service areas primarily served by FP hospitals grew more rapidly during the period 1989 and 1995 than did

per-capita expenditures for beneficiaries in hospital service areas primarily served by NFP hospitals.

In recent years empirical studies have compared FP and NFP hospitals' entrepreneurial practices intended to increase reimbursement. Horwitz (2005) concluded that FP hospitals were more apt to shift their service line offerings in accordance with shifts in reimbursement. In a separate analysis, Silverman and Skinner (2001) analyzed the coding practices and reimbursement of FP and NFP hospitals for patients with pneumonia or related respiratory disease. They concluded that, relative to trends in the "upcoding" practices of average hospitals, for-profit ownership status was associated with larger increases in upcoding. These results were found to be robust to the authors' case-mix and health status adjustments as well as a variety of sensitivity other supplemental analyses, though these analyses may not have been sufficient to rule out alternative hypotheses such as changes in patient mix. In general, the difficulties of measuring upcoding and other like entrepreneurial practices impede researchers' efforts to draw empirical conclusions about them and their differential use by ownership status and other factors of interest.

Ownership Status and Entrepreneurial Practices. Hospitals differ in the extent to which they bill Medicare for services not bundled under DRGs. Such services could include outpatient imaging services (e.g., for pre-operative evaluation), rehabilitative services, and other convalescent care. One study that addressed this issue in part was led by Silverman and colleagues (1999). As mentioned above, their main finding was that per-capita expenditures by Medicare for beneficiaries in hospitals service areas primarily served by FP hospitals were higher than in those areas primarily served by NFP hospitals. They were also able to examine expenditures by service type and concluded that the most significant differences were in hospital services and, notably, home health care, both of which were significantly higher in FP markets. Less noteworthy results were also derived for physician services (not statistically significant) and services at non-hospital facilities (small measure of difference). Galeener and Langabeer (2009) offered some countervailing evidence when they found that, controlling for hospital referral region, there was no difference in length of stay—a key driver of hospital operating costs (which can then be transferred to private payers in the form of higher prices) and indicator of patient care intensity—between hospitals by ownership status.

Similarly, it is not clear to what extent the use of emergency department and observation unit services for patients that might not need to be admitted—thereby potentially reducing costs—might be used to different extents by hospitals differing by ownership status or other important differentiating factors. Schuur and colleagues (2010) established that the use of observation care varied substantially by hospital and market but provided limited discussion of hospital-level factors contributing to this variation, their primary focus being patient-level factors.

Ownership Status of Other Facilities. The methods used by Silverman and colleagues (1999) for evaluating the effects of ownership status in mixed markets were relatively simple. The authors focused their analyses on hospital service areas served exclusively by FP or NFP hospitals and dedicated less discussion to regions served by at least one of each. Grabowski and Hirth (2003) expanded on this approach for nursing home care (which we discuss in greater

detail below) by analyzing the effects of FP and NFP ownership mix on market outcomes. The authors' outcomes of interest are quality of care and patient health outcomes. Their key idea is their emphasis on the effects of market concentration by ownership status, assessing competitive spillovers from FP providers to NFP providers and vice versa. Grabowski and Hirth's observations are consistent with the hypothesis that the presence of NFP providers in a nursing home market consisting predominantly of FP providers will bring about material increases in quality among FP providers as a result of competition. Their observations were also consistent with the hypothesis that those markets with increasing proportions of FP providers should see decreasing health care expenditures, even for care provided by NFP providers. These ideas are also consistent with Kessler and McClellan (2001), who found that the presence of a FP hospital in a market was associated with a 2.4 percent decrease in NFP hospital expenditures versus in markets consisting entirely of NFP hospitals, largely because FP hospitals would react more forcefully than NFP hospitals to adverse market conditions as by limiting wage increases. These findings present perhaps the strongest evidence that hospitals' ownership status and mix in a market may be important factors to consider in assessments of regional variation in health care expenditures, given that the mix of FP and NFP providers in a market can vary regionally (Silverman et al. 1999).

19. SUPPLY-SIDE FACTORS: EFFECTS OF VARIATION IN HOSPITAL BED SUPPLY ON USE AND EXPENDITURES

As with individual physicians, the supply of hospitals and hospital beds has been commonly cited as a driver of variation in use and expenditures, particularly for inpatient services (Roemer 1961; Cutler and Sheiner 1999; Wennberg and Cooper 1999; MedPAC 2003; Barnato et al. 2007b; CBO 2008). This idea is consistent with the classical economic models in which increases in the supply of a good—in this case inpatient care—are associated with increases in the volume of the good traded and a reduction in the good's price, all else equal.

A common finding across numerous studies is a positive association between hospital bed supply and the probability of admission, holding demand-related factors constant (see, for example, Fisher and colleagues 2000). We are not aware of any studies assessing geographic variation in average rates of hospital bed occupancy as distinct from raw use and expenditure outcomes—the focus of related studies has typically been on the related use, expenditures, and patient health outcomes instead. Assuming the demand-side factors that might be correlated with variation in hospital occupancy rates (e.g., public perceptions of care quality) are equal across regions, several studies argue that geographic variation in use and expenditures persist because of variation in the supply of hospital beds.

“Roemer's Law” is commonly cited by studies observing the associations between hospital bed supply and use (see, for example, Clark 1990). Simply stated, Roemer's Law refers to a principle expressed through the work of Milton Roemer that in hospitals “a bed built is a bed filled.” Several studies have generated findings consistent with this principle (Knickman and Foltz 1985; Wennberg et al. 1989; Ashton et al. 1999; Fisher et al. 2000). Yet rarely have studies explored specific mechanisms to explain this apparent relationship. As Taroni (2001) pointed out, “Roemer's call for sound research to find out the true need for hospital beds has just

not delivered. No established methods for defining a population's need for hospital beds exist.” This can be attributed to some extent to the intimate relationship between physician decision-making and trends in inpatient care use. Much of the findings of Wennberg and Cooper (1999), Fisher and colleagues (2003a), and others regarding the linkages between physician supply and use patterns may be transferable to this discussion of hospital care markets. This has been suggested by, among others, Deyo and colleagues (2006) in a study of variation in spinal surgery use.

Such studies are complicated by adjusting for a study population's health status, which is generally agreed to be the most important determinant of care use patterns (please see the early sections of this report focusing on demand-side factors). While this issue of endogenous determination with health status confounds attempts to identify the determinants of health care use in all care settings, there may be opportunities for isolating a few key determinants (e.g., health status) through examination of the use of services for which there are few, if any, viable alternatives. However, in these cases, such as patients with a hip fracture who are virtually universally recommended for hip repair surgery, the geographic variation in the use of the service is greatly diminished (Weinstein and Birkmeyer 2000). This is consistent with the findings of Fisher and colleagues (2004) who found that use of major procedures for patients with hip fractures, colorectal cancer, or acute myocardial infarctions was not significantly different between academic medical centers in areas otherwise very differentiated in terms of the levels of use of physician services, including with respect to minor procedures. These findings are particularly convincing given the lack of any adjustment for patient risk in the presented results. Chassin and colleagues (1987) observed that while significant proportions of care that could be considered inappropriate, even the frequency of inappropriate care provision did not vary between high-use hospitals and low-use hospitals. However, given the large number of studies conducted on variation in the appropriateness of procedures performed at hospitals since the work of Chassin and colleagues was published (see, for example, Tobacman et al. 1996), it is clear their findings are disputable. In any case, because of these meaningful impediments and for other reasons, it remains to specify and weigh the pathways that can explain the relationship between hospital bed supply and use and expenditures as well as the overall impact of this relationship on observable geographic variation in the latter.

Hospital Service Mix, Specialization, and Competition. Geographic variation in use and expenditures may also be influenced by hospital service mix. All else equal—including the purposes for which offered services maybe used—regions offering a costlier mix of services would be expected to have higher average expenditures than regions offering a less expensive mix of services. Like expenditures and use directly, service mix in a region can be jointly determined with patient case mix. It may also be driven by the reimbursement schedules and other policies put in place by local payer organizations. For example, Eldenburg and Kallapur (1997) showed that, in response to changes in Medicare payment policies, hospitals changed their service mix (and, thereby, their patient mix) to maximize cash flows. Particularly for those hospitals with a greater percentage of revenue coming from local employers and in-state health plans (e.g., Blue Cross), it can be interpreted that differences in these local payer organization's payment schedules and policies could then drive geographic variation in service mix and therefore use patterns. Furthermore, just as with physicians as we discussed earlier, a growing

proportion of these service-mix driving policies may be in the form of quality- or efficiency-based incentive programs such as pay-for-performance in the coming years.

In recent years, the sharp growth of specialty hospitals, which offer markedly different mixes of services than general hospitals, has offered new opportunities to examine the effects of differences in service mix on a market's expenditures and use patterns. Recent work by Barro and colleagues (2006) shed light on this issue in their study of markets following the entry of a cardiac specialty hospital. As they note, specialty hospitals could drive increased efficiency and welfare to local patients through the exploitation of increased technical economies of scale through higher patient volumes of specific types, improved quality as a spillover benefit of surgeon specialization and experience, increased incentives on the part of hospital ownership to drive efficiency in care processes (cardiac hospitals are often for-profit and may be owned at least in part by physicians (GAO 2003)), and, indirectly, increased efficiencies at other local hospitals through as a result of competitive forces. Barro and colleagues found evidence that cost growth was significantly slower in markets with a cardiac specialty hospital entrant, and there were no significant impacts on patient outcomes, both findings assuming the continuation of pre-entry trends in costs and quality. They also highlighted evidence that specialty hospitals have disproportionately located in areas with healthier, more profitable patients and have provided "additional intensive treatments of questionable cost-effectiveness."

A report produced by The Lewin Group on behalf of MedCath Corporation, which owns and operates several cardiac specialty hospitals, also provided evidence of increased efficiencies at MedCath specialty hospitals attributed to incentives of for-profit, physician ownership as well as some limited evidence of improved mortality outcomes (Dobson 2003). On the other hand, Carey and colleagues (2009) found, like other studies had previously (see, for example, Mitchell 2007), that within regions where cardiac specialty hospitals had entered, the use of cardiac services was significantly higher than in regions without such entrants. Carey and colleagues added to this the observation that more *hospitals* in addition to the specialty hospitals were providing cardiac services in these regions than were providing these services in peer regions. It is not clear whether these increases in cardiac services were offset by lower use and expenditures for other services in these regions.

In their models of patient-driven and payer-driven competition among hospitals, Dranove and colleagues (1993) noted that prices (and therefore expenditures) may be driven higher by hospitals—such as specialty hospitals—that differentiate themselves to patients or payers on the basis of their service specialization, demonstrated quality of outcomes, or other visible factors, such as access to high-tech medical care innovations. Athey and Stern (2000), in a related example, noted that in some counties, hospitals have increased their emergency room bed and technology capacity in response to the structures of emergency response systems that would send patients to those hospitals best equipped to handle patients' needs. (The push to adopt newer high-tech medical care innovations in response to market competition has been referred to, at times, as a "medical arms race.") Each of these may increase the prices payers may be willing to pay in order to keep these hospitals in their networks and thereby maintain or increase their plans' attractiveness to current and potential members. Similar models were supported by the research of Teplensky and colleagues (1995). In addition to their findings discussed above, Carey and colleagues (2009) identified that regions with specialty hospitals demonstrated

significantly higher growth in the use of high-technology services relative to regions without such hospitals. The inconsistent adoption of such high-technology services may increase variation in regional health care expenditures as a result of the increased use of higher-priced services and potentially new incentives on the part of the hospitals (and their physician-owners, if relevant) to realize returns on their technological investments in those regions where the technology is adopted. Evidence that acquiring such technologies increases use and expenditures is relatively strong (see, for example, Bryce and Cline 1998 for a study of the effects on use of hospitals' acquisition of MRI technology, or Baker 2010 for a study of the effects of MRI technology acquisition by physician practices). Moreover, the inconsistent adoption of health information technology, in the forms of electronic medical records, computerized physician order entry systems, or decision-support tools, for examples, may also increase variation by reducing expenditures and use, particularly inappropriate or redundant use and expenditures, for those hospitals adopting them. Evidence on the effects of these technologies on efficiencies, use, and expenditures, while of great interest given recent political efforts to expand their use, remains limited (Chaudhry et al. 2006; Lubick, Goldzweig, et al. 2009).

As we have seen, the effects of hospital competition on regional expenditures and use measures often manifest through non-price competition with respect to service mix, technology adoption, administrative and care efficiency improvements, and in other areas. Price competition also occurs, if indirectly, through negotiations with commercial payer organizations. In particular, less competitive hospital markets have long been understood to have higher prices per service (see, for example, Woolley 1989) and vice versa. This is a key issue in the assessment of the legality of any merger of nearby hospital units and may become an even greater concern moving forward as hospitals increasingly integrate with local non-hospital provider organizations as in the formation of an accountable care organization (Fisher and Shortell 2010). Hospital chain membership can also result in improved administrative efficiencies, improved coordination of care, and increased economies of scale and scope (Menke 1997) for reasons similar to those we have discussed above in the context of the differences between for-profit and non-for-profit hospitals. As the distribution of investor-owned and chain hospitals varies geographically (CBO 2006b), it can be expected that regional variation in inpatient costs can in part be attributed to chains and chain membership.

20. SUPPLY-SIDE FACTORS: POST-ACUTE CARE USE AND EXPENDITURES

Inpatient Hospital and Nursing Home Care in Variation Studies. While inpatient hospital care accounts for a larger percentage of health care expenditures nationally, there has been increasing attention paid to variation in post-acute care, such as long-term care, step-down inpatient care (e.g., rehabilitation hospitals), and home health care (Franzini et al. 2010; Brennan 2011). Still, much of the research that has been conducted in this area is descriptive and methodologically unsophisticated. Moreover, much of observed variation in the use of these services may be explained largely by variation in hospital practices, hospital characteristics similar to or the same as those we have discussed in this section, and many of the issues we raise in the context of physician decision making and agency. Dalton and Howard (2002) noted, for example, that up to 12 percent of nursing homes are operated as sub-units within hospitals, further supporting the idea that hospital practices may strongly influence the use of nursing and other post-acute services. Demand-side factors, naturally, are also important (see, for example,

Wallace et al. 1997 in the context of nursing home use). Below, we discuss a few prominent findings regarding the effects of supply-side factors on post-acute service use.

Among the important research conducted on variation in use and expenditures for nursing facility care was the work of Grabowski and Hirth (2003), which we discussed previously in the context of hospital ownership status. As we indicated, their work offered strong evidence that, at least in the nursing facility care market, the increasing presence of one or more for-profit nursing facilities in a market dominated by non-for-profit nursing facilities tended to increase the efficiency of average nursing facility care in the market. Moreover, the primary focus of their work was a small collection of outcomes-related quality measures, and their findings suggested that the presence of one or more non-for-profit nursing facilities in a market dominated by for-profit nursing facilities (which is more common in the U.S.) tended to increase the quality of average nursing facility care in the market. Because of regional variation in the ownership mix of nursing facilities, we may expect corresponding variation in the use and costs of nursing facility care.

Nursing Homes and Medicaid Policy Variation. Another active series of efforts exploring variation in the use of nursing facility care and associated expenditures has pertained to the determinants of hospitalization among nursing home residents. One such effort has been conducted recently by Cai and colleagues (2011). The external validity of their findings may be limited because they observed patient admission data only in the State of New York. Still, their central finding that Medicaid payer status was significantly associated with the probability of hospital admission among nursing home patients. This is significant for studies of geographic variation in expenditures given the high costs of hospitalizations for such elderly and disabled patients and the meaningful variation in Medicaid eligibility and benefit structures by state.

That such variation in Medicaid policies and reimbursement rates materially affect nursing home expenditures has also been the conclusion reached by several others studies, including that by Intrator and colleagues (2007). They observed in a cross-sectional study of 48 contiguous U.S. that Medicaid reimbursement rates were associated with increased staffing ratios and reduced odds of hospitalization. Furthermore, “bed-hold” policies (by which hospitalized nursing home patients’ beds in the nursing homes are reserved for them) were associated with increased odds of hospitalization. They also described wide variation in rates of hospitalization among nursing home residents by state. We note the results of Intrator and colleagues’ research were generated controlling for select supply-side and supply-demand interactive factors that are certain to vary geographically (in addition to demand-side factors such as case mix), including the medical wage index, nursing home bed count, chain membership, RN-to-total nurse ratios, market competition, and hospital bed supply. Several of these variables may interact significantly with the independent variables of interest or one another, and so including such interactions might dramatically affect results. Yet this study represents strong evidence that reimbursement levels, state policies, and other exogenous local factors can cultivate meaningful variation in use and expenditures among nursing home residents. We discuss other regionally variant policies that may affect use and expenditures below in this section.

Nursing Homes and Care Quality. In a similar study, Mor and colleagues (2010) conducted a largely descriptive study to illustrate the margins of difference in readmission rates

among states' nursing home residents. They found a correlation between readmission rates and physician visit rates during the last two years of life. The latter of these was interpreted to be largely discretionary, and so Mor and colleagues concluded that the differences in observed readmission rates might be largely supply-side driven. A recent report by Goodman and colleagues (2011) also highlighted regional variation in readmission rates for patients previously admitted for select diagnoses and procedures, focusing less on the nursing home care setting. Its authors asserted that two important factors correlated with elevated readmission rates were overall inpatient hospital expenditures (see our discussion above of geographic variation in inpatient hospital use and expenditures) and failures of post-discharge "care coordination." While care coordination has no standard analytic definition, Goodman and colleagues identified it with rates of primary care or specialist physician office visits within two weeks of the initial discharge. While they posited that these factors were significant in driving variation in readmission rates, their analytic methods, largely limited to descriptions of regional variation and correlation, did not support for causal inference. Further explorations of this phenomenon are needed to identify the key driving factors of this variation and to distinguish high readmission rates attributable to patient illness or morbidity from high readmission rates attributable to supply-side factors. We do acknowledge that, as a MedPAC report noted (2009), conducting large-scale research on such issues can be difficult at times because of limitations of some Medicare claims data extracts in distinguishing nursing home care claims from those for other resource use, particularly given that Medicaid is more commonly the primary payer for nursing facility costs than Medicare.

Nursing Homes and Competition. Other information about geographic variation in the use of nursing facility services emerged from a relatively recent surge in research on nursing home supply. This research followed on the implementation of a prospective payment system (PPS) for nursing home and home health services by the Health Care Financing Administration (now the Centers for Medicare and Medicaid Services or CMS) in 1998, as mandated by the 1997 Balanced Budget Act. Among a collection of related studies, Dalton and Howard (2002) conducted a review of the entry and exit patterns of skilled nursing facilities between 1985 and 2000, associating changes in nursing home counts during the post-PPS period at the county level with various county-level characteristics such as income, age, race, population size, nursing home bed-to-unit population ratio, and market dynamics immediately pre-PPS, though no detailed discussion is given as to why we might expect these variables to be predictive of nursing home entry and exit patterns. Their multivariate regressions suggested that pre-PPS nursing home count increase were the strongest predictor of post-PPS nursing home count declines, while none of the counties' sociodemographic characteristics were significantly predictive. They emphasized that the changes in service prices put in place for the Medicare program in 1998 appeared to be by far the strongest reason to expect a region's nursing home volume to decline, observing also that urban counties, where competition was fiercer, were more likely to observe nursing home declines.

Similar work by Castle (2005)—also at the county level and relying on principally the same data source—used a different model specification method and additional covariates and found more local sociodemographic characteristics, such as median income, to be significantly predictive of nursing home closures. More recent work by Castle and colleagues (2009) offered perhaps the most robust framework for analyzing factors both specific to nursing homes and in

general locally that have impacted nursing home closures, including among their analyses a literature review, bivariate comparisons, and logistic regression analyses that featured the informed selection of covariates based on previous research. Among the covariates found most predictive of nursing home closures were several factors that tend to vary geographically: market competition, in-hospital operations, and, importantly, the proportion of patients enrolled in Medicaid. Given these findings and other research correlating provider supply with use and expenditures with respect to other health care services, it would be natural to expect geographic variation in nursing facility service use due to these factors.

Variation in Home Health Care. Welch and colleagues (1996), among others have documented wide geographic variation in the use of home health care services. Welch and colleagues provided little discussion of this variation, however, broadly attributing it to a general lack of “consensus on whether and, if so, how often to use home health care.” FitzGerald and colleagues (2009) observed that numerous studies had confirmed there was substantial reduction in the use of home health services following the implementation of CMS’s prospective payment system for such services, including for those patients who had recently had a major operation. Their own contribution consisted of confirming these previous findings and also specifying the extent to which the industry’s responses to the Balanced Budget Act of 1997 varied regionally. In general, they found that while the absolute differences between regions in most measures of home health use did decline, the percent average declines in home health use did not vary significantly by region.

Relatively few articles have been published exploring the determinants of regional variation in home health services’ use, particularly recently. McCall and colleagues (2001) noted that some of this variation could be attributable to urban-rural variation in home health care use as well as the variable existence by state of Medicaid “buy-in” programs for select populations who might benefit from home health care, though neither of these assertions was explored in significant depth. Similarly, FitzGerald and colleagues (2009) cited one potentially significant factor is variation in the extent of fraud and abuse among providers ordering the provision of home health services, but likewise they did not explore the influence of this factor in detail. In 1992, Kenney and Dubay computed two-stage least squares regressions at the metropolitan statistical area (MSA) level to investigate the impact on the probability of home health service use and the number of home health visits received of numerous independent variables. These included hospital use variables, nursing home and home health market variables principally reflecting the supply of post-acute service providers and capacity, supply variables reflecting the availability of other services that may be considered complements or substitutes for home health services, and various demand-side and other geographic indicator variables. It is remarkable that many of their parameter estimates (including for the region’s average number of hospital discharges per enrollee, average hospital length of stay, nursing home bed supply, home health agency capacity, the level of the local Medicare ceiling on skilled nursing home health visits, and proprietary ownership share, in addition to most of the region-specific dummy variables) were found to be significant despite the large number of covariates. Still, because the study’s key independent variables were measured at the level of the MSA and higher and followed different theoretical frameworks, it is difficult to synthesize Kenney and Dubay’s findings and draw actionable conclusions from them.

Kitchener and colleagues (2003) conducted a similar study at the state level using longitudinal data to analyze determinants of participation in and expenditures through Medicaid home and community-based services (HCBS) programs (i.e., waiver programs, personal care programs, and home health services programs). Although they found demographic and other demand-side factors to be more significant in determining HCBS participation and use than Kenney and Dubay (1992) had found looking at home health services, Kitchener and colleague's findings were largely similar in nature to Kenney and Dubay's. Neither study dedicates meaningful discussion to the mechanisms by which the findings translate into regional variation in home health service use.

21. SUPPLY-SIDE FACTORS: PRESCRIPTION DRUG USE AND EXPENDITURES

One health care service type that has been, until recently, understudied in the context of geographic variation has been prescription drugs. The lack of such studies is attributable in part to the absence—until the advent of the Medicare Part D benefit—of large datasets reflecting patterns of prescription drug use across regions as well as historically significant gaps in prescription drug coverage among the elderly (Schur et al. 2004). Now that Medicare Part D data are more widely available, it is expected that further assessments of geographic variation in this population's drug use and expenditures will proliferate. The problem of data limitations persists for non-Medicare beneficiaries, and so studies of geographic variation in prescription drug use for the non-elderly, privately insured and uninsured populations have often been (and may continue to be) limited in sample and scope.

Moreover, for the same reasons researchers studying this variation have focused on the somewhat smaller population of insured individuals whose insurance benefits include pharmacy coverage. In recent years the proportion of insured, working-age individuals whose benefits include drug coverage has increased significantly – insurance benefit packages without prescription drug coverage are now rare (Kaiser Family Foundation 2011), though historically this has not been the case (Schur et al. 2004). Because prescription drug coverage, like health insurance broadly, is not universal and is correlated with other determinants of health care use and expenditures (Davis et al. 1999; Schur et al. 2004; Folland and Reschovsky 2009), we are concerned that existing studies may oversell the external validity of their findings (based on non-representative datasets) in the U.S. population as a whole. This concern holds for studies relying on Medicare Part D data as well, since Part D coverage varies geographically due to individual- and market-level variation in demand-related factors (Miller and Weissert 2007; Levy and Weir 2009; Heiss et al. 2011; MedPAC 2011). The structure of Part D benefits also varies with Medicare Advantage Part D plans, which themselves have penetrated different regional markets to varying degrees (Levy and Weir 2009; CMS 2012).

An early study of geographic variation in drug utilization and spending by Dubois and colleagues (2002) examined patterns of prescribing and use among the beneficiaries of three health plans in California. They assessed the extent of variation in the prescription of select drug classes overall and for subpopulations with specific diagnoses at the region (aggregates of small numbers of adjacent counties) level. Dubois and colleagues found relatively little variation within drug classes across regions both in terms of population-level use and diagnosis-specific

use in comparison with a select set of procedures used to treat patients with the same set of diagnoses. However, since these analyses were purely descriptive, limited to enrollees in a select set of health plans, and limited to one state, it was neither clear to what extent these results would be transferable to other populations nor to what the relatively limited variation in this sample could be attributed.

Zhang and colleagues (2010) generated similar findings based on a descriptive analysis at the HRR level of Medicare Part D per-capita drug spending and use across a national sample. In particular, they likewise observed that interregional variation in drug spending was significantly less than variation in medical spending. Additional findings included that there was only a weak correlation between medical spending and pharmaceutical spending across regions, though the correlation was higher in the absence of controls for health status. Zhang and colleagues noted reasonably that drugs are substitutes for medical care in some cases and complements to medical care in others and that many sicker patients make greater use of both. More generally, we may infer that the mechanisms driving non-drug health care use and expenditures are not always effective jointly with those driving drug utilization and expenditures or, perhaps, that the mechanisms driving them are inconsistent.

The most detailed examination of patterns of drug spending and use to date was prepared by MaCurdy and colleagues (2009), contracted by CMS, with the purpose of identifying any sources of variation in drug prices and use among Medicare Part D enrollees that should be corrected with geographic adjustments in reimbursement. Like Dubois and colleagues (2002) and Zhang and colleagues (2010), MaCurdy and colleagues assessed the extent of geographic variation in use and expenditures on prescription drugs across regions—prescription drug plan regions, in this case. They found relatively limited variation in drug prices by region but more substantial variation in drug utilization and expenditures. They also found that intraregional variation in drug utilization was greater in magnitude than interregional variation, which is consistent with similar patterns observed in medical spending by other researchers, including Epstein and Nicholson (2009). MaCurdy and colleagues' findings are somewhat in conflict with those presented recently by Donohue and colleagues (2012), who found that regional variation in drug spending should be attributed more to variation in prices than to variation in prescription volume. These discrepancies may be attributed to differences in the studies' levels of analysis or the latter study's focus on three particular drug categories.

Effects of Incentives on Drug Utilization and Expenditures. One would expect prescription drugs to be influenced by managed care incentive and reimbursement structures, which in broad terms, as we discuss elsewhere, vary geographically. The use of formularies, prior authorization requirements, and other constraints may significantly affect consumer's demand for drugs and physicians' prescribing patterns (Joyce et al. 2002), and so regional variation in the use of such mechanisms can translate into regional variation in correspondingly reduced drug prescription and spending (Tseng et al. 2007). Pharmacy benefit managers (PBMs)—which may employ strong incentive structures to curtail and direct pharmacy drug use and expenditures—contracted by ERISA employers and health plans may also drive geographic variation in drug spending to the extent their formularies and management mechanisms vary. However, to the extent the formularies used by health plans and PBMs are relatively similar, we may observe less variation in prescription drug expenditures than in other service types. This

hypothesis has been put forth by several authors, including those whose work we discuss below. Relatively few studies have examined specifically the effects of variation in such managed care incentives on geographic variation across all services, let alone with respect to drugs. Yet the studies of geographic variation in prescription drug use and expenditures have regularly observed less variation than in other types of service; this is among the key contributions of early work on geographic variation in drug expenditures and use.

MaCurdy and colleagues (2009) extended assessments of variation by conducting regression analyses of the relationship between these measures of drug utilization and expenditures and a collection of demand-side risk factors, including age, gender, a range of health conditions, age-disease interactions, institutional status, and disability-based eligibility status. Their results showed that, among non-institutionalized individuals, by adjusting for regional demographic composition and prices as much as one third of interregional dispersion in mean drug expenditures (and one fifth of median drug expenditures) could be explained. MaCurdy and colleagues attributed a significant proportion of the unexplained variation in drug utilization patterns to variation in managed care penetration rates and other omitted variables that likewise explain variation in Medicare Advantage participation rates (since MA members experience lower expenditures per-capita on prescription drugs). They did not substantiate these claims empirically.

A variety of studies have examined the effects of managed care structures and incentives on drug utilization. They have regularly found strong evidence that the impact of formularies, preferred drug lists, prior authorization, and other like managed care practices on drug utilization and expenditures – both per-capita and with respect to specific drugs and drug classes – can be significant (Huskamp et al. 2003; Hamel and Epstein 2004; Morden et al. 2008; Siracuse and Vuchetich 2008; Adams et al. 2009). These studies have implications for geographic variation insofar as the studied policies are systematically more commonly implemented in select regions. This information is not commonly available for non-elderly, working populations, though Tseng and colleagues (2007) documented such variation among Medicare Part D Plan formularies.

22. LOCAL- AND STATE-LEVEL DETERMINANTS OF GEOGRAPHIC VARIATION IN USE AND EXPENDITURES

As we have discussed previously when identifying Medicare and Medicaid as important demand-side factors affecting variation in use and expenditures, the government's role in affecting the health care system and variation within it is very important. Its role is highly complex as well, mediating demand-side and supply-side factors alike.

Public Health Systems. At the local- and state-level public officials and agencies may play a significant role in determining regionally variable patterns of use and expenditures. Two of the most important government-influenced factors driving use and expenditures are the public health system's funding and organization. As the Institute of Medicine noted in its 1988 report "The Future of Public Health," the extent of public investment in public health efforts varied considerably across the U.S. There was documented important variation in the effectiveness and skillfulness of public health administrators, the local political landscape and its inclination

toward supporting public health interventions, and other mediating factors. We are not aware of any large-scale studies that have assessed regional variation in the public health system's funding and effectiveness in the intervening years. Still, we expect that the variation described by the Institute of Medicine is not likely to have diminished over the last twenty years, though the extent of this variation may vary by intervention or program (e.g., smoking cessation programs may be more consistently supported nationally than exercise promotion programs).

Related work on a smaller scale by Mays and colleagues (2006), among others, has demonstrated similar variation in public health system performance associated with a variety of institutional, economic, and community-specific factors. In particular, Mays and colleagues conducted a cross-sectional multivariate regression analysis with random effects accounting for county-level variation unaccounted for in their model's independent variables. Among other findings, they observed that the size of the public health system (i.e., in terms of the number of individuals served in the community) was the strongest predictor of effective delivery of most public health services, perhaps due to economies of scale. Yet public health systems' effectiveness waned as population size approached and exceeded 500,000. The most consistent, if not the strongest, predictor of effective delivery across all ten services assessed was local health department funding. While the size of the population served is largely beyond the control of local authorities, the level of local health department funding is strictly within their control. As such, local politics and economics, which vary considerably from county to county, will significantly affect the effectiveness of the local public health system, which, in turn, will affect health care use and expenditures. Moreover, numerous studies have also been conducted regarding the key determinants of success in the implementation of specific interventions. We discuss a few of these studies later in this section.

Investment in Public Health Interventions. Public health interventions may generate regional variation in use and expenditures insofar as they are cost-effective, reducing the use of acute care and other costly health care services. At an aggregate level, it may be difficult to measure the effects of such interventions, as they can be difficult to attribute and compare across regions. Among the individual public health interventions for which cost-effectiveness researchers have typically found support are smoking cessation (Cromwell et al. 1997), at least for select populations (Ladapo et al. 2011), pediatric vaccinations (see, for example, Shepard et al. 2005; Trunz et al. 2006), and contraceptive use (Trussell et al. 2009), provided the programs promoting them are effectively organized and deployed. On the other hand, numerous public health interventions have been deployed regularly without robust evidence of cost-effectiveness, including the HPV vaccine, particularly for men and adult women (Kim et al. 2009; Chesson et al. 2011) and exercise promotion programs (see discussion in Valente 2006).

A growing body of evidence promises to identify specific choices local governments and organizations can make to increase the effectiveness and cost-effectiveness of these initiatives. Brownson and colleagues (2001), for example, identified that physical activity in a local population was likely to be improved with increased availability of sidewalks and other facilitative neighborhood features, though it may be noted that the causal inference of this study may be strained by its basis in cross-sectional data analysis. Another study assessing factors that affect the appropriate use and delivery of public health services was produced by Faulkner and Schauffler (1997) who found a consistent positive association between levels of insurance

coverage and preventive care-seeking behavior, such as would be expected for most medical care services. Furthermore, due to state-level variation in insurance company offerings, variable benefit and incentive structures, and variable deployment of managed care practices that may hinder the use of these services, it is reasonable to expect that the use of preventive services would vary geographically. In highlighting an exception to this conceit, Li and colleagues (2010) noted that several states' laws mandate similar benefit structures to support the provision of care intended to prevent incidence of diabetes. Depending on the effectiveness of the preventive care service in question, this variation would translate into variation in the use of acute care services, emergency department services, and other costly services. Similar studies about other public health interventions have generally supported this same conclusion—that while preventive care services have the potential to reduce costs, they are provided inconsistently. One study, for example, demonstrated that the quality of preventive care services provided varied greatly across a small sample of outpatient clinic sites (Dresselhaus et al. 2004).

Emergency Response and Infrastructure. One final local-level factor that could affect use and expenditures of health care services, though perhaps not as significantly as other factors discussed in this section, is variation across communities' emergency response capacity and performance. Yet relatively few studies have evaluated the extent of variation in emergency response services and associated expenditures geographically, perhaps owing to relatively few available datasets for use.

One report produced by Athey and Stern (2000) described such variation in multiple ways. Among the most important variation observed was regional variation in the level of service provided: while in 1995 most counties offered “Enhanced 911” services (between 60 and 74 percent of counties), other counties offered “Basic 911” services or no 911 services at all. In their review of a national (though not necessarily representative) survey of counties, Enhanced 911 services tended to be offered by more densely populated, urban, and wealthier counties. By contrast, those counties offering Basic 911 services or no 911 services tended to be more rural, sparsely populated, or lower income, and there did not appear to be many significant factors useful for identifying whether a county would be more likely to adopt basic 911 services or none. Athey and Stern also demonstrated that the structural variation observed had significant consequences for the productivity and effectiveness of the system at the county-level. Again, however, these results were not demonstrated to be of equal importance to the variation observed by studies in other areas of the health care system. Finally, Athey and Stern noted that this system-level variation would have consequences for local hospital market dynamics, as discussed earlier in this section.

At the level of the individual emergency response team, effectiveness in responding in a timely fashion, providing the most effective and cost-effective treatments to their patients, and ensuring patients survive until they reach the hospital may also vary geographically for several reasons including variation in the emergency response team personnel and organization. For example, these teams may consist of varying mixes of paramedics, emergency medical technicians, and first responders. Moreover there may be significant regional variation in the training and scope of practice restrictions for each of these personnel types, and these teams may be equipped with different medical technologies and equipment in the field (Cummins 1993). Likewise Braun and colleagues (1990) described how across a sample of 25 mid-size cities, there

was meaningful variation in the organizational structure of emergency response teams as well as the average population size served by different structural types and the timeliness of their responses. Such variation can reasonably be expected to have some effect on regional variation in use and expenditures, particularly with respect to acute care services.

Provider Staffing Mandates. One set of state-level policies and laws that can directly affect the provision of health care services are mandated minimum and maximum care thresholds for particular patient groups. For example, many states, such as California, have implemented minimum nurse-patient staffing ratios (often coupled with staff skill mix minimum thresholds as well) in acute, rehabilitation, and psychiatric hospital settings, which therefore we would expect to observe increased operational costs on the basis of staff expenses. One recent study by Serratt and colleagues (2011) drew the important conclusion that California hospitals out of compliance with the minimum staff requirement had, in fact, appreciably increased their medical/surgical registered nurse and registry nurse staff levels following the enactment of the state's mandate, while support staff and other non-nurse staff levels did not change meaningfully at those hospitals. This finding supports the presumptions that such laws are associated with hospital responses of altering the mix of services they provide and thereby increased hospital operational costs. As of 2005, 14 states had introduced legislation describing such mandates (Rothberg et al. 2005). Unfortunately, the findings of Serratt and colleagues (2011) may not be transferrable to other states where such laws may be applied to select patient types or subgroups, where the enforcement of these laws is nonequivalent, or where for other reasons local hospitals are less inclined to alter their staffing mix in the short term.

Lang and colleagues (2004) completed a systematic review of research on the effects of varying nurse-patient ratios on a variety of other outcomes related to patient health, staff accidents and burnout rates, and hospital-level operational and financial outcomes. They found inconsistent evidence with respect to many patient outcomes, though there was strong, if somewhat mixed, evidence suggesting inverse relationships between nurse-patient ratio and the probability of failure to rescue (i.e., death within 30 days for patients experiencing a complication) as well as between nurse-patient ratio and mortality. Subsequent, econometrically sophisticated research generated by Sochalski and colleagues (2008) supported the findings of Lang and colleagues (2004) regarding mortality but found no significant effects of increased staffing ratios on failure-to-rescue outcomes. A literature review on the subject by Kane and colleagues (2007) observed consistently significant effects of increased staffing ratios on patient outcomes, but the interpreted magnitude of these effects may be overstated because of the use of odds ratios to report results. Overall, these results appear to suggest that increased nurse-patient ratios would likely lead to increased use and expenditures of health care services on behalf of surviving patients (since resource use terminates at death), though this outcome was not explicitly measured.

Among those research efforts attempting to associate minimum staffing ratios with use and expenditures more directly, Lang and colleagues (2005) also noted that evidence suggests an inverse relationship between nurse-patient ratio and average admission length of stay. Although the authors were unable to identify studies with strong evidence of effects on financial outcomes that did not predate the advent of relevant managed care policies, the nine relevant studies they did find on the subject showed that "better staffing was either cost neutral or cost saving." These

conclusions were also reached in one cost-effectiveness study by Rothberg and colleagues (2005), in which the authors found that increasing nurse-patient ratios could result in relatively high cost savings in line with other commonly accepted, cost-effective interventions. Much of the benefit Rothberg and colleagues observed was attained through improved patient mortality outcomes. To our knowledge, no studies have measured the effects of elevated staffing ratios or minimum staffing ratios on per-capita expenditures or related service use.

Minimum Length of Stay Mandates. A related policy that can affect local use and expenditures is the minimum length of stay guaranteed to many newborns under the Newborns' and Mothers' Health Protection Act of 1996. Newborns of parents whose insurance is not through a self-insured employer but includes coverage for hospitalizations postpartum are entitled to 48 hours' hospital stay following the birth. Almond and Doyle, Jr. (2008) assessed a large sample of records of births in California (where the regulations are applied more broadly) to examine the effects of changes in length of stay along the lines of what would be expected as a result of the law's provisions. Using a regression discontinuity design, they observed that these increases in length of stay were not significantly associated with decreases in readmission or mortality for either the newborn or the mother. As a result, they concluded that the increased length of stay resulting from mandated minimum lengths of stay was likely to substantially increase acute care costs for these cases without compensating health benefits. While the results of this study are compelling when considering patient outcomes, the magnitude of the implications of the mandate for expenditures and service use may require further study.

Certificate of Need Regulations. States also have different certificate of need (CON) regulations, which require that state agencies approve both new hospitals' entry existing hospitals' "large" capital expenditures, and enforce them to different degrees. These laws, in place in 36 states as of 2002 (Rivers et al. 2010), have been intended to reduce costs by limiting unnecessary increases in bed supply, which historically has been correlated with increased use and expenditures, and unnecessary large-scale investments in medical technologies (e.g., magnetic resonance imaging technologies). The costs of these, it has been argued, would be inevitably passed along to consumers either through increased service use of marginal value or increased prices. Rivers and colleagues (2010) recently conducted an analysis of the effects of CON regulations and variations in the stringency of these regulations on quality of care, access, and system-level costs in metropolitan areas. The key findings were somewhat surprising: that CON laws had no significant or appreciable effect on inpatient costs per adjusted admission and that CON regulation stringency was positively associated with the same cost outcome. The authors attributed these results to competitive effects, since CON regulations, if enforced, restrict competition among providers of care. We note that the measure Rivers and colleagues employed in quantifying a state's CON regulation stringency is not wholly transparent and is therefore difficult to validate. However, similar findings had been reached by Lanning and colleagues (1991) among others. As Rivers and colleagues (2010) noted, the majority of prior studies addressing these questions may be less relevant today due to the prospective payment system for hospital care or the rise of managed care, and so with this limited evidence it is difficult to draw conclusions.

Malpractice Threats and Defensive Medicine. Another state-level policy issue that may affect health care use and expenditures is malpractice tort reform and the extent of defensive

medicine in response to local litigiousness. The magnitude of this concern as a proportion of medical care expenditures has been a hotly debated issue in recent years (though, by contrast it has been generally agreed, at least among health services researchers, that defensive medicine only explains a few percent of the growth in health care expenditures in recent decades (Newhouse 1992; Baicker and Chandra 2005)). Physicians tend to believe that it is a significant driver of health care expenditures—this can be explained by noting that large proportion of physicians have claimed to have ordered “more tests than they would if based only on professional judgment of what is medically needed”: 91 percent according to the CBO (2006a). The CBO also noted, on the other hand, that a large percentage of physicians had also reported avoiding certain procedures out of concerns of malpractice lawsuits. Still, estimates of defensive medicine’s contribution to health care costs can vary significantly.

In large part this variance can be attributed to the fact that defensive medicine is difficult to measure empirically. Estimates of the magnitude of its effects can be influenced by the measurement structures employed. Yet it is likely that the practice of defensive medicine and its associated effect on use and expenditures is highly correlated with the volume of malpractice lawsuits, claims, and awards, as well as average malpractice premiums, all of which can be measured more easily. As such, efforts to quantify existing variation in defensive medicine have often relied on these data. In one study, for example, Zuckerman (1984) documented regional variation both in malpractice claims and in physicians’ operational responses to their perceptions of increased litigiousness among patients. Baicker and Chandra (2005) likewise observed variation by state in malpractice premium trends over time. These two studies, among others, indirectly support the hypothesis of regional variation in defensive medicine.

Efforts to estimate the effects of defensive medicine on expenditures, quality, and other outcomes have relied on the use of similar variables. Baicker and Chandra (2005) applied regression analysis to a large, cross-sectional, state-level database to assess the relationship between malpractice payments or case volume and several outcomes of interest. Among these outcomes of interest was the concept of defensive medicine, as measured through total Medicare expenditures, expenditures on select health care service types, and use of individual procedures. In particular, Baicker and Chandra found that a ten percent increase in malpractice cases was associated with a 1.3 percent increase in total Medicare expenditures, including a 2.9 percent increase in health care imaging expenditures and little if any change on the use of major surgical procedures. At the individual procedure level, such increases in malpractice cases were significantly associated with increases in the rates of computed tomography imaging and cardiac catheterization, while it was not statistically significantly associated with rates of hip replacement and bypass surgery. Baicker and Chandra used similar techniques to assess the effect of malpractice premiums on the number of physicians per capita and found very little effect. A recent, similar study by Thomas and colleagues (2010) assessed the effects of physicians’ exposure to medical liability, as approximated using average state-level medical malpractice premiums, on health care episode-level costs using a large, regionally diverse commercial insurance claims database. They computed results for thirty-five distinct physician specialties and observed that average per-episode costs were positively associated with malpractice premiums for most of these across a variety of different patient conditions. Still, across these specialties, the average effect of reducing malpractice premiums by ten percent was estimated to be less than one percent of total medical costs, a relatively small effect. While the

observed associations between premiums and per-episode costs could potentially be explained by other factors, the authors employed a variety of econometric techniques (e.g., inclusion of state-level random effects and adjustments for clustering at the physician and region level for the reappearance of patients with the same episode multiple times) that increase the credibility of their findings.

Tort reform legislation is another popular tool of researchers interested in assessing the effects of malpractice liability and associated defensive medicine, as it permits the construction of pre-post quasi-experimental studies. Naturally, tort reform legislation is also a popular subject of health policy researchers in its own right. Kessler and McClellan (1996) constructed individual-level difference-in-difference models, including state and year fixed effects, of the effects of tort reform legislation on several relevant outcomes of interest. Among their findings was that between 1984 and 1990, the effect of direct tort reform legislation on average medical care expenditures was a decline of 5.3 percent. Moreover, they estimated similar cost declines when estimating the effects of direct reform on expenditures for patients with specific conditions: 9.0 percent for patients with ischemic heart disease and 4.0 percent within two years of adoption (5.8 percent three-to-five years out) for patients with acute myocardial infarction. Some of these results may be somewhat exaggerated in magnitude, given that expenditure growth was lower in states where reform was implemented than in states where it was not.

In a similar study, Hennesy (2004) used linear models estimating the effects of different tort reform structures (e.g., mandatory versus voluntary arbitration, contingency fee caps, caps on noneconomic or total judgments) on inpatient hospital expenditures. Hennesy's analyses revealed strongly significant associations between nearly all tort reform structures and expenditures, though the signs on the estimated coefficients were not all the same or as expected. Of note, however, is that Hennesy's findings, while they were often in agreement with those of Kessler and McClellan (1996), were not all in agreement with theirs: while they had found that damage caps reduced expenditures, Hennesy's (2004) estimate of the effect of damage caps had the opposite sign. The CBO (2006a) constructed very similar models of the effects of tort reform legislation at the state level on total, hospital, and physician expenditures and likewise found mixed effects of different tort reform structures. Given these mixed results it is clear that tort reform legislation, taken as a whole, has not had a consistent effect on state-level health care expenditures. However, the evidence appears relatively strong that the effects of certain tort reform structures (e.g., caps on noneconomic damages, the elimination of joint-and-several liability) are understood and can help to explain some proportion of geographic variation in use and expenditures at the state level.

Effects of Non-health Policy. Medical sociologists have noted that socioeconomic status is a significant predictor of health outcomes—it is often thought that health care use, by comparison, a considerably smaller proportion of observed variation in health. Accordingly, several authors have remarked that a wide variety of public policies, ostensibly affecting individuals' socioeconomic status, will also affect their health and health care use downstream (see, for example, Adler and Newman 2002). Relevant policies include those pertaining to taxes, housing, environmental exposures, education, law enforcement, and the economy broadly, among others. Because of the long time over which these policies may effect significant changes in a region's health and health care use, few studies exist that measure these effects. However,

given the magnitude of relevant effects as theorized, it is conceivable that much of the unexplained regional variation in health and health care use and expenditures can be attributed to these underlying upstream factors. Several are discussed in greater detail in early sections of this report focused on demand-side factors.

23. PRICE VARIATION

In the context of geographic variation research, the prices of medical services are of great interest as the link between a region's use and its expenditures. Yet the geographic variation literature often takes use and expenditures to be interchangeable; any variation in prices is overlooked. It has been argued that in debates over geographic variation in expenditures, the emphasis on volume has often come at the cost of a discussion of prices. This unbalanced emphasis is detrimental to efforts to reduce this variation (White 2011). As we discuss in the following sections, studies of variation in prices have shown that such variation is non-trivial and should not be discounted.

While there is undeniably variation in prices, the amount of variation is probably far less than the geographic variation in use. Gottlieb and colleagues (2010a) concluded that differences in prices account for a meaningful proportion of observed variation in expenditures based on an examination of variation statistics for unadjusted and price-adjusted samples of Medicare claims data, analyzed at the HRR level. However, in most markets, this proportion is relatively small – their reported correlation coefficient between unadjusted and price-adjusted expenditures at the market level is 0.84, and the magnitude of this expenditure measure's standard deviation declines 18.2 percent with price adjustment when regions are population-weighted (5.1 percent when unweighted). They did observe that the importance of a region's prices in driving its mean expenditure levels to vary materially across regions, but their primary conclusion—that price variation accounts for a small, but non-trivial, fraction of per-capita expenditures among Medicare beneficiaries—stands.

Nevertheless, research regarding geographic variation in prices is less common than research on geographic variation in use and expenditures. There are many reasons why this is the case. First, there is considerable emphasis in the geographic variation literature on the Medicare population, in part because of its standardized payment structures and price list. Geographic adjustments to Medicare prices have been theoretically designed to minimize imbalances in revenues relative to the costs of providing care across regions as well. Thus the reasons for geographic variation in Medicare prices are well-established, and there has been little interest in studying them. Second, in other markets, such as care for the privately insured, there may have been less interest in studying geographic variation in prices due to the numerous factors that are simultaneously determined in the marketplace, prices among them. Thus the challenge of predicting prices in such a market is heightened due to the need to simultaneously model numerous outcomes—service quantity, provider volume and mix, provider and service quality, and consumer demand, among others—and to identify a large and effective set of exogenous determinants for them.⁶

⁶ The multitude of factors that may be considered relevant exogenous determinants of these outcomes of interest—even with the context of a single health care service—include the market shares of local health

Such efforts are further confounded by the dearth and inconsistency of price data. Data on non-Medicare, non-Medicaid prices are rarely available or estimable due to the general lack of uniform pricing even for the same provider (across different payers) or the same payer (across different providers); divergence from uniform price lists in such markets has only increased over the last several decades. Many of the arrangements determining a given provider-payer dyad's prices are confidential. Moreover, when such data are available, which is rare, the prices given may not be of the same form as would be available in another, similar context. For example, raw charges may be available for services rendered by one provider, while negotiated price schedules, adjusted paid amounts, final paid amounts, or patient cost-sharing amounts may be available in other contexts. Still, we may not have access to more than one of these forms of prices in any one context.

Finally, across studies of geographic variation in expenditures, particularly in privately insured populations, there is considerable variation in the methods used to identify and apply price values; there is a surprising amount of variation in the methods used even by those researchers relying on data from Medicare or other public payers. As others have noted (Chernew et al. 2002), the heterogeneity of these methods impedes the consistent, precise, and unbiased estimation of variation in expenditures across providers, payers, and regions and the development of new research reliant on such estimates. Such data difficulties can be significant barriers to conducting reviews of prices within a single geographic region or market, let alone across markets.

Determinants of Prices for the Privately Insured. Conditional on the parameters specifying a region's health care providers, the price of a given medical care service is governed substantially by the identity of the payer involved and the process by which its prices are determined. For private insurance, prices are typically negotiated with providers (Reinhardt 2006). Several factors influence the prices to which these negotiations resolve, including market competitiveness (Zwanziger and Melnick 1988; Melnick et al. 1992; Bamezai et al. 1999; Keeler et al. 1999; Zwanziger et al. 2000; Capps and Dranove 2004), the extent to which insurers employ managed care principles locally (Melnick et al. 1992; Bamezai et al. 1999; Zwanziger et al. 2000), and perhaps to a lesser extent hospital input costs (Keeler et al. 1999; Hay 2003). Moreover, there is evidence that each of these factors varies geographically. We discuss evidence that hospital market competitiveness and managed care vary geographically elsewhere in this report; geographic variation in hospital input costs has been demonstrated empirically by Sloan and Elnicki (1980) among others and is the basis for Medicare's wage adjustment factor. Taken together, these studies present strong evidence of systematic variation by region in the medical care prices negotiated by private insurers.

The type of private insurer may also influence the prices determined through the negotiations of private insurers and care providers. For example, self-insured (ERISA)

plans and providers, the relative substitutability of one provider's service for another's, projected consumers' demand for the service, the dynamics of input factor markets (e.g., for staff labor), and the presence of any payer-driven programs awarding ancillary fees to providers for using health information technology and other factors influencing physician practice patterns. The extent to which researchers have access to these data varies considerably with the intended level of analysis.

employers may be able to negotiate lower prices with providers than third-party insurers if ERISA regulations exempt the self-insured employers from regulatory restrictions on selective contracting and other managed care tactics (Marsteller et al. 1997). We are not aware of empirical evidence regarding this question.

Transparency and Medical Care Prices. While there exists moderate evidence regarding the effects of various factors in determining private insurers' negotiated prices, more extensive and robust empirical evidence is difficult to obtain as a result of the lack of meaningful transparency in hospital-payer price negotiations and even in the prices themselves—the details of many of these arrangements are confidential or protected by contract terms. In particular, while California's hospitals are required to make public their list prices (Reinhardt 2006), this may have little bearing on the average prices paid for care at their facilities.⁷ Reinhardt (2006) also observed that when Aetna announced plans to make public the payment rates that it had negotiated with physicians in Cincinnati in the summer of 2005, the event made national news.

This lack of meaningful price transparency was a major focus in a recent GAO report (2012). GAO researchers examined the extent hospitals were able to exact low prices on medical devices from their manufacturers. Lack of transparency in hospitals' prices was the principal reason for these researchers' difficulties; refusals to provide this information were often required by confidentiality clauses in hospital-manufacturer contracts. The report suggested that this lack of price transparency also inhibited hospitals' efforts to negotiate device prices. Thus, the lack of price transparency for some hospital services may be contributing to escalation of those prices over time. To our knowledge, there has been no empirical analysis of this phenomenon. It is not clear to what extent the use of such confidentiality clauses or other important factors influencing hospital-device manufacturer negotiations may vary regionally.

A related consideration is the fact that, due to managed care organizations' different incentives and administrative systems, it is often more difficult to ascertain "prices" for individual services rendered or paid for by these organizations than it is in the context of more prototypical FFS-paying insurers and self-insured ERISA employers. Because of their different overhead and operating structures, it is reasonable to assume that managed care organizations' prices would differ from those of other organizations. Thus we expect there may be a systematic bias (of unclear direction) in measures of average region- or market-level prices as managed care penetration varies geographically.

Determinants of Prices for Medicare. Medicare prices are determined nationally based on algorithms that vary by the type of service provided. In the determination of physician and hospital payment rates, Medicare seeks to set rates equitably and to adjust those rates in accordance with its policy objectives. Whether or not these adjustments and rates (as currently structured) are an accurate representation of what it costs to provide a particular medical service, serve a particular group of beneficiaries, or serve other community needs (e.g., medical

⁷ Since 2006, numerous states have proposed or enacted legislation increasing the transparency of the prices of health care services at hospitals and other facilities, in part due to increasing demand for such information from supporters of consumer-directed health care initiatives. See "State Actions Relating to Transparency and Disclosure of Health and Hospital Charges," National Conference of State Legislatures. Updated Oct. 2011. <<http://www.ncsl.org/issues-research/health/transparency-and-disclosure-health-costs.aspx>>.

education) in a given region is the subject of a recent IOM report edited by Sloan and Edmunds (2011). Their committee is also producing a second report that will evaluate the effects of the adjustment factors on the distribution of the health care workforce, quality of care, population health, and the ability to provide efficient, high-value care, issues relevant in the discussion of geographic variation in use and expenditures. We provide commentary of a similar nature but more specifically focused on the implications of these rate-setting processes for geographic variation. We limit our discussion to the processes by which physician and hospital payment rates are determined. We would make similar observations in consideration of Medicare's payment policies for skilled nursing care, home health care, dialysis care, and other services.

For a given physician service, Medicare's physician fee schedule determines a payment rate based on a work relative value unit (RVU), which is an estimate of the resources and effort required to provide the service independent of any market factor, relative to other services. This value does not vary geographically. Subsequently, geographically varying adjustments—called geographic practice cost indices or GPCIs—are applied to this base rate to account for estimated geographic differences in the price of the physician's own work (i.e., cost of living), local medical care input prices, and local liability insurance costs (IOM 2011). These GPCIs have been the subject of much political controversy and research, including some with possible implications for geographic variation in use and expenditures. Because we discuss the issues raised in this research (e.g., implications for physician practice location decisions) elsewhere in this report, we do not explore them further here. However, the determinants of these GPCI adjustments are exogenous to the physician and her practice, and we have fewer concerns about diminished social welfare as a result of provider choices.

Additional adjustments are made to physician fees due to policy initiatives intended to attract physicians to designated health professional shortage areas and physician scarcity areas; the Medicare Modernization Act of 2003 also instituted a floor on the physician work GPCI so as to benefit rural physicians further. For the period 2005 to 2007, the two fee supplements—excluding the GPCI floor—raised rates as much as 15 percent (GAO 2005). As Reschovsky and Staiti discussed in a 2005 issue brief, however, it was not clear that these fee changes would achieve their intended ends of increasing rural physician supply, since rural physician incomes, when controlling for specialty mix and cost of living differences, were actually higher than urban physician incomes. In the context of studies of geographic variation in expenditures among Medicare beneficiaries, the most significant effect of these fee adjustments may be to lead researchers to underestimate differences in per-capita expenditures between urban and rural areas—urban areas having been shown to have higher per-capita expenditures on average (LaVela et al. 2004; Klein et al. 2009; Kelley et al. 2011).

Some of the geographically varying adjustments applied to Medicare's inpatient rates are endogenous to individual hospital strategic decision-making. For a given admitted patient, the base rate is based on the patient's reported condition and severity, which determines the Medicare severity diagnosis related group (MS-DRG). The MS-DRG's value is weighted with a hospital wage index intended to account for variation in hospitals' local labor input costs. Subsequent policy-based adjustments are also made to account for the degree to which the hospital serves as a teaching hospital (indirect medical education or IME), serves a disproportionately low-income population with limited ability to pay for their care

(disproportionate share hospital or DSH), transfers the case to another acute care facility, incurred an unusually large cost burden in treating the patient (outlier), or is a rural hospital (critical access hospitals, sole community hospitals, and Medicare-dependent hospitals) (MedPAC 2009).

Recent evidence suggests that the current IME and DSH rate adjustments may be too large relative to corresponding true hospital costs (Nguyen and Sheingold 2011). These findings further substantiate the conclusions regarding IME adjustments of Dalton and colleagues (2001), who found that “operating costs are positively related to teaching activity, but the association shows a decline in strength over time” and that unmeasured patient severity appears to be an important contributor to differences in costs between academic and non-academic medical centers. Further commentary regarding the accuracy of these adjustments as reflections of true cost differentials is provided by the IOM (2011). If these adjustments are larger than actual cost differentials then these overpayments exacerbate differences in per-capita expenditures between urban and rural areas among Medicare, because the IME and DSH adjustments are concentrated in select cities and states (CHCF 2001). This bias is in the opposite direction of that attributable to the policy-based fee supplements to rural physicians. Medicare’s geographically varying adjustments to payments for inpatient hospital care may have other distortionary effects contributing to observed geographic variation in use and expenditures (e.g., with respect to bed supply in academic medical centers). As was the case with studies of the effects of geographically varying physician fee adjustments, we do not explore them further here because the issues are raised and discussed in detail elsewhere in this report.

Determinants of Prices for Medicaid. State Medicaid programs’ payment rates, like Medicare’s, are set prospectively. However, there is considerable variation across states in the generosity of the payment rates established, and year-to-year changes in the payment rates for a given service may also differ considerably across states (for discussion of state-by-state variation over time in primary care physician reimbursement rates, for example, please see Norton 1999; Menges et al. 2001; Zuckerman et al. 2004; Zuckerman et al. 2009; CMS 2010). There are relatively few studies of geographic variation in expenditures among Medicaid beneficiaries in part because of difficulties in acquiring data from fifty-six different Medicaid programs. However, these significant differences in Medicaid fees make clear the importance of accounting for differences in Medicaid fees in such studies.

These differences in Medicaid rates may be attributed to different budgetary constraints, differences in the generosity of allotted Federal Matching Assistance Percentages (FMAPs), benefits covered, the number of beneficiaries covered, underlying population health, local lawmakers’ preferences for different health care services or provider types, or broad differences in politics (e.g., with respect to perspectives on welfare programs, and recently different responses to the states’ financial status). Medicaid fee schedules may change irregularly over time because some of these underlying factors may change unpredictably. Still, we do not explore the relationships among these factors further here because many of these factors are correlated with average demographic characteristics (e.g., education, income), which we discuss elsewhere in this report, and because research regarding geographic variation in Medicaid expenditures per beneficiary is limited.

Comparing Geographic Variation in Prices across Public and Private Insurers.

Philipson and colleagues (2010) recently analyzed differences between the patterns of variation in use and expenditures among Medicare beneficiaries versus among privately insured individuals. They also analyzed variation in unit prices of the care both populations received. This study was similar in scope to Chernew and colleagues (2010), which we discuss elsewhere in this report. However, Philipson and colleagues analyzed patterns of use as well as expenditures, which allowed for further investigation of price variation. Because Philipson and colleagues conducted their analyses at the MSA level (HRRs and five-digit zip codes were not reported in the authors' datasets), the variation they observed in all statistics was probably less than in other studies at HRR and finer levels.⁸

The authors found different patterns between their analyses of use and their analyses of expenditures. In particular, they observed that while there was considerably more variation among Medicare beneficiaries than among privately insured individuals with respect to service use—this is not surprising given the strong likelihood that the authors' methods would not account for substantial variation in the Medicare population due to unmeasured health status variation—there was *less* variation among Medicare beneficiaries with respect to expenditures.⁹ This suggests that price variation was considerably greater within the privately insured population, which is not especially surprising given the price restraints effective in the Medicare program. The variation in prices observed among the privately insured is at least as significant as the variation in use observed among Medicare beneficiaries. The potential selection issues identified by the authors (e.g., that their privately insured sample was drawn from larger, Fortune 500 companies) may well have contributed to these observed levels of price variation, but the findings are noteworthy nonetheless.

Alternative Reimbursement Structures. One issue for geographic variation in prices and expenditures, particularly for privately insured populations, is the extent to which this variation is driven by varying use of alternative reimbursement structures. Broadly, such structures may be described on a continuum of risk-sharing between purchasers and providers, with fee-for-service (FFS) structures on one end, and capitation on the other. The evidence, while neither unanimous nor conclusive, suggests that FFS structures are associated with higher costs, and capitation structures are associated with lower costs (see, for example, Ellis and McGuire 1986; Epstein 1986; Murray et al. 1992; Gaynor and Gertler 1995; Hurley and Labelle 1995; Goodson et al. 2001; Grytten and Sorensen 2001; Finkelstein 2007). As such, documentation of geographic variation in the extent to which markets rely on reimbursement structures at one end of the risk-sharing continuum or the other is of use in studies of geographic variation in use and expenditures as well.

⁸ Since most HRRs are largely contained within MSAs, the MSA may subsume some of the HRR-to-HRR variation one would otherwise observe in a HRR-level analysis.

⁹ Kane and Friedman (1997) conducted a similar analysis focusing on the Medicare population and found that the vast majority of regional variation in Medicare expenditures could be attributed to variation in the volume of services used rather than prices. Kane and Friedman's primary focus was on identifying the most significant market-level determinants of per-capita Medicare expenditures; these findings were largely in line with those discussed elsewhere in this report, though their empirical approach was somewhat less rigorous than those in other studies we have highlighted.

Until the early 1980's, FFS structures were the most common, but reimbursement structures have shifted towards capitation, with the advent of prospective payment systems, shared savings arrangements, and other structures (Reinhardt 2006). Alternative reimbursement structures an important role in determining medical care prices in a given region. The extent to which the use of alternative reimbursement structures and risk-sharing arrangements varies geographically has not been the subject of a great deal of empirical research, however, largely due to the same data limitations hindering other studies of geographic variation in prices.

One important trend is the growth of hospitals and other providers directly integrating with insurers, who can aid the providers with risk management, patient engagement, and other traditional responsibilities of risk-bearing payer organizations. Greenberg and Goldberg (2002) estimated the effects of various market characteristics on the probability of local hospitals becoming affiliated with HMOs. They found that among the market-level factors correlated with such affiliations' formation were average hospital occupancy, the number of hospitals in the area, the number of HMOs in the area, the physician-to-population ratio, the hospital's ownership status, the hospital's teaching status, and location in an MSA. As we discuss elsewhere in this report, many of these factors have been shown to vary geographically. In particular, observed regional variation in these factors is correlated with the distribution of HMOs (Mukamel et al. 2001) and integrated delivery systems (Gleave 2009) nationally, with increased presence on the West Coast, in the Northeast, and in pockets of the northern Midwest.¹⁰ Rosenthal and colleagues (2006), among others, have also presented evidence on geographic variation in the use of pay-for-performance incentives and other related supplements to more standard reimbursement models.

McClellan (2011) and McKethan and colleagues (2009) recently reviewed the theory and the limited evidence of the effects of different reimbursement systems, particularly focusing on FFS- and capitation-based systems. McClellan and McKethan and colleagues observed that some payers have implemented "gainsharing" or "shared savings" arrangements. It has been thought that physicians, who may benefit from reduced use of health care services by all of their patients under these arrangements, might no longer be incentivized to discriminate against potentially high-risk patients in their practices. Unfortunately, the empirical evidence surrounding the effects of such arrangements on resulting access, use, and expenditures, relative to other payment systems, is limited.

McClellan (2011) highlighted evidence that pay-for-reporting programs, which provide supplemental payments to physicians in return for the submission of data related to the quality of their care, may improve quality in some dimensions and harm quality in others (e.g., access). Still, there is little evidence that pay-for-reporting programs alone substantially affect a region's use or expenditures. McKethan and colleagues (2009) also discussed empirical evidence regarding pay-for-performance programs, which reward physicians who score highly or improve substantially on a standardized set of quality measures. Assessing small and large-scale cross-

¹⁰ Further evidence of the relevance of delivery system integration to geographic variation in prices and expenditures was presented in a recent study by Afendulis and Kessler (2011). They found that the institution of Medicare's prospective payment system (PPS) for skilled nursing facilities (SNFs), was associated with reduced spending on average across all markets, but the decline in spending was larger in markets featuring a larger share of vertically integrated providers and provider systems.

sectional demonstrations, they recorded that pay-for-performance programs were associated with “mixed results or small positive quality improvements.” They did not remark on any longitudinal studies of this payment structure. In general, where McClellan (2011) and McKethan and colleagues (2009) found reliable evidence of reformed payment structures’ effects on use and expenditures, those effects were modest in size.

Methodological Considerations in Studies of Alternative Payment Structures. A few potentially important issues about alternative payment systems were not discussed by either McClellan (2011) or McKethan and colleagues (2009). First, the external validity of most studies assessing the effects of a pilot or demonstration of an alternative payment system may be questioned due to a selection bias. Unmeasured local factors, such as legal constraints or market power dynamics, may have contributed to a region’s success or failure in operationalizing the payment structure. Second, there is considerable variation across markets in the specific terms of these payment structures as implemented. Few studies have examined whether these components are necessary or significant in determining the intervention’s effects in a market.

As we alluded to previously, many studies of alternative payment systems’ effects on use, expenditures, and quality compare these outcomes between the intervention group (i.e., where the alternative payment system has been deployed) and a “control group” exposed to a FFS system. However, it is important to recall unadjusted FFS systems are largely a thing of the past—negotiated or “adjusted” FFS reimbursement structures are more common—and so the so-called “control group” is not, in fact, fully immune to the effects of market power, negotiation, and managed care structures. Furthermore, Ginsburg (2011) noted that there was evidence in some markets, providers’ (adjusted) FFS rates may have been negotiated higher than expected on account of their market power. This provides additional evidence that comparisons to markets where “FFS” reimbursement systems are in place are not straightforward—even controlling for a payer’s market share would not be sufficient. It is rarely noted in these quasi-experimental studies that the “control group” has been exposed to many of the same market dynamics influencing the intervention group’s rates (if to a lesser extent); this leads to underestimation of the average effects of introducing alternative payment systems.

Payer Mix, Insurance Status, and Average Unit Prices. In this section, we highlight determinants of variation in unit prices at the market level (i.e., average prices). Several of the concepts that we present have been raised by Reinhardt (2006) and others.

As we discussed in early sections of this report, regions vary with respect to payer mix, managed care penetration, and the proportion uninsured due to a variety of reasons. That regions vary in these population-level metrics is important for the discussion of geographic variation in prices because such studies typically consider average prices. These average prices reflect a weighted average of those prices charged to uninsured individuals and those experienced by their insured peers. List prices reflect the initial amount charged by a provider for each service rendered; they identify the prices paid by uninsured individuals (if they are able to pay), except in case of discounts or charity. Such prices typically exceed by a considerable margin the amounts paid by insurers for their beneficiaries’ care for several reasons, some of which we discuss below. Thus, while there may be little correlation between list prices and amounts paid within a single payer, let alone across payers (Reinhardt 2006), payer mix and the proportion of

the population that is uninsured, which vary geographically, are critical in determining a region's measured average prices when prices are measured as the total cost of care per service. A similar statement could be made regarding average prices when measured as out-of-pocket costs, again since the "price" of care experienced by insured individuals is considerably less than that experienced by uninsured individuals.

Variation in the proportion uninsured and payer mix also influences measures of average prices as a result of insured beneficiaries' moral hazard. Insured consumers experience prices as only the portion of gross "charges" per unit passed through to them by their insurers via premiums or cost-sharing.¹¹ The more removed patients are from the full prices of health care services, the more distorted their patterns of service use will be. Thus the proportion of a market's population that is insured may promote variation in use through the mechanism of price distortion and moral hazard. Measures of average local prices, which are typically weighted by local patterns of service use, are also affected, in turn.

Effects of Competition on Prices. Virtually all studies of geographic variation in prices emphasize the importance of competition. In the neoclassical economics of competitive markets, an increase in the supply of providers, all other things equal, should reduce prices. Likewise a decrease in providers' supply should lead to reduced competition and higher prices. The same hypotheses generally also hold in oligopolistic or monopolistically competitive markets, such as we believe to be the case for inpatient hospital care. Many studies have examined the effects of competition under these frameworks with most focusing on prices, costs, and expenditures as outcomes.¹²

Several empirical estimations of the effects of competition on medical care prices can be found in discussions of prescription drug prices. Duggan and Scott Morton (2010; 2011) tested competition-based hypotheses in their recent studies of drug prices following the implementation of Medicare Part D. They found that, within a given drug class, increases in the number of available drugs have been associated with reduced drug prices to Part D plans, presumably due to competitive pressures. This finding may meaningfully contribute to regional variation in drug prices within the privately insured population as well if insurance companies' and pharmacy benefit managers' formularies vary with respect to the levels of competition they induce within drug classes.

To that end, researchers have demonstrated that formularies vary geographically. Dranove and colleagues (2003) provided evidence that HMOs' organizational characteristics, such as ownership status, relationships with pharmaceutical manufacturers, and intensity of local market competition—which themselves have been shown to vary geographically—were important drivers of their decisions to adopt (or not adopt) several new and expensive drugs.

¹¹ The payment structures defining these cost-sharing arrangements can vary geographically, including the extent of insurers' use of FFS, capitation, and shared-savings arrangements, such as we discuss previously in this section.

¹² One study by Kessler and Geppert (2005) expanded this framework to investigate the effects of competition on intensity of treatment and quality for heart attack patients. They found that average intensity of treatment was significantly determined by patient severity, the intensity of market competition, and, most interestingly, the interaction of the two. Follow-on work in the context of other types of care and conditions may be necessary to clarify the internal and external validity of the authors' frameworks and findings.

Tseng and colleagues (2007) examined Medicare Part D formularies and likewise found considerable variation across plans within California and Hawaii, though it could reasonably be expected that at least one drug in each of the drug classes they assessed would be covered by a given plan. Moore and Newman (1993) not only demonstrated that there was considerable variation across states' Medicaid programs in terms of whether or not formularies had been adopted but also that there was considerable variation among those formularies implemented. Moore and Newman also showed that these formularies had, on average, been effective in reducing pharmaceutical prices. These findings are particularly consistent with those presented by Duggan and Scott Morgan (2011) and suggestive that geographic variation in formularies' use and structures may contribute to overall geographic variation in drug prices.

Insurers and pharmacy benefit managers have often successfully driven down drug prices by leveraging competition through the establishment of preferred drug lists and other mechanisms and, perhaps most importantly, encouraging increased use of generic drugs over brand-name alternatives (Berndt and Aitken 2011). We are not aware of any empirical studies documenting geographic variation in the adoption of preferred drug lists or their effectiveness, though Baicker and Goldman (2011) provided a detailed discussion of the theoretical connections between patient cost-sharing (one of the key levers applied by insurers and benefit managers) and several important medication-related outcomes. In a recent study of geographic variation in Medicare Part D expenditures, Donohue and colleagues (2012) examined variation both in spending and in brand-generic mix across three major drug classes at the HRR level. They—unlike Gottlieb and colleagues (2010a) who conducted a similar analysis of per-capita expenditures—found that on average more than three-quarters of the variation in expenditures they observed was attributable to variation in costs per drug rather than variation in use¹³ and that average prices were very strongly correlated with brand-generic prescription ratios. Unfortunately, the scope of this study was relatively limited. It would be valuable to ascertain more information about the brand-generic drug mixes of other drug classes, to explore what would lead certain regions to persistently prescribe brand or generic drugs more often, and to investigate these questions at finer levels of analysis.

Related strategies such as pooling, coordination, and collective bargaining have been employed by payers and purchasers (e.g., large ERISA employers) in efforts to rein in the prices of medical care services provided by hospitals, physicians, and other providers. Robinson (1995) presented one example of such activities and their effects by examining the effects of large California employers engaging in collective bargaining through the Pacific Business Group on Health (PBGH) and other activities with local HMOs. In his discussion, Robinson pointed out that these strategies were not intended to extract significant premium decreases from the HMOs (such that would cause the HMOs to raise premiums for other employers) but rather to promote improved administrative efficiency and reduced costs for all stakeholders over time. Yet over the course of one year, Robinson documented “enrollment-weighted average HMO premium reduction of 9.2 percent that is attributable to collective negotiations” on top of upgrades in

¹³ As we discuss elsewhere in this report, Zhang and colleagues (2010) observed that variation in drug expenditures was less than for other types of medical care. As such, it may be more appropriate to conclude that variation in use is significantly less for drugs relative to variation in use for other medical services than to conclude that variation in drug prices is significantly greater than it is for other services, which is expected in a market that is more national than most.

select components of the benefit package. While the activities of the Pacific Business Group on Health have not been replicated nationwide, this study presented evidence of welfare gains and price reductions that could be achieved through collective bargaining and other like measures in markets for health care services. We should expect, then, that those regions that have adopted such tactics will have realized corresponding gains. Unfortunately we are not aware of any empirical studies that have tracked this relationship within a geographic variation framework.

In a 2005 study, the GAO assessed variation in prices paid for medical care services—hospital care in particular—through the FEHBP program that highlighted the importance of competition in determining prices. GAO researchers reviewed relevant literature and recorded that most (though not all) studies had found links between measures of market competitiveness and health care service prices; in a few studies, the conclusions were ambiguous. Observations of reduced prices where competition was reduced competition were attributed to reductions in administrative costs—a subject to which we return below. In its own analysis, the GAO found very large variation in inpatient hospital care prices. Variation in physician service prices was also significant and correlated with hospital care price variation, though the variation in physician service prices was not as wide. These findings persisted when the GAO researchers employed robust specifications adjusting for variation in local input costs, population health status and demographics, and other factors. Most other scholars considering this question in recent decades (Zwanziger and Melnick 1988; Melnick et al. 1992; Bamezai et al. 1999; Keeler et al. 1999; Capps and Dranove 2004) have reached similar conclusions. In addition some of these have shown that price competition among hospitals has intensified over time, holding market shares constant, leading to continued increases in hospital prices. However, because of limitations in available data, most of these studies have historically been based on data samples featuring California hospitals.

A related issue is the effect of varying managed care penetration at the market level on geographic variation in prices. Because managed care organizations often leverage competitive forces through modified reimbursement structures, provider network restrictions, and incentives for patients to seek care selectively within those networks (i.e., “tiering”), we expect regions with higher levels of managed care penetration to have lower health care prices. Unfortunately, as with use and expenditures, the relationship between geographic variation in these practices and variation in prices is not well understood in part because of managed care researchers’ focus on other outcomes of interest (Petersen et al. 2006; Rosenthal 2007; Baker et al. 2010; Herring and Adams 2010; Peckham and Wallace 2010). Lack of documentation and transparency of managed care organizations’ per-unit “prices” also inhibit such work.

Cost-shifting and Provider Inefficiency. The GAO (2005) highlighted a few additional issues that have been raised elsewhere as potential drivers of patterns of price variation, including cost-shifting—price discrimination by providers such that private insurers are charged more to recoup margins lost due to lower payments from Medicare and Medicaid. Several important questions on cost-shifting have been the subject of debate for many years. Among these questions are whether providers—hospitals in particular—engage in the practice, the extent to which cost-shifting has taken place, and the degree to which it should be considered relevant in setting public insurers’ payment policies. The implications of the cost-shifting issue for geographic variation in medical care prices are clear: because take-up and generosity of public

insurance benefits varies geographically (as we note earlier in this report), the mechanism of cost-shifting could lead to substantial variation in prices facing private insurers.

For a thorough discussion, we refer readers to Frakt (2011), who reviewed the last fifteen years' theoretical and empirical work on the subject. Among Frakt's (perhaps controversial) findings were that much of the recent literature on cost-shifting has been theoretical or descriptive, and most of them have merely implied that cost-shifting has been pervasive and highly problematic. Moreover, Frakt observed that the most robust empirical evidence that has been gathered has typically supported the idea that while cost-shifting is observable; the scale of the issue is not nearly as large or worrisome as others had suggested. In its 2005 report—not discussed by Frakt—the GAO itself found no evidence that payments to hospitals made by FEHBP's PPO plans were not higher in markets with higher proportions of Medicare and Medicaid beneficiaries. Also, with respect to physician services, the GAO found that PPO rates were lower in markets where average Medicaid fees were lower or the proportion uninsured was higher, controlling for other factors. However, because the covariates used by the GAO to investigate cost-shifting (proportion uninsured, proportion in Medicaid, proportion Medicare, and average Medicaid payment) were gathered at the level of the metropolitan area or state rather than the market or hospital levels, which would be more likely to drive individual hospitals' behaviors, the GAO's findings are less informative.

In a recent study not available at the time of Frakt's (2011) review, Robinson (2011) conducted an empirical analysis testing the cost-shifting hypothesis in a sample of California hospital discharges for select procedures. He also used these results to test an alternative hypothesis proposed by MedPAC, which suggested that hospitals receiving relatively low Medicare and Medicaid case rates would respond by cutting costs and increasing efficiencies rather than shifting costs to private insurers. Robinson's (2011) empirical approach consisted of a multivariate statistical analysis at the patient-level that included a variety of market-level (HRR-level), hospital-level, and patient-level controls. He found that observed differences between hospital revenues and direct expenses (i.e., margins) were substantially higher for privately insured patients than for Medicare patients, which supported the cost-shifting hypothesis. However, interestingly, margins for Medicare patients were significantly less in more concentrated markets than in more competitive markets, which supported the MedPAC hypothesis.¹⁴ That both hypotheses found supportive evidence in this study suggested to Robinson that hospitals may vary in their relative capacities to improve operating efficiencies or to negotiate for higher prices from private insurers, and hospitals would pursue both alternatives. Dobson and colleagues (2006) had argued similarly, though they expressed greater concern that hospitals, increasingly unable to improve efficiencies, could become more unstable financially.

¹⁴ Robinson (2011) also found that, for privately insured patients, per-patient revenues were considerably higher in concentrated markets versus in more competitive markets (i.e., where there are more hospitals and payers may more credibly threaten to exclude a hospital from its network). This evidence lent further support to the importance of competition in determining hospital prices. In this work, Robinson was able to conduct analyses of hospital revenues and costs at the patient level (and avoid limitations common among most studies of geographic variation in prices) through the use of an unusually detailed dataset consisting of California hospitals' experience. All of his sampled hospitals were participants in a value-based purchasing initiative of the Integrated Healthcare Association.

While these results and conclusions appear reasonable, it would be valuable to investigate them further and identify what hospital characteristics predict whether a given hospital would pursue cost-shifting or cost reduction more vigorously. Such investigations would be useful in predicting geographic variation in cost-shifting and, through this mechanism, medical care prices. A review of the extent to which hospital financial performance itself varies systematically and geographically may also be useful in such investigations.

24. VARIATION IN QUALITY OF CARE

Although the geographic variation literature has focused largely on health care expenditure and use, quality of care is also an important outcome. The goal of health policy is not to minimize health spending or use, but rather to achieve the greatest value given constrained resources. Therefore, it is important to understand geographic variation in quality of care and how that variation is correlated with expenditures and use.

Yet, while quality is often studied by health services researchers, considerably less work has been done on geographic variation in quality, with few studies estimating causal effects. Geographic variation in quality of care has been studied using a variety of identification and research methods. We provide a brief synopsis of literature on variation in quality, measured as mortality, morbidity, various measures of the process of care, and patient satisfaction. To the extent variation in use and expenditures is associated with measures of quality of care—causal inference has rarely been shown—we will discuss such measures later in this section.

Mortality Measures. Measures of mortality are frequently used in geographic variation research because data measuring mortality are highly valid and the outcome is of great interest. Mortality has been commonly used in comparative effectiveness research. The main limitation to using mortality measures is that for many populations mortality is relatively rare and difficult to link directly to explanatory variables of interest. Drawing significant conclusions based on all-cause mortality measures may be especially difficult, and controls for underlying health status and sociodemographic variation will be critical for models at most levels of analysis. Geographic variation studies that use mortality measures often do not satisfactorily address endogeneity (e.g., regarding physician location and local health care needs and health status). We also find that condition-specificity in mortality measures may be important in distinguishing studies with regard to the validity and precision of their estimates. And so, while the conclusions presented across studies of geographic variation in mortality are relatively consistent, it is difficult to infer causality here as often as the studies' authors might advocate.

One early example of the use of mortality measures in geographic variation research was produced by Baker and colleagues (1987), who analyzed county-level variation in mortality due to motor vehicle crashes. They found vehicle crash-associated death rates varying by a factor of more than 100, with the variation reduced somewhat through controls for age, income, and population density. The authors argued higher death rates could be potentially explained by both supply- and demand-side factors, including reduced proximity and access to trauma centers, poorer road quality, higher speed limits, and local preferences in car models and seat belt use. They did not explore these factors empirically, and so this study must be considered primarily an

effort to whet the appetite of federal authorities and other scholars for further analyses using their new datasets. Baker and colleagues' methods, however, including mapping and the introduction of sociodemographic controls in explanatory models, persisted across much of this literature, extending well beyond the context of automotive accidents.

Physician Supply and Mortality. Baicker and Chandra (2004b) assessed the association between physician supply measures and mortality and a variety of other outcomes at the HRR level. In particular, they examined areas' primary care presence and degree of physician specialization—measured as the number of physicians of primary care and select specialties per 100,000 population, respectively. They found only small associations of either measure with all-cause mortality, controlling for the density of other specialists as well as select, condition-specific health status and demographic indicators. Baicker and Chandra theorized that increases in the volume of the physicians could increase quality by improving access to needed services, but the same increases could decrease quality if effective coordination of care and communication declined correspondingly. Their results suggested that these effects netted each other out. Importantly, Baicker and Chandra do not discuss their results in terms of causation, which they admit they cannot infer due to the endogeneity between their outcomes of interest and physician supply measures. Also, it is possible, given the authors' relatively limited set of covariates, that unmeasured factors at the region or individual level biased their results toward zero and caused their observed associations to be small. Some biases might have been better addressed through patient-level analyses that would permit accounting for variation in individual health status.

A related study by Chang and colleagues (2011) evaluated the relationship between two measures of primary care workforce presence and patient health outcomes, namely mortality and ambulatory care sensitive condition (ACSC) hospitalization rates, as well as expenditures. They found that mortality and ACSC hospitalization rates were both somewhat lower in areas with higher primary care workforce levels, though the magnitude and significance of these results changed varied materially depending on the primary care workforce measure evaluated. It is difficult to ascertain how the authors' empirical specifications may have affected these results in light of the fact that they claimed to adjust for "specialty mix," while it is not clear how this was accomplished. Still, their results are reasonably consistent with those of Baicker and Chandra (2004b), suggesting that the magnitude of a region's primary care workforce and its specialty mix may affect its mortality rates, but the difference associated with changes in these variables are difficult to measure without large samples.

Moreover, all-cause mortality is much less sensitive to variation in specific medical specialist volume than condition-specific mortality. For example, Ayanian and colleagues (2002) found that among Medicare patients who had an AMI and survived at least three months, those patients who saw only primary care providers in an ambulatory care setting following their initial hospitalization were less likely to survive at two years than like patients who saw a cardiologist. Several other outcomes were also significantly different between the two groups. While many of the factors that determine access to care, such as various socioeconomic status measures, were unmeasured, Ayanian and colleagues presented strong, consistent evidence that the population of patients who received care from a cardiologist were materially different from their peers who did not receive care from a cardiologist in terms of their demographics,

underlying health status (as measured by quantifying comorbid conditions), and mortality outcomes. This supports the conclusion that key determinants of access, which vary by population and by region (we provide more in-depth discussion of these issues elsewhere in this report), may significantly affect condition-specific mortality outcomes and drive regional variation in them. We caution, however, that the authors' use of propensity scores to match patients who did and did not receive care from a cardiologist leads to some concerns of omitted variable bias. In light of the significant differences in state of residence between matched and unmatched samples, we have concerns about unmeasured market-level factors driving the behaviors they attribute to specific determinants of access.

From a methodological perspective, we also observe that the results generated by Ayanian and colleagues were likely stronger than those of Baicker and Chandra (2004b) because of their condition-specificity; all-cause mortality measures may wash out a considerable amount of meaningful variation in mortality outcomes by region. However, like Baicker and Chandra, Ayanian and colleagues are not able to address endogeneity concerns regarding the location of physicians and the needs of these patients, and this may bias their results.

Medical Care Intensity and Mortality. Some have argued that a proportion of observed variation in mortality measures should be attributed to increased intensity of medical care, including Skinner and colleagues (2010). They produced an analysis in response to criticisms of their work and the work of many of their Dartmouth colleagues on intensity of care (i.e., health care expenditures) during the last six months of life. We present some of these criticisms in the sections of this report focused on supply-side factors. The authors conducted a simple correlation of expenditures, measured two ways (a dummy for quintile of spending, and a continuous spending variable), and one-year mortality due to acute myocardial infarction (AMI) at the hospital level. They found their measures of expenditures, which served as proxies for overall medical care intensity, were “modestly but positively associated” with mortality. While, as the Congressional Budget Office (CBO) noted (2008), this result has been consistently shown across studies, the extent to which this association should be attributed to the intensity of medical care services or to unmeasured variation in health status is unclear. Skinner and colleagues (2006) correctly pointed out that no adjustment for risk, health status, or other like measures using known data sources and methods could fully satisfy us that a regression of mortality on expenditures would be free of bias from unmeasured health status. Nevertheless, this does not allow us to infer causality from their results.

Similarly, Fisher and colleagues (2003b) examined the relationship at the region level between overall intensity of local care practice, as measured using end-of-life spending, and several key outcomes, including mortality. They found that intensity of care was positively correlated with mortality among patients with AMIs and negatively correlated with mortality among patients with hip fractures. However, since Fisher and colleagues' results were inconsistent across some conditions, and no significant results emerged for cohorts with other conditions, general conclusions could not be drawn. Skinner and colleagues (2006) later showed that trends in similar, though less complicated, analytic results for AMI patients varied over time and differentially by region as well.

As we discuss elsewhere in this report, there are reasons to be skeptical of the analytic approach Fisher and colleagues (2003b) applied in this and a companion paper considering a different set of dependent variables. Despite the authors' arguments that their approach minimized the need for managing variation in populations' health status, it is not clear whether the unmeasured variation associated with underlying population health status would bias their results upward or downward. As such, it is difficult to put stock in even the authors' potentially more reliable condition-specific results.

Morbidity Measures. By contrast, relatively few geographic variation studies have evaluated variation in patient health outcomes other than mortality, thus making it difficult to draw general conclusions about the relationship between geographic variation in health care and patient morbidity outcomes. In large part, this should be attributed to greater uncertainty, concerns about validity and sensitivity, and probably most importantly lesser availability of measures of morbidity and other dimensions of health. Still, some research has attempted to capture variation in such measures—studies by Fisher and colleagues (2003b) and Sloan and colleagues (2001) will be those discussed. The limitations of these studies are similar to those of the mortality variation studies discussed above.

Fisher and colleagues (2003b) studied how a region's overall intensity of care practice affected patient functional status. They found no significant relationship between care intensity and patient functional status. While the authors clearly demonstrate geographic variation in their key explanatory variables, this analysis identified practice intensity using a measure of spending on patients in the last six months of life. We discuss our skepticism of this approach elsewhere in this report. The concern is that the population's underlying health status may explain both expenditures during the last six months of life and reported functional status, which is likely to bias the results in the Fisher and colleagues' paper. There may also be other biases in this analysis due to variation in the quality of health care provided in certain regions: low-quality regions may produce higher expenditures and more rapid declines in functional status among care-seeking Medicare beneficiaries.

Sloan and colleagues (2001) investigated the effects of hospital ownership status (for-profit, non-for-profit, or government) on quality of care and financial outcomes. In addition to survival (mortality), the authors relied on measures of quality of care found in the National Long-Term Care Survey (NLTCS), which assessed participants' functional status (based on changes in the number of ADLs and IADLs reported to be performed with difficulty or limitation), cognitive status, and whether participants were living in the community (as opposed to living in an institutional setting). The authors found their ownership status covariates held no explanatory power in predicting each of these quality outcomes. The sample size of their study is somewhat small, particularly when limited to those who received care in hospitals with a particular ownership type and when assessing variation in low-frequency outcomes such as mortality; as such it is unlikely that they would be able to pick up statistically significant effects on their quality-related outcome variables. Still, the authors' work is otherwise robust. We provide further discussion of this study as well as the geographic variation of hospital ownership elsewhere in this report.

Temel and colleagues (2010) analyzed the effects of integrating palliative care with standard oncological practice on a population of patients with metastatic non-small cell lung cancer. The authors found that their study's intervention patients reported higher quality of life, showed fewer symptoms of depression, and even experienced longer median survival than their control patients. Given the high specificity of the identified study population, the external validity of Temel and colleagues' study is not clear. However, this study did provide meaningful evidence of improved patient quality of life and reduced morbidity given the introduction of palliative care. Given known geographic variation in the use of palliative care (see, for example, Goodman et al. 2010), this study also provided evidence of geographic variation in patient morbidity attributable to medical care.

Process of Care Measures. Numerous studies, both condition-specific and across conditions, have been conducted regarding variation in the quality of health care as measured through one or more process of care measures. Such measures assess whether or not providers rendered specific care processes that have been identified as high-quality practices when data suggested there were opportunities to do so (or, conversely, whether services identified as low-quality practices were provided in place of other services—or none at all—given opportunities). They are generally regarded as reflective of the appropriateness of the care provided. Results in studies relying on these measures have been reported at the level of the individual measure or in aggregate (including the use of weighted or unweighted average values) across multiple measures. There are important concerns about many studies in this literature including omitted variables bias and that the measured dimensions of quality neither reflect the full spectrum of an area's provided health care services nor acknowledge all of the benefits patients may derive. Still, these process-of-care measures are used widely because of the ease with which many can be calculated and analyzed using available data sources, the ease with which they can be conveyed to readers, and the belief that providers are accountable for their associated results.

McGlynn and colleagues (2003) is one of the most commonly cited studies of variation in the quality of health care in the U.S. This study assessed provider performance on over four hundred process of care measures, finding only approximately 55 percent of patients received recommended care overall. Unfortunately, while the authors referred to analyses assessing how their quality measure results varied across their twelve sampled metropolitan areas and how that variation might be correlated with other market and individual characteristics known to vary geographically (e.g., managed care penetration), they did not discuss findings from this work. Thus this work cannot be taken to inform discussion of geographic variation in quality.

Many studies like that of McGlynn and colleagues but more limited in scope have been conducted regarding variation in provider performance on process of care quality measures, including a few studies more explicitly assessing geographic variation in results. In one such paper, Jencks and colleagues (2000), writing with the support the then Health Care Financing Administration (HCFA), adopted or developed 24 process-of-care measures across six conditions. In subsequent years, these measures studied by Jencks and colleagues (2000) were evaluated and used by HCFA in comparisons of the quality of health care provision nationally and at the state-level across geographies and over time. Certainly, as Jencks and colleagues acknowledge, their set of 24 measures presents an incomplete picture of the whole of medical care provision—omitting any discussion of post-acute care management for their six conditions

of interest, long-term care, and mental health care, and underrepresenting ambulatory care and surgical interventions—and so drawing conclusions that extrapolate from this “unbalanced” set to all of an area’s medical care would be perilous.

Through their study, Jencks and colleagues demonstrated that their measures varied significantly across states and that, when considering measure score mean and median values, there were regional trends in this variation. According to these aggregate measures, quality of care was consistently higher in New England, the North Plains, the Mountain West, and the Pacific Northwest, while quality was most consistently lower in California and the South. A few years later, Jencks and colleagues (2003) produced a similar analysis using 22 of the previous study’s 24 measures, measuring the differences in measure rates nationally between the two cross-sectional assessments of performance. Appropriately, they did not extend these analyses to draw conclusions about individual providers of care within a given region, since the authors’ data were collected at levels no finer than the state-level. See our discussion of variation across studies in the level of analysis and the hazards of assimilating evidence from across studies at different levels of analysis earlier in this report. Jencks and colleagues also noted that the improvement they observed in most metrics was in the same areas emphasized by Medicare Quality Improvement Organizations (QIOs), but argued reasonably it would be difficult to conclude the relationship was causal. Jencks and colleagues did not explore the key underlying causes for this observed variation or their observed trends of improvement in further detail in either study.

Baicker and Chandra (2004b) also examined the effects of increased primary care specialists per capita and medical specialists per capita each on an aggregate measure of care process quality, which they termed an “effective care index.” This index was comprised of 11 measures of individual processes of care identified as of high quality. The authors found that this measure was not associated with increases in the number of specialists per capita at the HRR level, nor were the individual measures composing the index when assessed separately. The authors provided two related explanations for this finding: market distortions and coordination externalities associated with increased volumes of specialist physicians. They also raised the idea that, when the concept of “quality” is being interpreted broadly, patient utility ought to be considered among quality’s unmeasured and potentially important dimensions. For example, patients in an area with a high volume of urologists should have reduced travel times and wait times. We discuss this facet of increased provider volume effects elsewhere in this report, and we also discuss select aspects of it below in the context of patient satisfaction measures of quality. Baicker and Chandra’s study is noteworthy for its clear analysis of one potentially important issue – physician supply – in determining the quality of health care. It will be important to evaluate an area’s level of physician specialization with a broader set of quality measures encompassing a wider range of medical care services and conditions before ruling it out as a principal driving factor.

In a separate study, Baicker and colleagues (2004) divided a set of medical procedures into low-cost, high-benefit services (overlapping with those conditions evaluated by Jencks and colleagues (2000; 2003)), and higher-cost, questionable-benefit services for analysis. They then plotted simple measures of use and disparities in use between blacks and non-blacks for these services. Results were assessed both within and between HRRs. Baicker and colleagues found

that there was significant variation in quality measures by region, consistent with others' findings (e.g., Jencks and colleagues 2000). Their new contribution was the finding that, across several process measures of quality, disparities in quality of care also varied significantly across regions; prior studies had obtained similar findings for fewer measures or for care pertaining to select conditions (see, for example, Zaslavsky 2002). Perhaps more strikingly, they also found that "disparities are not clustered in particular geographic areas (such as the South) and thus cannot easily be attributed to historical regional patterns of discrimination." They found that overall performance on quality measures was not a strong indicator of reduced disparities in quality. Similar results were also generated for measures of disparities in per-capita expenditures by region. The most significant limitation of this study was its lack of controls for numerous important demand- and supply-side factors (e.g., income) that, if included in an analysis, could significantly affect the observed disparities as well as more effectively examine the sources of these findings. Without further clarification here, it is not clear to what extent the observed variation should be attributed to differences in health markets, labor markets, linkages between them (e.g., inter-racial disparities in insurance status), or other factors.

In a more targeted study, Landrum and colleagues (2008) studied geographic variation in the use of chemotherapy for colorectal cancer patients. The authors found that higher-spending regions make more appropriate and inappropriate use, both, of chemotherapy treatments, but their results with respect to other treatments and tests, both appropriate and inappropriate, were mixed. This suggested that in some cases regions with increased propensity to provide more intensive care may benefit some patients but harm others by doing so (see Stukel and colleagues (2005) and Chandra and Staiger (2007)). In general, their results are consistent with those of previous studies concluding that there is not a consistent relationship between average expenditures or intensity of care and quality measures at the region level, though there may be reason for concern that areas with increased expenditures and intensity of care may realize poorer quality of care than regions with lower expenditures and intensity. The Congressional Budget Office (2008) reached a similar conclusion in its review of this literature, including studies by Fisher and colleagues (2003a; 2003b), Baicker and Chandra (2004a; 2004b), and Sirovich and colleagues (2006); we discuss these articles elsewhere in this report. For further discussion of this and similar condition-specific findings, see the work of Chassin and colleagues (1987), Leape and colleagues (1990), and Guadagnoli and colleagues (2001).

Care Intensity and Patient Satisfaction. Patient satisfaction has been shown to vary geographically. Many of the results from these studies rely on patient satisfaction and experience information collected in one of a small group of datasets, primarily the Medicare Current Beneficiaries Survey (MCBS) and Consumer Assessments of Healthcare Providers and Systems (CAHPS) surveys. In general, researchers have concluded that a wide variety of factors may influence patient care experiences and satisfaction, ranging from system-level factors, such as the overall intensiveness of care, to individual physician-level factors. Furthermore, the most important factors to account for may depend on the population of interest, the condition or care setting, and how one measures patient satisfaction.

Fisher and colleagues (2003b) included patient satisfaction among the quality-related outcomes they considered in their analysis of the effects of variation in overall area care intensity. This study relied on a measure of end-of-life expenditures as a marker of exposure to different levels of overall care intensity in the region, and, as we discuss elsewhere, we have

reservations about this approach. Following the work of others, Fisher and colleagues aggregated patient satisfaction survey data from MCBS into two general measures of patient satisfaction with care and three measures of satisfaction with the patient's usual physician, if one was reported. Descriptive analyses and linear regressions of these satisfaction measures on the end-of-life care intensity measure—including limited controls for patient and region characteristics—identified meaningful variation in patient satisfaction by region but no consistent relationship between regional care intensity and the five aggregated measures of satisfaction.

In a related study using the same explanatory variables, Fisher and colleagues (2003a) also showed that with respect to several measures of access to care, regions with higher intensity of care fared somewhat worse, though not always significantly. Mittler and colleagues (2010) conducted another similar study using ten aggregated CAHPS measures rather than MCBS but also relying on the end-of-life care index used by Fisher and colleagues. These CAHPS data allowed the authors to investigate the effects of exposure to a region's intensity of care on both overall patient satisfaction and patient reflections on other specific aspects of the care process not previously investigated in this context, such as the effectiveness of the doctor's communications and the helpfulness and courtesy shown by staff. Mittler and colleagues also used a more comprehensive set of control variables than had Fisher and colleagues (2003a; 2003b). In general, Mittler and colleagues found that there was a strong negative relationship between their case-mix adjusted mean CAHPS scores and a market's intensity of care. That is, survey respondents in regions with greater care intensity were more likely to report lower satisfaction on several dimension of their care. Results were particularly strong for measures of access, interactions with office staff, and the doctor's effectiveness of communication.

Wennberg and colleagues (2009) obtained a similar finding when considering the effects of care intensity in the inpatient setting, "measured by days spent in the hospital and inpatient physician visits," on inpatients' care experience. They assessed patients' satisfaction with their inpatient care using Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey data, focusing particularly on the proportion of patients who reported a negative care experience overall. Correlation coefficients between this measure of patient dissatisfaction and hospital care intensity were reported as strongly positive. Given the straightforward nature of the authors' explanatory variable and its demonstrated geographic variation across the sample, this study's identification strategy was more credible than prior results relying on the end-of-life care intensity index. Overall results may be questioned somewhat, however, given the authors' relatively limited comments on efforts to estimate the effects of the HCAHPS survey's 74.4 percent response rate (in terms of discharges).

In general, the work of Fisher and colleagues (2003a; 2003b), Mittler and colleagues (2010), and Wennberg and colleagues (2009) could lead us to conclude that the relationship between a region's intensity of care and the patient's satisfaction is negative or, at best, indeterminate. However, in addition to our concerns about the use of the end-of-life intensity of care index in these studies, that they were conducted at the HRR level is also potentially problematic. As Mittler and colleagues noted, very different results might be observed if analyses were conducted at the patient or physician level. This might be very important if, as we suspect, most of the mechanisms by which the patient's experiences would change would be

effective at these finer levels. It is also not clear to what extent unmeasured variation in health status and other variables might bias these results. For these reasons, we are not able to draw meaningful conclusions about the relationship between these measures.

Variation in Quality as a Determinant of Geographic Variation in Expenditures and Use. Not only is quality of care an important outcome that varies geographically, quality of care may indirectly affect expenditures and use. The mechanisms by which a region's quality of care might drive expenditures include undertreatment, overtreatment, inappropriate care, untimely care, unskillful care, and ineffective accounting for patient safety. Each of these mechanisms has been linked to poor patient outcomes and higher per-episode expenditures (see, for examples in the context of overtreatment, Hellström et al. 2004; Paré et al. 2006; Goodman et al. 2010; Temel et al. 2010). However, these mechanisms are also mutually determined. Determining the direction of causality between expenditures and quality of care is far from straightforward. A region's quality of care and expenditures are both jointly determined by each other and by other factors, such as those we highlight in our earlier discussions of the determinants of geographic variation in use and expenditures. Because of these complexities and the indirectness of the potential mechanisms by which quality variation may drive variation in expenditures, discussion of these effects is beyond the scope of this report.

Conclusions on Geographic Variation in Quality. Quality of care is as important an outcome as expenditures and use, which have comprised the majority of our review in this report. In the preceding section we have provided a brief review, first of literature identifying geographic variation in quality itself, and then of how quality of care may be indirectly related to health care use and expenditures. Quality of care remains an area of key interest to researchers, providers, and policymakers, and key questions remain to be investigated. In the context of geographic variation research, we have reviewed numerous articles linking variation in patient outcomes with various market and demographic characteristics known to vary geographically, although these links have generally not been shown to be causal. The literature on geographic variation in quality may still be broadly characterized as incomplete.

25. CONCLUSION

Regardless as to whether demand-side factors may drive a greater proportion of observed geographic variation in health care use and expenditures, supply-side factors are commonly of particular interest to researchers pursuing the subject. In part, this can be attributed to the wide variety of factors that may play a role, from the individual physician-level to the state- and system-levels. Extensive research has been dedicated to exploring many of these factors individually, yet there remain many unanswered questions and concepts about which more robust results could yield significant insights. Under most circumstances, we may conclude that supply-side factors do play a meaningful role in determining local patterns of use and expenditures. The relative significance of these factors, relative to demand-side factors, demand-supply interactions, and other factors, however, may well vary significantly by condition, setting, and region. Moreover, there is a serious concern that these results are merely associations or partial correlation in nature, not causal, or adequately reflecting clinically and policy relevant

differences that vary geographically but which were omitted in many of the individual studies we reviewed.

A. APPENDIX

Note on R^2 . In the course of our review, we have encountered a number of papers reporting R^2 measures, especially in discussing the importance of area or state effects on outcomes. We want to caution readers about both the pitfalls of such reports and about the misinterpretation of such findings. But before we provide any details, we need to summarize some of the types of remarks that we think are problematic in using these measures to discuss geographic variation when the ultimate interest is in how area, state, or other contextual variables affect the behavior of populations of interest such as potential and actual patients, providers, and even payers. What raises a problem is that the work is rarely done at the most *micro* level of concern: members of the population of potential or actual patients or patient-provider dyads. Instead, much if not the majority of what we reviewed was rolled up to a more *macro* or higher level of aggregation such as an institution, an HRR, or some geopolitical unit like an MSA or state.

Among the most important items to remember in this context (but which are often overlooked) are:

- A summary of explained variation at an area level—versus more micro summaries—may not say much about micro-level behavior.
- In this context, statements about levels or changes in R^2 capture only area effects (implicit or explicit).
- As always, partial adjustment does not deal with endogeneity, nor does it permit the researcher to draw inferences that can be interpreted as causal. Partial correlation is just that, that which can be explained by the model as specified, whether the estimator is level appropriate for the model as specified (or not).

We start with a simple linear model (in both the sense of linearity in coefficients and additivity in the error as well as in the sense that the dependent variable y is analyzed by linear estimator). Then at the *micro* level of person i in state s (or any other encompassing group), the equation of interest for outcome y is

$$\begin{aligned}
 y_{si} &= x_{si} \beta_x + z_{si} \beta_z + \mu_s + \varepsilon_{si} \\
 i &= 1, \dots, N_s \\
 s &= 1, \dots, S \\
 N &= \sum_s N_s
 \end{aligned}
 \tag{1}$$

where the x 's are the included observable characteristics, z are the potentially observable covariates that are not included, μ are the unobserved state or aggregate level summary variables, and ε is an independent and identically distributed (*iid*) patient or consumer level measure. At the micro level, the total population is N . At this point, we are not saying anything about

whether the state level error μ is *iid* or not; much of what follows is more in the vein of a random intercept model because a fixed effects model would sweep out the geographic variation of interest. All variables are taken as deviations from their grand mean for the population as a whole.

Now we roll the analysis up to the state or higher level by taking averages for that state for the dependent variable y , the included covariates x , the excluded covariates z , the state error term μ , and the *iid* individual error term ε . Let the overbar with a subscript of $\{s\bullet\}$ indicate the state level mean. Equation 2a is the true model averaged at the state level, while equations 2b and 2c are stated in terms of the observables and errors that belong to the model as estimated. That is, these last two equations correspond to what is often done in practice.

$$\bar{y}_{s\bullet} = \bar{x}_{s\bullet} \beta_x + \bar{z}_{s\bullet} \beta_z + \mu_s + \bar{\varepsilon}_{s\bullet} \quad 2a$$

or

$$\bar{y}_{s\bullet} = \bar{x}_{s\bullet} \beta_x + \theta_s \quad 2b$$

where

$$\theta_s = \bar{z}_{s\bullet} \beta_z + \mu_s + \bar{\varepsilon}_{s\bullet} \quad 2c$$

This reformulation leads to the following conclusions:

1. If there is some covariate x which has a major impact on individual behavior in equation 1, it need not in equation 2. If most of the variation of x is within states, but not across states, then it may have very little explanatory power in equation 2b, because there is little variation in $\bar{x}_{s\bullet}$ across states. The classic example of this would be gender.

The precision at the second level (and hence its R^2) is directly related to the variation in that x at the state level. In contrast, age, racial and ethnic effects do come through because they do vary across states more than gender does. However, they often explain less at the state level than they do at the individual level because the variation between states is less than the variation between individuals.

2. As long as the population at the state or HRR level is large enough to ignore the small sample properties of $Var(\bar{\varepsilon}_{s\bullet}) = \left(\frac{\sigma_\varepsilon^2}{(N_s - 1)} \right)$, then the variance of the error in equation 2a ($\mu_s + \bar{\varepsilon}_{s\bullet}$) is about the same as variation in the state level of μ , and we can ignore the cross-person variation at the individual level.¹⁵ Thus,

$$Var(\mu_s + \bar{\varepsilon}_{s\bullet}) \ll Var(\mu_s + \varepsilon_{si})$$

¹⁵ We may ignore the term since we are taking variables as deviations from the grand mean. In small samples of cases nationally, such as rare diseases, this term may be more significant.

3. The actual error in equation 2b includes not only μ , but also the part of the covariates z that was not included in the analysis and varies across states. So, if the health variables are less than complete or are subject to notable measurement error, then the associated variance will be added to the variability due to μ alone. How much this matters will depend critically on what is in the x list and what is in the z list, as well as to what degree they vary by state or region, as well as within state.
4. If we were to move an included covariate to the excluded list, the explanatory power of a remaining included x would tend to increase by the usual omitted variables argument; the magnitude of the increase would depend on the covariance between the excluded and included variables and the signs of their β 's in the true equation 2a. This effect is offset by the fact that the variance of θ will also increase.

Finally, we remind our readers that increments in the percent of variance explained depend on what is already in the model and the order with which sets of variables are either entered or deleted. Unless these variables are uncorrelated and strictly additive (no interaction effects), one cannot do a unique decomposition of variance that is not dependent on path and order.