**Physician-Induced Demand**

EM Johnson, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2014 Elsevier Inc. All rights reserved.

**Introduction**

Physicians are often blamed for the high cost of healthcare in the US. Physicians dupe patients into consuming too much care, the story goes, driving up costs without producing commensurate gains in health. This line of reasoning derives from the physician-induced demand (PID) hypothesis, which is a long-debated topic in health economics. Under the PID hypothesis, physicians influence patient demand to suit their own interests. They are able to do this because their patients know relatively little about the type or quantity of treatment they need. Faced with payment systems that reward quantity of care on the margin, the inducing physician provides care beyond the level that objective clinical judgment and patient preferences would dictate. In short, inducing physicians create their own demand rather than reacting to market demand.

The idea that doctors create their own demand is often used to make the case for healthcare reform, particularly changes to provider payment systems. Peter S. Orszag, the director of the Office of Management and Budget from 2009–10, was paraphrased by the New York Times as saying, ‘the supply of hospitals, medical specialists, and high-tech equipment appears to be driven by the practicing physicians’ (Evans, 2009). Induced demand is a leading explanation for the geographic variation in utilization that has been documented across the US (Fisher et al., 2003a, b). Atul Gawande has argued in the New Yorker magazine that induced demand combined with differences in the ‘culture of money’ across areas explains regional variation in Medicare costs per capita (Gawande, 2009). Although there has been no rigorous test of this relationship (as noted in Fuchs (2004)), policymakers have latched onto the idea that altering physician incentives in high cost areas can reduce costs without sacrificing quality of care. ‘The Economic Case for Health Care Reform’ of the White House states, ‘large variations in spending suggest that up to 30 percent of health care costs (or about 5 percent of GDP) could be saved without compromising health outcomes’ (http://www.whitehouse.gov/administration/eop/cea/The Economic Case for Healthcare Reform).

This article reviews the empirical evidence on PID from the health economics literature. In the next section, induced demand is defined and the evidence on the topic reviewed. The following section brings evidence to bear from related literatures. Finally, the concluding section discusses policy implications and areas for future research.

**Empirical Evidence**

The concept of induced demand is first attributed to Evans (1974). The precise definition of McGuire (2000) follows: Under induced demand, a physician takes an action to shift the patient’s demand curve in the direction of the physician’s own interests. Physicians can affect such a shift, because they have more information regarding the patient’s condition and treatment options than the patient, an example of the market failure known as asymmetric information. In theoretical models of induced demand, the action taken by the physician is unobserved and is limited at the margin by its costs. Typically inducement is itself costly for the physician, but in some models inducement negatively impacts patient flows (e.g., Pauly, 1980; Rochaix, 1989) or physician reputation (e.g., Dranove, 1988).

Two aspects of this definition merit clarification. First, induced demand does not include actions that influence demand in the best interest of the patient. Indeed, moving demand toward the patient’s optimum is a responsibility of physicians. Second, the definition leaves room for treatment to vary across patients and providers. The benefits and risks of treatments vary with patient characteristics, and it is the job of the physician to tailor care to individual patients. Moreover, differences in physician practice styles, practice environments, and experience mean the true costs and benefits of treatments vary across physicians.

In the researcher’s ideal world, the quantity of care the physician views as being optimal from the patient’s perspective would be observable. The econometrician could then compare actual treatment with this benchmark, taking any difference as evidence of inducement. By comparing inducement across incentive environments, she could then estimate the physician’s objective function. Many of the characteristics of patients and doctors that determine appropriate treatment are unobserved. For this reason, empirical work on induced demand has used alternative identification strategies. This review groups papers according to the empirical approach and reviews each group sequentially. First are studies that use shocks to physician incomes, and especially physician-to-population ratios, to test for induced demand. Next are studies that use changes in physician fees or variation in patient information to identify inducement.

Before turning to empirical results, it is helpful to briefly clarify the predictions of PID models and compare them with alternative models of physician behavior (see McGuire (2000), for more detail). Under PID quantity is determined in equilibrium by physicians equating the marginal cost of induction with its marginal benefit. Physician incomes, fees, and patient information are all predicted to affect the quantity of care. Physician incomes affect the quantity of care through the income effect. A negative (positive) income shock increases (decreases) the marginal utility of income and increases (decreases) the returns to induction. If quantity of care is reimbursed at the margin, physicians then respond by increasing the quantity of care. An important caveat is that this prediction applies only when there is an income effect; otherwise inducing doctors induce equally at different levels of income (McGuire, 2000).
Changes in physician fees also affect quantities under PID. As in the case of an income shock, a fee reduction increases care quantities through the income effect. In addition, if the fee reduction differentially affects one area of the physician’s practice (e.g., certain treatments or patients) relative to others, then there is a substitution effect (McGuire and Pauly, 1991). The relative returns to inducing decrease in the more affected area, and quantities are shifted to less affected areas. Thus, in response to own-fee reductions quantity can either increase or decrease, depending on the relative strength of the income and substitution effects. The prediction for less affected areas is unambiguous: quantity will increase. A final prediction that has been tested empirically is that physicians should have less ability to induce demand among more informed patients, where the asymmetry of information is lessened. This arises as long as the costs of inducement to informed patients are higher.

All of these predictions are in contrast with the perfect agent benchmark, in which only patient preferences and clinical factors matter for treatment, but it is more interesting to contrast the predictions of PID with models of physicians under symmetric information. When patients are informed, profit-maximizing physicians cannot shift demand, but they can affect healthcare quantities by making take-it-or-leave-it offers of nonretradable services or by altering their choice of quality or effort (McGuire, 2000). This means observing quantities that depart from the patient’s optimum is not sufficient evidence of PID. More relevant for evaluating the existing empirical work on PID, observing substitution in response to a fee change is not informative on PID, as physicians with informed patients can also be expected to shift quantities in this manner (McGuire, 2000).

However, profit-maximizing suppliers differ from inducers in that they do not adjust quantity in response to income shocks. This explains the focus of the empirical literature on using shocks to physician incomes and large fee changes to test for PID, though this approach amounts to jointly testing PID and the income effect. Because profit-maximizing physicians have no reason to treat informed and uninformed patients differently, studies using variation in patient information are also informative as to the underlying model of physician behavior. With this in mind, for some policy questions, one only needs to understand the reduced form relationship between physician incentives and utilization, so evidence in this category is considered as well.

**Income Shocks**

Many empirical studies of induced demand use variation in physician incomes to test for inducing behavior. The earliest studies of induced demand fall into this category. For the most part, these studies examined the relationship between market-level physician-to-population ratios and utilization. This approach is rooted in the idea that an exogenous increase in the number of physicians in a local practice area should spread patients more thinly, lowering physician incomes. Healthcare utilization is then increased if inducing physicians respond through the income effect and treat patients more intensively. Termed the ‘availability effect’ by Pauly (1980), these studies come the closest to directly testing the proposition that healthcare supply creates its own demand.

The first paper in this vein, Fuchs (1978), runs cross-sectional two-stage least squares (TSLS) regressions on the number of surgeons per capita at the market level. To identify supply shifts, Fuchs instruments for the number of surgeons using characteristics of metropolitan areas that should affect surgeons’ location decisions, but not local demand (e.g., metropolitan status, hotel receipts, and percent white). He finds a 10% increase in the surgeon-to-population ratio increases surgery by 3%, which he interprets as evidence of induced demand. Cromwell and Mitchell (1986) use a similar methodology with more data and finer geographic markets and find a 1.3% increase in elective surgery. Rossiter and Willensky (1983, 1984) relate healthcare utilization to physicians-per-capita using physician-level data and find even smaller effects.

These studies were highly influential, but there is concern that the instruments employed to isolate supply shocks do not satisfy the exclusion restriction. For example, Gruber and Owings (1996) suggest that results are biased toward inducement, because the average coinsurance rate, which is unobserved, is likely correlated with demand and with the included measures of attractiveness of an area to physicians. An additional concern is that supply shocks may reduce the price or time cost of services, causing patients to move down a static demand curve. Omitting these factors would bias the results toward finding inducement. Dranove and Wehner (1994) provide a powerful critique of the empirical methodology. Employing a method similar to Cromwell and Mitchell (1986), they show that increasing the number of obstetricians increases utilization on a dimension clearly out of the physician’s control: the number of births.

Gruber and Owings (1996) avoid many of these problems and provide some of the best evidence to date on PID. In this paper, the authors instrument for state-level changes to the physician-to-population ratio, using the secular decline in the fertility rate from 1970 to 1982. They then look for evidence that physicians respond to the income shock by increasingly performing highly reimbursed Cesarean sections (C-sections) in lieu of less profitable vaginal deliveries, and they find a modest effect: obstetricians replaced approximately 10% of their income by increasing C-sections. By studying a plausibly exogenous shock to income, this approach is not subject to the criticisms of the previous literature. There are also fewer concerns about changes in time-cost in this context. However, it is difficult to compare the size of the estimated effect with previous studies, as obstetricians may have also recovered income on other margins.

**Changes in Physician Fees**

There is also a large empirical literature that uses changes in physician fees to identify inducement. The main advantage of this approach is the availability of large, exogenous fee changes for study, and most studies have used Medicare fee changes. Medicare fee changes are also appealing for testing PID because Medicare patients make up a significant fraction of physicians’ practices. This is important because only fee changes that affect physician incomes have differential...
predictions for utilization under PID and models with symmetric information (McGuire and Pauly, 1991).

Rice (1983) studies a large fee change enacted by Medicare in Colorado in 1976. Consistent with an income effect, Rice finds increases in Medicare volume in the Denver metropolitan area, where fees are lowered, relative to the surrounding areas where fees are raised. Point estimates suggest that a 10% decline in reimbursement led to a 6.1% increase in medical services and a 2.7% increase in surgery. However, it is possible that patient demand was affected by changes in patient responsibility over the time period. The short panel also prevents the author from assessing whether urban and rural areas were affected by differential trends over the period of study.

Nguyen and Derrick (1997) also study the impact of a Medicare fee change on Medicare volumes. They study the 1990 Medicare fee change, legislated in the Omnibus Budget Reconciliation Act of 1989, which reduced reimbursement for procedures deemed to be ‘overpriced.’ Using physician-level data, the authors find physicians experiencing fee reductions increase Medicare volumes. The volume response is similar in magnitude to Rice (1983), but it is significant only for the 20% of physicians who experienced the largest price reductions. While results are again consistent with PID, the study suffers from limitations similar to Rice (1983).

Yip (1998) also studies the 1990 reform, additionally considering the effect of Medicare fee changes on non-Medicare volumes. The paper focuses on thoracic surgeons, whose reimbursement rates were significantly reduced by the 1990 Medicare fee change. In this context, fee cuts led to increased volumes to both Medicare and private payers, with providers recouping, on average, 70% of income lost due to price reductions. The paper also convincingly demonstrates that the income effect is driving results by showing that physicians whose incomes were hit hardest by the reform have the largest volume responses. Jacobson et al. (2011) exploit a more recent Medicare fee change. The authors study the 2005 change in Medicare’s reimbursement of outpatient chemotherapy drugs, and they find that physicians responded to reduced fees by increasingly administering chemotherapy. They also show that physicians substituted toward drugs that were less affected by the fee reduction in their prescribing behavior.

Gruber et al. (1999) is another important paper on fee changes and quantities. This paper studies Medicaid fee changes, specifically changes in Medicaid’s reimbursement for C-sections. The authors expect the policy to have only a small income effect since Medicaid patients are a small fraction of providers’ practices, and in fact, they find that the substitution effect dominates in this context: a 10% increase in the Medicaid fee led to an 8.4% rise in C-sections in the Medicaid population. While this result is consistent with models of PID, it is also consistent with physicians setting quantity under symmetric information.

Variation in Patient Information

There is also a long-established literature that uses variation in patient information to test for induced demand. These studies are motivated by the idea that informed patients should resist doctors’ attempts at moving them away from their optimum consumption level. When inducing physicians are reimbursed for treatment on the margin, one expects utilization to be lower among more informed groups. The first study in this vein, Bunker and Brown (1974), compares rates of surgery for lawyers, businessmen, and ministers with those of physicians, who they view as informed consumers of medical services. Contrary to the prediction, they find self-reported surgery rates to be equal or higher among physician families when compared with other professional families in the same county. They conclude that physicians must have unobservably higher demand for medical services.

The conclusion of Bunker and Brown (1974) highlights the main weakness of the approach. When comparing utilization across patient groups, any omitted factors that are correlated with utilization will bias results. For example, prices, care quality, and health status may all differ across the professionals considered in this study. Hay and Leahy (1982) adopt the same approach using survey data with more extensive controls, including income, insurance coverage, and self-reported health status, but they also find higher use among physicians.

Domenighetti et al. (1993), in a more recent survey in Switzerland, find that the average person’s probability of receiving one of seven major surgical interventions is one-third above that of a physician or a member of a physician’s family. Ubel et al. (2011) survey physicians and find they want less intensive treatment for themselves than they would recommend to patients in two fatal disease scenarios. Again, results are difficult to interpret as patient characteristics influencing demand may differ across groups. Currie et al. (2010) address this weakness by conducting a patient audit study, which allows them to ensure comparability across informed and uninformed groups. Fake patients visited physician offices in China, where physicians have a financial incentive to prescribe medication. They then compare prescription rates of patients who verbally signal their understanding of appropriate prescription behavior with those who do not. It is found that prescription rates for the uninformed patient are higher by 25%. However, the physician could also have interpreted the information signal as a signal of patient preferences.

Related Literature

In this section, evidence on PID from related empirical literatures is considered. First, the empirical literatures on medical malpractice and defensive medicine are reviewed briefly, and results from the growing literature on physician incentives in managed care are summarized. The literature on physician self-referral, which considers whether physicians respond to the private incentive to use resources they partially own more intensively, is also reviewed. Finally, the literatures on pay-for-performance programs and studies of physician convenience factors are discussed.

First consider the literature on medical malpractice. So far, it has been assumed that the incentive physicians respond to in inducing demand is financial, but physician response to private liability risk is also consistent with the definition of
induced demand. Kessler and McClellan (1996, 2002) show that tort reform reduces medical expenditures on Medicare heart patients without affecting patient outcomes. They interpret this as evidence that doctors practice ‘defensive medicine,’ providing care that does not benefit patients in order to reduce their liability risk. More recently, Currie and MacLeod (2008) show that malpractice pressures increase the utilization of procedures that reduce liability risk, such as diagnostic testing, but decrease the use of risky treatments, such as the performance of C-sections in delivery. More research is needed to explore the relationship between financial and malpractice incentive systems.

The discussion has so far presupposed that it is financially rewarding for physicians to provide more healthcare to patients. Although this is true in fee-for-service payment systems, physicians paid by capitation have incentives to provide less treatment (Ellis and McGuire, 1986; McGuire, 2000). In fact, researchers have shown that physicians paid by capitation spend less time with patients (Mechanic et al., 2001; Tai-Seale et al., 2007; Glied and Zivin, 2002; Melichar, 2005) and provide less care to each patient (Epstein et al., 1986; Safran et al., 2002; Stearns et al., 1992; Greenfield et al., 1992). Salaried doctors and doctors with bonuses tied to utilization measures also appear to respond to incentives for providing less care (see e.g., Hickson et al., 1997), Barro and Beaulieu (2003), Gaynor et al. (2004), and Hemenway et al. (1990)). These results cannot be interpreted as conclusive evidence of PID, however, as symmetric information models also predict this behavior. Further complicating interpretation is the fact that even perfect agents may reduce care if resources are rationed under managed care.

There is also a large empirical literature on self-referral practices by physicians. This literature studies treatment decisions when physicians have an ownership stake in some part of their practice. Reimbursement is typically higher when resources are owned by the physician, and studies find that physicians respond to this incentive by increasingly recommending patients for treatment. Mitchell (1992) and Hillman et al. (1992) study ownership incentives and referrals to diagnostic testing facilities, Yee (2011) studies ambulatory surgery centers, Barro et al. (2005) study specialty hospitals, and Baker (2010) studies the utilization of imaging devices. Iizuka (2012) also contributes evidence by showing that physician prescription behavior responds to pharmaceutical markups in Japan, where physicians dispense as well as prescribe drugs. Afendulis and Kessler (2007) studies a related conflict of interest. They observe that integrated cardiologists, who can both diagnose and perform interventional procedures to treat heart disease, have stronger incentives to recommend patients for intervention compared with non-integrated cardiologists, who must refer patients for treatment. They find that patients of integrated cardiologists are, in fact, more likely to receive percutaneous interventions.

Finally, the recent literature on pay-for-performance programs, which tie physician reimbursement to observable quality measures, suggests that performance incentives can affect care (Campbell et al., 2007; Rosenthal and Frank, 2006; Mullen et al., 2010). Studies of labor and delivery also suggest that obstetricians sometimes perform C-sections for their own convenience (Burns et al., 1995; Spetz et al., 2001). These results are inconsistent with the perfect-agent model. However, symmetric information models also predict substitution toward more highly reimbursed and away from more costly treatments. Therefore, it is again difficult to disentangle distortions due to financial incentives from those due to principal-agent concerns.

**Suggestions for Future Research and Policy Implications**

There is a large and growing body of empirical evidence that physicians’ treatment decisions are influenced by factors beyond their patients’ needs. Convincingly identified studies have shown that obstetricians do more C-sections in response to declining fertility, cardiac surgeons treat more intensively when their incomes are impacted by fee reductions, and physicians in China prescribe more medication to ‘un-informed’ patients. This evidence is inconsistent with the model of physicians as perfect agents, and it supports PID as one avenue through which physicians affect quantities of healthcare. In addition to this direct evidence of PID, physicians respond to private malpractice incentives and financial incentives for self-referral. Physicians also appear to respond to the incentives in managed care plans. Finally, there is some evidence that pay-for-performance programs and physician convenience factors affect healthcare choices. Taken together, these studies suggest that physician incentives, broadly defined, are important determinants of both healthcare costs and the distribution of health resources in the United States.

However, more work is needed before one can make statements about the economic importance of PID. Although empirical research has provided estimates in several contexts, there is reason to believe that the effect will differ across incentive environments, physician specialties, patient groups, and even across treatment categories within physician–patient pairs. Future research exploring this heterogeneity should also aim to bridge our current understanding of PID with claims made in the health policy arena. How much of the variation in utilization across geographic areas can be explained by demand inducement? Have physician incentives or constraints on inducement changed, such that PID has contributed to growth in health spending over time?

More theoretical work is also needed. Exploring the impact of competition on physician behavior is a promising area for research (and one that may produce new testable implications for PID), though this first requires refining our understanding of the sources of physician market power. It would also be interesting to theoretically explore the interplay of the various incentive systems that physicians face, for example, by studying physicians who are contracted with both HMOs and PPOs. The PID research agenda is important for policy. The general direction of health policy in the US and other countries is to push some financial risk to physician groups, as accountable care organizations (ACOs) do in Medicare. If PID is pervasive and powerful, creating an interest among physicians in providing less care may work well to reduce healthcare costs, but not without raising concerns about access and quality. The extent of induced demand also has implications for physician workforce and training policies; and gaining
clarity into induced demand behavior has implications for health insurance design – inducement affects the interpretation of parameters that are central in the optimal insurance literature. Finally, the literature on PID can help us to understand the impacts of patient empowerment policies, for example, patient ownership of their own medical information.


References


**Further Reading**


