How Do Doctors Behave When Some (But Not All) of Their Patients are in Managed Care

Sherry Glied, Columbia University
Joshua Graff Zivin, Columbia University

Available at: http://works.bepress.com/josh_graffzivin/8/
How do doctors behave when some (but not all) of their patients are in managed care?

Sherry Glied, Joshua Graff Zivin

Department of Health Policy and Management, Joseph Mailman School of Public Health, Columbia University, 600 West 168th Street, Room 608, New York, NY 10032, USA

Received 1 October 2001; received in revised form 1 November 2001; accepted 9 November 2001

Abstract

Most physicians today treat a variety of patients within their practices where a variety of insurance arrangements co-exist. In this paper, we propose several theoretical explanations for physician treatment patterns when the patient population is heterogeneous at the practice level. Data from the 1993–1996 National Ambulatory Medical Care Survey (NAMCS) are used to test how practice level managed care penetration affects treatment intensity. Practice composition has strong effects on treatment. Visit duration appears to be constant across patients within a practice, while medications prescribed appear to be converging as managed care penetration increases. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: I11; D2

Keywords: Capitation payments; Heterogeneous practice; Physician treatment

1. Introduction

Physician practices are rarely comprised of entirely managed care patients. Indeed, most physicians today see a range of patients whose insurers pay providers using a variety of payment levels and methods. Remler et al. (1997) surveyed 2000 physicians and found that the mean physician practice received capitation payments for about 13% of all patients, and 41% of all practices included at least some capitated patients. These heterogeneous conditions may help explain the relatively modest impact managed care appears to have had on physician office practice.

In this paper, we propose several distinct theoretical models of patient treatment in a stylized heterogeneous environment with two types of payers: traditional fee-for-service and...
managed care. At the practice level, we propose three models of physician behavior: one in which managed care organizations pay marginal costs for excess physician capacity; one in which physicians induce demand for services; and one in which there are both fixed and variable costs associated with different types of physician effort. In this mixed payment environment, these models yield distinct predictions about the behavior of physicians toward patients who pay with each type of coverage. The excess capacity model suggests that an increase in managed care penetration reduces excess capacity within a practice and therefore increases the negotiated fees that physicians receive for treating these patients. This implies a convergence in treatment intensity to fee-for-service levels as managed care penetration in a practice increases. The demand inducement model suggests a divergence in effort intensity; fee-for-service patients receive more treatment when managed care enrollment is high. The implication of the final model has two parts. First, some components of effort will be characterized by fixed costs and will not vary across patients within a practice. As managed care penetration increases, this fixed cost effort should decrease. Second, the variable cost efforts to both patients will increase. Now that fixed cost effort is lower, physicians must increase variable cost effort to managed care patients, thus providing them with sufficient levels of aggregate effort to retain them as customers. Fee-for-service patients receive additional variable cost effort to compensate for the diminished fixed cost effort they now receive. Whether an increase in practice level managed care penetration yields a convergence or divergence in these variable cost efforts, depends on the rate of increase in intensity.

These theoretical hypotheses are tested using data from the National Ambulatory Medical Care Survey, an annual nationally representative survey of randomly sampled physicians in office-based practice. For completeness, we report data on Medicaid and Medicare patients as well. We find that at an individual patient level, average visit duration is indeed lower under managed care, while all other dimensions of visit intensity are roughly equivalent. Medicaid patients look very similar to managed care patients, except they tend to receive more medications. Medicare patients, as one might expect given the health of this group, are treated more intensively than all other patients.

Practice composition is shown to have a large and statistically significant effect on practice intensity. Visit duration is constant across all patients within a practice and decreases as the percentage of managed care patients, Medicaid patients, or Medicare patients in a practice increase. Medications prescribed increase for both managed care and fee-for-service patients as managed care penetration increases. The increase for managed care patients is faster, suggesting a convergence in this dimension of treatment intensity. Similarly, we observe convergence toward the Medicare level in Medicare-dominated practices and, although the results are not statistically significant, convergence toward the Medicaid level in Medicaid-dominated practices. In the case of tests ordered and return visits specified, the data are quite noisy and no hypothesis can be ruled out. Overall the data appear to be most consistent with the fixed cost hypothesis of treatment in a heterogeneous practice. In summary, visit duration appears to be a fixed cost and there is some evidence of treatment convergence in medication prescription.

Our conclusions speak to effects at the practice level. In our empirical specification, however, what we are calling practice level effects may incorporate some market area effects. While we include regional × metropolitan area controls, these are imperfect proxies for market level managed care penetration. In addition, physicians are not randomly assigned
to insurers; they choose them. Our ability to control for physician selection into networks most consistent with their practice style is limited. The basic tests we can perform, however, do suggest that physician selection alone is not sufficient to explain our empirical findings.

The paper is structured as follows. Section 2 presents theoretical models for physician behavior in a mixed payment environment. Section 3 describes the data and methods employed in this paper. Section 4 provides empirical results and the final section concludes.

2. Models of physician behavior in a mixed payment environment

Economic theory suggests that, holding all else equal, patients who pay using capitation rates or who visit salaried physicians will receive less intense services than those who pay using fee-for-service rates. The simple economics of capitation, salary, and fee-for-service payment, however, provide no guidance about how these alternative payment mechanisms might be affected by the behavior of other patients in a practice. The industry literature on physician and hospital practice has begun to explore the problems of management in a mixed payment system, but this literature is very limited (Krohn and Broffman, 1998; Manheim and Feinglass, 1994). In the economics literature, most analyses of economies of scope contemplate the sale of two related products using similar pricing strategies, not two (nearly) identical products using different pricing strategies.

Physicians may respond to changes in practice characteristics by changing the quality (or intensity) of services provided to an individual patient through changes in effort. Models of quality choice among monopolistically competitive firms show that firms will choose practice intensity such that the marginal revenue achieved through increased intensity is equal to the marginal cost of producing output at that intensity (Sheshinski, 1976). In the context of physician practice, HMO pricing behavior will mean that the marginal revenue associated with an increase in service intensity for an HMO patient will be smaller than that for an indemnity patient. If the marginal cost of production is increasing in effort, physician firms serving HMO patients will choose a lower level of effort than will firms serving indemnity patients. In such a model, the effect of practice characteristics on the treatment of individual patients will depend on the structure of the service intensity production function and the economic rewards for effort.

In the simplest case, each payer would purchase his own ‘package’ from the physician, independent of other payers, and receive services based on payment rates at the margin. Under capitation, where payment at the margin is zero, this model would imply zero service provision. As no patient would purchase such a plan, it is necessary to impose a minimum service constraint that is sufficiently high to ensure customer participation. Thus, service intensity by payer would simply reflect how well that payer pays subject to the minimum quality constraint. Treatment intensity would be completely independent of practice composition. Section 2.1 describes several more complex models of physician behavior in a heterogeneous practice, where treatment intensity is not independent of practice composition. Practices are comprised of two types of patient: traditional fee-for-service and managed care. For simplicity, we ignore other payers such as Medicare and Medicaid where, due to important patient characteristics, physician incentives to provide effort are less clear. They are, however, included in our empirical work for completeness.
2.1. Excess capacity model

In this model, we assume that HMOs seek to pay marginal costs for excess physician capacity (Gold et al., 1995a). If part of the incentive for physicians to participate in managed care plans is to ensure a large client base, excess capacity will likely be larger when HMO concentration is low. Thus, HMO bargaining power will be highest when HMO concentration is low. If bargaining of this type exists, then the fees paid for HMO patients will rise the larger HMO payment is as a share of a physician’s practice. As HMO penetration increases, excess capacity in the practice decreases, and HMOs can no longer extract such low prices from physicians.¹ In this case, practices with larger HMO shares will be paid higher reimbursement rates, which, in turn, will lead to greater effort levels to managed care patients, while leaving effort levels to indemnity patients unaffected. Given the economic assertion that, ceteris paribus, effort level to managed care patients is lower than that for fee-for-service patients, patterns of care for indemnity and HMO patients will converge as the share of HMO patients in a practice increases.

**Hypothesis 1.** In a model where HMOs pay marginal costs for excess physician capacity, treatment intensity of fee-for-service and HMO patients will converge to the fee-for-service level as HMO penetration increases within a practice.

2.2. Demand inducement model

A second model in which physician practice characteristics can affect the treatment of individual patients is one that incorporates demand inducement. If physicians can induce demand for services and income effects are important, then changes in the fees paid by one payer in a multi-payer firm can generate demand inducement for the other payer (McGuire and Pauly, 1991; for a more comprehensive treatment, see McGuire, 2000). In the context of alternative financing arrangements, physicians may profit from inducing behavior from indemnity patients but may not profit from inducing demand from patients paid through HMOs. If HMOs also offer physicians lower returns than indemnity plans, then an increase in the HMO share of a practice will lead to a reduction in physician income. If abstention from demand inducement (or ethical behavior) is a normal good, this increase in HMO share will lead to an increase in demand inducement from fee-for-service patients. Effort to managed care patients will remain unchanged. Thus, this framework suggests a divergence in effort intensity by payment type as HMO concentration increases, with fee-for-service intensity increasing.

**Hypothesis 2.** In a model of demand inducement, treatment intensity of fee-for-service patients will increase as HMO penetration increases and intensity to HMO patients will be unchanged.

¹ If, on the other hand, HMOs were able to significantly reduce service utilization, then excess capacity could grow with HMO prevalence in a practice. The literature suggests that no such reduction in utilization has occurred, and if anything shows a slight increase in utilization (Glied, 2000). As such, we do not formally address this scenario in the paper.
2.3. Fixed cost model

In this model, we assume that there are both fixed and variable costs of effort production. The fixed cost investment can be thought of as durable equipment or office capacity (intellectual or physical) that cannot vary for the two types of patients. Thus, a physician must assess the needs of their practice, based partly on practice composition, and make an unalterable investment accordingly. More generally, we can think of the fixed cost investment as simply one that is prohibitively costly to change in the short-run. Anticipating our empirical work, it may be reasonable to expect the duration of a patient visit, given the need for advanced scheduling, to be costly to vary by patient and therefore fixed for all patients.

Physicians can, however, alter the variable components of effort to each type of patient and thus incur variable costs as well. A practice that treats predominantly fee-for-service patients will choose a level of fixed cost investment that is optimal for fee-for-service patients. In such a practice, fee-for-service patients receive variable cost effort that reflects marginal reimbursement rates for this effort. Any HMO patients seen in this practice, however, will receive lower variable cost effort to compensate for the “excessive” fixed cost effort with which they are provided. Variable cost effort will be reduced to the point where total effort just satisfies the minimum service constraint. In a practice that treats mainly managed care patients, the optimal fixed cost investment will be smaller. HMO patients will then receive higher levels of variable cost effort than in FFS-dominated practices; again ensuring that total effort is large enough to satisfy the participation constraint. Fee-for-service patients receive even more variable cost effort to compensate for the diminished fixed cost effort with which they are provided. Thus, there exists a tension when investing in fixed costs that will be greatly influenced by practice composition.2 When a practice is predominantly fee-for-service patients, the fixed cost investment will be geared more towards fee-for-service patients. As the HMO population within a practice increases, physicians will make fixed cost investments more appropriate for a managed care style practice.

The fixed cost investment response function with respect to practice composition may be smooth or stepwise depending on whether there exists a continuum of optimal investment levels or a discrete set. In the latter case, one would expect to find key investment tipping points at specific practice composition levels. In so far as composition changes are large enough to span a tipping point, results in either case will be the same.

In sum, under a fixed cost model, as the HMO population increases, variable cost effort to HMO and fee-for-service patients will increase. Convergence or divergence will depend on the relative rates of increase. If effort to HMO patients increases faster than effort to fee-for-service patients they will converge.3 If fee-for-service effort rises faster, effort levels will diverge. The fixed cost of production theory also suggests that firms will tend toward more homogeneous payment configurations over time (Glied, 1998).

2 The use of the term practice composition here could describe either the share of visits by payer or the share of enrollees by payer. If insurance status does not greatly influence the frequency of office visits, these should be equivalent. If it does influence visit frequency, we expect fixed cost investments to reflect composition by visits, as it better captures the returns to effort.

3 Note that the assumption that, in the absence of practice effects, fee-for-service patients receive more effort than HMO patients coupled with the fact that effort to fee-for-service patients is increasing in HMO penetration, implies that the effort levels will never cross.
Hypothesis 3. In a model with fixed costs of effort type, treatment intensity of fee-for-service and HMO patients will increase as HMO penetration increases within a practice. If effort to HMO patients increases faster than effort to fee-for-service patients they will converge. If fee-for-service effort rises faster, effort levels will diverge.

2.4. Data and methods

We test these hypotheses using data from 1993 to 1996 National Ambulatory Medical Care Survey (NAMCS), an annual nationally representative survey of randomly sampled physicians in office-based practice. Over 1100 physicians are sampled in each survey year. The NAMCS asks physicians to report characteristics of about 30 randomly sampled visits that take place within a sample week. In total, the surveys include information on about 35,000 patient visits each year, including the expected source of payment for the visit, the geographic location of the visit, and the physician’s specialty. The NAMCS payer categories vary slightly across years. In particular, PPO visits are separated from other privately insured non-HMO visits only after 1994. The NAMCS data through 1996 do not contain detailed information about the type of payment. For example, the data do not indicate whether a physician received capitation or fee-for-service payment.4

The NAMCS data contain considerable detail about patient characteristics and about the treatment provided during a visit. These data include demographic information, up to three diagnoses, services provided at the visit, medications ordered or provided, and duration of visit. Information is available on the geographic location of the visit (region and metropolitan area) and on the physician’s specialty.

We code insurance across all years using a consistent metric. The NAMCS permits physicians to code multiple insurance types for a single visit. We omit all visits where there was “no charge” or paid by “unknown” (including visits paid for by workers’ compensation and non-Medicaid and -Medicare government insurance) or “other” insurance (these comprise about 8% of visits in each year). We then rank insurance categories so that each individual falls into only the highest ranked category. From lowest to highest, these categories are: self-pay (which may include traditional indemnity coverage), private insurance, Medicaid, Medicare, PPO (after 1994), HMO. In most of the analysis, we group together self-pay, private insurance, and PPO and call this group traditional fee-for-service (TFFS). Separate analyses suggest that these three groups closely resemble each other in both patient and visit characteristics. We repeated the analyses treating self-pay as a separate group, but the results were generally insignificant, in part because the population is so small. We classify beneficiaries enrolled in Medicaid or Medicare HMOs as HMO patients. Treating these patients as Medicaid/Medicare instead has no substantive effect on our results.

Next, we use information on all visits to a specific physician and compute the percentage of those visits paid by Medicaid, Medicare, HMOs, and PPOs (after 1994). We exclude the index patient from this analysis in order to calculate the payment composition of all other patients seen in the practice. We omit from this analysis any physician with fewer than 15

---

4 The 1997 data, which are in several other respects different from the preceding data, do include this information. Using these data, we find that about 34% of visits paid by HMOs are capitated.
Table 1
Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th>Mean</th>
<th>S.D.</th>
<th>25th percentile</th>
<th>75th percentile</th>
<th>TFFS mean</th>
<th>HMO mean</th>
<th>Medicaid mean</th>
<th>Medicare mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>100195</td>
<td>17.82</td>
<td>10.50</td>
<td>10</td>
<td>20</td>
<td>18.28</td>
<td>16.95*</td>
<td>16.10*</td>
<td>18.52*</td>
</tr>
<tr>
<td>Tests</td>
<td>109895</td>
<td>0.83</td>
<td>0.92</td>
<td>0</td>
<td>1</td>
<td>0.77</td>
<td>0.84*</td>
<td>0.77</td>
<td>0.95*</td>
</tr>
<tr>
<td>Medications</td>
<td>109895</td>
<td>1.33</td>
<td>1.37</td>
<td>0</td>
<td>2</td>
<td>1.20</td>
<td>1.23*</td>
<td>1.36*</td>
<td>1.72*</td>
</tr>
<tr>
<td>Return specified</td>
<td>109895</td>
<td>0.62</td>
<td>0.49</td>
<td>0</td>
<td>2</td>
<td>0.59</td>
<td>0.57*</td>
<td>0.61*</td>
<td>0.75*</td>
</tr>
<tr>
<td>HMO patients (%)</td>
<td>109895</td>
<td>24.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicaid patients (%)</td>
<td>109895</td>
<td>8.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significantly different from TFFS mean ($P < 0.05$).

visits (observations) in the data. Finally, we compute interaction terms for TFFS insurance multiplied by the HMO share of practice category.

In all analyses that follow, we control for age, age-squared, female, black, other race, Hispanic, a set of dummies for metropolitan area residence $\times$ region of residence $\times$ sample year and a set of dummies for patient diagnosis. After excluding visits with “no charge” or paid by “unknown” or “other” insurance, the sample has 109,895 visits. It is important to recognize, however, that the extent of actual variation in the data is more limited than suggested by the sample size. The practice share coefficients are identified from information on 4116 distinct practices. Since we observe information on only about 30 visits per physician, roughly 20% of the average physician’s caseload, our estimates of practice characteristics are measured with error. This error may reduce the precision of our practice level results.

Standard measures of case mix do not reflect the demands of different types of visits on outpatient visit characteristics. Instead, we code diagnosis using two sets of variables. First, we include a variable for the number of separate diagnoses reported by the physician for the visit. The NAMCS allows physicians to enter up to three diagnoses. Second, we construct diagnosis categories for each outcome variable. The diagnosis categories divide diagnoses into deciles according to their mean effect on visit characteristics across the entire sample. For example, diagnoses are divided into those with mean visit duration of 13.75–15, 15.1–16.1 min, and so on for the analysis of duration. The same diagnosis classifications are used regardless of the source of payment. In some of the analyses, we omit these diagnosis categories in order to assess the potential effects of patient selection on the observed outcomes.

We examine four principal measures of visit intensity: duration of visit, number of tests ordered, number of medications ordered, and whether a return visit was specifically scheduled by the physician. Table 1 presents descriptive statistics for these outcome variables and

---

5 There are 3806 practices included for the analysis of visit duration since, for this purpose, we exclude from our calculation of visits per practice all visits with duration of zero.

6 After omitting practices with fewer than 10 HMO visits, correlations between TFFS decile and HMO decile were 0.72 for visit duration, 0.91 for test number, 0.79 for number of medications ordered, and 0.76 for return visit specified.
for the insurance variables that are the focus of this study. Duration is of interest because it is an easily observable measure of visit intensity from the perspective of the patient. Under either fee-for-service or capitated payment, increasing the duration of visits makes it difficult for physicians to see more patients. Under fee-for-service payment, however, the physician may be able to bill more for a longer visit. Thus, we expect HMO payment to reduce the duration of visits. Note that we omit from the analysis of duration all visits with duration equal to zero and all visits with duration greater than 1 h. In these data, on average visits last 17.8 min. Overall visit duration has increased slightly over the 1993–1996 period. HMO visits average nearly 1 min shorter than the sample average. Medicaid visits were more than 1.5 min shorter than average, while Medicare visits were 3/4 of a minute longer than average.

Ordering more tests is one way for physicians to increase the intensity of services—and billings—without using up much of their own time (McGuire and Pauly, 1991). Again, we expect HMO payment to reduce the number of tests ordered. Unfortunately, the tests included in the NAMCS change periodically. To keep the measure consistent over the time period we restricted the analysis to a subset of tests that were asked about specifically in each survey. The tests are blood pressure exam, HIV serology test, urinalysis, other blood test (including cholesterol), and other (non-blood) test. Over the whole period, the number of tests averaged 0.83 per person. HMO patients receive about the same number of tests as the sample average. It is plausible that HMO physicians might order less expensive tests than do other doctors. While the NAMCS data do contain information on the type of test, the information is not sufficiently detailed to address this possibility. TFFS and Medicaid patients received slightly fewer tests, while Medicare patients received more than the average number of tests.

The effect of HMO payment on medications is less clear. While reducing medication orders may save money for the MD under certain contractual arrangements, this is certainly not true for all HMO arrangements and is not generally true in fee-for-service arrangements. Lower rates of medication prescription may also be a sign of reduced overall service intensity. Some factors suggest that HMO patients are likely to be prescribed more medications than are indemnity patients. First, HMO patients are more likely to have insurance coverage for pharmaceuticals, so that demand side pressures to reduce intensity will, in this case, be stronger outside HMO practice. Second, providing patients with a prescription may be a way to reduce other more costly interventions. Third, providing patients with prescriptions may reduce their propensity to return for another visit, again saving the physician’s time. The average visit in these data included a new prescription order or medication continuation for 1.3 medications. Coding changes occurred between 1994 and 1995, with the maximum number of medications, a physician could list increasing from five to six. We capped the number of medications at five for the 1995 and 1996 data for consistency, which resulted in the number of medications ordered or continued to remain quite steady over the 4-year period. HMO and TFFS patients received slightly fewer medications than the sample average. Again, HMOS may change the types of medication ordered. In particular, they may substitute generic for brand name drugs, or less costly for more costly prescriptions. The data do not allow us to address this possibility. Medicaid patients received slightly more medications, and Medicare patients many more medications, than the sample average.

Finally, asking a patient to return at a specified time is the prototypical example of demand induction. Physicians paid capitation rates or salaries would prefer that their
patients returned seldom, while those paid fee-for-service rates would like them to visit often. A return visit was specified for over 60% of all visits. The share was slightly lower for HMO patients and higher for Medicare patients.

For each outcome variable, we run successive analyses examining the effects of individual insurance and practice composition. We then run models that also include interaction terms.

The fully saturated model with interactions is

\[ Y_i = \alpha X_i + \sum \beta_j \text{insurance}_{ji} + \sum \delta_j \text{share}_{j,i} \times \sum \text{insurance}_{j,i} + \phi \text{share}_{j,i} \times \text{TFFS}_i + \epsilon_i \]

where \( i \) indexes patients, and \( j \) indexes types of insurance (HMO, Medicaid, Medicare). \( X_i \) is a vector of individual and practice characteristics (demographic, regional, specialty, and diagnosis related), \( \text{insurance}_{ji} \) a vector of dummies for insurance status (HMO, Medicaid, Medicare) for the individual patient, \( \text{share}_{j,i} \) is the share of visits in the practice visited by patient \( i \) paid by each of these payers.

We perform linear regressions for analysis of visit duration, number of tests ordered and the number of medications ordered, and logistic regression for whether a return visit was specified.\(^7\) In each case, we estimate Huber–White standard errors. Finally, we adjust the standard errors for intra-practice correlation. This adjustment more than doubles the standard error on coefficients associated with individual insurance status. We adjusted all analyses using population weights so that results reflect the experience of a random sample of patients throughout the country. Initially, we examined the behavior of generalists separately from the behavior of specialists. Coefficients on variables of interest were very similar. We therefore grouped them together (controlling for specialty effects) in the analyses reported here.

3. Results

The bottom half of Table 1 describes the characteristics of the sample in terms of insurance. About one-quarter of visits in the sample were paid by HMOs.

To assess the distribution of patients, we examine payer characteristics across practices. We measure the characteristics of the average practice visited by a TFFS patient, an HMO patient, a Medicaid patient, and a Medicare patient. We report results in Table 2.

While HMO patients constituted one-fourth of the sample, they made up only 13% of the patients in the median practice. The average HMO patient visited a practice where nearly 60% of the patients were HMO members. For each type of insurance, the average patient visited a practice dominated by her type of coverage. HMO patients were more segregated than either Medicaid or Medicare patients, however.

Despite the high degree of segregation, the average non-HMO patient visited a practice that also treated at least some HMO patients. The median TFFS patient visited a practice at which about 5% of all patients were HMO paid. The share HMO was lower for Medicaid and

\(^7\) We repeated the estimates using negative binomial regression for tests and medications ordered. The results were qualitatively similar to those reported here. Since a negative binomial regression is much more computationally costly and more difficult to interpret, we used OLS regression instead.
Table 2
Practice composition by patient insurance status

<table>
<thead>
<tr>
<th>Patient insurance status</th>
<th>TFFS share</th>
<th>Medicaid share</th>
<th>Medicare share</th>
<th>HMO share</th>
<th>25th percentile HMO share</th>
<th>Median HMO share</th>
<th>75th percentile HMO share</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFFS</td>
<td>0.60</td>
<td>0.07</td>
<td>0.18</td>
<td>0.15</td>
<td>0</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>HMO</td>
<td>0.29</td>
<td>0.04</td>
<td>0.10</td>
<td>0.57</td>
<td>0.33</td>
<td>0.59</td>
<td>0.83</td>
</tr>
<tr>
<td>Medicaid</td>
<td>0.36</td>
<td>0.39</td>
<td>0.14</td>
<td>0.12</td>
<td>0</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Medicare</td>
<td>0.41</td>
<td>0.06</td>
<td>0.42</td>
<td>0.11</td>
<td>0</td>
<td>0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>Total</td>
<td>0.46</td>
<td>0.08</td>
<td>0.21</td>
<td>0.25</td>
<td>0</td>
<td>0.13</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Medicare patients. More than one-fourth of TFFS patients visited a practice at which over 20% of all patients were HMO paid. Similarly, the average HMO patient visited a practice at which over one-fourth of all patients were TFFS paid and a further 15% were publicly paid. Only 15% of all HMO patients visited practices that treated only HMO patients. These results confirm that there is substantial heterogeneity within physician practices.

3.1. Visit characteristics

Table 3 reports the results of analyses that examine the effect of patient insurance on treatment patterns. We report results for all independent variables (except 15 dummy variables for specialist type). In general, the results for all independent control variables are as expected. Older patients receive treatment of greater intensity than do younger patients. Their visits are longer, they get more tests, and more medications are ordered for them. However, a return visit is specified less frequently. Quadratic terms in age are also positive indicating that intensity increases faster at older ages. Black patients have shorter visits, more tests, and are more likely to be told to return than are white patients. Patients with more diagnoses are treated more intensively than those with fewer diagnoses. Generalists, including pediatrics, internal medicine, and general and family practitioners, have shorter but slightly more intensive visits than do generalists, but they are less likely to specify a return visit (not reported in Table 4).

Turning to the effect of health insurance on intensity, HMO visits appear to be slightly less intensive than TFFS visits. They average over 1 min shorter in duration and are less likely to have a return visit specified, but they are somewhat more likely to include tests and medications. After adding controls, Medicaid and Medicare visits are shorter than TFFS visits, but include more tests and medications and specify more return visits.

It may be that patients seen in HMO-dominated practices are healthier than those seen in TFFS-dominated practices. If selection of this type were an important explanation of differences between HMO and TFFS intensity, we would expect the results of these analyses to be sensitive to exclusion of observable measures of severity—here, diagnosis codes. We repeated the analyses excluding the dummy variables for diagnosis intensity. Controlling for diagnosis substantially increases $R^2$ in these analyses, but it has little or no effect on the HMO or other insurance coefficients. While this result by no means conclusively indicates that HMOs do not select healthier patients, it does suggest that selection at the patient level is unlikely to fully explain observed intensity differences.
Table 3
Effect of individual insurance status on treatment

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Duration</th>
<th>Count of tests</th>
<th>Medications ordered</th>
<th>Return specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-practice correlation</td>
<td>16.04** (0.832)</td>
<td>0.271** (0.048)</td>
<td>-0.491** (0.077)</td>
<td>-2.005** (0.145)</td>
</tr>
<tr>
<td>Age</td>
<td>0.026** (0.010)</td>
<td>0.014** (0.00090)</td>
<td>0.00233* (0.0014)</td>
<td>-0.0076** (0.0026)</td>
</tr>
<tr>
<td>Age-squared</td>
<td>-0.00008 (0.00011)</td>
<td>-0.00012** (0.00001)</td>
<td>0.000003* (0.00002)</td>
<td>0.00014** (0.00003)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.081 (0.090)</td>
<td>0.0068 (0.0070)</td>
<td>-0.083** (0.011)</td>
<td>0.019 (0.021)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.668* (0.347)</td>
<td>-0.021 (0.020)</td>
<td>0.0078 (0.032)</td>
<td>0.131** (0.062)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.827** (0.314)</td>
<td>0.068** (0.026)</td>
<td>0.041 (0.036)</td>
<td>0.285** (0.065)</td>
</tr>
<tr>
<td>Other race</td>
<td>-0.576 (0.514)</td>
<td>0.076** (0.038)</td>
<td>-0.074* (0.043)</td>
<td>0.056 (0.094)</td>
</tr>
<tr>
<td>Year × region × metro</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Seen before</td>
<td>-4.477** (0.231)</td>
<td>-0.095** (0.013)</td>
<td>0.124** (0.020)</td>
<td>0.340** (0.036)</td>
</tr>
<tr>
<td>Number of diagnoses</td>
<td>1.091** (0.096)</td>
<td>0.079** (0.0081)</td>
<td>0.376** (0.013)</td>
<td>0.233** (0.020)</td>
</tr>
<tr>
<td>HMO</td>
<td>-0.926** (0.198)</td>
<td>0.018 (0.014)</td>
<td>0.016 (0.020)</td>
<td>-0.094** (0.037)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>-0.986** (0.333)</td>
<td>0.016 (0.022)</td>
<td>0.139** (0.033)</td>
<td>0.169** (0.068)</td>
</tr>
<tr>
<td>Medicare</td>
<td>-1.040** (0.200)</td>
<td>0.025* (0.013)</td>
<td>0.084** (0.027)</td>
<td>0.095** (0.044)</td>
</tr>
<tr>
<td>15 specifications</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Diagnosis codes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.221</td>
<td>0.390</td>
<td>0.248</td>
<td>0.164</td>
</tr>
<tr>
<td>No. of observations</td>
<td>100195</td>
<td>109895</td>
<td>109895</td>
<td>109895</td>
</tr>
</tbody>
</table>

* Standard errors in parentheses. Data are from 1993 to 1996 National Ambulatory Medical Care Survey. Standard errors are corrected for clustering and heteroskedasticity. Results are weighted by NAMCS sample weights. Duration, tests, and medications are OLS. Return is logistic regression.

* $P < 0.05$.

** $P < 0.01$. 
Table 4
Effect of practice composition on treatment

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Duration</th>
<th>Count of tests</th>
<th>Medications ordered</th>
<th>Return specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMO</td>
<td>−0.161 (0.104)</td>
<td>−0.0055 (0.0091)</td>
<td>−0.028** (0.014)</td>
<td>−0.032 (0.029)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>−0.490** (0.149)</td>
<td>−0.0043 (0.015)</td>
<td>0.086** (0.023)</td>
<td>0.130** (0.044)</td>
</tr>
<tr>
<td>Medicare</td>
<td>−0.662** (0.164)</td>
<td>0.018 (0.012)</td>
<td>0.056** (0.023)</td>
<td>0.027 (0.040)</td>
</tr>
<tr>
<td>Practice HMO (%)</td>
<td>−2.724** (0.525)</td>
<td>0.081** (0.034)</td>
<td>0.171** (0.052)</td>
<td>−0.063 (0.094)</td>
</tr>
<tr>
<td>Practice Medicaid (%)</td>
<td>−2.297** (0.964)</td>
<td>0.088 (0.058)</td>
<td>0.225** (0.084)</td>
<td>0.131 (0.200)</td>
</tr>
<tr>
<td>Practice Medicare (%)</td>
<td>−2.997** (0.770)</td>
<td>0.059 (0.045)</td>
<td>0.204** (0.084)</td>
<td>0.508** (0.138)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.224</td>
<td>0.390</td>
<td>0.249</td>
<td>0.165</td>
</tr>
<tr>
<td>No. of observations</td>
<td>100195</td>
<td>109895</td>
<td>109895</td>
<td>109895</td>
</tr>
</tbody>
</table>

*a Standard errors in parentheses. Data are from 1993 to 1996 National Ambulatory Medical Care Survey. Analyses also controls for all variables in Table 3 and for diagnosis. Standard errors are corrected for clustering and heteroskedasticity. Results are weighted by NAMCS sample weights. Duration, tests, and medications are OLS. Return is logistic regression.

** $P < 0.01$. 
3.2. Practice characteristics

The next set of analyses control for patient insurance and practice composition. Practice composition has a large and statistically significant effect on most measures of practice intensity (see Table 5). In each case, F-tests for the group of practice composition variables strongly reject the null hypothesis that practice composition does not affect outcomes.

Visits in TFFS-dominated practices are longer than are those in public-dominated practices or HMO-dominated practices. The effects are sizable. Moving from a practice without any HMO patients (the 25th percentile of practices) to one with 50% HMO patients (slightly above the 75th percentile) reduces visit length by 1.3 min, or 7%. Once practice composition is held constant, visits by HMO patients are no longer any shorter than are those of TFFS patients.\(^8\) Increased Medicaid and Medicare practice shares also reduce the length of visits. In addition, Medicaid and Medicare patients have shorter visits after controlling for practice composition.

Results for tests ordered and medications prescribed are quite similar. Practices with large fractions of Medicaid, Medicare, or HMO patients prescribe significantly more medications than do other practices, and practices with large HMO shares also order significantly more tests. The effect size for moving from a practice at the 25th percentile to one at the 75th percentile of HMO share is just over 4% of the mean for tests ordered and 8% for medications prescribed.

After controlling for practice composition, HMO-paid visits are less likely to include medications than are TFFS visits, while Medicaid and Medicare paid visits are more likely to include medications. This suggests that medication prescription may be a sign of the intensity of a visit. Tests no longer appear to vary by insurance status.

The pattern for return visits is somewhat different. At the practice level, HMO-dominated practices do not differ from other practices in their return policies, while Medicare-dominated practices have more return visits specified than other practices. Medicaid patients are more likely to be asked to return than other patients, after controlling for practice characteristics.

3.3. Interactions between practice and payer

Our hypotheses have implications for the interaction between visit and practice characteristics. In Table 5, we report interaction results for patients of a given insurance type × percentage of patients in that practice who have the same type of insurance (e.g. HMO patient × percentage HMO in practice), and for TFFS patients for each measure of practice composition (e.g. TFFS patients × percentage HMO in practice). All other interactions are included in the regression but not reported here. Note that sample sizes for some of the interaction cells are quite small (e.g. there are very few Medicaid patients in predominantly HMO practices).

The difference between HMO and TFFS coverage and practice composition interaction effects in the visit duration equation is small (and statistically insignificant) suggesting that

---

\(^8\) One reason for this finding might be a tendency for physicians to report a constant length of visit for all patients. We repeated the analysis omitting all practices where the reported visit length was the same for all patients. The basic results were not affected. For the sample as a whole, over 40% of the variation in duration is within practices.
Table 5
Effect of practice composition on treatment with interactions

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Duration</th>
<th>Count of tests</th>
<th>Medications ordered</th>
<th>Return specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMO</td>
<td>−0.333 (0.531)</td>
<td>0.016 (0.039)</td>
<td>−0.150** (0.054)</td>
<td>−0.012 (0.106)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>0.076 (0.661)</td>
<td>−0.074* (0.43)</td>
<td>−0.039 (0.066)</td>
<td>−0.148 (0.152)</td>
</tr>
<tr>
<td>Medicare</td>
<td>−0.821 (0.531)</td>
<td>0.025 (0.033)</td>
<td>−0.184** (0.064)</td>
<td>−0.392** (0.104)</td>
</tr>
<tr>
<td>HMO × share HMO</td>
<td>−2.652** (0.732)</td>
<td>0.071 (0.054)</td>
<td>0.273** (0.082)</td>
<td>−0.188 (0.141)</td>
</tr>
<tr>
<td>FFS × share HMO</td>
<td>−2.835** (0.721)</td>
<td>0.052 (0.048)</td>
<td>0.018 (0.060)</td>
<td>−0.122 (0.131)</td>
</tr>
<tr>
<td>Medicaid × share Medicaid</td>
<td>−2.754* (1.410)</td>
<td>0.219 (0.093)</td>
<td>0.334 (0.135)</td>
<td>0.445 (0.346)</td>
</tr>
<tr>
<td>FFS × share Medicaid</td>
<td>−2.706** (1.217)</td>
<td>0.043 (0.073)</td>
<td>0.084 (0.105)</td>
<td>−0.244 (0.217)</td>
</tr>
<tr>
<td>Medicare × share Medicare</td>
<td>−2.427** (1.054)</td>
<td>0.067 (0.059)</td>
<td>0.508** (0.134)</td>
<td>1.110** (0.207)</td>
</tr>
<tr>
<td>FFS × share Medicare</td>
<td>−3.038** (0.903)</td>
<td>0.074 (0.052)</td>
<td>0.039 (0.086)</td>
<td>0.287** (0.155)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.225</td>
<td>0.391</td>
<td>0.250</td>
<td>0.165</td>
</tr>
</tbody>
</table>

No. of observations: 100,195

* Standard errors in parentheses. Data are from 1993–1996 National Ambulatory Medical Care Survey. Analyses also controls for all variables in Table 3, for all remaining interaction terms (e.g. Medicaid × HMO, %), and for diagnosis. Standard errors are corrected for clustering and heteroskedasticity. Results are weighted by NAMCS sample weights. Duration, tests, and medications are OLS. Return is logistic regression.

* $P < 0.05$

** $P < 0.01$
Visit length is set at the practice level and does not vary by insurance status. The inability of physicians to vary individual visit duration may be partly due to the scheduling difficulties associated with variable visit lengths. These results provide strong evidence for a model of physician behavior with fixed costs in the duration of a visit.

Visit duration decreases as HMO penetration (or Medicaid or Medicare penetration) increases for both HMO (Medicaid, Medicare) and TFFS patients. The rate of decline is similar for all payer types, consistent with strong fixed costs in duration. For medications, most of the effect of being in an HMO-dominated practice accrues to HMO patients. The number of medications prescribed to TFFS patients is increasing, but only slightly, and insignificantly. This pattern suggests that the convergence in treatment intensity for medications occurs because HMO patients are treated much more intensively in HMO-dominated practices while TFFS patients are treated only slightly more intensively. The effects for tests and return visits are very imprecisely measured. Tests results and return visits show no significant pattern in HMO penetration for either HMO or TFFS patients. The number of tests is increasing for both types of patients and the return visit rate is decreasing for both types of patients. Medicare patients are more likely to be asked to return in Medicare practices; to a lesser extent, so are TFFS patients who visit those practices. Overall, the empirical evidence appears to support the fixed cost hypothesis of treatment, where duration is fixed and medications are variable.

3.4. Sensitivity analyses

Many practices that treat only managed care patients are located in staff or group model HMOs, which may be systematically different from other practice settings. In that case, the estimated effect of increased managed care penetration may just be a proxy for whether the practice is a staff or group model HMO. To test this possibility, we repeated the analyses using discrete measures of HMO penetration at the practice level in place of the continuous variable reported. We categorize practices by their HMO share: none (0%), some (between 0 and 49.9%), most (between 50 and 95%), and all (greater than 95%). For duration, where practice level effects are strongest, we find that the effects of managed care penetration at the practice level increase monotonically and are nearly linear up to 100% (each dummy is statistically significant). There is an additional large and significant effect of being a >95% HMO practice. For medications, the effects also increase monotonically (but not all the dummies are significant) and there is a large and significant additional effect for being in a >95% HMO practice. These results suggest that there is a substantial effect of increased managed care penetration on behavior even in practices that are not staff or group model HMOs. For tests, the effects of HMO practice level are the same for both practices with most (significant) and those with all managed care patients. There is no effect of practice composition on the specification of a return visit at any penetration level.

One important limitation of our results is that physicians themselves choose their level of participation in each of the insurance programs considered here. Physicians whose practice style, ceteris paribus, most closely resemble the HMO model are those most likely to accept HMO patients. We cannot directly correct for this potential selection. In order to assess the potential importance of selection, we conducted an analysis using information on managed care penetration by region. We expect selection effects to be stronger at times and in areas
where penetration is low. Thus, if selection is very important, we would expect to find that practice level effects are weaker in areas of high managed care penetration.

As a crude measure of managed care penetration, we used NAMCS data on the sample year, region of physician’s practice, and whether the practice was in a metropolitan area or not. We limited the sample to the top and bottom quartile of managed care penetration by year × region × metropolitan area cells. In the top quartile, managed care penetration averaged 30%. In the bottom quartile, managed care penetration averaged 6%. We repeated the analyses above including terms for interactions between HMO, practice share HMO, and TFFS × practice HMO with being in a high penetration cell. 9

We find that the duration effects measured above are much stronger in high penetration areas, suggesting that shorter HMO visits are not entirely a consequence of physicians with short visits choosing to join HMOs. There are no significant interaction effects for tests or return visit specified. Medication effects are stronger in low penetration areas than in high penetration areas, but both areas show significant effects of HMO payer share on medications.

A second implication of the pure physician selection story is that effort should vary only with practice type and interaction effects should be uniformly zero. In fact, we find that medication prescribing patterns vary within a practice, as well as between practices. Physician selection alone cannot explain this finding.

4. Conclusions

Most physicians treat a variety of patients within their practices where a variety of insurance arrangements exist. While economic theory suggests that, at an individual level, managed care patients will receive less intense treatment than fee-for-service patients, the ability of physicians to modify their behavior when the patient population is heterogeneous may be severely limited. This paper develops several theoretical models of physician effort intensity in a mixed payment environment. When managed care organizations pay marginal costs for excess physician capacity, increases in managed care penetration lead to a convergence in effort levels to each patient within a practice. When physicians induce the demand of fee-for-service patients to compensate for low payments from managed care, expansion of managed care at the practice level leads to a divergence in effort levels. When there are both fixed and variable cost components of effort, the fixed cost components are constant across patients within a practice and decrease as managed care penetration increases. Variable cost components increase as managed care penetration increases, implying either a convergence or divergence of intensities, depending on the relative rates of increase.

Data from the National Ambulatory Medical Care Survey are employed to test the theoretical hypotheses. We find that financial incentives do matter. Treatment intensity varies between patients, according to the method by which their insurers pay physicians. But variations in treatment intensity depend as much—or more—on the insurance status of other patients treated by the same physician as they do on the insurance status of an individual patient.

9 Note that we cannot conduct this analysis for Medicaid and Medicare because there is less variation in penetration by these programs.
The analysis suggests that physicians do not make up losses in income from their HMO practice by increasing the intensity with which they practice TFFS medicine. Nor does it support a model where managed care pays physicians marginal cost for excess capacity in the practice. Instead, doctors with large HMO caseloads seem to change the way they do business altogether. Physicians who treat mostly HMO patients appear to adopt a practice style that offers equivalent treatment intensity along most measurable dimensions. Visit duration appears to be constant across patients within a practice, while medications prescribed appears to be converging as managed care penetration increases. Together, these results are most consistent with a fixed cost hypothesis of treatment in a heterogeneous practice.

Our controls for the effects of patient composition at the market level are imperfect, suggesting some regional effects may, in part, be captured in our practice level measures. Further, our ability to control for physician selection into networks that best match their practice style is limited. While neither of these concerns is sufficient to explain the patterns described by the data, the empirical results derived here should be interpreted with some caution.

The differences in practice patterns do have important implications for our understanding of the physician marketplace. If it is difficult to efficiently respond to HMO incentives in non-HMO practices (and vice versa), practices are likely to become increasingly specialized in the treatment of one or another group of patients. In turn, such segregation is likely to lead to increased vertical integration in the physician services market.

Acknowledgements

The authors would like to thank David Cutler, two anonymous referees and seminar participants at the New York Federal Reserve Bank, Cornell University, Columbia University, and the Mount Sinai School of Medicine for many helpful comments.

References