Reprinted with Permission
MetroDoctors Journal
September/October 2008

A Conversation With
David Satin, M.D.

David J. Satin, M.D. received his medical degree from the University of Western Ontario, London, Canada and participated in a medical education student fellowship at the University of Western Ontario, Faculty of Medicine and Dentistry, London, Canada. He completed a residency in family medicine at the University of Minnesota Medical Center (Smiley's) Residency Program, Minneapolis and is currently a postdoctoral fellow in Bioethics at the University of Minnesota, Center for Bioethics. He is board certified by the American Board of Family Medicine; and is a fellow of the Royal College of Physicians and Surgeons of Canada. He is an assistant professor, Department of Family Medicine and Community Health, University of Minnesota, University of Minnesota (Smiley's) Residency Program and is a Compliance Officer and Risk Management Leader for the Department of Family Medicine and Community Health, University of Minnesota Physicians, Minneapolis. Dr. Satin also serves as a Family Physician/Ethicist, on the American Medical Association Physician consortium for quality improvement: Geriatric Pay for Performance Work Group. He developed Pay for Performance measures for submission to the Centers for Medicare and Medicaid, National Committee, coordinated through AMA headquarters, Chicago. He is a member of the Minnesota Medical Association and Minnesota Council of Health Plans P4P Measurement Alignment Work Group.

Questions were provided by Drs. Peter Bornstein, Ronnell Hansen and Robert Geist; David Allen and Becky Schierman.

Pay for performance (P4P) programs often encourage reimbursement for hitting clinical targets based on process or outcome measures. If possible, what metrics of ethical behavior could be included in P4P programs?

I would not recommend any P4P metrics of ethical behavior. Judging behavior as ethical or unethical requires knowing many contextual factors that cannot be accounted for in even the most sophisticated P4P system. Ultimately, it is only the moral agents themselves that can truly know if their behavior is ethical. We may ask for metrics that might reflect patients' perception of a given clinician's professionalism, but that becomes more of a patient satisfaction survey. Indeed, many P4P programs are including patient satisfaction surveys as a factor for determining clinician reimbursement. But again, this is not a metric of ethical behavior. Other metrics that try to approximate ethical behavior are those that measure compliance with laws and institutional policies. These are interesting in that they help us understand how ethics, law and policy are distinct yet related concepts.

On a personal level, there is something perverse about the notion of an economic incentive for ethical behavior. Nevertheless, one can view the entire enterprise of striving for better patient outcomes in moral terms. In this light, all clinical P4P measures are measures of ethical behavior. I would caution us against viewing P4P in this light for reasons I began with above.

P4P programs are predicated on the notion that better outcomes should be rewarded with better pay. What are the ethical conflicts physicians face when evaluating whether to participate in such programs?

The premise of this question, that "P4P programs are predicated on the notion that better outcomes should be rewarded with better pay" is not universally accepted. There is a competing view that "P4P programs fund successful quality improvement projects." This is a subtle yet critical point. If one views P4P as personal financial reward for better outcomes, then one is more likely to take the payers judgment personally, as if P4P is measuring ethical behavior as described in question #1. In contrast, if one views P4P as a reflection of your quality improvement projects, positive or negative judgments become objective benchmarks for continuous quality improvement. In reality, the working premise of P4P programs is often a mixture of these two distinct views.

Now to answer the question, the most salient ethical conflicts physicians face under P4P are conflicts of interest. Should you recommend that your 50-year-old patient with pancreatic cancer get a screening colonoscopy? Should you terminate, dismiss, or fire a diabetic patient who fails to adhere to his or her medication regimen? Should you refuse new diabetic patients who smoke? Facilitating non-coerced, informed consent will be a greater challenge when your P4P bonus rides on what

(Continued on page 12)

(Continued from page 11)

your patient chooses. There is a finer line than one might think between teaching with a keen awareness of what is on the test and teaching to the test. P4P programs are not yet sophisticated enough to be completely fair. The great moral challenge for clinicians under P4P is to accept the occasional bad P4P outcome when you know it's the right thing for your patient.

P4P criteria tend to measure clinical statistics that are believed to make a difference in the average health of a population. But, to an individual, quality probably has much more to do with whether the physician really listened and understood, whether the physician gave actionable treatment guidance, and such things as affordability, convenience, service, etc. If P4P criteria are at significant divergence to what individuals seek in terms of quality, then how can P4P ever lead to quality that individual's value?

When I travel by commercial airliner, I assume that the airline is diligent and that the pilot is sufficiently skilled to get me to my destination safely. I rate the quality of an airline according to their timeliness, friendliness, and whether I got peanuts on the flight. Now what if I were to discover that airlines actually do differ significantly with regard to safety? That is, for better or for worse, what patients are starting to discover about health systems.

For years, patients have assumed that their clinics are diligent and that their doctors are sufficiently skilled. The public reporting of outcomes is revealing that there are differences in the more important criteria we took for granted. The current data tells us that patients will typically not change doctors in response to the outcomes. Indeed, they say quite sensible things like, "Well that doctor must have sicker patients!" But the current data also tells us that at least some patients choosing a doctor for the first time will take the outcome data into account. Finally, P4P programs are increasingly taking into account the items listed in the question such as affordability, convenience, service, etc.

The Massachusetts health plan while reducing the rate of uninsured by half (14 percent to 7 percent), has had the state medical society file a lawsuit to block or change a ranking program it says harms doctors and patients. Are such P4P methods really useful in actually evaluating and improving systems versus their overhead cost and complexity?

First, I think it's great that the medical society stood up for its doctors who felt they were being unfairly rated. I have heard story after story from clinicians whose "efficiency" rating was poor. Upon further investigation, "efficiency" often means "how much you cost in comparison to your peers." I have not yet seen a program that calculates efficiency in terms of cost per outcome, taking into account the baseline illness and complexity of the patient population. But most importantly, what your efficiency means according to a given health plan is typically not conveyed to patients in a fair and transparent fashion.

That said, the question asks, "Are such P4P methods really useful in actually evaluating and improving systems versus the overhead cost and complexity?" My answer is, "I don't know, but I know that it is not known." That is, *nobody* knows yet. I do know that P4P programs (not necessarily ranking programs) have been shown to improve intermediate level markers of health such as blood pressure, blood sugar, daily aspirin use, smoking cessation, and lipid levels in patients with Type 2 diabetes. It has not demonstrated similar success with hospital based end-points such as repeat myocardial infarction and all cause mortality. So the jury is out on even the benefits side of the equation. Much of my work concerns the burden side of the P4P equation. What we know about the benefits is 1,000 fold what we know about the burdens. So it's too early to sensibly predict if the benefits are worth the burdens. Given the international scope of P4P, I do predict that some form of P4P will have net worth. I see the United States' 150-plus P4P programs as a giant experiment. Unfortunately, we didn't all consent to be research subjects in this great experiment.

Is Minnesota ready to link the incentives from P4P more directly to the physicians whose behavior is being measured (individual physician level measures)? What are the ethical concerns when moving toward this level of granularity?

Personally, I think this would be a mistake, especially for public reporting. There are too many unsettled questions before we are ready to purport we can judge the quality of a physician's care based on these measures. Imagine treating this as a research question. Randomizing patient populations to negate confounding variables would only get us halfway there. We would still be left to answer questions like, "are these measures representative of this clinician's overall care?" and "What is the ideal rate of Chlamydia screening given that this is ultimately a patient choice?" Different styles of doctoring will get different results. As someone who participates in both medical school and residency admissions, I can say with confidence that even physicians are not unanimous on what constitutes the ideal style of doctoring. Moreover, different styles of doctoring can be a good thing given the heterogeneity of patient styles.

That said, there is certainly a role for individual physician level measures. I think they can tell us who the outliers are and from a quality improvement perspective, that can be helpful. I enjoy seeing where I rank on various measures, knowing that perhaps I ought to pay more attention to whether or not I prescribe aspirin to my patients with diabetes. I enjoy it more when the rankings are private because I know how hard it is to resist seeing a rank list as anything other than a ranking of how good a doctor you are. The science of this kind of individual level measurement is too young. Its shoulders are not yet broad enough for us to stand on and dole out bonuses or rank clinician quality. So my ethical concerns here turn out to be primarily pragmatic concerns.

How can we improve data collection methods to account for noncompliant patients and to eliminate confounding variables?

Britain has two fairly elegant solutions. I describe them in the April 2006 issue of *Minnesota Medicine.* www.minnesotamedicine.com/PastIssues/April2006/CommentaryApril2006/tabid/2386/Default.aspx.

In brief, Britain risk adjusts and allows for specific exceptions. Their risk adjustment formula is based on the average household income of your clinic's postal code. With economic status as the greatest single predictor of outcomes, the British decided that some adjustment was necessary in order to be fair to clinics serving poorer neighborhoods. Britain also allows for specific exceptions. For example, one P4P measure of quality is to perform an annual in-person medication review with each patient taking psychiatric medications. But rather than losing out if you serve a transient population, British P4P allows for an alternative process measure of due diligence. Clinicians attempting to recall their patients to clinic via registered mail and telephone, demonstrating a quality system, earn the same bonus afforded to their colleagues lucky enough to have their patients show up.

In addition to these suggestions for improving quality and fairness, New Zealand has some tricks we can learn. These include bonuses for "case finding" of patients who have not had a recommended screening test such as a PAP smear within five years.

There will always be a trade off between maximal fairness and minimal complexity in a P4P system. Nevertheless, we need only look abroad and at one another's P4P programs to see great innovations available to us all.

As care becomes more coordinated between primary care and specialty care, how will P4P programs determine to which provider the attribution of patient outcomes goes? Will they pay incentives to both providers for improved outcomes?

Let's begin with what has not worked. If one attempts to assign responsibility for a patient outcome to the clinician who has seen that patient most frequently within a year, oftentimes you get unintuitive and inappropriate assignments, such as an ophthalmologist being responsible for a diabetic patient's A1c and cholesterol. If one assigns responsibility based on primary clinic designation, as is typically the case, then you get so-called "invisible" patients whose outcomes are the responsibility of a doctor whom they have never seen.

One solution that has been proposed is a quasi capitated system in which clinicians are reimbursed per-patient per-month (PMPM) to be responsible for that patient's outcomes. That is, clinicians are reimbursed on a PMPM basis, over and above any fee-for-service and P4P arrangements. This small degree of capitation is meant to reimburse clinicians for the added administrative expenses required to reach out to the "invisible" patients. Additionally, the PMPM capitation reimburses for the added difficulty of caring for patients who rarely attend the clinic and who are cared for frequently in non-traditional ways such as phone calls with doctors and nurses.

Speaking specifically to the issue of primary care and subspecialty care, the Centers for Medicare and Medicaid, as well as many private P4P programs, have designated certain measures to be specialty specific. Once such a designation is made, I believe the PMPM capitated approach may or may not be appropriate depending upon the circumstances of the measure and the specialty. I am open to the possibility of a split bonus between two specialties, but I am not aware of a model for this complex arrangement.

There is not transparency in the formula used to determine tiering decisions, and in the relative weight of quality and cost more importance is placed on cost than quality. Should we seek to separate cost and quality measures to more easily allow assessment of the clinical outcomes by patients?

Yes—see Question #4.

Nevertheless, we should consider that there may be a legitimate role for an additional measure of "efficiency" that recognizes clinicians who achieve similar outcomes as other clinicians with similar patient populations, but at half the cost.

Does the new Minnesota health care reform statute mean that the "medical home" will become a corporate gatekeeper house for carve out "packages" of services (e.g., for all orthopedic services including hospital services) with fixed annualized capitated prices?

I don't know.

As an educator, I think it's important that my students occasionally hear me say, "I don't know." Someone who never says, "I don't know" can't be trusted because no one knows everything. The unfortunate consequence of appearing to know everything is that bright students will eventually treat everything you say as equally suspect. A good teacher and a good physician, since all physicians are teachers, occasionally says, "I don't know, but I'll find out."

So check in with me again. I'll find out. ✦