Rational Noncompliance with Prescribed Medical Treatment

Douglas O. Stewart
Joseph P. DeMarco

Kennedy Institute of Ethics Journal, Volume 20, Number 3, September 2010, pp. 277-290 (Article)

Published by The Johns Hopkins University Press

For additional information about this article
http://muse.jhu.edu/journals/ken/summary/v020/20.3.stewart.html
ABSTRACT. Despite the attention that patient noncompliance has received from medical researchers, patient noncompliance remains poorly understood and difficult to alter. With a better theory of patient noncompliance, both greater success in achieving compliance and greater respect for patient decision making are likely. The theory presented, which uses a microeconomic approach, bridges a gap in the extant literature that has so far ignored the contributions of this classic perspective on decision making involving the tradeoff of costs and benefits. The model also generates a surprising conclusion: that patients are typically acting rationally when they refuse to comply with certain treatments. However, compliance is predicted to rise with increased benefits and reduced costs. The prediction that noncompliance is rational is especially true in chronic conditions at the point that treatment begins to move closer to the medically ideal treatment level. Although the details of this theory have not been tested empirically, it is well supported by existing prospective and retrospective studies.

Patient noncompliance with physician prescriptions, especially in nonsymptomatic chronic diseases, is frequently characterized in the literature as harmful and economically costly (Miller 1997). Nancy Houston Miller views patient noncompliance as harmful because noncompliance can result in continued or new health problems leading to hospital admissions. Further, she places the annual monetary cost of noncompliance at $100 billion.

Patient noncompliance with prescribed treatment is considered the least understood form of health behavior (Coons 2001). Despite the plethora of attention in journal articles, the issue of noncompliance remains unresolved even in terms of understanding the main correlates to noncompliance, such as age, sex, and race, which may be used as predictors of noncompliance (Vermeire et al. 2001; Lassen 1998; Bissella, May, and Noyce 2004).
The usual response when patient noncompliance is observed is a call for better physician communication (Goldberg, Cohen, and Rubin 1998). However, without better understanding of when and why patient noncompliance exists, improved communication is unlikely to have more than a minimal effect (Roter et al. 1998). Lack of understanding of noncompliance has filtered into physician perceptions, where there is little correlation between belief about noncompliance and actual noncompliance among individual patients (Phillips and Jones 1991). A recent editorial even suggests the need for a paradigm shift away from “Me doctor, you patient, take pill” to one in which there is recognition that doctors and patients are human, imperfect, and fallible (Powsner and Spitzer 2003).

Missing from the literature on noncompliance is a theory or model that can be used to understand more fully why patients are noncompliant, and the extent to which greater compliance is desirable. If greater compliance is found to be desirable, then the theory should help to understand what it may take to gain greater compliance (Miller 1997; Vermeire et al. 2001). Such a theory could also indicate the extent to which noncompliance might be accepted as fully reasonable (Donovan and Blake 1992).

Given the attention that noncompliance has received, it is surprising that a microeconomic approach to analysis of noncompliance has not been attempted. Patient decisions about medical treatment involve an evaluation of its costs and benefits, just as all other economic transactions do. In many cases, medical care decisions may involve disability or even death, so the importance of these decisions means not that economic analysis is irrelevant, but rather that the benefits and costs involved may be greater than the purely monetary considerations involved in typical economic decisions. Microeconomics includes a highly respected and fruitful theory of individual decision making. Hence, we propose the use of microeconomic analysis when examining patient decision making and noncompliance.

The microeconomic analysis we propose goes beyond the basic textbook approach by construing costs and benefits in very broad terms. “Benefits” are not simply monetary payments received, and “costs” are not simply monetary outlays. Rather “benefits” include other sources of “utility” as well, and “costs” also include other sources of “disutility.” Sources of utility perceived by a patient include such benefits as feeling better, sleeping better, and added income, if more work is possible. Sources of disutility might include adverse side effects of treatment, discomfort from adhering to a strict treatment regimen, out-of-pocket treatment expenses, and reduced income, if less work is possible.
The results of using microeconomic analysis may be surprising because the analysis shows that patient refusal of some medical care is generally rational, not foolish, and thereby is worthy of full respect. Actually, we show that acceptance by the patient of medically ideal treatment is more likely to indicate a foolish decision than does the rejection of such treatment.

**THEORY OF PATIENT DECISION MAKING REGARDING TREATMENT**

Our theory of patient decision making regarding treatment is as follows: Patients will comply with treatment instructions if and only if the perceived marginal benefit of treatment (MTB) is greater than or equal to the perceived marginal cost of treatment (MTC).\(^3\) “Perceived marginal benefit” is the increase in well-being of the patient from additional treatment when the increase in well-being is evaluated by the patient. “Perceived marginal cost” is the increased cost the patient bears from additional treatment when the increased cost is evaluated by the patient.

We use Type 2 diabetes and its treatment to illustrate the workings of the theory in the context of a patient’s choice of a goal for treatment. We selected Type 2 diabetes for the illustration because nearly everyone has heard of the disease, many people have knowledge of the criteria for diagnosing the disease and for measuring the success of treatment, and there is widespread attention presently directed at the disease. For many patients with Type 2 diabetes, treatment involves attention to diet and exercise, perhaps supplemented by oral medication, and insulin is not necessary. (Hypertension, or high blood pressure, is another disease that meets the criteria to illustrate our theory.)

We define a “medically ideal treatment level” as the level of treatment at which all symptoms of the diagnosed disease are controlled.\(^4\) We also will assume that the medically ideal treatment level for diabetes is the level that yields a hemoglobin A1c (HbA1c) level of 6 percent as the percentage of glycated hemoglobin in the patient’s blood. The ideal result of 6 percent on the HbA1c test is not chosen arbitrarily. In 2009 the International Expert Committee, formed by the American Diabetes Association, the International Diabetes Federation, and the European Association for the Study of Diabetes, recommended 6.5 percent as the minimum glycated hemoglobin level at which diabetes should be diagnosed—i.e., a patient should be diagnosed with Type 2 diabetes if the HbA1c test result is 6.5 percent or greater (International Expert Committee 2009). The International Expert Committee (2009, p. 1332) further stated that “individuals
with an A1c level [greater than or equal to] 6 percent but [less than] 6.5 percent are likely at the highest risk for progression to diabetes.”

Suppose a patient presents with a HbA1c level of 10 percent. The patient’s physician may propose a course of treatment designed to lower the patient’s HbA1c level to 6 percent. The question we seek to answer is whether it is rational for the patient to select a HbA1c level of 6 percent as the treatment goal. The answer typically is “no.”

The standard microeconomic theory of consumer choice includes the assumption that consumers make decisions so as to maximize the net benefit derived from their choices. The value of net benefit is the value of total benefits resulting from the consumer’s decision less the value of the total costs borne by the consumer as a result of the decision. The decision-making process for consumers involves weighing benefits versus costs as they are valued by the consumer. In the present context, costs to the patient are such treatment costs for diabetes as the value of time and money spent in treatment, the value of discomfort suffered during treatment, and the value of any treatment complications or side effects suffered by the patient. Benefits are the value of expected reductions in both long- and short-term symptoms of the disease. We assume here that patients have perfect information about all relevant benefit and cost relationships.

We assume that as treatment reduces the patient’s HbA1c level close to the medically ideal treatment level, the cost of a further reduction in the level is relatively high, and the benefit derived from a further reduction is relatively low. After weighing treatment benefits versus treatment costs, the conclusion is that a patient might reasonably not select the medically ideal treatment level as his or her treatment goal. The conclusion follows because at a certain treatment level, the patient faces treatment costs that exceed the benefits to be derived from additional treatment.

If the medically ideal treatment level is not a rational treatment goal for the patient to pursue, then perhaps the patient should continue treatment to lower his or her HbA1c level just as long as the total treatment costs are less than total treatment benefits. Under this criterion, the stopping point would occur when the total costs equal the total benefits—i.e., when the net benefit is zero. This treatment level might seem rational because the patient seemingly does not lose by continuing to pursue treatment that is reducing symptoms and is physically beneficial. However rational it might seem, choosing this treatment level as the goal of treatment does not make good sense; the patient does lose compared to an alternative goal
for treatment, that which maximizes net benefit rather than that which makes net benefit equal to zero.

How can we identify the treatment level at which net benefit is maximized? Standard microeconomic analysis provides the answer: a patient can maximize net benefit by pursuing treatment until the marginal benefit of treatment equals the marginal cost of treatment (Mankiw 2004). The advantage of using microeconomics is that it makes clear how best to evaluate the relationship between benefits and costs when both are broadly defined.

Basically stated, when additional treatment produces greater additional benefit than additional cost, it should be pursued, and when, from additional treatment, marginal benefit is less than marginal cost, additional treatment should not be pursued. (We offer a fuller explanation of the theory in Stewart and DeMarco (2005).) The result that the patient chooses a level of treatment at which marginal benefit is equal to marginal cost is important because the rationally pursued level of treatment is lower, and might be substantially lower, than that required to achieve the medically ideal HbA1c level. As mentioned, we assume that the medical ideal for diabetes treatment is a HbA1c level of 6 percent. Treatment regimens are designed to help patients achieve that level. Based on the patient’s goal of maximizing net benefit, typically patients will not fully comply with treatment recommendations.

Reference to Figure 1 may help to illustrate why a rational patient pursues treatment until the marginal benefit of treatment equals the marginal cost of treatment. It also demonstrates that in typical cases the treatment level a rational patient pursues is below the medically ideal treatment level and yields a HbA1c level that is greater than the assumed medical ideal.

The curve labeled Total Treatment Benefit (TTB) in Figure 1 (a), represents the relationship between the value of total treatment benefits and the HbA1c level resulting from a certain treatment level. As previously described, benefits are the value to the patient of expected reductions in symptoms of the disease. The subscript $i$ in the labels for various curves in Figure 1 indicates that the benefits and costs of treatment vary from patient to patient as patient preferences and other patient factors change; the curves in Figure 1 are drawn for the $i$-th patient. The patient’s valuation of benefits and costs is measured in utility terms as indicated by “Value (utility)” on the vertical axis in both panels of Figure 1. Total treatment benefits are assumed to increase at a decreasing rate as the HbA1c level is reduced from the patient’s presentation level of 10 percent toward the
medically ideal level of 6 percent. The shape of the curve labeled TTB reflects this assumption. The medically ideal level for HbA1c, the level at which TTB is maximized and the patient exhibits no symptoms of diabetes, is labeled $G_i$ in Figure 1.

The curve labeled Total Treatment Cost (TTC) in Figure 1 (a) represents the relationship between the value of total treatment costs and the HbA1c
level resulting from a certain treatment level. As described earlier, costs to the patient are such treatment costs for diabetes as the value of time and money spent in treatment, the value of discomfort suffered during treatment, and the value of any treatment complications or side effects suffered by the patient. Total treatment costs are assumed to increase at an increasing rate as the HbA1c level is reduced from the patient’s presentation level of 10 percent toward the medically ideal level of 6 percent. The shape of the curve labeled TTC reflects this assumption.

The curve labeled Treatment Net Benefit (TNB) in Figure 1 (a), represents the relationship between the value of treatment net benefits and the HbA1c level resulting from a certain treatment level. The magnitude of treatment net benefits at any given HbA1c level is equal to total treatment benefits minus total treatment costs at that HbA1c level. Based on the shapes of the TTB and TTC relationships and beginning at the HbA1c level of the patient’s presentation level of 10 percent, TNB first increases and then decreases as the HbA1c level is reduced by additional treatment.

The curve labeled Marginal Treatment Benefit (MTB) in Figure 1 (b), represents the relationship between the addition to total treatment benefits per 1 unit decrease in the patient’s HbA1c level—i.e., MTB or the marginal treatment benefits—and the HbA1c level resulting from a certain treatment level. The shape of MTB reflects the assumption made concerning the relationship between TTB and the patient’s HbA1c level that total treatment benefits increase at a decreasing rate as the HbA1c level is reduced from the patient’s presentation level of 10 percent toward the medically ideal level. MTB is said to be positive but decreasing as the HbA1c level is reduced below 10 percent.

The curve labeled Marginal Treatment Cost (MTC) in Figure 1 (b), represents the relationship between the addition to total treatment costs per 1 unit decrease in the patient’s HbA1c level—i.e., MTC or the marginal treatment costs—and the HbA1c level resulting from a certain treatment level. The shape of MTC reflects the assumption made concerning the relationship between TTC and the patient’s HbA1c level that total treatment costs increase at an increasing rate as the HbA1c level is reduced from the patient’s presentation level of 10 percent toward the medically ideal treatment level. MTC is said to be positive and increasing as the HbA1c level is reduced below 10 percent.

Now consider the patient’s appropriate treatment goal while making reference to Figure 1. As shown in (b), at the patient’s presentation HbA1c level of 10 percent, instituting a modest treatment plan that lowers the
HbA1c level somewhat below 10 percent produces marginal treatment benefits greater than marginal treatment costs. As treatment intensifies, the additional treatment further lowers the HbA1c level, and with MTB greater than MTC, treatment net benefit, represented by TNB in (a), increases. At the treatment level yielding the HbA1c level at which MTB equals MTC, TNB is maximized. This HbA1c level is \( G \) in both (a) and (b) in Figure 1.

Below the HbA1c level \( G \), additional treatment decreases TNB because at HbA1c levels below \( G \), MTC is greater than MTB—i.e., the additional treatment adds more to cost than it does to benefit. At \( G_N \) the treatment net benefit of lowering the patient’s HbA1c level from its presentation level to \( G_N \) equals zero. The HbA1c level \( G_N \) is not rational.

The HbA1c level \( G_I \) in Figure 1 is the level we have termed “medically ideal” in that it is most likely to eliminate or maximally reduce all symptoms. Figure 1 shows that \( G_I \) is below the HbA1c level \( G \), at which MTB equals MTC and TNB is maximized. The medically ideal treatment level, \( G_I \), is not chosen by the patient as the goal for treatment. The treatment level \( G_I \) should be the optimum treatment level from the perspective of the patient.

THE THEORY IS NORMATIVE AND POSITIVE

Our theory shows that it is not rational for a patient to pursue further treatment whenever the marginal treatment costs are greater than the marginal treatment benefits. This claim is a normative one based on an evaluation of the patient’s best interests from the patient’s broad perspective when taking into account all costs and benefits associated with treatment. As a normative claim, the theory stands whether patients actually act in a way that equalizes marginal benefit and marginal cost.

Although the theory is normative, it is also positive and is empirically testable. Microeconomic theory in general has been successful at explaining and predicting actual consumer behavior. Nevertheless, such factors as those to be described make empirical testing of our theory difficult; empirical tests would need careful construction. We should note that even if our theory were falsified empirically, its normative aspect would remain compelling.

EMPIRICAL OBSERVATION AND THE THEORY

Casual observation of patient behavior is not sufficient to reject or support the theory. For example, a situation could be observed in which
a patient is not receiving the level of treatment that the observer would expect a rational patient to accept. A violation of one of the assumptions made in our theory could explain the discrepancy between theory and practice: (1) the patient may be acting irrationally or (2) the patient may be acting on false information about benefits and costs. Either factor could account for a patient not accepting the level of treatment at which the marginal treatment benefits equal marginal treatment costs. It would be a mistake to reject the normative requirement that medical treatment is rational whenever marginal benefit equals marginal cost simply because a given patient rejects a level of treatment from which he or she could derive a high level of marginal direct benefit. Such a case may fail to take into account the patient’s costs. That same patient may also have high marginal expense or some other cost, such as suffering related to treatment. Unless both sides of the equation are taken into account, a judgment that the observed behavior of the patient conflicts with our proposed theory is myopic. Conversely, a patient might accept additional treatment that an observer would see as conferring only limited marginal direct health benefits. In this case, the observer might think that the patient was not acting rationally, but has failed to account for the complete marginal treatment benefits for the patient or to recognize that the patient had correspondingly limited marginal treatment costs.

An empirical investigation could be designed to test whether information that exaggerates benefits typically increases the level of compliance. Also, those who exhibit unusual costs or benefits curves will act in a way that is nontypical, but predictable if departures from the usual nature of costs or benefits are known. We have not conducted such empirical studies, but extant studies offer strong support for our view.

There is considerable evidence that patient behavior is consistent with the main prediction of our theory: when perceived costs are high and/or perceived benefits are low, patients are likely not to comply fully with medical treatment recommendations. For example, evidence supporting our theory comes from the fact that interventions designed to increase compliance, such as telephone reminders, have low success rates and are unlikely to offset fully the noncompliance effect of a patient’s evaluation of the relationship between benefits and costs. A meta-analysis of 153 studies of interventions designed to improve compliance found that the typical correlation between attempts to induce compliance and actual (or at least perceived) compliance are low (Roter et al. 1998). The meta-analysis found that individuals show the greatest compliance with drug use instructions.
Prescription drug therapy is often the least costly treatment, at least if patients have full or partial prescription drug insurance.

Also well tested is the notion that patients tend not to comply with a treatment regimen when it becomes burdensome in relation to perceived benefits. For example, in a recent study, Benner and colleagues (2005) found that patients who perceived greater benefit from compliance with a prescription of statin therapy, as measured by the extent of reduction in low-density lipoprotein (LDL) cholesterol, adhered more closely to the prescription.

Furthermore, studies have been conducted to test the Health Belief Model, a sociological theory that incorporates perceived benefits and burdens. The model examines patient behavior in relation to four intuitively appealing factors: (1) Perceived susceptibility, a patient’s perception of vulnerability to an illness; (2) Perceived severity, a patient’s perception of the seriousness of an illness; (3) Perceived benefits, a patient’s perception of the effectiveness of various treatment options; and (4) Perceived barriers, a patient’s perception of the gamut of costs involved in treatment, such as side effects and inconvenience. A critical review of 46 prospective and retrospective studies of the model, including several related to diabetes, found a significant association between each of the four factors and patient behavior. When susceptibility, severity, and benefits are perceived to be high, compliance is greater, and when barriers are viewed as high, compliance is lower (Janz and Becker 1984).

Further support for our theory comes by way of a survey of 11 studies that question whether financial incentives increase compliance with medical treatment plans. Ten of the studies found that financial incentives do improve compliance (Giuffrida and Torgerson 1997). We are not arguing for financial incentives to improve compliance, but the studies do show that additional benefits from treatment—e.g., monetary compensation—produce an increase in compliance.

Although extant studies support our theory, we know of no study of compliance that has followed our theory’s detailed outline. The theory we present indicates that empirical studies of compliance with diabetes treatment need to examine compliance in at least two additional ways: (1) the correlation between compliance and the degree to which a patient’s HbA1c level departs from the medical ideal, and (2) the correlation between compliance and the perceived burdens of various treatment components, such as taking medication and monitoring blood glucose.
In our approach to compliance, we examine the decision that a fully informed, rational patient would make. By assuming full information and full rationality, we make assumptions that do not describe any actual patients. Our claim is that any such patient would, under typical circumstances, fail to comply fully with treatment prescriptions and that this failure constitutes a rational decision. Given that patients are not fully informed, what is the value of our conclusion?

A fully informed patient represents an ideal case. Ideal cases have been presented by philosophers to indicate a benchmark, or a point from which to judge less than ideal circumstances. For example, in *A Theory of Justice*, John Rawls (1971) evaluates basic social arrangements by way of considering what hypothetical people in an original position would decide. Among other things, these hypothetical people, as assumed by Rawls, do not know their gender or intelligence. However, Rawls assumes that they do have full knowledge of science, including all the social sciences. The conditions Rawls proposes as offering the optimal perspective are unrealistic. Yet if one agrees that these assumptions about the original position provide the best vantage point from which to judge basic social arrangements, then one should be instructed by conclusions about what such hypothetical people would accept.

We use assumptions in our theory in a way similar to Rawls, but our assumptions are less controversial. Some of Rawls’s assumptions are made to address special concerns about the theory of justice; none of our theory’s assumptions are made with a special concern in mind. For example, the assumption by Rawls that members in the original position are risk adverse speaks against a utilitarian perspective. We do not depend on a claim that is similarly slanted. As with Rawls, our assumptions indicate the benchmark from which we can draw conclusions about proper decisions. We present the view of a fully informed patient knowledgeable about his or her value structure and other relevant decision factors. Given other realistic assumptions, such as the declining effectiveness and increasing cost of additional treatment, we draw deductively supported conclusions about such a patient’s degree of compliance with prescribed treatment. We show that it is rational for such a patient to reject additional treatment at a certain level. This prediction provides a clear benchmark from which to judge the behavior of actual patients, those who may lack perfect information about treatment costs and benefits. Without perfect
information a rational patient may accept more or less treatment than the perfectly informed patient would accept.

CONCLUDING THOUGHTS

Some implications about patient care follow from our theory. First, patients need accurate and clear information to decide on their proper treatment level. If treatment is overrated and/or the costs understated, patients will not be able to make the decision that coincides with the actual benefit and cost conditions they face. Second, patients who are thought to be noncompliant in fact may be exhibiting fully rationally behavior. Third, whenever patients accept treatment at the medically ideal level, they likely are acting without full knowledge or consideration of costs and benefits. Their decision may result from misplaced trust in, or misunderstanding of, medical advice. Fourth, one way to encourage the rational acceptance of treatment levels nearer to the medical ideal is to decrease the costs of treatment relative to the benefits. Fifth, efforts to increase compliance that ignore the rational basis for rejecting an ideal level of treatment have produced weak results and are likely to continue to do so (Donovan and Blake 1992). Given the general strength of the microeconomic analysis and support from existing studies, health care providers should keep these implications in mind when prescribing and monitoring patient compliance with treatment regimens.

NOTES

1. We use the term “compliance” rather than “adherence” throughout our article despite the appearance that “compliance” indicates a patient’s subservient role. We think that “compliance” more accurately portrays the dominant view in the literature, the view that patients ought to follow prescribed treatments.

2. Some readers may prefer to think of our analysis as “benefit-cost analysis” rather than microeconomic analysis. Benefit-cost analysis has its foundation in welfare economics, a component of microeconomics. When benefit-cost analysis is used to evaluate projects, benefits and costs are construed broadly as we do in our analysis of patient decisions. We think about our approach as microeconomic analysis because it provides the reasons a patient should select a particular level of treatment.

3. The view we present here is an adaptation from Douglas O. Stewart and Joseph P. DeMarco (2005).
4. Our definition of “medically ideal treatment level” is consistent with a statement in an American Diabetes Association Clinical Education Series publication that “Glycemic control that approaches the nondiabetic state postpones or slows the progression of the retinal, renal, and neurologic complications of diabetes” (American Diabetes Association 1998, p. 2).

5. Recent empirical studies have shown a statistical relationship between the percentage of glycated hemoglobin, or HbA1c, and the incidence of retinopathy, a prevalent symptom of Type 2 diabetes (International Expert Committee 2009). Until recently, diagnosis of Type 2 diabetes was based on the result of a patient’s fasting plasma glucose (FPG) test. A patient was diagnosed with Type 2 diabetes with an FPG test result equal to or greater than 126 mg/dl (Expert Committee 1997). Between 1997 and 2003, patients were said to have “impaired fasting glucose” if the FPG test result was between 110 and 125 mg/dl (Expert Committee 1997). Subsequently, the lower limit of the FPG range for a designation of impaired fasting glucose was reduced to 100 mg/dl (Expert Committee 2003). Using the criterion in effect between 2003 and 2009, our “medically ideal treatment” level for Type 2 diabetes is a fasting plasma glucose level of 100 mg/dl.

REFERENCES


