

**A Model and Empirical Analysis of Patient Compliance and Persistence in
Pharmaceuticals**

Scott A. Neslin
Tuck School of Business
Dartmouth College
Hanover New Hampshire
Scott.Neslin@Dartmouth.edu

Edward E. Rhoads
New England Consulting Group
Westport, Connecticut
eer@necg.com

Paul Wolfson
Tuck School of Business
Dartmouth College
Hanover, NH
Paul.Wolfson@Dartmouth.edu

September 1, 2009

The authors express their appreciation to Catalina Health Resource, LLC, for providing the data used in this work. We thank Asim Ansari, Eric Bradlow, Peter Fader, Raghu Iyengar, and Ka Lok Lee for invaluable comments and suggestions regarding the model and its estimation.

Abstract

We develop and estimate a stochastic model of patient compliance and persistence regarding pharmaceutical drugs. Persistence refers to whether a patient continues with therapy, i.e., whether the patient continues to obtain refills of the drug. Compliance refers to whether the customer obtains a refill on time, given he or she does in fact obtain a refill. We develop a simple stochastic model of these behaviors, drawing on models of customer value developed by Schmittlein, Morrison, and Colombo (1987) and Fader, Hardie, and Lee (2005). We estimate the model using patient-level data for 253 drugs and illustrate two applications: (1) We show how changes in drug characteristics would influence the number of days of therapy lost either through lack of persistence or lack of compliance, and (2) Show how the model could be used to identify patients who are at higher than average risk for losing therapy days due either to noncompliance or nonpersistence. We discuss implications of the work for researchers and practitioners.

A Model and Empirical Analysis of Patient Compliance and Persistence in Pharmaceuticals

Introduction

The tendency of patients not to take their prescribed medications on a consistent and continuing basis has emerged as an issue of significant concern. This behavior has been reported to contribute to 125,000 premature deaths each year in the United States and an additional \$100 billion in costs to the economy in terms of increased health care costs and lost productivity (Loden and Schooler 2000; Johnson and Bootman 1995). The result has been an active effort on the part of the health care industry to remedy the situation. These efforts include mechanical devices to monitor individual patients (Pharmaceutical Executive 2006) as well as better packaging to encourage adherence to the drug regimen. In a highly publicized move, CVS Pharmacy intends to use its data and marketing muscle to “get patients to take their pills” (Boyle 2009).

The financial implication for pharmaceutical drug manufacturers is clear – to the extent that patients take their medications more consistently and continually, these companies stand to earn millions of dollars more in revenues. From a customer relationship management standpoint, increased patient diligence in taking their drugs increases customer retention and thus the lifetime value of the customer (CLV). Thus, from either a patient health standpoint or a company prosperity perspective there is a huge impetus to improve customer’s adherence to the instructions of their drug regimens.

A first step in examining the issue is to note that “patient diligence in taking their drugs” consists of two distinct behaviors – “compliance” and “persistence”. Persistence means that the patient continues to refill the prescription, while nonpersistence means

that the patient stops refilling at some point. For example, the patient may get the initial prescription filled, but fail to refill it again. So long as the patient persists, compliance becomes relevant. Compliance means that the patient refills the prescription in a timely fashion. For example, the patient obtains a 30-day supply of a drug and refills the prescription by the end of that 30-day period. Noncompliance means that the patient is late and misses at least one day of therapy. The patient's last script might have been for 30 days, but is renewed after 40 days— the patient was without therapy for 10 days.

A few academic papers have made important progress in studying compliance and persistence. Bowman, Heilman, and Seetharaman (2004) examined the determinants of patient compliance. They found for example that for certain market segments, direct-to-consumer (DTC) drug advertising improved compliance. Wosinska (2004) however found that in economic terms the impact of DTC on compliance was rather small. Lee, Fader, and Hardie (2007) found that persistence could be predicted using a simple stochastic model of patient behavior.

While these papers have made significant contributions, they have not examined compliance and persistence as two distinctive behaviors, yet there are practical and behavioral reasons to do so. From a practical perspective, total days of therapy lost can be decomposed into days lost due to noncompliance and days lost due to nonpersistence. Which component dominates this decomposition influences the managerial action called for to rectify the situation. If noncompliance is the problem, the health provider might provide more effective packaging. If nonpersistence is the problem, the health provider might work with the pharmacy to remind the patient to come in for a refill. From a behavioral perspective, compliance and persistence are obviously related. However, it

isn't clear whether this relationship is positive or negative. It may be that there is a general behavioral tendency to follow instructions, and therefore compliance and persistence would have an underlying positive relationship. On the other hand, it may be that patients struggling to follow instructions may either be able to be compliant or be persistent, but find it difficult to be both. In addition, better compliance may mean the customer gets "cured" and no longer needs the medication. Hence there would be an innate negative relationship between compliance and persistence.

The purpose of this paper is to develop, estimate, and test a patient-level stochastic model of compliance and persistence. We aim to measure the impact of both patient and drug descriptors on compliance and persistence, and show how the model can be used to leverage these insights to devise marketing efforts for addressing the problem. The model draws on work in customer profitability models (Schmittlein, Morrison and Columbo 1987; Schmittlein and Peterson 1994; Fader, Hardie and Lee 2005; Lee, Fader, and Hardie 2007) to devise an integrated model of compliance and persistence. We expand upon this work by applying the model to compliance and persistence rather than customer interpurchase times and retention, and by explicitly modeling patient heterogeneity as a function of patient characteristics.

We derive estimates of the model for 253 drugs using de-identified patient-level data. The estimates show the impact of patient characteristics. We then show how the model can be used to decompose total days potential therapy into days on therapy and days lost due to noncompliance and nonpersistence. Finally, we use the model to determine how drug descriptors influence this decomposition, and how "high-risk" patients can be identified using the model.

The paper proceeds as follows: First we present the model. Second, we discuss estimation. Next we estimate the model and show the overall results, including coefficient estimates and the overall decomposition into therapy, noncompliance, and non persistence. Then we demonstrate two applications: providing guidelines on how changes in drug characteristics can improve compliance and persistence, and identifying patients who are at risk for noncompliance or nonpersistence. We close with an overall discussion of research and managerial implications, as well as limitations and avenues for future research.

Model

As mentioned above, Lee, Fader, and Hardie (2007) were the first to apply a stochastic model to the study of patient persistence. Their work shows the feasibility and promise of patient-level stochastic models. Schmittlein, Morrison and Colombo (1987), Schmittlein and Peterson (1994) and Fader, Hardie, and Lee (2005) studied the question of customer value. These authors were concerned with whether the customer is still a customer, and if so, his or her purchase rate. Analogous to the still-a-customer issue is the question of persistence – is the customer still active in refilling the drug? Analogous to purchase rate is the question of compliance – how promptly does the customer obtain a refill once he or she runs out of medicine? The analysis is challenging because of the ambiguity, particularly at the end of the period of observation, of whether a patient who has not yet refilled has given up on the drug (i.e., nonpersistent), or will eventually refill but is late (i.e., is noncompliant).

This is illustrated in Figure 1. Compliance is directly measurable if the script has been refilled. By knowing the date of the previous fill and the days supply obtained on

that refill, we can calculate the expected date of the next refill (see Bowman, Heilman, and Seetharaman 2004). If the patient refills 10 days after that expected refill date, we have observed 10 days of noncompliance – the patient has refilled but refilled late. We define C_1 as days lost to noncompliance, given a specific observed refill. In Figure 1, we can see $C_1 = 0, 15,$ and 20 for the observed three refills for this patient. However, the ambiguity occurs after the third refill, when the data period ends and the patient hasn't refilled. Is this due to the patient being noncompliant or has the patient given up on the medicine altogether (nonpersistence)? We define C_2 as noncompliance that may have occurred after the last observed refill, and I as nonpersistence. We can see in Figure 1 that either $C_2 = 15$ or $I = 15$. However, the problem is that C_2 and I are not directly observed, and hence need to be estimated by a model. The situation is completely analogous to problems confronting the customer value research mentioned above (e.g., see Fader, Hardie, and Lee 2005).

[Figure 1 Goes Here]

In addition to the challenge in disentangling unobserved noncompliance from nonpersistence, it is likely that patients are heterogeneous in their compliance and persistence behaviors. We expand on the Schmittlein et al. and Fader et al. approaches by directly modeling patient heterogeneity, specifically, the relationship between patient characteristics and persistence and compliance, and the correlation between these two behaviors due to unobserved patient characteristics.

The model requires two simple definitions:

p_i = “persistence”, the probability patient i refills again after the previous refill.¹

¹ Note: Persistence and compliance are relevant to patients who have filled an initial prescription. We are concerned with what happens after that first fill. Hence there always is a “previous refill.”

c_i = “compliance”, the probability patient i obtains refills on each day, given the patient will eventually refill but hasn’t refilled yet.

We assume each refill occasion is independent of the previous occasions, so essentially we have a geometric distribution with respect to persistence (with parameter p_i), and a geometric distribution with respect to compliance (with parameter c_i). The model could be expanded to include time-varying persistence or compliance propensities, and a lack of fit or validation accuracy would suggest introducing this complexity. We will indeed inspect the predictive ability of the model. But for our initial model specification, we follow Schmittlein et al. and Fader et al. and focus on cross-sectional heterogeneity. In particular we assume the propensity to be persistent or compliant differs across patients (i.e., patients are heterogeneous).

We observe for each customer the following over a finite time interval T:

- R_i = the number of refills for patient i during T.
- X_{ij} = the number of days patient i was “late” for the j^{th} refill, where $X_{ij} = 0$ means the patient renewed by the end of the first day after he or she had run out of medicine.
- V_i = the number of days “late” observed for patient i after the last refill observable in the data, given the patient would have run out of medicine by this time.²

We can then compute the following probabilities:

$$\begin{aligned} \text{Prob}(X_{ij}) &= \text{Prob}(\text{Patient refills the } j^{th} \text{ time, } X_{ij} \text{ days after running out of drug | refilled } \\ &\quad j-1 \text{ times)} \\ &= p_i(1 - c_i)^{X_{ij}} c_i \end{aligned} \tag{1}$$

² Note we can only observe lateness if the patient should have refilled by the end of the data. E.g., if the patient obtained a 30-day supply at the last refill and that refill was 40 days before the end of the data, we know the patient was 10 days late, although we don’t observe whether this is due to nonpersistence or noncompliance. If however, the patient last obtained a 30-day supply and that refill was 20 days before the end of the data, this last observation provides no information. We therefore treat it as missing in the estimation.

$$\begin{aligned}
\text{Prob}(V_i) &= \text{Prob}(\text{Patient does not refill after the last observed refill} \mid \text{customer has run out of medicine}) \\
&= p_i(1 - c_i)^{V_i} + (1 - p_i)
\end{aligned} \tag{2}$$

Equation (1) follows because in order to observe the patient being X_{ij} days late for the j^{th} refill, the patient must have decided to be persistent (with probability p_i), and then not refilled for X_{ij} days after running out of medicine, but refilled on the $X_{ij} + 1^{\text{th}}$ day.

Equation (2) follows because if we observe that the patient ran out of medication but had not refilled V_i days after he or she needed a refill, the patient could have decided to refill but is V_i days noncompliant (with probability $p_i(1 - c_i)^{V_i}$), or has decided not to refill (with probability $(1 - p_i)$) (see Fader et al. 2005).

We model heterogeneity in p_i and c_i as a function of observed patient characteristics and an unobserved “error term.” We will utilize three observed characteristics: patient age, whether or not the customer usually pays for prescriptions using cash (an indication of the lack of insurance), and gender. Heterogeneity is thus modeled as follows:

$$\text{logit}[p_i] = b_0 + b_1 \text{age}_i + b_2 \text{cashpay}_i + b_3 \text{gender}_i + e_{1i} \tag{3}$$

$$\text{logit}[c_i] = g_0 + g_1 \text{age}_i + g_2 \text{cashpay}_i + g_3 \text{gender}_i + e_{2i} \tag{4}$$

$$[e_{1i} \ e_{2i}] \sim \text{BVN}[0, \Sigma] \tag{5}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \quad (\sigma_{12} = \sigma_{21}) \tag{6}$$

where “logit” is the logit transformation ($\text{logit}(x) = \ln(x/(1 - x))$) used to ensure the persistence and compliance parameters will lie between zero and one.

The b and g coefficients are important for establishing a “baseline” level of persistence or compliance, as well as the impact of age, cash-payer (vs. non-cash payer), and gender on these quantities. We do not have formal hypotheses with respect to the signs of these coefficients, except we expect that the signs of b_2 and g_2 to be negative, meaning that cash-payers are less persistent and less compliant (they are likely to lack insurance and therefore face budget issues). The correlation between unobserved factors affecting persistence and compliance could be positive or negative – persistence and compliance after all are different behaviors, and as discussed earlier, it could be that unobserved patient characteristics affect each differentially.

Data

We have data on over 400 drugs, each with de-identified patients observed over a one-year (365-day) time horizon. This panel database was generously provided by Catalina Health Resource, LLC. Several of the drugs are used almost exclusively for acute conditions (e.g., antibiotics) and we eliminated these from the data. This resulted in 262 drugs. We will estimate the model described above separately for each drug, yielding a set of coefficients for each drug. We used up to 2000 patients for each drug model; with the remainder of patients held out for validation testing. Given the number of patients available per drug, this yielded 104 validation drugs. There was an average of 1,248 patients per drug in the calibration sample and 4,658 patients per drug in the validation sample.

The panel data provide measures of patient age, “cashpay”, and gender. Age was coded as age/100 to put it on the same scale as cashpay and gender, which were dummy

variables. Cashpay equaled 1 if the patient paid with cash; 0 otherwise. Gender equaled 1 if the patient was female; 0 otherwise. We found some inconsistencies in the data and eliminated patients for example whose age changed by more than one year over the course of the data, or whose gender or cashpay status changed during the year.

Estimation

Each of the 262 drug models was estimated using Bayesian estimation programmed in WinBugs. Estimating the model over 262 drugs was a logistical challenge given the sheer volume of data and time required to estimate each model. This was remedied by writing a SAS program that iterated over drugs and “called” WinBugs as a subroutine to estimate the model for each drug (Smith and Richardson 2007). SAS handled the preparation of the data and the storage of the results. Exploratory runs suggested that the estimation converged quickly, undoubtedly due to the relatively small number of parameters. Given this, and the desire to conserve computer time, we used 20,000 iterations for burn-in and the next 30,000 for estimation.

After obtaining the estimates, we noticed that nine drugs had slope coefficients ($b_1, b_2, b_3, g_1, g_2, \text{ or } g_3$) that were clearly outliers (several standard deviations away from the mean across drugs) and we eliminated these drugs from further analysis. Seven of these cases involved either relatively small sample sizes (< 1000) or were predominantly one gender ($> 95\%$ of its patients were the same gender). However, that there were other drugs that were predominantly one gender that did not yield outlying coefficients. We also estimated these models without the gender variable, and found the slopes for age and cashpay correlated highly ($> +0.95$) with those estimated for the full model that included

gender. Hence we retained the full model for these drugs. We therefore report the results for 253 drugs, each of which is estimated with age, cashpay, and gender covariates.

Results

Table 1 provides descriptive statistics of the estimated coefficients. There are several important observations to note. First, the patient characteristics are statistically significant³ in several instances. For example, age is statistically significant for 50.2% of drugs as a determinant of persistence. Gender is significant noticeably less often, only 12.7% of the time as a determinant of persistence, and 12.3% of the time for compliance. Also, in general, there are fewer significant coefficients for the compliance equation than for the persistence equation. This suggests that we have had a more difficult time measuring heterogeneity in compliance than in persistence.

[Table 1 Goes Here]

A second observation regards the signs of the coefficients. Generally, age is positively related both to persistence and compliance – older patients are more diligent. This is an interesting finding and could be due to the fact that the consequences of poor health – of not following doctors' orders – are more real to older patients. However, there are various other explanations – it may be a generation effect (older people were imbued at an earlier age with the importance of listening to the doctor) or an indication of severity of disease.

Cashpay is negatively related to persistence and compliance. This is an important public policy finding – lack of insurance results in less compliance and less persistence. Pushing this further, it means that patients without insurance are less likely to be diligent

³ The 95% credible interval excludes zero.

in taking their medicine and hence are at greater risk for negative health consequences. Gender is not significant as often as the other two variables, but it is interesting that the signs are different for compliance and persistence. The positive sign on average for b_3 suggests that females are more likely to be persistent, while the negative sign on average for g_3 means that females are less likely to be compliant.

A third important finding is the overall negative relationship between the error terms connecting persistence and compliance. This suggests that on average, unobserved characteristics that increase compliance tend to decrease persistence, and *vice versa*. This could relate to psychographic traits such as a “better late than never” approach to drug-taking. It also could mean that one unobserved factor causing a patient to be compliant is the patient taking better care of him or herself. Therefore, the compliant patient gets “cured” faster and does not have to be as persistent. It is interesting to speculate on this negative correlation, and it is certainly an important area for future research, however the large standard deviation across these across drugs suggests that this correlation is highly drug-dependent.

Overall, Table 1 reinforces the value of modeling heterogeneity directly as a function of observed variables, and that the underlying behaviors of compliance and persistence are related, which is why they need to be modeled jointly.

Table 2 shows the ability of the model to recover what we directly can observe about compliance and persistence. In particular, we directly can observe the % of the year the patient is on therapy (e.g., if the patient purchased 73 pills and the prescription calls for taking one pill per day, % therapy is $73/365 = 20\%$). We can also observe C_1 , days lost due to noncompliance associated with a specific refill. We cannot observe C_2

(days lost due to noncompliance after the last refill) separately from I (days lost due to nonpersistence), as discussed regarding Figure 1. However, we can observe the C_2 or I (depicted as “%IC2” in Table 2) simply by counting the number of days between the last refill and the end of the sample period.

[Table 2 Goes Here]

Table 2 shows the % of total potential days on therapy, % of total potential days lost due to observed noncompliance (“%CI”), and % of total potential days lost due to end-data noncompliance or nonpersistence (“%IC2”) both for the actual data and as predicted by the model, for the calibration and validation samples. The predictions were made using simulation. For each drug, we took its estimated coefficients and simulated each patient’s compliance/persistence behavior over a one-year period, using the model (equations 1-4) with uncertainty included via the error terms as well as by the probabilistic nature of the model itself (captured by p_i and c_i). The across-patient variation in age, cashpay, and gender ensured that the p_i ’s and c_i ’s varied systematically across customers. We used 1000 replications to predict average behavior for each patient. We monitored days on therapy, days lost due to noncompliance, and days lost due to nonpersistence. Note we can distinguish C_2 from I in the simulation because we take a random draw (using p_i) to determine if after the end of the data, the patient will not refill again, or will refill again but is just late. The results were then averaged over all patients for the drug, either in the calibration or validation samples, to yield the results summarized in Table 2.

Table 2 shows that observed noncompliance and the observed combination of end-data noncompliance or nonpersistence are recovered nicely by the model. This holds

both for calibration and validation data. The correlation across drugs between the calculated percentages and the actual percentages is quite high. We do note a slight upward bias in % C1 and % Therapy, and a corresponding downward bias in %IC2. This is to be expected – there is no guarantee that nonlinear models will generate the original average data, as is the case with linear models. However, clearly the percentages are quite close in absolute terms, and the high correlations suggest we are able to discriminate very nicely between brands that have high versus low compliance or persistence.

We next used the simulated results to decompose the end-data days lost into nonpersistence versus noncompliance. Again, this required distinguishing between C2 and I, but this was available from the simulation as described above. Although we cannot validate this decomposition directly, we were encouraged by the results of Table 2, which showed we could recover the total end-data days lost. We therefore decompose the total potential therapy days into % on therapy, % days lost due to noncompliance (C1+C2), and % days lost due to nonpersistence (I). Figure 2 shows that for the average drug over a year's time, the customer is on therapy for 29.51% of the days, but 17.65% of days are lost due to noncompliance and 53.44% of days are lost due to nonpersistence. Note that if we just used observed data (C₁ or IC₂), we would have calculated 11.27% days lost due to noncompliance and 59.82% days lost due either to end-data noncompliance or nonpersistence, without providing any detail of how that 59.82% is split. The results show, perhaps not unexpectedly, that most of the 59.82% is attributable to nonpersistence. The model knows the customer did not refill at the end of the data, and he or she is so many days late that the model infers it is due to nonpersistence. However,

the end-data noncompliance is nontrivial in magnitude (6.38%) so it is important that one is able to calculate it.

[Figure 2 Goes Here]

Applications

Relating Drug Descriptors to Compliance and Persistence

An important issue for drug manufacturers is what actions they can take to improve compliance and persistence. For example, Figure 2 suggests that for the average drug, roughly 70% of the potential days therapy over a year is lost either due to noncompliance or nonpersistence. This amounts to $0.70 \times 365 = 256$ days off therapy and 99 days on therapy. If the 70% could be reduced just to 65%, this would mean an additional $0.05 \times 365 = 18$ days of therapy, an increase in revenues of $18 / 99 = 18\%$. Thus the stakes for improving compliance and persistence are high.

We can use the model to help manufacturers in this regard by doing the following: (1) Calculate % of days on therapy (T), % of days lost due to noncompliance (C), and % of days lost due to nonpersistence (I) for each drug. This can be done using the simulation results summarized in Table 2 and Figure 2. (2) Note that $T + C + I = 100\%$ or 1 in decimal terms. (3) Formulate and estimate a regression model that uses T, C, and I as dependent variables and drug descriptors as independent variables. (4) Use this model to examine T/C/I scenarios for particular brands when the brands change these descriptors.

The model we estimate for this purpose is a differential effects multinomial attraction model (see Cooper and Nakanishi 1988). This is a regression model that

allows drug descriptors to have a differential impact on therapy, noncompliance, and nonpersistence, and is designed so that predictions of these three percentages add to one, as they should to be logically consistent. The model is stated as follows:

$$A_{ij} = e^{\left(\alpha_i + \sum_{k=1}^K \beta_{ki} X_{kj} + \varepsilon_{ij} \right)} \quad (7)$$

$$S_{ij} = \frac{A_{ij}}{\sum_{m=1}^3 A_{mj}} \quad (8)$$

where:

A_{ij} = “Attractiveness” term representing the tendency of drug j to experience outcome i (therapy T, noncompliance C, or nonpersistence I).

S_{ij} = “Share” of outcome i (therapy T, noncompliance C, or nonpersistence I) for drug j , i.e., the fraction of potential days therapy decomposed into therapy, noncompliance, and nonpersistence for drug j .

X_{kj} = Value of drug characteristic k for drug j .

i = 1, 2, or 3 indicating three outcomes: therapy ((1=>T), noncompliance (2=>C), and nonpersistence (3=>I)).

j = 1, 2, ..., J indicating drug ($J = 253$ for our data).

k = 1, 2, ..., K indicating drug characteristic.

Equation (7) says, through β_{ki} , that drug characteristic k has a differential effect on the three outcomes – therapy, noncompliance, and nonpersistence. This is a differential effects attraction model (Cooper and Nakanishi 1988) and in fact is necessary for our application because the explanatory variables do not vary across outcomes within drug. This aspect is in turn analogous to the case with customer-level multinomial logit models

where the independent variable (e.g., demographics) does not vary across alternatives (see Guadagni and Little 1983; Thomas and Sullivan 1995).

The model can be estimated using the following equation:

$$\ln \frac{S_{ij}}{\tilde{S}_j} = \alpha_i^* + \sum_{k=1}^K \beta_{ki}^* X_{kj} + \varepsilon_{ij}^* \quad (9)$$

where

$$\tilde{S}_j = \left(\prod_{i=1}^3 S_{ij} \right)^{1/3} = \text{geometric mean of outcomes for a drug } j.$$

$$\alpha_i^* = \alpha_i - \bar{\alpha} = \text{differential main effect for outcome } i.$$

$$\beta_{ki}^* = \beta_{ki} - \bar{\beta}_k = \text{differential effect of drug characteristic } k \text{ on outcome } i.$$

$$\varepsilon_{ij}^* = \varepsilon_{ij} - \bar{\varepsilon}_j = \text{differential effect on outcome } i \text{ of unobserved factors for drug } j.$$

$$\bar{\alpha}, \bar{\beta}_k, \bar{\varepsilon}_j \text{ average } \alpha_i, \beta_{ki}, \text{ and } \varepsilon_{ij} \text{ over } i.$$

Equation (9) is estimated by OLS using dummy variables for each outcome interacted with each drug characteristic, and no intercept. After estimation, one can estimate the fraction of each outcome for each brand (T, C, and I), for a given set of drug descriptors, by substituting the appropriate values into the estimated equation (9). This produces:

$$\hat{Y}_{ij} = \text{prediction of } \ln \frac{S_{ij}}{\tilde{S}_j} \quad (10)$$

and the inverse log-centering transformation yields (Nakanishi and Cooper 1982):

$$S_{ij} = \frac{e^{\hat{Y}_{ij}}}{\sum_{m=1}^3 e^{\hat{Y}_{mj}}} \quad (11)$$

Table 3 shows 17 drug descriptors used in estimating equation (9). These fall into three classifications: (1) Average statistics for the *patients* taking the drug, (2) Descriptors of the *drug* itself, and (3) Descriptors of the types of diseases the drug *treats*. The *patient* variables include age, cashpay, and gender. The *drug* descriptors include its side effects, how long it takes the drug to work, how the patient knows it is working, whether the drug is branded or generic, whether it uses (TV, radio, or print) DTC, its cost, and whether the predominant form of the drug is a “pill” (tablet or capsule) or another form (e.g., cream, drops, or solutions for injection). The *treatment* variables include prevention/promotion orientation of the drug (“Motivation”), the life preserving/lifestyle orientation of the drug (“Lifestyle”), and the degree to which the drug is prescribed for chronic versus episodic ailments.

[Table 3 Goes Here]

The Motivation and Lifestyle variables merit more description. Examples of drugs classified as Motivation-prevention include cholesterol-lowering medications such as Lipitor that prevent heart disease. Motivation-promotion would be drugs that promote the quality of life without curing or preventing a disease *per se*, for example an anti-depressant such as Zoloft. Lifestyle drugs are related to the idea of motivation-promotion but deal with what are deemed to be less life-threatening conditions. These include diet suppressants such as Dexedrine or anti-acne medication such as Accutane (see for example http://en.wikipedia.org/wiki/Lifestyle_drug , accessed August 13, 2009).

The average age, cashpay, and gender for each drug were calculated from the patient data for each drug. The side effect, time-to-work, awareness-of-working, motivation, lifestyle, and chronic usage measures were compiled by an independent

Board Certified Pharmacotherapy Specialist (BCPS) with a Doctor of Pharmacy (Pharm. D.), based on known side effects, clinical tests, indications (diseases) for which the drug could be prescribed, and prescription data. The DTC and cost variables were obtained from industry sources (Verispan PSA and Redbook), and the drug form was available directly from the patient-level data.

In addition, the drugs were members of various drug “categories” (e.g., statins, anti-hypertensives, etc.) and dummy variables were coded to indicate these categories so category effects would be controlled for in the model. Some categories had only one drug, so could not be coded as a separate dummy. However, most of the categories had more than one drug, and a total of 53 category dummies were coded. In addition, there are the three differential behavior effects (the α 's) which we coded as dummy variables. To accommodate the differential effects aspect of the model, these dummies were multiplied by the drug descriptors and dummy variables, so that the final model was estimated on $253 \times 3 = 759$ observations, with $3 + 3 \times (53 + 19) = 219$ independent variables. The results for the 17 drug descriptors (β_{ki} 's) are shown in Table 4. The p-values reflect whether the β 's differ across behaviors. We see some interesting results:

Patient Characteristics: Note these results pertain to cross-drug variation, whereas the patient variables used directly in the compliance/persistence model pertain to within-drug/ across-patient variation. Older patient populations result in a higher fraction of days on therapy, and less nonpersistence and less noncompliance. Cashpay has the opposite effect, with drugs with a larger proportion of their patients paying with cash especially suffering more noncompliance. Average gender is not as statistically

significant, but suggests drugs with more female patients incur less noncompliance and more nonpersistence.

Drug Characteristics: Side effects are found not to have a huge impact on outcomes – they are significant at the 10% level but not the 5% level for mild and severe effects. Interestingly, in these cases, side effects increase noncompliance but improve persistence. This may be due to side effects serving as a proxy for disease severity. The patient has to “take a break” now and then, so is late in refilling, but does keep refilling the prescription. Brand, Cost, and Pill are statistically significant predictors of the drug’s noncompliance/nonpersistence/therapy profile. Branded drugs enjoy better compliance but worse persistence. Higher cost drugs enjoy more therapy; we conjecture this may be related to cost being associated with more serious diseases. Pills enjoy much better compliance, although not as good persistence. It could be that due to the enhanced compliance, patients do not need to stay on therapy for as long a time.

Treatment Characteristics: Motivation is an important predictor, with drugs whose purpose is to prevent disease enjoying more therapy than those meant to increase the quality of life. Preventive drugs enjoy much better compliance and hardly differ from quality of life drugs on persistence. A similar pattern holds with Lifestyle, where life-preserving drugs realize more therapy through less noncompliance and also less nonpersistence. The chronic results suggest a positive impact on therapy due to much less nonpersistence. This makes sense in that drugs for chronic ailments undoubtedly need to be taken for prolonged periods. Finally, it is noteworthy that DTC advertising has only a weak positive, and statistically insignificant relationship with therapy, consistent with Wosinka’s findings.

[Table 4 Goes Here]

We can use these results to analyze scenarios where the drug manufacturer is considering making changes in the drug in the hopes of improving compliance and persistence. For example, Table 5 shows the case of Nexium, a drug used to decrease stomach acid. The drug is now classified as motivated to promote quality of life rather than prevent disease. The average age of its patient base is currently 54 years, and 18% of them pay cash for the drug. Table 5 shows that our multinomial regression model predicts a profile of 13.3% noncompliance, 59.2% nonpersistence, and 27.5% therapy (the actual numbers are 13.7% / 58.5% / 28.4%). In Table 5, we examine the implications for this profile of changing motivation, patient age, and percent paying cash

[Table 5 Goes Here]

Case 1 shows the implications of changing the motivation for taking the drug from promoting quality of life to preventing disease. This change would occur if for example clinical trials suggested a new “indication” for the drug, say preventing gastric ulcer. Experts such as the person who coded our data would notice this approval and that more prescriptions were for preventing gastric ulcer. They would then classify the drug as “prevention” since its primary use had shifted in that direction. As suggested by the coefficients for the Motivation variable in Table 4, this will increase the % days therapy for this drug up to 32.2%, and Table 5 indicates that the profile improves through decreases in both noncompliance and nonpersistence. In Case 2, we assume that through better drug insurance coverage, the percent paying with cash is decreased from 18.3% to 5%. This results in still more therapy days, up to 35.5%. Finally, we assume through either targeting or simply the aging of its patient base, the average patient age increases to

60 years from the current 54 years. This provides still further gains, so that 37.3% of days are now therapy days.

In summary, the successive steps of broadening the applications of the drug to disease prevention, decreasing the percentage of patients who have to pay cash for the drug, and increasing the patient base age profile, days therapy increases from the current 27.5% to 32.2%, then 35.5%, then 37.3%. Table 5 shows that this means an increase in revenues for the manufacturer of 17.1% ($32.2 / 27.5$), 29.1%, and 35.6% in total.

The illustration with Nexium is just one case but illustrates how the model can be used to evaluate the impact of changes in drug characteristics on therapy, noncompliance, and nonpersistence outcomes.

Identifying High Risk Patients

Another important use of the model would be more in the realm of customer relationship management, i.e., using the model as a “predictive model” (cf Blattberg, Kim, and Neslin 2008) to identify customers who are at greater risk for noncompliance and/or nonpersistence. The procedure is conceptually simple: Use the simulations we used earlier for Table 2 and Figure 2 to calculate, for each patient, for each drug, days lost due to noncompliance, and days lost due to nonpersistence. Then, sort patients in order of the three outcomes and identify the patients at highest risk.

Figure 3 shows the results for four drugs: Ritalin, Clonazepam, Zyprexa, and Lipitor. In particular, the patients were simulated, sorted, and then divided at the midpoint. In all cases, the patients predicted to be in the top half experience more days lost than those that are predicted to be in the lower half. Since this is a test of the model, the results are shown for observed noncompliance (C_1) and observed end-data

noncompliance or nonpersistence (C_2 or I). The patients were ordered based on the predicted values of these variables, but the results shown are the actual observed values.

[Figure 3]

For example, Figure 3a shows that those predicted to be in the low-risk half for Clonazepam actually lost 40 days due to observed noncompliance, while those predicted to be in the high-risk half lost 46 days. This is a 15% increase. Figure 3b shows that for Clonazepam, those predicted to be in the low-risk half for end-data noncompliance or nonpersistence lost 190 days, while those predicted to be in the high-risk half lost approximately 235 days. This is an increase of 45 days, or more than the equivalent of one refill per year (usually one refill is a 30-day supply). So in both cases (Figures 3a and 3b), patients identified to be at high risk experience more days lost to either noncompliance or nonpersistence.

Efforts to subdivide patients into finer groupings such as deciles for particular drugs resulted in “bumpier” graphs, although on average, the trend was monotonically increasing over deciles. Part of this could be due to smaller sample size per “tile” when constructing finer groupings (e.g., dividing a sample of 2,000 into deciles means 200 patients per decile). But it also suggests that the model would be even better able to identify at-risk patients if it contained more customer characteristics, for example, health status, other medications prescribed, etc. Unfortunately, we did not have access to these data. However, the model could certainly incorporate them, and Figure 3 suggests that just based on the three customer characteristics age, cashpay, and gender, we can identify a managerially meaningful risky half. Health professionals could use these data to target the high-risk half with information or devices for improving compliance/persistence.

Such efforts are expensive, so targeting them at patients for whom the risk is higher, and hence there are more days therapy to recover, makes sense.

Summary

We have developed, estimated and demonstrated the potential application of a patient-level stochastic model of patient drug compliance and persistence. The stochastic model portrays compliance and persistence as two correlated behaviors, each described by its own patient-specific parameter and each modeled as a geometric process. The patient-specific parameters are in turn functions of patient characteristics plus unobserved factors. The model draws on the rich literature on customer profitability (Schmittlein, Morrison, and Colombo 1987; Schmittlein and Peterson 1994; Fader, Hardie and Lee 2006), and in fact the behaviors of noncompliance and nonpersistence are related to the issues of interpurchase times and “alive” customers investigated in those articles. Our model expands upon this work through the unique application to pharmaceutical patient behavior, the new method for modeling patient (customer) heterogeneity explicitly as a function of observed characteristics, and allowing for explicit correlation between the behaviors. The model also expands upon the efforts of Lee, Fader, and Hardie (2007), who model persistence, and upon Wosinska (2005), Bowman, Heilman, and Seetharaman (2004), who conduct empirical analyses of compliance.

We have estimated the model for 253 brands encompassing more than 250,000 patients. We have learned much from our analysis, as follows:

- Patient characteristics, specifically age, tendency to use cash rather than insurance to pay for prescriptions, and gender have significant impacts on compliance and

- On average, there is a negative correlation between unobserved factors influencing compliance and persistence. It is as if the patient focuses either on compliance or persistence, but has a hard time focusing on both. Or perhaps by focusing on compliance, the patient's health is improved so does not have to take the drug for as long a period.
- The model is able to recover patient behavior for both calibration and validation samples. Specifically, after estimating the model, simulated behavior at the patient level, aggregated up to the brand level, correlates highly with the actual breakdown of potential days of therapy into actual days on therapy, days lost due to observed noncompliance, and the combination of days lost due to end-of-data days lost through noncompliance and nonpersistence (Table 2).
- For the 253 drugs analyzed, we infer an average decomposition of [30%,18%,53%] for percentage of days on therapy, percentage of days lost through noncompliance, and percentage of days lost through nonpersistence. While there is ample variation among brands, this suggests that nonpersistence is in general a larger source of lost days than noncompliance.
- The models for the 253 brands in turn can be “meta-analyzed” using a multinomial logit differential effects attraction model to distill the impact of drug characteristics

on compliance and persistence. We find for example that drugs used primarily for chronic health problems experience especially less nonpersistence, a little less noncompliance, and more days on therapy.

- The meta-model can be used to analyze drug strategies for improving compliance and persistence. The results of this analysis can be managerially, and presumably, medically meaningful. For example, we found that changing the motivation for taking a drug from promoting quality of life to preventing disease would increase days on therapy by 17.1%.
- The model can be used to identify patients who are at risk for noncompliant or nonpersistent behavior. Comparing the high-risk half to the low-risk half can identify differences in days lost up to the equivalent of one script per year.

Our work has several implications for researchers: First, from a modeling standpoint, we have demonstrated that simple but powerful models of customer profitability can be augmented by explicit consideration of observed causes of customer heterogeneity. Second, we have demonstrated the power of these models in an entirely new context – patient compliance and persistence rather than customer profitability – showing the applicability, relevance, and predictive accuracy of these models. Third, our particular empirical findings suggest avenues for future research. For example, we find the correlation between unobserved patient characteristics influencing compliance and persistence to be negative. We need to see whether this generalizes, and if so, why. As another example, we find that DTC advertising has only a small impact on compliance.

This supports Wosinska's findings but is less consistent with Bowman, Heilman, and Seetharaman's findings. Further work is needed to reconcile these results.

Our work also has important implications for pharmaceutical retailers, manufacturers, and public policy officials. Perhaps most paramount is that patient-level data on drug usage behavior is valuable for diagnosing the causes of noncompliance and nonpersistence, both customer-specific causes and drug-specific causes. This means that collection of such patient-level data is important. Obviously this data collection needs to be done in a way that protects patient privacy, but the benefit is in learning how to identify at-risk patients and in evaluating the impact of changes in drug characteristics on compliance and persistence.

A basic yet not-to-be-overlooked finding, reflected in Table 2 and Figure 2, is that that indeed, noncompliance and nonpersistence are problems, with nonpersistence being more of a problem. The term "problem" is value laden. These behaviors are certainly a problem from the perspective of drug manufacturers, because they decrease revenues. Whether it is a problem from a public policy perspective depends on the impact of drugs on patient health. However, as alluded to earlier, lack of diligence in patient drug-taking behavior is said to result in 125,000 premature results per year.

The final implication for managers is there appear to be three benefits of the modeling exercise: (1) Insights on the determinants of noncompliance and nonpersistence, (2) Guidelines on how to change drug characteristics in order to remedy noncompliance and nonpersistence, and (3) the ability to distinguish high from low-risk patients in terms of lack of diligence in taking medications.

As a first foray into developing and estimating an integrated patient-level model of compliance and persistence, we are pleased with the results. However, we believe the results could be improved through better data and more detailed modeling. For example, we observed only three patient characteristics. These were enough to identify the differences in risk observed in Figure 3. However, data on more patient characteristics would yield even stronger results. There are three ways in which the model might be improved. First, the model is very simple, modeling compliance and persistence essentially as constant hazard phenomena. This is consistent with the patient profitability models from which our model evolves, but the model might be more accurate if a non-constant hazard were used. Second, the model does not contain observable measures of dynamics. For example, it does not contain state dependence – one would think that noncompliance for refill r would predict noncompliance for refill $r+1$. The model would be improved by including such dynamics. Third, the ultimate goal would be an integrated model of drug-specific and across-drug effects. We estimated 253 separate models of compliance and persistence. We considered the models to be independent. This could be partially justified on the pragmatic basis that the data did not tell us whether a patient with a particular ID was taking multiple drugs, but one could imagine an integrated model where the multinomial attraction model, or its equivalent, would be estimated jointly with the drug-specific model. This would require careful specification of brand-specific error terms, etc. In addition, the effort would face nontrivial logistical issues in estimation. E.g., 253 drugs \times 1,248 patients per drug would mean 315,744 patients, \times say five refills per patient, would yield more than one and a half million observations. It isn't clear whether such an effort would collapse under its own weight (it

would certainly consume weeks of computer time), but the potential benefit would be more accurate predictions due to correlations between the various error components of the model.

As stated above, we believe our efforts have generated rich and useful results, but we hope that future research efforts will improve these efforts, as outlined above.

References

1. Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*, New York: Springer.
2. Bowman, Douglas, Carrie M. Heilman, and P.B. Seetharaman (2004), "Determinants of Product-Use Compliance Behavior," *Journal of Marketing Research*, 41 (3), 324-338.
3. Boyle, Matthew (2009), "CVS Dispensing Drugs – and Health Reform," *Business Week*, February 29, 42-44.
4. Cooper, Lee G., and Masao Nakanishi (1988), *Market-Share Analysis*, Norwell, MA: Kluwer.
5. Fader, Peter S., Bruce G. S. Hardie, and Kok L. Lee (2005), "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 21 (2), 275-284.
6. Guadagni, Peter M., and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 342-355.
7. Johnson, Jeffrey and Lyle Bootman (1995), "Drug Related Morbidity and Mortality in the United States," *Archives of Internal Medicine*, 155 (October), 1949-58.
8. Lee, Ka Lok, Peter Fader, and Bruce Hardie (2007), "How to Project Patient Persistency," *Foresight*, Issue 8 (Fall), 31-35.
9. Loden, John and Caroline Schooler (2000), "Patient Compliance," *Pharmaceutical Executive*, 20 (7), 88-94.
10. Nakanishi, Masao, and Lee G. Cooper (1982), "Simplified Estimation Procedures for MCI Models", *Marketing Science*, 1 (3), 314-321.

11. *Pharmaceutical Executive* (2008), "Drug Compliance Monitor", November, p. 56.
12. Rubin, D. (2006), "Patient Retention Strategies," *DTC Perspectives*, December 26-30.
13. Schmittlein, David C., and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (1), 41-65.
14. Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next," *Management Science*, 33 (1), 1-24.
15. Thomas, Jacquelyn S., and Ursula Y. Sullivan (2005), "Managing Marketing Communications with Multichannel Customers," *Journal of Marketing*, 69 (4), 239-251.
16. Smith, Michael K. and Richardson, Helen (2007), "WinBUGSio: a SAS Macro for the Remote Execution of WinBUGS." *Journal of Statistical Software*, 23 (9) (December), 1-10
17. Wosinska, Marta (2005), "Direct-to-Consumer Advertising and Drug Therapy Compliance," *Journal of Marketing Research*, 42 (3), 325-352.

Table 1: Descriptive Statistics of Estimated Model Coefficients

Coefficient	Mean^a	Median^a	Std Dev^a	% Significant^{a,b}
b ₀ (baseline / persistence)	0.018	0.176	1.762	41.9%
b ₁ (age / persistence)	1.686	1.396	2.421	50.2%
b ₂ (cashpay / persistence)	-0.452	-0.467	0.698	44.3%
b ₃ (gender / persistence)	0.128	0.054	0.936	12.7%
g ₀ (baseline / compliance)	-2.880	-2.685	1.288	94.9%
g ₁ (age / compliance)	0.399	0.694	1.709	43.1%
g ₂ (cashpay / compliance)	-0.298	-0.298	0.494	37.2%
g ₃ (gender / compliance)	-0.012	-0.069	0.559	12.3%
σ ₁ (persistence)	6.020	5.448	3.823	100.0%
σ ₂ (compliance)	3.194	2.337	2.824	100.0%
correlation	-0.184	-0.166	0.327	45.5%

a. Calculated across n=253 drugs

b. 95% credible interval excludes zero.

Table 2: Model Fit and Validation Recovery of Compliance and Persistence

	Calibration Sample			Validation Sample		
	% Therapy	% C1	%IC2	% Therapy	%C1	%IC2
Actual	28.73%	10.59%	60.68%	29.33%	10.77%	59.90%
Predicted	29.51%	11.27%	59.82%	30.31%	11.47%	58.86%
Correlation	0.993	0.976	0.988	0.993	0.952	0.986
# Drugs	253	253	253	104	104	104

Notes: % Therapy = % of days average customer is on therapy

% C1 = % of days lost due to noncompliance (excluding censored observations)

%IC2 = % of days lost due to nonpersistence and/or noncompliance on censored observations

% 's are averages across n=253 drugs for calibration; 104 drugs for validation

Table 3: Definitions of Drug Descriptor Variables

Variable Type	Variable	Description
Patient	Age	Average age of prescribed patients
Patient	Cashpay	% of prescribed patients who pay cash for filling script
Patient	Gender	% of prescribed patients who are female
Drug	Volrank	Sales rank of the brand
Drug	Side-Mild	1-5 scale indicating likelihood of experiencing mild side effects
Drug	Side-Moderate	1-5 scale indicating likelihood of experiencing moderate side effects
Drug	Side-Severe	1-5 scale indicating likelihood of experiencing severe side effects
Drug	Time	1-4 scale indicating how long it takes drug to benefit patient
Drug	Exam	=1 if exam can tell whether drug is working; 0 otherwise
Drug	Experience	=1 if patient can tell by him/herself whether drug is working; 0 otherwise
Drug	Brand	=1 if branded drug; 0 if generic
Drug	DTCUse	=1 if drug uses Direct-to-Consumer advertising; 0 if not
Drug	Cost	Cost per script
Drug	Pill	=1 if drug predominantly in tablet or capsule form; 0 if not
Treatment	Motivation	=1 if drug prevents disease; 0 if promotes quality of life
Treatment	Lifestyle	=1 if life preserving drug; 0 if lifestyle drug
Treatment	Chronic	Extent to which drug used for chronic ailments (0-1 scale)

Table 4: Multinomial Attraction Model: Estimated Coefficients for Drug Descriptors

Variable Type	Variable	Noncompliance	Nonpersistence	Therapy	p-Value ^a
Patient	Age	-0.069	-0.616	0.685	0.001
Patient	Cashpay	0.853	0.072	-0.925	0.000
Patient	Gender	-0.263	0.303	-0.040	0.066
Drug	VolRank	0.0001	0.000001	-0.0001	0.451
Drug	Side-Mild	0.059	-0.062	0.003	0.083
Drug	Side-Moderate	-0.022	0.018	0.004	0.850
Drug	Side-Severe	0.118	-0.104	-0.013	0.065
Drug	Time	-0.068	0.084	-0.016	0.401
Drug	Exam	0.138	-0.126	-0.012	0.209
Drug	Experience	0.080	-0.127	0.046	0.407
Drug	Brand	-0.078	0.100	-0.022	0.039
Drug	DTCUse	0.013	-0.055	0.042	0.436
Drug	Cost	-0.0002	0.0001	0.0001	0.005
Drug	Pill	-0.628	0.342	0.286	0.000
Treatment	Motivation	-0.213	0.012	0.201	0.005
Treatment	Lifestyle	-0.260	-0.062	0.322	0.070
Treatment	Chronic	-0.005	-0.324	0.328	0.067

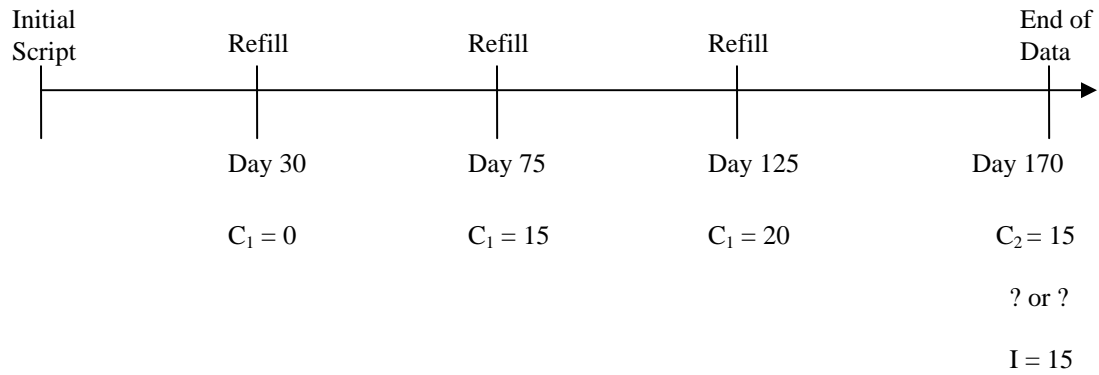
Note: p-value tests the null hypothesis that all three coefficients for a given variable are equal to zero.

Table 5: Impact of Changing Brand Characteristics on Days Therapy – The Case of Nexium

Variable Type	Variable	Base Case	Case 1	Case 2	Case 3
Patient	Age	0.539	0.539	0.539	0.600
Patient	Cashpay	0.183	0.183	0.050	0.050
Patient	Gender	0.616	0.616	0.616	0.616
Drug	Volrank	16	16	16	16
Drug	Side-Mild	2	2	2	2
Drug	Side-Moderate	2	2	2	2
Drug	Side-Severe	2	2	2	2
Drug	Time	1	1	1	1
Drug	Exam	0	0	0	0
Drug	Experience	1	1	1	1
Drug	Brand	1	1	1	1
Drug	DTCUse	1	1	1	1
Drug	Cost	152.123	152.123	152.123	152.123
Drug	Pill	1	1	1	1
Treatment	Motivation	0	1	1	1
Treatment	Lifestyle	1	1	1	1
Treatment	Chronic	0.626	0.627	0.627	0.627

	Base Case	Case 1	Case 2	Case 3
Noncompliance	13.3%	10.3%	9.0%	9.0%
Nonpersistence	59.2%	57.5%	55.5%	53.7%
Therapy	27.5%	32.2%	35.5%	37.3%
Change in Therapy vs. Base Case		+17.1%	+29.1%	+35.6%

Figure 1: Compliance and Persistence – Defined and Inter-twined



- C_1 = Days lost due to noncompliance, given observed refill.
- C_2 = Days lost due to noncompliance, after the last observed refill.
- I = Days lost due to nonpersistence
- T = Days on therapy
- D = Total days of observation

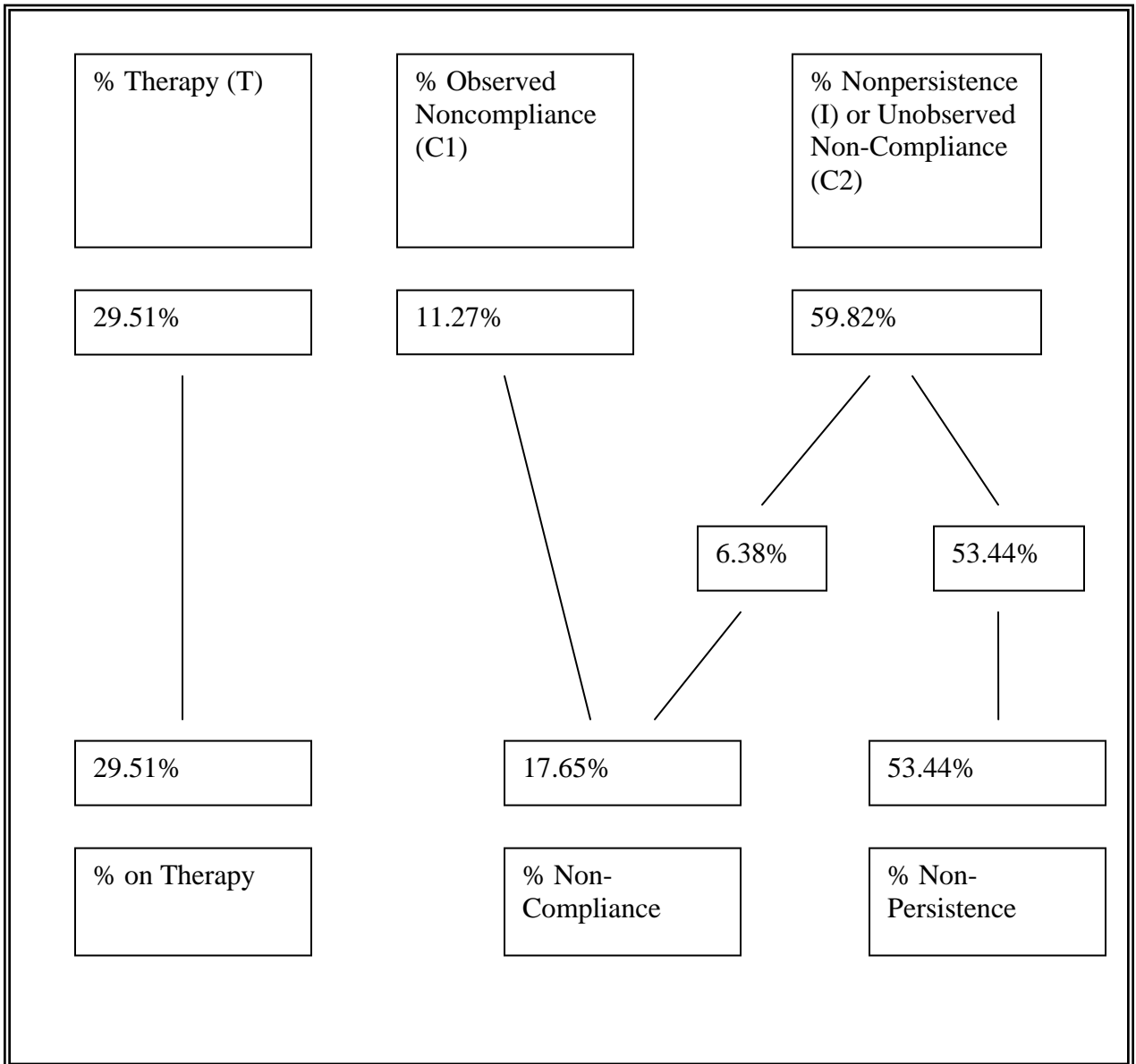
$$D = T + C_1 + C_2 + I$$

We observe:

- $D = 170$
- $T = 120$
- $C_1 = 35$
- $C_2 \text{ or } I = 15$

Figure 2

Calculating Average Percentage of Time on Therapy and Percentage of Time Lost Due to Non-Compliance and Non-Persistence



Note: Average results across 253 brands.

Figure 3: Identifying at Risk Patients

Figure 3a: Observed Noncompliance (C_1)

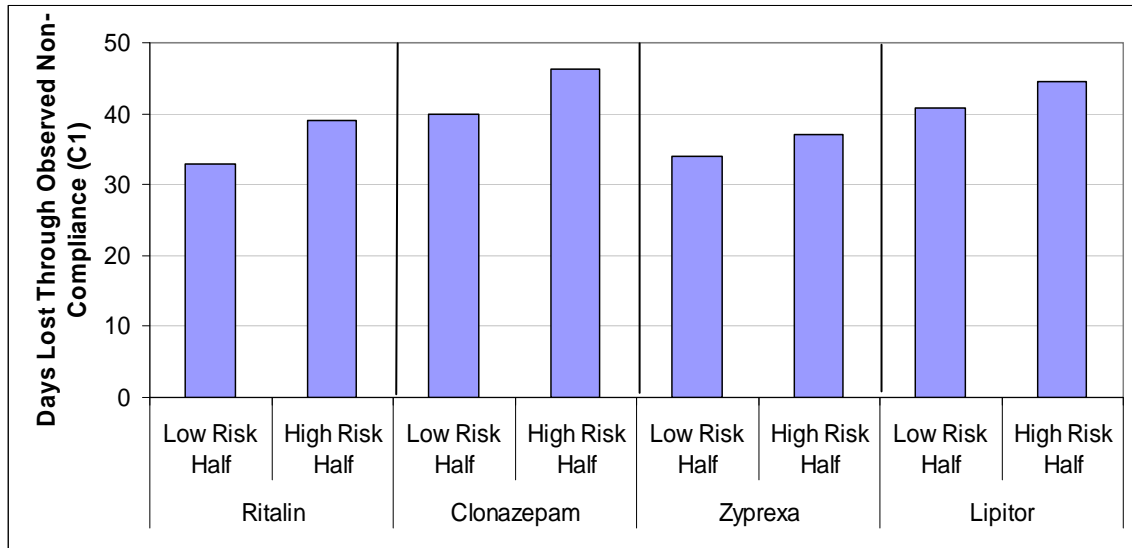


Figure 3b: NonPersistence or Unobserved Noncompliance (IC_2)

