Evolving Patient Compliance Trends: Integrating Clinical, Insurance, and Extrapolated Socioeconomic Data

Joseph J. Klobusicky, PhD, Arun Aryasomayajula, Nicholas Marko, MD
Geisinger Health System, Danville, PA

Abstract

Efforts toward improving patient compliance in medication focus on either identifying trends in patient features or studying changes through an intervention. Our study seeks to provide an important link between these two approaches through defining trends of evolving compliance. In addition to using clinical covariates provided through insurance claims and health records, we also extracted census based data to provide socioeconomic covariates such as income and population density. Through creating quadrants based on periods of medicine intake, we derive several novel definitions of compliance. These definitions revealed additional compliance trends through considering refill histories later in a patient's length of therapy. These results suggested that the link between patient features and compliance includes a temporal component, and should be considered in policymaking when identifying compliant subgroups.

Introduction

In the field of pharmacoeconomics, compliance (or adherence) to a medication is defined as the degree to which a patient follows instructions from a health care provider[1]. Compliance measures span over all fields of treatment, including preventative measures such as exercise and diet. However, the most intensely studied form of compliance is with prescription drug consumption, due to its major economic impact. Approximations for annual costs of noncompliance due to medication related hospital visits are as high as $100 billion[2,3,4,5], while the consulting group Capgemeni Worldwide recently published the striking estimate of $564 billion of annual costs suffered globally by pharmaceutical companies, of which American companies lose $188 billion[6]. Along with staggering economic losses, patient compliance also presents a major hurdle to patient health. For instance, compliance rates below 90% for HIV patients can cause viral replication and disease progression[7,8], while for diabetics, proper compliance is essential in preventing hypertension and myocardial infarction[9].

Given a specific prescription, compliance can be further classified with respect to the potential ways a patient can deviate from a provider's instructions. Primary compliance is defined as a patient's fidelity of filling and refilling prescriptions. Secondary compliance refers to whether a patient actually consumes their medication. While direct observation and blood tests can validate secondary compliance, these methods are mostly infeasible, and instead rely on methods such as surveys[1]. Our data focuses on an aspect of primary compliance which observes patterns of drug intake conditioned on the existence of a patient refill history. Our concern, therefore, is not the event that a patient has taken a medication, but rather the discrepancy between recommended and observed fill dates.

Studies based on claims data most commonly use MPR, or medication possession ratio, as a means for quantifying patient compliance. This ratio is defined as the total number of medication supply days divided by the total length of therapy. To illustrate, a patient with a prescription history spanning 60 days receiving 2 prescription fills for 15 days is assigned an MPR of 30/60, or .50. Studies have linked high MPR with commonly used covariates such as age, race, and sex[10], as well as insurance related quantities including out-of-pocket payments[11,12]. Our study diverges significantly from other investigations through a novel analysis of evolving patient trends. Rather than testing changes in compliance under some intervention (e.g. smart pill counters or medication reminders), we define notions of general, initial, eventual, and consistent compliance, and compare patient trends across these different features. These notions of compliance can address questions of policy, which seek not only to identify high risk patients, but also those patients most likely to respond to intervention.

Our main dataset is insurance claims data of prescription orders. While not accounting for secondary compliance, fill records provide an objective history of intake that is resistant to biased responses, which can arise in survey data. Insurance claim data is natural for questions concerning consistency of medication refills, as the dataset includes prescription fill dates, supply lengths, and refill numbers. This approach was taken by Soloman and colleagues[13] and Gibson and colleages[14] for studying the relation of cost sharing and compliance. Our statistical approach is similar to Rolnick and colleagues[10], who also use prescription fill data to examine trends, including census derived

covariates of education, income, and poverty. For our study, we use extrapolated data from the US Census and American Community Survey (ACS) to examine compliance with median family income and population density.

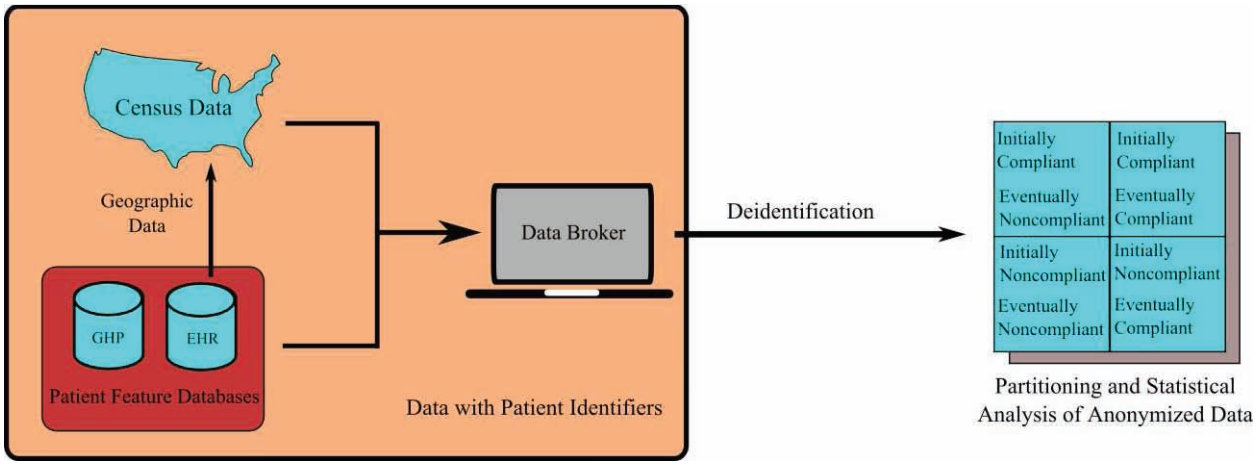

Figure 1: Data flow. A data broker combines GHP (insurance) and EHR (clinical) datasets, available from GHS database (lower left). Zip codes from the EHR dataset are used to obtain extrapolated census data (top left). The data broker (center) then deidentifies all data for statistical analyses (right).

Data Sources

All clinical and insurance claims data is aggregated through available databases of Geisinger Health System (GHS), a hospital located in central Pennsylvania which serves approximately 2.6 million people, mostly residing in surrounding rural areas. To collect clinical features such as body mass index (BMI), we restrict attention to patients who have both insurance and clinical data. The use of city-based geographic data, an instance of PHI (Protected Health Information), in addition to other deidentified data for this study was approved through an IRB.

Using zip code information from EHR data, we use census data to create an enhanced dataset containing socioeconomic factors, with coarseness at the city level. For the purposes of anonymization, we use the following scheme. A certified data broker deidentifies GHP and EHR data through random date shifting and random patient number identifier assignments. All PHI sensitive data is restricted to a work computer inaccessible to other investigators. From this computer, the data broker collects web-based data, and returns deidentified data, along with socioeconomic data, to an analyst for running statistical models. An illustration of the data flow between agents is given in Figure 1.

The chief dataset which contains information about prescription fills is collected through the Geisinger Health Plan (GHP), a subsidiary HMO of GHS. For each patient in the GHP database, there exists a collection of prescriptions taken, each with its separate history. For each of these drugs there are lists of dates corresponding to refill occurrences and supply length of medication. As the purpose of this study is to investigate changes in compliance over a series of refills, we will require at least three fills and a minimum of 180 day length of therapy. Each separate prescription for a person taking multiple prescriptions will be treated individually. Thus, in the sequel, to avoid confusion, we will use the term compliance of prescriptions, rather than compliance of patients, to denote scores for a single prescription. Through taking averages over all prescriptions, we can calculate patient compliance scores in a straightforward manner.

For a prescription with N refill dates $d_1, \dots, d_N$, with corresponding supply of prescription $s_1, \dots, s_N$ given in days, we can define the medical possession ratio (MPR) as

$$MPR = \frac{\sum_{i=1}^{N-1} s_i}{|d_N - d_1|},$$

where $|d_N - d_1|$ denotes the length in days between d_N and d_1. Note that the final fill is not taken into account when calculating MPR, as we have no way of deducing a patient's compliance if there is no subsequent refill. We will use the conventional definition of a compliant prescription as one with an MPR of greater than .80, a commonly used threshold.

Electronic health records (EHR) were obtained from patients who were also members of GHP. The entirety of data available from EHR is massive in both magnitude and variety, containing millions of records of patient medications, events such as surgeries, complications such as occurrence of disease, and lengths of stay. While this rich dataset is certainly of future interest for more specific investigations, for this study we have selected a restricted set of covariates, consisting of age, sex, race, and BMI. Socioeconomic data was provided through the 2010 US Census and 2006-2010 ACS, which provide specific socioeconomic features. For our additional variables, we have chosen median income and population density.

Methods:

Definitions of Evolving Compliance. We now define different ways of describing compliance based on changes in behavior in time. To do so, we separate each prescription with a length of therapy of L days into an initial period of day range 1 to $\lfloor L/2 \rfloor$, and a final period of day range of $\lfloor L/2 \rfloor + 1$ through L, where $\lfloor x \rfloor$ denotes the floor value of x. For each period and prescription, we calculate the corresponding MPR to give us initial and eventual compliance scores. This split defines each prescription as belonging to one of four quadrants Q1-Q4:

Q1: Consistently Compliant. A prescription is compliant in both initial and final periods.
Q2: Compliant to Noncompliant. A prescription is compliant in the initial period, and noncompliant in the final period.
Q3: Consistently Noncompliant. A prescription is noncompliant in both initial and final periods.
Q4: Noncompliant to Compliant. A prescription is noncompliant in the initial period, and compliant in the final period.

There are seven ways of dividing quadrants, either through comparing one quadrant against three quadrants, or two quadrants against another two quadrants. Our study focuses on three such divisions with conditions (see Figure 2):

1. A prescription is **consistently compliant** (Q1 vs. Q2, Q3, and Q4).
2. A prescription is **initially compliant** (Q1 and Q2 vs. Q3 and Q4).
3. A prescription is **eventually compliant** (Q1 and Q4 vs. Q2 and Q3).

We refer to a patient who is compliant over the entire length of therapy as **generally compliant**. Thus, each prescription was given four classifications based on which types of compliance are realized.

Statistical Design. For each variety of compliance, we ran chi-square tests of independence and logistic regression models through binarizing inputs. For EHR, we classified prescriptions as white, male, obese (BMI >30), or above median age (54 years old). For census data we classified by median household income ($43,837 per year) and population density (245 residents per square mile). We repeat the fact that individual prescriptions were taken as separate data points, meaning that a patient with N prescriptions provided N separate data points with repeated inputs and varied outputs of MPR. The full dataset consists of 75940 prescriptions for 8889 patients. The median MPR is 65.5%. This number is smaller than other studies which take MPR as a compliance measure for specific chronic conditions[10]. However, our study was over the space of all prescriptions, including those which were non-chronic or with more irregular drug intake patterns.

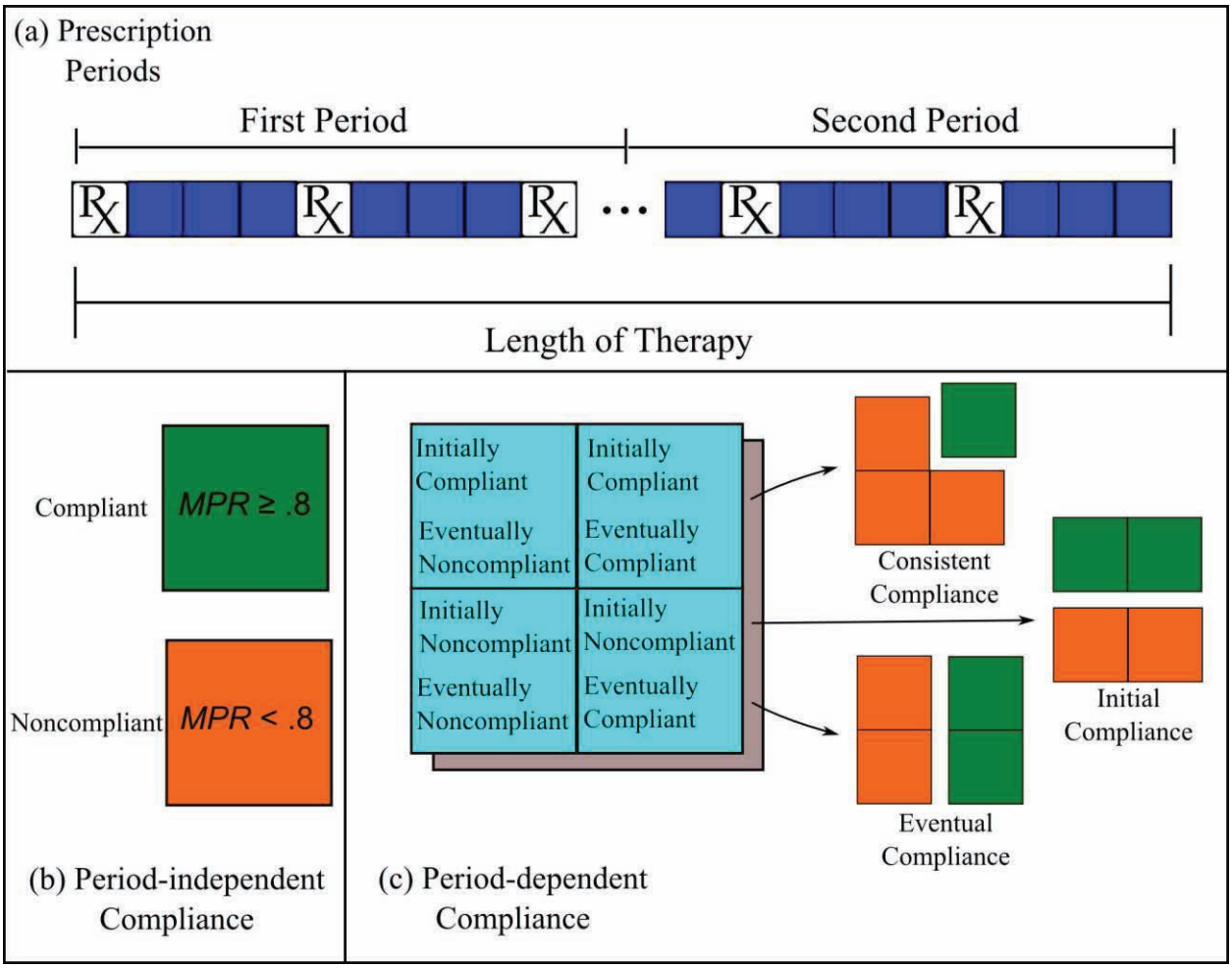

Figure 2: Variations of compliance. (a) A prescription record over a length of therapy consists of an initial first period and subsequent second period. Tiles denote individual days, and prescription symbols define fills. (b) Considering *MPR* only over the entire length of therapy, patients can be partitioned as compliant or noncompliant based on a threshold score of .80. (c) **Left:** Considering *MPR* over both first and second periods, patients can fall into one of four types of compliance behaviors. **Right:** Separating quadrants creates several definitions of compliance, including compliance over first period (initial), second period (eventual), and both periods (consistent).

Results

Summaries of chi-square tests for independence and multivariate logistic regression models are presented for the four compliance types in Tables 1-4 of the Appendix. The p values for logistic regression are based on a Wald test with the null hypothesis that odds ratios are equal to 1. Age and race are the strongest predictors of compliance, as white and older patients have higher scores across all definitions of compliance. While race has a greater difference of compliance scores than age, p values are consistently lower, as the sample population was largely homogenous (about 90% white). Males and obese patients are also more likely to have higher compliance scores, although differences in compliance are less pronounced than those for age and race. Census related data, income and density, produce the smallest differences in compliance, suggesting that high income and low population density related to higher compliance. However, while compliance scores are slightly higher for those living in more rural areas, we are not able to reject the null hypothesis that population density is independent of compliance.

Consistent and eventual compliance produce the most statistically significant variables for $p<.05$, whereas initial compliance produce the least. In particular, a small, but statistically significant, difference in scores was reported with respect to income data when observing consistent and eventual compliance, while there no such evidence of a difference with initial and general compliance. Also, race produces statistically significant differences in all definitions of compliance except for initial compliance.

Discussion

Changing Compliance Measures Imply Changing Significant Populations

While we observed certain covariates as statistically significant in terms of general compliance, the focus of this study was to develop alternative metrics which can capture information not available from traditional measures of compliance. In this analysis, we create variations in scores by considering compliance as a function dependent upon time. When considering compliance over the entire length of therapy for a patient, the more random nature of initial compliance can hide trends that more clearly occur after a patient has received several refills. A direct method of removing initial noise would consider either consistent or eventual compliance. While consistent compliance does, in fact, capture new socioeconomic trends, requirements are more stringent, leading to a smaller set of testable patients. The same trends are captured with eventual compliance, while providing a larger percentage of compliant patients. Thus, it is reasonable to condition medication compliance scores after several refills have taken place, especially in the case of studying chronic conditions spanning over several years. The question of how many refills we should exclude before testing would be specific to the type of condition of question, both in the inherent compliance statistics, as well as the relative important for early compliance to medication.

New perspectives of analyzing quadrants can also shed light on different topics of interest. One particularly interesting approach is to investigate behavior in quadrants Q2 and Q4, which describe patients who experience a change in compliance. In terms of policy, a practical option when considering target audiences might focus on those who are susceptible to changes. While our current set of covariates was too generic to provide a definitive profile of such patients, we have provided a metric that can measure such subgroups for new studies with more specific datasets.

New Directions for Compliance Metrics

Our extensions to medication compliance can be used for providing a notion of "persistence" for claims related studies, which is considered as a distinct measure[15]. Medication persistence is defined as the act of continuing a treatment for a prescribed duration. While claims data only offers information on prescriptions which have been filled, notions such as consistent compliance may be seen as a type of persistence with respect to the lag between drug consumption and refill dates. Notions of initial and eventual compliance should also be taken into account when conducting studies which focus on temporal changes, as the population may have tendencies toward compliance change. We also note that this type of analysis is an initial step for targeted intervention of patients who are likely to experience changes in compliance.

A natural extension of our study would consider MPR as a continuous function of time. One such approach for comparing MPR can be understood through comparing cumulative functions of total missed medication days. Through an appropriate normalization, we can then compare cumulative distributions for different prescription histories. While the common method for comparing such functions would be through calculating a Kolmogorov-Smirnov distance, the promising field of quantile regression, popular in econometrics, offers more relevant information on quantiles conditioned on certain feature sets[16]. While studies have used methods from quantile regression[17], they use general compliance rates as the object of study, without considering variations in time. Another point of research would examine the differences between adherence and a patient's medication portfolio, including focusing on frequency of intake effects on coupling certain drugs.

Another future direction of study involves the almost limitless source of data available from textscraping or data aggregation services. Strategies of utilizing social media tend toward intervention, where findings are currently mixed. A comprehensive study from Scheurer and colleagues[18] of over 5000 articles reached the conclusion that online social support networks, including social media, can increase compliance. However, the same study showed that "practical social support", such as offering transportation to pharmacies, offered the greatest increase. A less studied aspect uses alternative data sources for compliance prediction. Examples of success stories are from social feeds such as Twitter and Yelp restaurant reviews, which have been used to track the progression of influenza[19] and post-partum depression[20]. For socioeconomic issues, data aggregators can provide more specific public information which may relate to compliance, such as housing and criminal justice records. This additional inclusion of data may allow for a more advanced analysis of compliance scores through machine learning techniques.

Conclusion

Estimates of patient compliance are sensitive to temporal considerations. Trends toward eventual compliance are similar to those of overall compliance, while compliance near the beginning of a prescription tends to be more uniformly random over all features. This suggests that studies should consider conditioning compliance studies after a patient receives several refills. Future work will focus on more extrinsic data collection and partitioning methods that more accurately capture compliance trends.

References

1. Osterberg, L, Blaschke T. Adherence to medication. New England Journal of Medicine. 2005; 353.5: 487-497.
2. Dischler J, Wagner DJ, Raia JJ, Palmer-Shevlin N. Medication compliance: a healthcare problem. Vol. 27. Whitney Books Company; 1993.
3. Levy, G, Zamacona MK,, Jusko WJ. Developing compliance instructions for drug labeling. Clinical Pharmacology & Therapeutics. 2000; 68.6: 586-591.
4. Senst BL, Achusim LE, Genest RP, Cosentino LA, Ford CC, Little JA, Raybon SJ, Bates DW. Practical approach to determining costs and frequency of adverse drug events in a health care network. American Journal of Health-System Pharmacy. 2001;58.12: 1126-1132.
5. McDonnell PJ, Jacobs MR. Hospital admissions resulting from preventable adverse drug reactions. Annals of Pharmacotherapy. 2002;36.9: 1331-1336.
6. Forissier T, Firlik K. Estimated annual pharmaceutical revenue loss due to medication non-adherence. Capgemini Consulting & HealthPrize Technologies. [updated 12 November 12 2012]. Available from http://www. adherence564.com
7. Catz, SL, Kelly JA, Bogart LM, Benotsch EG, McAuliffe TL. Patterns, correlates, and barriers to medication adherence among persons prescribed new treatments for HIV disease. Health Psychology. 2000; 19.2: 124.
8. Hinkin, CH, Castellon SA, Durvasula RS, Hardy DJ, Lam MN, Mason KI, Thrasher D, Goetz MB, Stefaniak M. Medication adherence among HIV+ adults: Effects of cognitive dysfunction and regimen complexity. Neurology. 2002; 59.12: 1944-1950.
9. Rainer D. Overcoming barriers to effective blood pressure control in patients with hypertension. Current Medical Research and Opinion. 2006; 22.8: 1545-1553.
10. Rolnick SJ, Pawloski PA, Hedblom BD, Asche SE, Bruzek RJ. Patient characteristics associated with medication adherence. Clinical medicine & research 2013; cmr-2013.
11. Curkendall S, Patel V, Gleeson M, Campbell RS, Zagari M, Dubois R. Compliance with biologic therapies for rheumatoid arthritis: Do patient out-of-pocket payments matter? Arthritis Care & Research. 2008; 59.10: 1519-1526.
12. Piette JD, Heisler M, Wagner TH. Problems paying out-of-pocket medication costs among older adults with diabetes. Diabetes Care. 2004; 27.2: 384-391.
13. Solomon MD, Goldman DP, Joyce GF, Escarce JJ. Cost sharing and the initiation of drug therapy for the chronically ill. Archives of internal medicine. 2009; 169.8: 740-748.
14. Gibson TB, Mark TL, McGuigan KA, Axelsen K, Wang S. The effects of prescription drug copayments on statin adherence. The American journal of managed care. 2006; 12.9: 509-517.
15. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, Wong PK. Medication compliance and persistence: terminology and definitions. Value in Health. 2008; 11, no. 1: 44-47.d
16. Koenker R. Quantile regression. No. 38. Cambridge University Press, 2005.

17. Gebregziabher M , Lynch CP, Mueller M, Gilbert GE, Echols C, Zhao Y, Egede LE. Using quantile regression to investigate racial disparities in medication non-adherence. BMC medical research methodology. 2011; 11, no. 1: 88.
18. Scheurer D, Choudhry N, Swanton KA, Matlin O, Shrank W. Association between different types of social support and medication adherence. The American journal of managed care. 2012; 18, no. 12: e461-7.
19. Cassandra H, Jorder M, Stern H, Stavinsky F, Reddy V, Hanson H, Waechter H, Lowe L, Gravano L, Balter S. Using online reviews by restaurant patrons to identify unreported cases of foodborne illness—New York City, 2012–2013. MMWR. 2014;63:441–5.
20. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013; ACM

Appendix: Tables of Results

Table 1: **Statistics for General Compliance.** Chi-square and multivariate logistic regression tests are performed with respect to general compliance, or an $MPR \geq .80$ over length of therapy. Odd ratios are computed for the first vs. second entry in each category (age $\geq$ 54 years vs. age < 54 years, white vs. nonwhite, etc.) against probabilities of compliance.

Chi-Square Test for Independence		**Logistic Regression**	
Feature	**% Compliant**	**Feature**	**Odds Ratio**
Age $\geq$ 54 years	39.23***	Age	.78***
Age $<$ 54 years	32.50	Race	.70***
White	35.60**	BMI	.84***
Nonwhite	26.35	Sex	.88**
Obese (BMI $\geq$ 30)	37.20*	Income	.98
Not Obese	34.98	Density	.99
Male	37.04**	***: p value < .001 ** : p value < .01 * : p value < .05	
Female	34.24		
Family Income $\geq$ $43,837 per year	35.59		
Family Income $<$ $43,837 per year	34.57		
Pop. Density $\geq$ 245 per sq. mile	34.61		
Pop. Density $<$ 245 per sq. mile	35.51		

Table 2: **Statistics for Consistent Compliance.** Chi-square and multivariate logistic regression tests are performed with respect to consistent compliance, or an $MPR \geq .80$ over both first and second periods of therapy. Odd ratios are computed for the first vs. second entry in each category (age $\geq$ 54 years vs. age < 54 years, white vs. nonwhite, etc.) against probabilities of compliance.

Chi-Square Test for Independence		**Logistic Regression**	
Feature	**% Compliant**	**Feature**	**Odds Ratio**
Age $\geq$ 54 years	25.84***	Age	.73***
Age < 54 years	19.39	Race	.56***
White	22.36***	BMI	.80***
Nonwhite	12.84	Sex	.86**
Obese (BMI $\geq$ 30)	23.88*	Income	.88*
Not Obese	21.68	Density	.95
Male	23.52***	***: p value < .001 ** : p value < .01 * : p value < .05	
Female	21.14		
Family Income $\geq$ $43,837 per year	22.69*		
Family Income < $43,837 per year	20.64		
Pop. Density $\geq$ 245 per sq. mile	21.53		
Pop. Density < 245 per sq. mile	22.18		

Table 3: **Statistics for Initial Compliance.** Chi-square and multivariate logistic regression tests are performed with respect to initial compliance, or an $MPR \geq .80$ over the first period of therapy. Odd ratios are computed for the first vs. second entry in each category (age $\geq$ 54 years vs. age < 54 years, white vs. nonwhite, etc.) against probabilities of compliance.

Chi-Square Test for Independence		**Logistic Regression**	
Feature	**% Compliant**	**Feature**	**Odds Ratio**
Age $\geq$ 54 years	52.71***	Age	.83***
Age < 54 years	47.12	Race	.86
White	49.67	BMI	.85***
Nonwhite	44.59	Sex	.88**
Obese (BMI $\geq$ 30)	51.37*	Income	1.01
Not Obese	49.08	Density	.99
Male	51.43**	***: p value < .001 ** : p value < .01 * : p value < .05	
Female	48.34		
Family Income $\geq$ $43,837 per year	49.54		
Family Income < $43,837 per year	49.29		
Pop. Density $\geq$ 245 per sq. mile	49.61		
Pop. Density < 245 per sq. mile	48.97		

Table 4: **Statistics for Eventual Compliance.** Chi-square and multivariate logistic regression tests are performed with respect to consistent compliance, or an $MPR \geq .80$ over the second period of therapy. Odd ratios are computed for the first vs. second entry in each category (age $\geq$ 54 years vs. age < 54 years, white vs. nonwhite, etc.) against probabilities of compliance.

Chi-Square Test for Independence		**Logistic Regression**	
Feature	**% Compliant**	**Feature**	**Odds Ratio**
Age $\geq$ 54 years	47.82***	Age	.80***
Age $<$ 54 years	41.98	Race	.74*
White	44.54**	BMI	.84***
Nonwhite	36.15	Sex	.94***
Obese (BMI $\geq$ 30)	46.29*	Income	.87*
Not Obese	43.86	Density	.91
Male	44.99	***: p value < .001 ** : p value < .01 * : p value < .05	
Female	43.81		
Family Income $\geq$ $43,837 per year	45.04*		
Family Income $<$ $43,837 per year	42.62		
Pop. Density $\geq$ 245 per sq. mile	44.88		
Pop. Density $<$ 245 per sq. mile	44.10		