Developing a Model of Psychiatric Visit Non-Adherence

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfilment

of the Requirements for the Degree

Doctor of Philosophy

By

PATRICIA E. ALAFAIREET

Dr. Grant T. Savage, Dissertation Co-Supervisor
Dr. Yang Gong, Dissertation Co-Supervisor

May 2010

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

Developing a Model of Psychiatric Visit Non-Adherence

presented by Patricia Alafaireet,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Grant T. Savage, PhD

Yang Gong, MD, PhD

Gregory F. Petroski, PhD

Lanis L. Hicks, PhD

Howard L. Houghton, MD

This dissertation is dedicated to Earl D. Curry, who taught me, by word and deed that working hard is essential to one's goals.

Special Acknowledgement to:

Pam Roddy and Ana Fernandez, who prayed me thorough the hard times.

Cathy Marshall, who took me out to dinner often and let me vent.

Christine Curry, who makes the Rock of Gibraltar look shaky by comparison

My children, Hussain, Layla, Alia, and Lamia, who "egged" me on, and cheerfully ate macaroni and cheese for dinner many nights.

Acknowledgements

This study could not have been completed without the contribution of a number of individuals who shared knowledge and encouragement throughout the process. These individuals include:

Dr. Yang Gong, who served as co-chair of my dissertation committee and academic advisor. His ability to see to the heart of the matter and make gentle suggestions for improvements is unparalleled.

Dr. Grant Savage, who also served as co-chair of my dissertation committee and publication editor par excellence. If I've learned to write without "zusammen-gehen", it's because of his tireless efforts in that direction.

Dr. Howard Houghton, who was my partner in investigating psychiatric visit non-adherence long before the dissertation process began. May we have many more years of fruitful research.

Dr. Greg Petroski, whose depth of experience in bio-statistical analysis is phenomenal and who supported this work with that experience. Thanks Greg, for not laughing at some of the questions I asked.

Dr. Lanis Hicks, who served as a role model and a mentor long before helping with this dissertation. Thank you for all the editing and all the support.

Dr. Sue Boren, whose unstinting efforts smoothed the path to success many times.

Dr. Gordon Brown, who listened all those years ago when I said that a Master’s in Health Administration wasn’t enough, and then did something about it. This PhD program would not be in existence without your vision.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii
LIST OF ILLUSTRATIONS
LIST OF TABLES
NOMENCLATURE
ABSTRACT

Chapter

1. INTRODUCTION 1

2. LITERATURE REVIEW 8

    Search Strategy 8

    Specific Aims 10

    Determinant Identification 14

    Analysis Strategy Identification 26

3. RESEARCH METHODOLOGY 34

    Preliminary Studies 36

    Current Work 48

4. FINDINGS AND DISCUSSION 86

    Regression Results 88

    Data Mining Results 152

    Model Comparison Results 161

    Limitations 162

    Discussion 168

5. CONCLUSIONS AND RECOMMENDATIONS 172

APPENDIX

1. DEVELOPMENT DATA DESCRIPTIVE TABLES 179
2. SAS COMMANDS 199
3. DEVLOPEMENT DATA ANALYSIS TABLES 203

BIBLIOGRAPHY 221

VITA 238

LIST OF ILLUSTRATIONS

Figure		Page
1.	Visit Non-Adherence Cycle	1
2.	Description of Research Project Flow	5
3.	Data Management and Analysis Cycle	52
4.	Testing Flow	86
5.	Statistical Analysis Process	88
6.	Distribution of Gender by Status	98
7.	Distribution of Age by Status	99
8.	Distribution of Marital Status by Status	100
9.	Distribution of Employment Status by Status	101
10.	Distribution of Patient Race by Status	102
11.	Distribution of Payer by Status	103
12.	Distribution of Relationship of Contact Person to Patient by Status	104
13.	Distribution of General (Primary) Diagnosis by Status	105
14.	Distribution of Secondary Diagnosis by Status	106
15.	Distribution of Travel Distance by Status	107
16.	Distribution of Wait Days by Status	108
17.	Distribution of Time of Day by Status	109
18.	Distribution of Day of Week by Status	110
19.	Distribution of Service Date (Season) by Status	111
20.	Distribution of Use of Non-MD Mental Health Services	112
21.	Distribution of Number of Cancelled Appointments (Numbercnx) by Status	113
22.	Distribution of Number of Appointment by Status	113
23.	Distribution of Maker by Status	114
24.	Distribution of Provider by Status	115
25.	Distribution of Referring Provider by Status	116
26.	Distribution of Same Visit Type by Status	117
27.	Distribution of Same Day of Week by Status	117
28.	Distribution of Same Time of Day by Status	118
29.	Distribution of Same Payer by Status	118
30.	Distribution of Same Provider Type by Status	119
31.	Distribution of Percent Non-Adherent (Percentnon) by Status	120
32.	Distribution of Previous Visit Non-Adherent by Status	120
33.	Distribution of Payer by Race	121
34.	Distribution of Provider by Race	122
35.	Distribution of Gender by Age	122
36.	Distribution of Payer by Employment	123
37.	Frequency Distribution of Age by Waitdays	134
38.	Frequency Distribution of Age by Season	135
39.	Frequency Distribution of Age by Day of Week	135
40.	Frequency Distribution of Age by Marital Status	136
41.	Frequency Distribution of Age by Prev Visit NOS	136
42.	Frequency Distribution of Age by Provider Type	137

43.	Frequency Distribution of Wait Days by Appt Day of Week	137
44.	Frequency Distribution of Wait Days by Marital Status	138
45.	Frequency Distribution of Wait Days by Prev Visits NOS	138
46.	Frequency Distribution of Season by Appt Day of Week	139
47.	Frequency Distribution of Season by Wait Days	139
48.	Frequency Distribution of Season by Marital Status	140
49.	Frequency Distribution of Season by Prev Visit NOS	140
50.	Frequency Distribution of Same Payer by Season	141
51.	Frequency Distribution of Same Payer by Wait Days	141
52.	Frequency Distribution of Same Payer by Marital Status	142
53.	Frequency Distribution of Same Payer by Day of the Week	143
54.	Frequency Distribution of Marital Status by Previous Visit NOS	143
55.	Frequency Distribution of Day of Week by Marital Status	144
56.	Frequency Distribution of Appt. Day of Week by Prev Visit NOS	144
57.	Frequency Distribution of Provider by Same Payer	145
58.	Frequency Distribution of Provider by Season	146
59.	Frequency Distribution of Provider by Wait Days	146
60.	Frequency Distribution of Provider by Day of Week	147
61.	Frequency Distribution of Provider by Marital Status	147
62.	Frequency Distribution of Provider by Prev Visit NOS	148
63.	Data Mining Process Flow	153
64.	Classification Tree	160
65.	Distribution of Referrals for Attention Deficit Disorder	163
66.	Distribution of Referrals for Anxiety	163
67.	Distribution of Referrals for Bi-polar Disorder	164
68.	Distribution of Referrals for Depression	164
69.	Distribution of Referrals for Dementia	165
70.	Distribution of Referrals for Psychosis	165
71.	Distribution of Referrals for Drug/Alcohol Abuse	166
72.	Distribution of Referrals for Other Diagnoses	166

LIST OF TABLES

Table		Page
1.	Literature Support for Specific Aims	10
2.	Determinant Selection	43
3	Summary of Data and Tools	47
4.	Data Completeness	58
5.	Initial Determinant Classification	60
6.	Data Formatting	70
7.	Comparison of Study Determinants to Determinants in Potentially Useful Predictive Models	84
8.	Description of Regression Success Rate in Initial Trial	91
9.	Frequency of Occurrence of Determinants in Initial Trial	92
10..	Results of Regressions sans Number of Non-Adherent Visits	93
11.	Frequency of Occurrence of Determinants sans Number of Non-Adherent Visits	94
12.	Means Procedure Results for Age	95
13.	Means Procedure Results for Travel Distance	96
14.	Means Procedure Results for Wait Days	96
15.	Means Procedure Results for Number of Cancelled Appointments (CNX)	97
16.	Means Procedure Results for Number of Appointments	97
17.	Intermediate Frequency Procedure Results for Status by Each Categorical Determinant	125
18.	Effect Entered for Bootstrap 1000 Replicates	128
19.	Effect Entered for Bootstrap 100 Replicates	128
20.	Summary of Effects Entered	129
21.	Effects Entered	130
22.	Analysis of Findings	130
23.	Effects Entered with Payer Removed	131
24.	Analysis of Findings with Payer Removed	131
25.	Effects Entered with Payer and Gen Diagnosis Removed	132
26.	Analysis of Findings with Payer and Gen Diagnosis Removed	132
27.	Analysis of Maximum Likelihood Estimates for Final Model	132
28.	Odds Ratio Estimates for Final Model	133
29.	Association of Predicted Probabilities and Observed Responses in Final Model	134
30	Classification Table for Final Model	134
31.	Analysis of Findings in Validation Model	149
32.	Analysis of Maximum Likelihood Estimates for Validation Model	150
33.	Odds Ratio Estimates for Validation Model	150
34.	Association of Predicted Probabilities and Observed Responses in Validation Model	151
35.	Model Summary for Complete Determinant Set	151
36.	Frequency Comparisons Between Development Data and Validation Data	152

37.	Model Summary for Complete Determianat Set	157
38.	Analysis of Accuracy with Complete Determinant Set	157
39.	Model Summary for Reduced Determinant Set	158
40.	Analysis of Accuracy with Reduced Determinant Set	158
41.	Model Summary for Testing Determinant Set	159
42.	Analysis of Accuracy with Testing Determinant Set	159
43.	Model Summary for J48 Classifier Determinant Set	160
44.	Analysis of Accuracy with J48 Classifier Determinant Set	160
45.	Comparison of Study Determinants to Determinants in Potentially Useful Predictive Models	162
46.	Status by Gender	179
47.	Status by Age	179
48.	Distribution of Marital Status by Status	181
49.	Distribution of Employment Status by Status	181
50,	Distribution of Patient Race by Status	181
51.	Distribution of Payer by Status	182
52.	Distribution of Relationship of Contact Person to Patient by Status	182
53.	Distribution of General (Primary) Diagnosis by Status	182
54.	Distribution of Secondary Diagnosis by Status	183
55.	Distribution of Travel Distance by Status	183
56.	Distribution of Wait Days by Status	186
57.	Distribution of Appointment Time by Status	190
58.	Distribution of Day of Week by Status	190
59.	Distribution of Service Date (Season) by Status	190
60.	Distribution of Use of Non-MD Mental Health Services	190
61.	Distribution of Number of Cancelled Appointments (numbercnx) by Status	191
62.	Distribution of Number of Appointment by Status	191
63.	Distribution of Maker by Status	192
64.	Distribution of Provider Type by Status	192
65.	Distribution of Referring Provider by Status	193
66.	Same Visit Distribution of Same Visit Type by Status	193
67.	Distribution of Same Day of Week by Status	193
68.	Distribution of Same Time of Day by Status	193
69.	Distribution of Same Payer by Status	194
70.	Distribution of Same Provider Type by Status	194
71.	Distribution of Percent Non-Adherent (Percentnon) by Status	194
72.	Distribution of Previous Visit Non-Adherent by Status	195
73.	Distribution of Payer by Race	195
74.	Distribution of Race by Provider Type	195
75.	Distribution of Gender by Age	196
76.	Distribution of Payer by Employment	198
77.	Distribution of Payer by Employment Continued	198
78.	Frequency Distribution of Age by Waitdays	203
79.	Frequency Distribution of Age by Season	205

80.	Frequency Distribution of Age by Day of Week	207
81.	Frequency Distribution of Age by Marital Status	209
82.	Frequency Distribution of Age by Prev Visit NOS	211
83.	Frequency Distribution of Age by Prev Visit NOS	213
84.	Frequency Distribution of Wait Days by Appt Day of Week	215
85	Frequency Distribution of Wait Days by Marital Status	215
86.	Frequency Distribution of Wait Days by Prev Visits NOS	215
87.	Frequency Distribution of Season by Appt Day of Week	216
88.	Frequency Distribution of Season by Wait Days	216
89.	Frequency Distribution of Season by Marital Status	216
90.	Frequency Distribution of Season by Prev Visit NOS	217
91.	Frequency Distribution of Same Payer by Season	217
92.	Frequency Distribution of Same Payer by Wait Days	217
93.	Frequency Distribution of Same Payer by Marital Status	217
94.	Frequency Distribution of Same Payer by Day of the Week	218
95.	Frequency Distribution of Same Payer by Previous Visit NOS	218
96.	Frequency Distribution of Marital Status by Previous Visit NOS	218
97.	Frequency Distribution of Day of Week by Marital Status	218
98.	Frequency Distribution of Provider by Same Payer	219
99.	Frequency Distribution of Appointment Day of Week by Prev Visit NOS	219
100.	Frequency Distribution of Provider by Season	219
101.	Frequency Distribution of Provider by Wait Days	219
102.	Frequency Distribution of Provider by Day of Week	220
103.	Frequency Distribution of Provider by Marital Status	220
104.	Frequency Distribution of Provider by Prev Visit NOS	220

NOMENCLATURE

ADT - Admission, Discharge and Transfer system. Commonly used in health care settings to manage scheduling and billing functions.

CPOE - Computerized Physician Order Entry. A software application that permits the digital transfer of physician's orders.

Determinant - a contributor to a result, sometimes, alternatively, known as an attribute, variable or correlate

EMR - Electronic Medical Record. Electronic charting system.

ICD-9CM - International Classification of Disease. Version 9 Clinical Modification was used in the study.

Visit - A visit is a set of characteristics of the individual, his or her diagnosis, and of the logistical aspects of service provision/acquisition.

New visit- For the purposes of this study an appointment is defined as "new" if the patient has not been seen in the clinic of study in the prior three years.

Non-adherent visit - A set of circumstances in which a combination of the socio-economic condition (of the patient), his/her disease state, and manner in which his/her care is assessed and provided, results in the non-occurrence of a planned meeting of the individual with a health care provider. From this perspective, visit non-adherence is not attributable directly to either a "bad patient" or a "bad schedule".

Return visit - For the purposes of this study an appointment is defined as "return" if the patient has been seen in the of study in the three years.

Visit - A visit is a set of characteristics of the individual, his or her diagnosis, and of the logistical aspects of service provision/acquisition.

WEKA - Waikato Environment for Knowledge Analysis is a popular suite of machine learning software written in Java, developed at the University of Waikato. WEKA is free software available under the GNU General Public License.

ABSTRACT

Non-adherence to psychiatry visits costs the US mental health care system more than one hundred billion dollars annually [1]. Non-adherent visits undermine improvements to patient care quality, erode patient well-being, and prevent the effective use of technology driven improvements to health care quality. Psychiatric visit non-attendance is often perceived as an intractable problem, because of the direction taken in previous studies of the problem. Previous research into the issue of visit non-adherence focus either on specific patient demographics or on redundant scheduling methods, neither of which addresses quality of care issues or the development of useful tools to decrease visit non-adherence. This formative study addressed the issue of visit non-adherence by leveraging readily available electronic billing and scheduling system data, as well as data from an EMR, to identify and analyze a set of determinants of visit non-adherence. Three strategies, statistical analysis, machine learning/data mining and model comparison, were utilized in the analysis. Results from this multi-phase study provide a parsimonious set of visit non-adherence determinants and a useful model based on those determinants capable of supporting the development of predictive tools suitable for use in ambulatory health care services delivery.

Key words: Adherence, adherent appointment, outpatient appointment, non-adherent visits, predictive model, predictive tool.

CHAPTER 1- INTRODUCTION

This study reflects an effort to address the need for a predictive model that can be used to identify non-adherent visits prior to their occurrence. An effort is made to develop a replicable and extensible process through which health care providers and their staff members, especially those providing ambulatory mental health services, can predict, prior to appointment scheduling, if the planned appointment is likely to be attended. Visit non-attendance (non-adherence to visits) creates an expensive and vicious cycle that typically requires additional care and logistical resources as part of a worsening disease process, which, in turn, increases the costs for care, results in loss of revenue (though decreased provider productivity), and decreases the quality of care delivered. [2-14] Please see Figure 1 below for a schematic of the cycle of non-adherence.

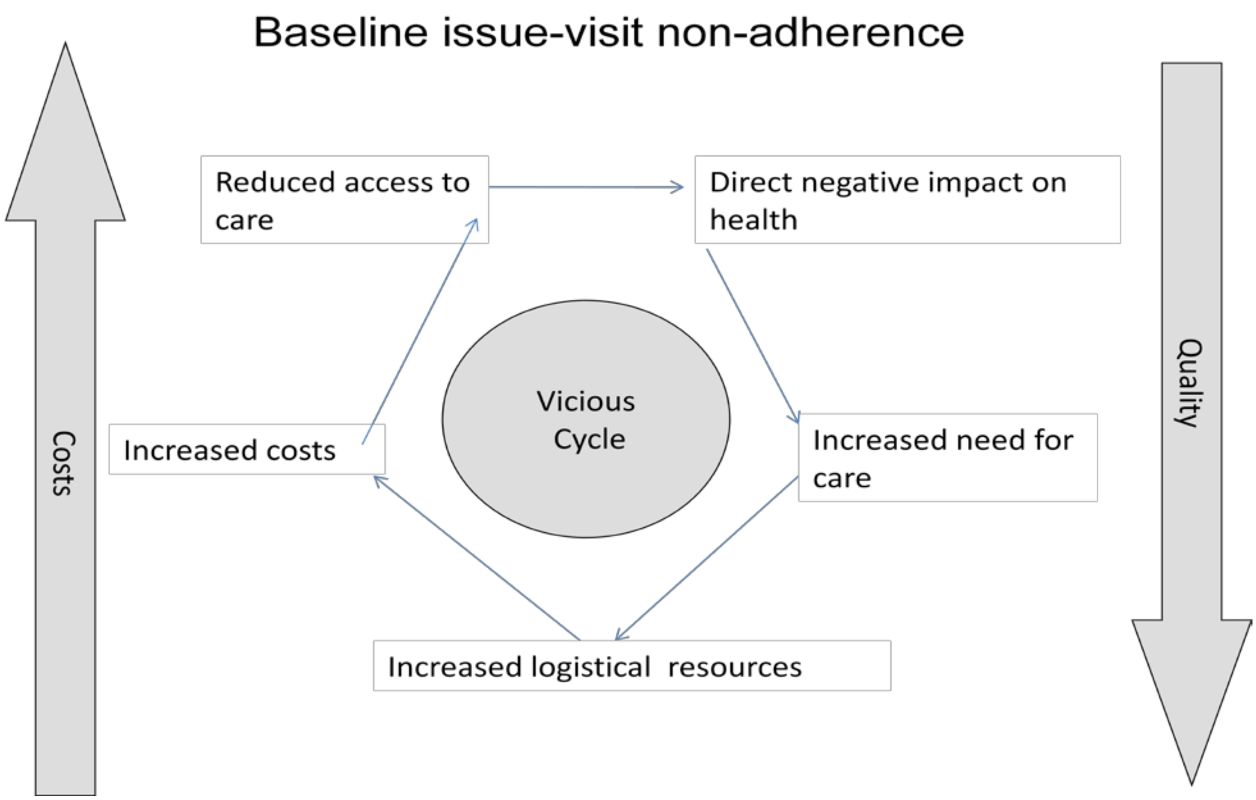

Figure1. Visit Non-Adherence Cycle

These increased costs and worsening disease states further limit the availability of, and access to, appropriate care and the vicious cycle continues, often until it is arrested by either extreme morbidity and/or by mortality [15-18]. Previous studies of visit non-adherence focus on either specific demographic characteristics of groups of patients or on designing scheduling strategies sufficiently redundant to compensate for the operational effects of non-adherent visits as a means to address quality of care guidelines [11, 16, 19-34].

Neither strategy encompasses the entirety of the circumstances that create visit non-adherence. Moreover, pursuit of these strategies has fostered the mindset that non-adherence to visits is an intractable problem, or that prediction of non-adherence is little better than a coin toss [35]. The lack of a gold standard for measuring visit non-adherence in health care has further complicated the issue [36]. Additional research is needed to provide useful methods of identifying the likelihood of visit non-adherence [37-41]. Research that includes socio-behavioral characteristics of patients, aspects of the disease and its therapy, and organizational features of the providing facility may prove to be more important in the prediction of visit non-adherence than research centering on patient demographics and scheduling interventions [42-45].

The true costs of visit non-adherence are multi-dimensional. In 2001, visit non-adherence represented a $100 billion drain on mental health care provided in the US, as well as a significant opportunity for improvement in the level of human suffering [1, 46-49]. For example, 40% of schizophrenic patient are re-hospitalized as a result of discontinuing their antipsychotic medications (for which, generally, attendance at outpatient visits is required) [6].

It is estimated that 75 percent of patients miss at least one appointment in 18 months of visits, and that 30% miss three or more over the period of one year [20]. Visit non-adherence creates a 3% to 14% revenue shortfall, even when the scheduling slots given to non-adherent visits are filled with walk-in patients [16]. Reminders, via phone calls for upcoming appointments, do not increase the short-term revenue of a practice [50]. Reminders also appear to have little effect on visit adherence [51].

Prevention of visit non-adherence, therefore, is an essential consideration in healthcare that directly affects the quality of care delivered and the availability of healthcare resources, including access to care [18, 52-54]. Additionally, preventative health care, patient centered care, and even EMRs are inherently sub-optimized as strategies to lower health care costs and increase access to care because the *structure of visit non-adherence* is insufficiently modeled.

This formative study draws upon research done on various aspects of medical adherence, including visit non-adherence in health care [3, 5, 33, 55-66]. But, this study draws also on similar types of work done in the airline (seat abandonment), hospitality (hotel room reservation abandonment), cinema (box office success prediction), and banking (credit scores) industries [67-74]. The phenomenon of visit non-adherence may have been best studied in airline and hotel reservations [69, 75-77]. Viable work in these industries has, for the most part, occurred only after the advent of sophisticated electronic data collection and the accumulation of data at the individual passenger (or hotel guest) level [67, 70]. Such work, while providing useful strategies for similar work in healthcare, is constrained by its motivation (i.e., increased profitability) and does not reflect the level of acuity represented in non-adherent visits in healthcare [78-79].

Little is known about the multiple determinants of visit non-adherence for a health care visit and less is known about how these determinants interact. This lacunae has further fostered the mindset that visit non-adherence is an intractable problem. Few studies have attempted to determine a model of visit non-adherence. The sole example, a recent study conducted by researchers at Purdue, adopts a similar strategy to that planned in this research, but is limited in both the range of determinants and in the population of patient visits used [80]. Selecting visit non-adherence determinants is an essential step toward developing a predicative algorithm to maximize visit adherence; therefore, this formative study is composed of three specific aims [23]. Please see Figure 2 below, which describes the research process flow and the goals of each stage. The first aim of this phase of study is to refine a group of determinants. The second aim is to use these determinants to build a useful model of visit non-adherence. And third, this phase of study seeks to compare the resulting model with other such models currently in use as a means of determining the potential utility of the new model.

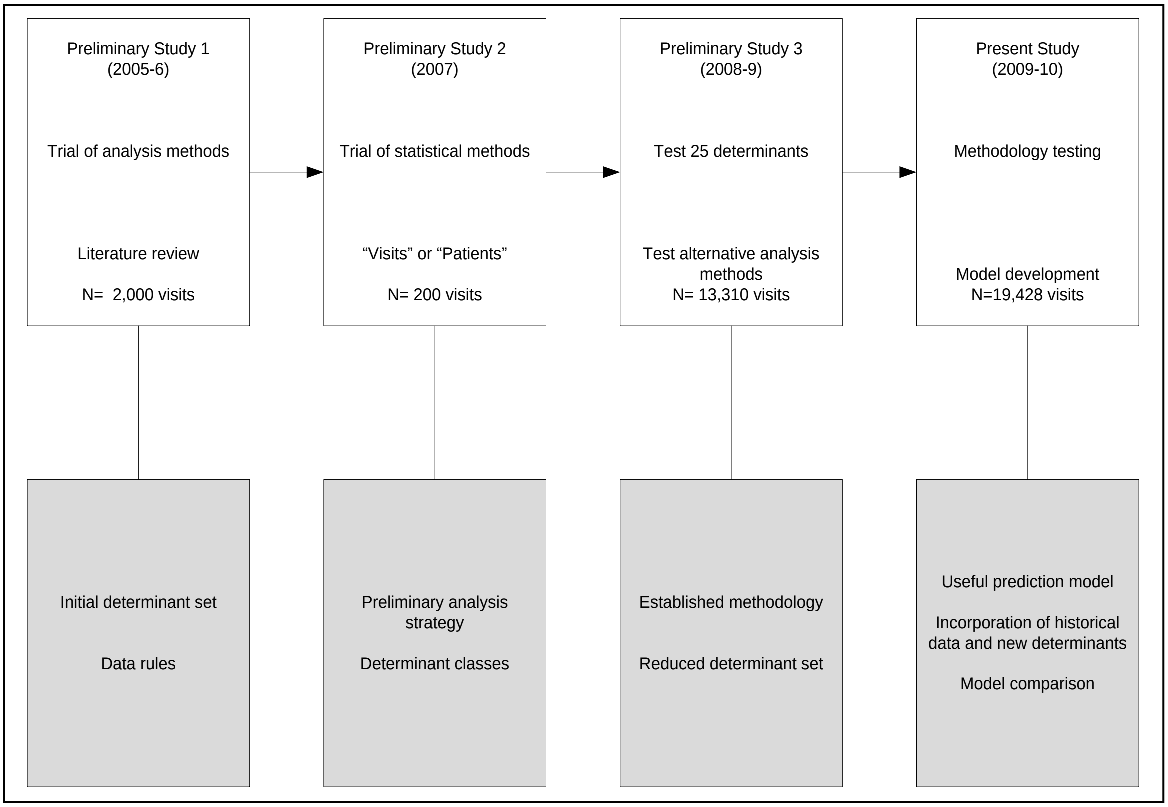

Figure 2. Description of Research Project Flow

Ambulatory psychiatry was selected as the laboratory for this study because mental disorders are common (about 30%) in the general population, and because psychiatry clinics typically function with an 11 to 19 percent non-adherence rate [42, 81-83]. Individuals with mental health conditions also increase the need for other health services, because patients with mental health problems frequently consume two or more times normal amounts of other medical services [6-7, 15-17, 47-48, 54, 84]. The presence of a psychiatric disorder, often depression or anxiety, is also significantly associated with missed *medical* appointments [85]. Untreated depression is also associated with increased costs (of approximately 50%) for patients who have a chronic illness [8, 17]. Additionally, mental health issues are contributing factors in the consumption of other non-mental healthcare resources [18, 52-53] and may affect the health of those individuals close to the patient [86-87].

The Psychiatry Clinic at the University of Missouri was selected as a test site because it had an interest in finding a solution to visit non-adherence and because it, typically, functioned with an eleven to nineteen percent visit non-adherence (no-show) rate. The Psychiatry Clinic at the University of Missouri had several other characteristics that enhanced its utility as a test site. As part of a larger system, it provided not only access to more than 13,000 annual outpatient visits, but also access to schedules of a variety of mental health care providers and to visits for patients with a wide range of mental health conditions. A maturing EMR and a mature billing and scheduling system provided electronic access to large amounts of data.

This research can create a significant paradigm shift, if knowledge gained from it results in a solution that replaces the current “bad patient” or “bad schedule” mentality with an evidence-based solution for minimizing non-adherent visits. The ultimate objective of this research is to develop a model of visit non-adherence, which, in turn, would allow for development of a decision support tool that health care providers and their staff members could use to reduce the rate of visit non-adherence to that defined in quality of care guidelines [25, 88-89]. Specifically, the intent of this study is to introduce a new, evidence-based model of visit non-adherence, to improve the prediction of ambulatory psychiatric visit non-adherence, and to support the development of visit non-adherence “aware” scheduling tools, as most of the knowledge currently available regarding visit non-adherence has yet to be incorporated into successful interventions [90-92].

The following document includes a chapter of review of the relevant literature, followed by a chapter that contains a description of the study methodology. The chapter on methodology is followed by a report of the findings. A final chapter discusses the contribution of the work to the science of health informatics and to the practice of medicine, as well as future work.

CHAPTER 2- LITERATURE REVIEW

Search Strategy

The first step towards understanding outpatient psychiatric visit non-adherence was an in-depth review of the relevant literature. Despite limitations imposed by the literature's focus on models consistent with the business practices of for-profit service providers, or on patient demographic characteristics, or scheduling as a means of visit non-adherence control, the body of relevant literature provided a starting point regarding the suitability of some determinants [22, 31, 41, 93-97].

The following databases and browsers were used to conduct the study:

- Ovid- Medline
- Ovid- Ie Compendex Plus
- Ovid-PsycInfo
- Ovid-CINAHL
- ABI-Inform
- ACM-Digital library
- Google, including Google Scholar

Search terms used included: Adherence, adherent, airline, appointment, broken, credit score, drug, Fair Isaac, GAIL, general additive, general linear, health, care, information, inventory control, loss control, loss management, medical, medication, missed, model, no-show, passenger, patient, patron, physician, plan, predictive, predicative, psychiatric, psychiatry, system, theater, treatment, vantage, visit.

Searches were also carried out with key words to further develop the set of literature. The key words used included *adherence, adherent appointment, outpatient appointment, non-adherent visits, predictive model, and predicative tool.*

The body of literature considered was further constrained to articles published in peer-reviewed journals, with the exception of articles dealing with credit scoring (as this information was restricted by its proprietary nature). Articles were further restricted to those that had a strong quantitative analysis, preferably a controlled trial. Preference was also given to articles published in the last ten years, but exceptions were made for seminal articles. All literature in the review was published in either English or German.

This literature review was investigative in nature. Initially, the concentration was on a review of all articles that provided insight into the problem of visit non-adherence (also known as no-show, or missed or broken appointments) in health care. Because of the obvious lack of truly successful prediction models and solutions to the visit non-adherence problem, and because the need for further research in this arena was clearly identified in the literature, the focus of the literature review was broadened to include the general issue of non-compliance with care in health [98]. This allowed incorporation of a body of literature encompassing adherence behaviors to medical regime, to medication, and to treatment plans to be considered. Further expanding the scope of understanding, literature was reviewed to identify studies (and conclusions) of visit non-adherence in service industries. Specifically, studies with models to predict non-adherence in the airline industry, the hotel industry (room reservations), in banking (in the form of credit scores) and the cinema (box offices success) were examined. In addition, significant literature review was done to identify potentially useful data collection and analysis methods.

Specific Aims

Practically, the literature review was carried out with two specific aims in mind: 1) to identify potential determinants of visit non-adherence; and 2) to identify and consider the utility of a number of analytic techniques and subsequent models that might be applied to the available visit non-adherence data. Because none of the articles currently available in the literature directly addressed the issue of visit non-adherence in the manner proposed in this study, it was necessary to draw information from a total of 384 articles to form a composite view of both potential determinants and of potential methods by which visit adherence data could be analyzed in a manner that would lead to the successful formulation of a model. Please see Table 1 below for an overview of the supporting literature.

Table 1
Literature Support for Specific Aims

Specific Aim 1- Potential Determinants	Supporting Articles	Specific Aim 2- Model Development	Supporting Articles
Patient Gender	140	Analysis Methods(General)	76
Patient Age	202	Multi-Variant Analysis	2
Patient Marital Status	3	General Additive Model	6
Patient Employment	11	Medical Adherence Analysis	32
Patient Race	26	Drug Adherence Analysis	15
Payer Type	11	Treatment Plan Adherence	4
Contact Person	6	Psychiatric Visit Adherence	156
Primary Diagnosis	91	General Clinic Visits Adherence	46
Secondary Diagnosis	91	Air Flight Reservation Adherence	11
Travel Distance	2	Hotel Reservation Adherence	5
Wait Days	18	Banking/Credit Score Creation	10
Appointment Type	14	Theater Attendance	11
Appt. Time of Day	4	General Prediction Methods	90
Appt. Day of Week	1	Medical Adherence Prediction	13
Appointment Date	3	Airline Reservation Analysis	8
Use of Non-MD Providers	33	GAIL Model Methods	3
Total Appointments	14	Fair Isaac Credit Scoring	4

Specific Aim 1- Potential Determinants	Supporting Articles	Specific Aim 2- Model Development	Supporting Articles
Total Non-Adherent Visits	5	Vantage Credit Scoring	2
Total Cancelled Appts	5	Box Office Success Prediction	3
Appointment Maker	4	Complexity of Problem	1
Provider Type	24		

The first focus was on defining visit non-adherence. While it was necessary to understand visit non-adherence, it was also necessary to establish the context of non-adherence in general, and then narrow it to visit non-adherence in outpatient psychiatry. Without this context, the study would be severely self-limiting. In pursuit of adequate context, a wide range of articles pertaining to non-adherence or compliance was reviewed. The terms "compliance" and "adherence" (or non-adherence) appear to be used in the health care literature as functional synonyms, although their meaning is somewhat different. For the purposes of this study, "non-compliance with" and "non-adherence to" were treated as the same concept.

One of the first conclusions supported in the literature is that adherence is multi-factorial in nature, but that measurable characteristics exist that can contribute to both understanding and interventions [7, 20, 91, 99-101]. The literature also provided guidance regarding the importance of the non-adherent visit as an opportunity for investigation. Non-adherent visits, then, whether initiated by patient behavior, physician behavior, or by nature, are best considered as events into themselves, rather than non events [102]. Further, the common practice of addressing the financial aspect of the non-adherent visit before, or instead of, addressing the motivations and feeling and circumstances surrounding the event is a clinical mistake [102]. Specifically, visit non-adherence in outpatient psychiatry can have far-reaching effects on the patient at hand, on other patients (in the form of accessibility), and on the cost of providing mental health services [11, 103].

From the literature, the realization was gained that it should be possible to predict non-adherent visits more accurately with a small set of determinants, and the occurrence of visit non-adherence can be improved by focusing on characteristics of how the patient interacts in the situation [21, 104-105]. In fact, two researchers state that the single most important predictor in visit non-adherence is previous visit keeping behavior [21, 106]. The difficulty, of course, is determining what makes up that behavior.

The second focus in the literature was on identifying a set of potentially useful determinants. A meta-analysis published in 1998 outlined medical adherence (compliance) indicators discovered in prior work in five broad categories, including health outcomes, direct indicators (blood tests, etc), indirect indicators (prescription refills, etc), subjective reports (patient's reports), and utilization (appointment making and keeping, etc.) [82]. This study provided direction as to the types of determinants that might be useful for further investigation, but lacked the specificity needed to develop a predictive model. Therefore, articles were located that investigated and promoted the use of a wide variety of visit non-adherence determinants. It quickly became apparent that researchers tended to focus either on patient demographics [103, 107] or on redundant visit scheduling techniques as their primary interest [11, 13, 19, 108-109]. Within those two primary focus areas, a large number of potential determinants were listed, many of which were similar (at least in intent), but not exactly identical. The scope and variety of determinants covered in the literature occasioned the need for an organization scheme to classify the determinants found in the literature. Observations of the determinants found in the literature, and a classification scheme outlined in Agras' work, led to a determinant classification scheme that included three classes of determinants (patient socio-economic, clinical diagnoses, and logistical) [99, 110]. Potentially useful determinants substantiated in relevant literature included:

Patient socio-economic class

- Patient gender
- Patient age
- Patient marital status
- Patient employment status
- Patient race
- Payer type
- Relationship of listed contact person to the patient

Clinical diagnosis class

- Primary (or first) diagnosis
- Secondary diagnosis

Logistical class

- Five digit zip code of patient origin
- Wait days to appointment
- Type of appointment
- Appointment time of day
- Appointment hour
- Appointment day of week
- Appointment date
- Patient use of non-MD mental health care providers
- Total number of appointments
- Total number of visit non-adherent appointments
- Total number of cancelled appointments
- Maker of appointment
- Provider type
- Referral source

Determinant Identification

In order to coordinate the individual studies into a composite view of each determinant, it was necessary to create a working definition for each of the determinants. It was also necessary to aggregate information found in a number of articles to create a baseline understanding of the potential importance of, and function of, each determinant in the proposed model.

Patient gender, for the purposes of this study, was defined as male, female, or indeterminate. Patient gender and patient sex are frequently considered interchangeable terms in the literature; however, sex can be used to describe an act in addition to a classification, hence the preference for the use of gender in this study. The nature of psychiatric practice, including care for patient for whom gender identification is an issue, renders the inclusion of indeterminate gender necessary. Gender is considered an important factor in visit non-adherence by a number of researchers, but the results of these studies yield mixed opinions on which gender is more likely to have non-adherent visits [59, 105, 107, 109]. Additionally, women may face more barriers to access to care, which may impact visit non-adherence [111-112]. And, women may also require longer visits, which may also impact visit non-adherence [113]. Age may be differentially studied in visits non-adherence because of its relative ease of acquisition. However, because of the strength of impact of age on visit non-adherence, as demonstrated in several of the studies, it is included as a determinant in this study.

For the purposes of this study, patient age was defined as the age, in round years, of the patient at the time of service. Several researchers found age to be an important factor in visit adherence [28, 59, 62, 107-109, 114-116]. The relevant literature supports the idea that older patients tend to be more adherent to medical treatment, including visits [110, 114, 117-119]. The lowest visit non-adherence rate was found in women over age 60 in one study [119]. Otero, on the other hand, found that age had no effect on visit non-adherence, and Oppenheim, found neither age nor gender important to prediction of visit non-adherence [26, 96].

Patient marital status was defined per patients' categorization of their own personal relationship status, including long-term, life-partner relationships where formal marriage has not taken place. Marital status is supported in the literature as a significant determinant of visit non-adherence [120]. The literature indicated that a divorces marital status is positively associated with visit (treatment) non-adherence [121]. Married women also experience greater difficulties with visit non-adherence [112]. Furthermore, women who are parents and unmarried tend to have greater difficulties with visit non-adherence [122]. Divorced marital status has been identified as a determinant in the accurate prediction of visit non-adherence 75 percent of the time [121].

Patient employment status was defined in this study as patient self-reported employment, including part-time and unpaid work. Student status was included as a valid employment status. Home maker was also included as a valid employment status. Evidence was found, substantiated in the literature, that an employed patient with a mental health condition is more likely to sustain treatment, including visits [123], and that patients who are unemployed tend to miss more appointments [124].

Patient race was defined in accordance with the race categories used by the US Census. Patient race appears to be independently correlated to visit non-adherence [59, 107-108, 116, 119, 125-126]. Minority status itself also appears to be a determinant of visit non-adherence [110, 127]. Both African American and Latino patients are almost 40 percent less likely to seek and obtain mental health care [124, 128-131]. Visits scheduled with female African American patients, who are being seen for mental health disorders, are more likely to be visit non-adherent than those scheduled for other medical problems [132]. Asian or white patients tend to have lower rates of visit non-adherence [23, 116]. Visit non-adherence among Native American women has been reported as high as 36 percent [119]. However, two studies of factors influencing visit non-adherence found race to be unimportant as a determinant [26, 133]. Patient race may also factor into the visit non-adherence equation, given that information regarding the rates of depression diagnosed in female African American patients is substantially missing [134].

Payer type was defined as the general type of payer, including commercial fee-for-service plans, managed care plans, State and Federally funded programs, and self pay options. Payer type is closely related to the cost of care patients incur, and cost of care is significantly associated with visit non-adherence [120]. Uninsured patients have a significantly higher visit non-adherence rate [119, 121]. Patients with private or managed care insurance have higher visit adherence rates, provided that coverage is extended to mental health care [116, 135]. Margolis showed that, overall, female patients who were insured had a visit non-adherence rate of 22 percent, while those who lacked insurance had a rate of 33 percent [119]. Patients with government-funded insurance based on income may also have a tendency towards visit non-adherence [136]. In a study at Virginia Commonwealth's Department of Orthodontics, 15.4 percent of patients for whom Medicaid was the primary payer had non-adherence visits, while only 8.3 percent of non-Medicaid funded visits failed to occur [93]. Employment, or lack of it, may be associated with the presence or absence of a particular type of payer, and unemployment is associated with increased visit non-adherence [123].

Relationship of listed contact person to the patient was defined, for the purposes of this study, as the relationship of the emergency contact person to the patient. The lack of familial relationships can negatively affect visit adherence [125], while social support from family members and others decreases the likelihood of visit non-adherence [99]. Lack of exposure of family members to a patient's true medical condition has also been shown to be a determinant of visit non-adherence [137].

Primary (or first) diagnosis was defined as the first diagnosis listed by the physician at a visit as coded by ICD-9 CM codes. Secondary diagnosis was defined as the second diagnosis, if any, listed by the physician at a visit as coded by ICD-9 CM codes. Several articles identified chief complaint as a determinant of visit non-adherence [36, 118, 138]. Further, several studies specified that depression or anxiety may predispose patients to visit non-adherence [20, 125, 139-140]. Depression is identified as of special concern by many researchers, because previous studies show that depression also adversely effects individuals other than the patient (indirectly), and, thus, raises costs for care for more than one patient [141-142]. Patients with a diagnosis of depression often require tailored interventions to prevent visit non-adherence, which also raises costs [143].

Visit non-adherence has been shown to be directly related to the patient's perception of their own morbidity [56, 85, 113, 120, 144-150]. Mood disorders, including depression, co-morbid mood disorders and anxiety, and mental disorders associated with suicidal ideation and or impairments in social role functions were most likely to be perceived by patients as increasing their own morbidity [85].

The presence of a personality disorder also increases the likelihood of visit non-adherence [1]. Patients with a diagnosis of schizophrenia tend to have more difficulty with visit adherence, as do patient with a diagnosis of psychosis or addiction [37, 121-123, 140]. Patients with less severe illness and milder levels of distress are also associated with greater visit non-adherence [95, 110].

For the purposes of this study, travel distance was defined as the distance from the five-digit zip code of patient origin (as defined by US Postal Service codes) to the five digit zip code of the Psychiatry Clinic. Travel distance and mode of transport are considered important determinants of visit non-adherence and are well studied [103, 120]. In one study, transportation problems accounted for 13 percent of non-adherent visits [151]. Research has demonstrated that patients with longer travel distances have a higher risk of visit non-adherence [116, 152], and that living less than 15 miles from the care delivery site is also significantly associated with visit non-adherence in the form of treatment termination [96, 121]. The perception of travel difficulties associated with urban and rural settings may also influence the perception of travel distance [23]. Longer travel distances also increase visit non-adherence, because many patients under estimate the true costs of travel to care, including child care costs [112, 145] and visit non-adherence increases as travel costs increase [112]. Travel distance may be especially important when a diagnosis of schizophrenia is present, possibly because these patients have more difficulty with self-transport [83].

For the purposes of this study, wait days to appointment was defined as the number of days that elapsed from the time of a request for an appointment to the date service was available. The number of wait days appears to be an important determinant of visit non-adherence supported in a number of studies [1, 7, 19, 28, 52, 118, 148, 153-156]. Most studies indicate that longer wait time leads to more non-adherent visits [95, 154, 157], and that wait time of more than two weeks also increases the risks of visit non-adherence [23, 26, 158-159]. But, one study showed no demonstrable relationship between wait time and missed appointments [37]. In two other studies, the number of wait days prior to a counseling appointment was not shown to be related to patient attrition (visit non-adherence by termination) [133, 160].

For the purposes of this study, type of appointment is defined as “new” if the patient had not been seen in the clinic used in the study in the prior three years, and as “return” if they had been seen in the previous three years and were returning for care. Differentiation between physicians or type of provider was not incorporated in the assignment of the patient to new and return visit types. Review of the literature suggests that new appointments may be subject to more visit non-adherence than those for return patients [154]. The literature suggests further that approximately 25 percent of new outpatient psychiatry visits are non-adherent [105].

In this study, as in most others, appointment time of day is defined as standard military time of the scheduled appointment. The time of the appointment has been significantly associated with visit non-adherence [28, 161]. Males and females appear to miss appointments at different times of the day [105], so gender may be a valid consideration in the scheduling of appointments to avoid visits non-adherence. The literature also suggests that an appointment scheduled at an inconvenient time for the patient may negatively affect visit adherence [162].

In this study, appointment hour was defined as the hour only of the appointment. For example, all appointments occurring from 8 am to 9 am were considered to have occurred in the 8 o'clock hour. The literature suggests that the hour in which the patient's appointment is scheduled impacts the visit non-adherence rate, although the distribution of visit non-adherence varies across the patient population [28].

Appointment day of week was defined as Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, or Sunday. As suggested by one study, Mondays appear to have a higher non-adherence rate [45].

Appointment date was defined as the date of service. Because of the large distribution of appointment dates and the relatively low visit counts for some days, the literature was consulted for ways to group service dates into a more usable form. Research suggested that there are seasonal patterns in visit non-adherence [105] and in the causes for which care is needed [113]. Subsequent to the discovery of the Purdue study, seasonality of visits entered into the discussion of potential determinants.

Therefore, service or appointment date was modified to include a measure of seasonality [80, 113].

Patient use of non-MD mental health care providers was defined as documentable use, by the patient, of mental health services provided by counselors, psychologists, social workers, nurses, and lay mental health workers outside on the ambulatory psychiatric clinic environment. This was done because of research that reported that previous instances of any mental health visit non-adherence are considered to an accurate measure of the likelihood of future psychiatric visit adherence [118]. Furthermore, research has shown that patient satisfaction with staff competence (other than physicians) is a determinant of visit non-adherence [123]. Our intent was to investigate further the possibility that non-physician mental health services delivery may impact patient attendance at subsequent appointments scheduled with psychiatrists.

Given the characteristics of the billing and schedule system that formed the basis of data collection for this study, total number of appointments, total number of visit non-adherent appointments, and total number of cancelled appointments are all interrelated. Total number of appointments was defined as the count of appointments in the clinic of study during the data collection period. Total number of visit non-adherent appointments was defined as the count of non-attended, but scheduled, appointments in the clinic of study during the data collection period when the patient did not inform the clinic at least 24 hours in advance that they would not be attending the appointment. Total number of cancelled appointments was defined as the count of non-attended, but scheduled, appointments in the clinic of study during the data collection period when the patient informed the clinic at least 24 hours in advance that they would not be attending the appointment. Conclusions from relevant literature about number of appointments include the conclusion that larger numbers of appointments may lead to decreased visit adherence [20]. The percent of non-canceled appointments to kept appointments was independently significant [108]. Also, the likelihood of visit non-adherence has been effectively measured by calculating and analyzing the percentage of adherent visits to scheduled appointments [36, 108, 163]. Patient's previous history of non-adherent or canceled appointments has also been demonstrated as an important determinant of non-adherence [108]. Additionally, one study documents that patients may be visit non-adherent because they find it difficult to cancel appointments [162]. This idea is further supported by work done by Hashim, which shows that directly calling patients one day prior to their appointment and offering the opportunity to cancel the appointment significantly reduced the number of non-adherent visits [30].

For the purposes of this study, maker of appointment was defined as the individual who entered the visit as an appointment in the Admission, Discharge, and Transfer system used by the clinic of study. Poor medical office staff-patient relationships have been demonstrated to negatively affect visit adherence [33]. When patients are not satisfied with the competence of staff, visit adherence suffers [123]. When service quality and responsiveness are perceived by patients as inadequate, visit non-adherence increases [164]. Conversely, when the patient's perception of the clinic is that the services provided are at least satisfactory, visit adherence increases [165]. When patients have their appointments scheduled by the referring physicians and/or their staff, the likelihood of visit non-adherence decreases [136]. The appointment maker's role in explaining to patients about the scheduling system influences visit adherence [94]. Scheduling inefficiencies are perceived by patients as a key factor in good staff-patient relationship [33].

Provider type was defined as the general class of physician/provider for the visit. The specific treating physician is associated with visit adherence [125, 166]. Supportive and confident providers have higher rates of visit adherence [117, 167]. Visit non-adherence rates appear to be higher in resident clinics [79, 116, 168]. Low patient confidence in their providers also appears to increase the non-adherent rate [123]. It is interesting to note that physicians and other care providers are not particularly adept at estimating the likelihood of visit adherence from review of the patient's characteristics [169-170]. The lack of, or the perceived lack of, a personal physician or perceived personal physician in a care environment also negatively impact visit adherence [26]. Conversely, two studies found that the termination of an established patient-physician relationship, even those that include resident physicians, do not affect visit adherence [171-172].

This dichotomy is of special interest in this study (an academic setting), because it is an opportunity to study the differences in visit non-adherence that are a function of care delivered by residents. Visit adherence, even with reminders and other patient support, tend to be impaired in resident clinics, though the level of visit non-adherence varies widely among such clinics [168, 173]. This is especially concerning, given that visit non-adherence means less exposure to non-didactic training opportunities for these residents, which may negatively impact their ability to manage particular diseases in the future [174]. Residency clinics with high visit non-adherence rates typically also experience negative effects in patient health, continuity of care, and in clinic productivity [175]. Interestingly, patients seeing counselors typically have lower rates of visit non-adherence than do resident clinics [79]. Physicians and other providers whose cancellation policies are unclear, excessively rigid, or excessively permissive may influence visit non-adherence in ways that render treatment unsafe [102]. Several researchers found that the patient-physician relationship did not influence visit non-adherence [176-180].

Providers of care, including physicians, are not especially capable of estimating the potential for visit non-adherence [163, 181-182]. The ability of a physician to recognize the likelihood of compliance, including compliance with visits, is directly related to patient mortality, in that the more inexact a physician's estimation the more likely the patient is to die [182-183]. Treatment duration and provider consistency did not appear to impact visit adherence; nor did race or language concordance between patient and physician [110, 184].

Referral source was defined as the type of person or entity that sent the patient to the clinic of study for care. Referral source impacts the rate of visit non-adherence [28, 118]. Patients who are self-referred may have better visit adherence [117, 133]. In instances where patients were unable to self refer, visit adherence was negatively impacted [129]. Visits scheduled for patients who are referred to what is seen as specialty care by another physician are more likely to be adherent, especially if the relationship of the patient and the referring physician has been of a longer duration [136].

As a result of attempting to collect data concurrently with selecting determinants, the uses of some determinants advocated in some studies were omitted. A requirement in this study involved the need to use data that were collected electronically from data sources available at the time of appointment scheduling. Also eliminated from consideration was any determinant that could not be managed or controlled by clinic staff.

Determinants not considered for further investigation in this study, but found in the literature, include:

- Physician-patient gender concordance [108, 145, 184-187]
- Physician-patient racial concordance [108, 145, 184-187]
- Patient linguistic capabilities [108, 145, 184-187]
- Native language spoken [119, 184]
- Patient health beliefs and their estimations of the correctness of their diagnosis [108, 145, 184-187]
- Patient educational level [1, 115, 122]
- Patient's involuntary legal status [7]
- Patient's resistance to care [95, 132, 188]
- Physician prediction of visit non-adherence [183]
- Whether the patient lived alone or not [1]
- The availability of quality home care for the patient [123]
- Patient's self–evaluated quality of life [123]
- Social stigma associated with mental health conditions [189-192]
- Flexible scheduling [193]
- Patient access to social services [193]
- Patient smoking status [122]
- Patient forgetting or getting the appointment date incorrect [27, 83, 162, 194-195]
- Role of the clinic/institution in reinforcing visit adherent behaviour [107]
- The concept that patients achieve their preferred outcomes by abstaining from care [196-197]
- Patient's participation in medical decision making [198]
- Patient use of "avoiding" coping strategies [146]
- Use of monetary incentive to increase attendance [199]
- Use of reminder letters and phone calls [98, 157]
- Time limits or expiration dates set for therapy [200-201]
- Patient time utilization skills [202]

Use of these determinants was declined because of their unavailability in electronic format within the study clinic.

Analysis Strategy Identification

In order to better understand both the concept of non-adherence and successful data analysis strategies and prediction models, several non-health care service areas were investigated where adherence is an issue. Examined in this study were “visit adherence” and the methods and models used to predict the likelihood of non-adherence in a number of other service industries [203].

There are several tools currently in use that hold potential for improving visit adherence. One such tool is the GAIL model used in rapid calculation of a women’s breast cancer risk [204]. The GAIL model is used in clinical practice to rapidly calculate the risk of an individual (woman) developing breast cancer. It requires the input of a small set of determinants into a simple formula to obtain results. Advantages to using a (modified) GAIL include its acceptance by the health care community, its use of limited patient demographics as input, and its structure, which is simple enough to be used in a rapid and automated fashion [205].

Tools that are used by the airline industry to predict when a passenger may not show up for his/her scheduled flight (designed to enable over-booking) also hold promise. These models typically include a core set of client information, along with information regarding the attribute of a particular flight (such as origin point, cabin class, and flight “leg”). In the airline industry, traveler's no-show behavior and standby behavior are well studied. Typically, these studies are undertaken as a means to increase revenue or "right size" overbooking to ensure full flights [206-207]. These studies have been carried out through a number of data analysis methods including data mining, machine learning classification trees, and logistics regression [69-70, 75-76, 208]. The accurate prediction of passenger no-show (after a valid ticket is held) is dependent on historical flight data, booking class level, past passenger history, flight leg, wait time, and seat allocation.

This set of mixed determinants perhaps can be considered corollaries for patient socio-demographic information and for logistical determinants, such as provider type, wait days, etc. Airline prediction algorithms typically do not include a corollary, however, for medical diagnosis. Management of the visit non-adherence phenomenon in the airline transport industry has traditionally been studied as a factor in revenue management (known as yield management), which includes forecasting, overbooking, seat inventory control, and pricing [77, 209]. Significant gains were made in the ability to forecast future non-adherent behavior when records kept at the passenger name level were available and employed [70]. The use of predictive models showed a significant improvement over the application of more traditional judgmental methods [206].

Yield management models developed in the airline industry have promise to address other problems, such as those in hotel or cruise booking [67, 210]. It was this association that led to an examination of models created in those industries. Carrying the idea for forecasting further, literature describing gross box office returns in the cinema industry was examined [211]. Box office prediction models include consideration for the attraction of client based on "star power", which may be similar in effect to that of provider type on health care visit adherence [71, 212].

As odd as it may initially appear to be, predicting the financial success of a motion picture may hold promise as a way to understand health appointment non-adherence [72, 213-214]. Like the model used in health care reimbursement, in the cinema industry, there is an assumed relationship between the budget expended on the film and its box office (or care quality) success [215]. Like the way the delivery of psychiatric services and attendant rates of visit non-adherence is dependent on the distribution of health ailments and health care seeking behavior, movie attendance is related to seasonality [216]. Conceptually, the link between new patient visit behavior and first week viewership of a newly released movie may exist [217-218]. In terms of solving the need to identify instances of potential visit non-adherence, work done in the cinema industry whereby the forecasting problem was converted to a classification problem (which resulted in improved prediction rates, especially when used with neural networks) may hold promise in the health care arena [71]. In the cinema industry, predicted rates of attendance have proven possible prior to movie release [71].

Work done in the cinema industry around the issues of endogeneity (correlation between a parameter or variable and the error term) and simultaneity (the property of two events happening at the same time in at least one frame of reference) may contribute to understanding non-adherence in the health care setting. The influence of these two items may, just as they influence the potential success of any movie, assert an influence on visit non-adherence. The role of film critics (similar to health provider evaluation) is another possible corollary between health care and the film industry that may affect visit non-adherence [219]. Seasonality in movie releases often influences theater attendance, and may also be a factor in psychiatric visit attendance.

The best fitting algorithm for health care visit adherence, however, may be in the models used for credit scoring. Much of the demographic information used in these tools is a one-to-one match with patient demographic determinants under consideration. Credit scoring tools also generally allow for inclusion of "past history" information that may line up with total visits, number of non-adherence visits, and number of canceled visits. Credit scoring, used by the financial industry to predict the risk of non-repayment of loans, also offers a potential model for visit non-adherence prediction. The type of determinants used to predict the likelihood of a credit receiver to repay the loan closely reflects some of those determinants that might be helpful in predicting visit non-adherence [220-224]. For example, the amount of the current transaction and its influence on the receiver's overall credit is a very viable indicator of repayment success, and a determinant that may compare, in influence, to the effect the percentage of non-adherence appointments has on visit non-adherence [225].

Reputation or credibility of service provider has been shown to lead to increased utilization of services and may be similar in influence to the effect of health care provider reputation on visit non-adherence [73]. This type of risk assessment is somewhat new to the medical field as a replacement for subjective evaluations [226]. However, risk scores are strong predictors of future behavior in health care arenas [226]. Risk scores can be analyzed for success using the Kaplan Meyer tests and the variables incorporated in the risk calculators can be compared using the Mantel-Cox test [227]. Specific risk assessment tools, such as the GAIL breast cancer risk assessment tool, have a proven track record [204-205].

All of these examples of non-health area tools that may be helpful in visit non-adherence prediction have one underlying assumption that may influence their utility in healthcare. All assume that there are consumer buying patterns and intentions, and these intentions and behavior patterns, including trust, are relatively constant [203, 228-236]. The underlying premise is that individuals have rational expectations and that their responses to questions about their intentions are the best predictors of subsequent behavior [237]. For example, one can look to Google's feature where users can select options based on the percent of positive feedback from other users as an example, as well as to applications that use Bayesian approaches to provide consumer preference ratings to potential customers [233]. The literature informs us that it is probably best to not expect too much of consumer intentions, because they do not identify the probability a person will behave in a certain way. To further confound the issue, it is also necessary to be careful how intent is measured, because measurement of intent showed that purchasers of goods and services were both effected by just being asked about their intention and that their behavior was less effected by the intent measure if they had had previous experience with the produce/service to be purchased [238-245].

The third literature review focus was on methods of data collection, data analysis, and model generation and the specific characteristics of each that might influence the success of this study. This focus was necessary because key issues in patient compliance, including visit non-adherence, include the influence of sources of information and focus of measurements done to evaluate the care process [66]. One of the first valuable lessons gained from this section of the literature review was the establishment of precedents for retrospective chart review and the use of data obtained from scheduling systems and paper charts [23, 116, 119, 175, 246]. The second valuable lesson obtained from the literature review was that clinical decision making, including that regarding visit adherence, often includes the estimation of the likelihood of a dichotomous outcome (for an given patient or visit) [247]. Further, this type of estimate may be, perhaps, best obtained by a logistic regression model, although stepwise selection results in the least discriminating models [247-248]. Another consideration developed from the literature was the idea that the coding of the variables and the use of a selection format can affect the analysis results [247]. Within the literature, there were numerous examples of support for the use of logistical regression [4, 7, 95, 108, 120, 128, 166, 202, 247, 249-253].

Support also existed in the literature for the use of linear regression and for the use of univariate and multivariate analysis with data similar to that collected for this study [4, 20, 120, 123, 254]. In general, the conclusion supported in the literature is that multivariate analysis is better for this type of data analysis [120]. The use of generalized additive models in health care were also investigated, especially in time series data, and the use of generalized additive models in non health care settings, again in time series data, and its use where discovery is an important feature of the analysis [255-259]. Other issues related to the appropriate use of data collection and analysis methods included investigation into the following:

- How well Receiver Operation Characteristics (ROC) scores can be integrated in prediction models [227, 260]. Specifically, the use of the area under the curve in the health care setting was investigated, as best representing the probability that a randomly selected subject (patient, for example) is correctly rated [260]. The probability of a correct rating is estimated by the well-studied (for comparison's sake) nonparametric Wilcoxin statistic [260].
- The use of Random Forests [261-263]
- Boot strapping and cross validation for accuracy estimation and model selection [264]
- Predicative nomo-gram use in health care [265]
- Use of neural and adaptive systems and rough sets [71, 246, 266-269]
- Development and validation data set use [108]
- Use of data mining tools [270]
- Effective use of recursive partitioning in health sciences [271]
- The ability of increasing estimated class probability accuracy through the use of exponents [272]
- Use of decretization of continuous values to increase predictive accuracy [273]
- Confirmation that calibrated probability estimates are obtainable from decision trees [274]
- Characteristics of predictive modeling including the advantages and disadvantages of segmenting data records and developing models for

each segment [275]. The dependence of such segmentation on pre-defined levels within the data is also a consideration [276].

- The potential of increasing the accuracy of forecasts by applying statistical methods that distinguish between dependent and independent variables (for example discriminate analysis) over applying simple direct clustering approaches(a priori segmentation) [275].

Based on this review of the literature, it is posited that patient visit non-adherence is affected by the characteristics of the individual patient, his or her disease process, and the logistics of providing health, and that a useful model for the prediction of visit non-adherence can be developed for use in the ambulatory care delivery. The principle research question is, therefore, ***"Can a predictive model be developed using combinations of data available in healthcare information systems that identifies psychiatric non-adherent visits in a timely fashion?"***

CHAPTER 3-RESEARCH METHODOLOGY

Selection of the appropriate data collection and data analysis methodology for this study was an investigative process. Because this study's approach to the issue of visit non-adherence prediction differed from the perspectives that had been previously used in studies of this issue, it was necessary to experiment with several data collection and analysis techniques to determine which would work best given the data available. There was an additional consideration of the need to develop a model suitable for use in the creation of decision support tools for visit scheduling by health provider staff members.

Within the body of literature that relates to visit non-adherence, there were numerous examples of support for the use of specific statistical analysis methods. Significant support existed for the use of logistical regression [4, 7, 95, 108, 120, 128, 166, 202, 247, 249-253]. Support also existed in the literature for the use of linear regression and for the use of univariant and multivariant analysis with data similar to that collected for this study [4, 20, 120, 123, 254]. There also existed support for the use of the generalized additive model, especially its use with time series data and its use where discovery is an important feature of the analysis [255-257, 259].

This experimentation also allowed the opportunity to develop a process for the effective application of billing and scheduling system data to a quality of care issue. As part of this study's focus (to develop a usable model for the prediction of visit non-adherence), it was important that a replicable and expandable methodology for the use of billing and scheduling data be developed that leveraged this readily available resource to create a descriptive picture of the effects of patient and clinic characteristics that create the non-adherent visit. In this study environment, as in many US healthcare delivery sites, the billing and scheduling systems are more stable and provide more structured data than the EMR or other clinical systems. This phenomenon is probably the result of systems designed to enable reimbursement for services rendered, an important aspect of health care delivery. The use of billing and scheduling data also supports the analysis of the effects of the determinants on the individual visit level, as opposed to the patient level, provided the requisite care is taken to adjust statistical methods used to appropriately manage over-representation.

Given the relatively non-specific method principle research question, "*Can a predictive model be developed using combinations of data available in healthcare information systems that identifies psychiatric non-adherent visits in a timely fashion?",* used to guide this study, it was necessary to establish several specific methodological aims from which a more structured investigation could be made. *This research study sought to test the set of potential determinants identified in the literature review to select those best suited for inclusion in the model (Specific Aim 1), to create a model for the prediction of the likelihood of visit non-adherence (Specific Aim 2), and to compare the resulting model to other health and non-health related prediction models. (Specific Aim 3)*

This formative study employed both quantitative and qualitative methods, and used data obtained retrospectively (with Institutional Review Board approved status) from the Admission, Discharge, and Transfer (ADT) system, the Electronic Medical Record (EMR), and, initially, from paper charts used by an ambulatory psychiatry clinic at an academic medical practice. A total of four data sets were collected and three preliminary studies were conducted. Bio-statistical staff and clinical data administrators supported all of the studies. The three preliminary studies investigated the availability of usable data, the potential utility of the data obtained to a prospective algorithm, and the utility of statistical analysis techniques to the investigators. Each preliminary study allowed an opportunity to explore both the data and the analysis methods available. Each successive study was a refinement of the previous one(s) toward developing a useful and replicable process.

The following chapter discusses the contribution of the preliminary studies to the proposed methodology first, and then discusses the specific methods used in some detail.

Preliminary Studies

Three preliminary studies informed the methodology used in this study. An overview of each preliminary study, and the lessons learned from each, including results, are covered below as a prelude to the methodology used for this phase of the research. These three pilot studies investigated the availability of usable data, the potential utility of the data obtained for a prospective algorithm, and the utility of statistical analysis techniques to the investigators.

In each of the three preliminary studies described, data regarding potential determinants was extracted from the Admission, Discharge and Transfer (ADT) system in use for clinical operations through the use of the COGNOS Analyzer product and exported to

Excel spreadsheets for review for accuracy and completion of empty fields through manual data retrieval from electronic system or from paper charts.

Preliminary Study 1 Overview (2005-2006)

Preliminary Study 1 consisted of 2,000 patient visits collected from January 2005 to June 2005 from patient visits in the Psychiatry Clinic or the Continuity Clinic (a clinic of last resort for many patients). The visits were a complete data sample for that time frame. In Phase 1, the concentration was on obtaining sufficient, usable data from electronic sources. Retrospective data were obtained (with IRB approval) from the Admission, Discharge, and Transfer (ADT) system, the Electronic Medical Record (EMR), and, initially, from paper chart review. Data regarding potential determinants were first extracted from the billing and collections module of the ADT system using an analyzer tool designed for revenue cycle analysis. The data thus obtained were reviewed for accuracy and completeness. Data for each of the determinants were deemed accurate if it corresponded to information found in the paper chart documentation and/or in the scheduling functions of the ADT system and if the value was believable. For example, the diagnosis code was deemed accurate if it corresponded to the gender of the patient, the fully written out diagnosis, and was represented as a valid ICD-9 code.

Completion of empty fields was accomplished with manual data retrieval from paper charts and from fields in the ADT system that could not be automatically extracted. Data accuracy and completeness varied by clinic site. Data completeness was greater than 90% for all determinants collected, but required nearly 6 months to collect.

The data analysis in Preliminary Study 1 was limited to regression and ANOVA analysis and was carried out using a SAS program employing a binary logistic model with stepwise variable selection to determine frequency, percent, Chi-square, and the odds ratio. A confidence interval of 95% was used.

Preliminary Study 2 (2007) Overview

Data in Preliminary Study 2 were obtained from a complete data sample of visits (200 visits in 2004) to the OB/GYN clinic that were followed by a scheduled visit to the Psychiatry Clinic for the purposes of treatment for post-partum depression.

In Preliminary Study 2, the objectives were to determine if the unit of analysis (visit) was appropriate and if a slightly different statistical method would be more informative than that used in Preliminary Study 1. Data were obtained and reviewed the same way as in Preliminary Study 1, except the data were obtained from the OB clinic this time. In response to questions about the utility of information collected and analyzed at the "patient" level rather than at the "visit" level, the use of a nested approach, whereby all visits for an individual patient were nested together and the "patient" rather than the "visit" was analyzed. Using data from the OB clinic allowed a re-check of the possibility that data availability and accuracy might vary from clinic site to clinic site and how big the variance might be.

Statistical analysis for Preliminary Study 2 (also using SAS) was accomplished with multinomial logistic regression using parameter significance tests, odds ratio, and analysis of deviance. The confidence interval was also set at 95%.

Preliminary Study 3 (2007-2008) Overview

Data for Preliminary Study 3 were obtained as a complete sample (every visit) from the Psychiatry Clinic in FY 2006-2007. Analysis in Study 3 considered each of the 13,310 visits (rather than the individual patient) as the unit of measure. This "per visit" unit of analysis permitted the development of a model of visit non-adherence that included three contributory classes of determinants (patient socio-demographic, diagnostic, and logistic). This type of structure was hypothesized to provide better predictive value than the use of any of the classes of determinants can singularly provide.

Preliminary Study 3 concentrated on developing a replicable process, utilizing what had been learned from Preliminary Studies 1 and 2. Clearly preliminary Studies 1 and 2 were limited by the small sample size that could be obtained and prepared. By 2007, the EMR contained sufficient historical data, so paper chart review was discontinued. The ADT and the analysis tool used to extract data from it had evolved to the point that all the determinants could be collected at the individual visit level. Data completeness was above 90% for each of the determinants without additional "fill in the blank" effort.

Based on the two initial pilot studies, a modified approach was taken with the third data sample and its analysis. To allow better sample stratification of the 13,000+ individual psychiatric clinic visits selected, a status determinant was established to class visits by past visit adherence history. Visits with a history of at least one non-adherent visit in the previous 12-month period were grouped together and visits associated with adherent visits only were placed in a second group. Knowledge gained from preliminary Studies 1 and 2 also indicated the need for refinement of the statistical analysis part of the study, so the use of additional analysis strategies and tools were trialed. In Preliminary Study 3, multiple data analysis methods were trialed, including the use of data-mining and the use of spatial mapping for certain determinants. For statistical analysis purposes, data were partitioned by randomly selecting a set of 2,000 visits to be retained for further testing, and by carrying out random selection with replacement to create 10 subsets of observations for analysis. Missing data elements in a visit record resulted in the omission of that visit when calculations were conducted of the data category that contained the missing element. The visit was included for each data category where the element was present. Logistic regression modeling was performed with forward variable selection**.** Because some of the determinants did not appear to be linear, the General Additive Model was also applied [258]. Additionally, the effect of various data coding and data aggregation strategies on the analysis of data was observed.

Additional information obtained regarding data collection strategies and data analysis strategies from the three preliminary studies included:

- In each of the three preliminary studies, data regarding potential determinants extracted from the ADT through the use of the COGNOS Analyzer product required review for accuracy and for completion of empty fields through manual data retrieval from either the electronic system or from paper charts. Over time, however, from the first to third preliminary study, the data source and data extraction capabilities matured to the extent that such review was no longer required, thus lessening the time and effort required for data pre-processing and raising the likelihood that a tool could be developed that would be easily employed in clinical operations.

- By the end of preliminary Study 3, all data elements were collected at the individual visit level. This level of data granularity was chosen as it allows the most expeditious use of the data sources and because any tool devised from this study would need to function at the point just prior to the scheduling of a visit, when those data sources would likely be the only two available.

- The samples used for all phases of this study are characterized by an unequal distribution by gender (approximately two-thirds female). The distribution is well supported in the literature as a function of typical help seeking behaviors between the genders, but is, by the time data collection occurred for preliminary Study 3, further acerbated by changes to Missouri's Medicaid policy. The policy changes led to the removal of large numbers of adult males from the Medicaid rolls, thus further limiting access to care for those patients.

Each preliminary study contributed to an understanding of the issues around data collection and analysis. Results from preliminary Studies 1, 2, and 3 suggest that the determinants of visit non-adherence are multi-factorial in nature and that a degree of correlation exists among the determinants. Statistical analysis of the determinants supported the use of determinants in all three proposed structural classes, including patient socio-economic (patient age, patient employment status, patient marital status, and patient race), diagnosis (primary diagnosis), and logistic factors (patient travel distance, number of appointments, number of previous non-adherent appointments, and appointment time, appointment date, type of appointment, payer type, and type of provider).

As a result of the preliminary Studies 1 and 2, two determinants initially considered (referral source and presence of supported housing) were discarded due to lack of reliable data even through their use is well supported in the literature. Referral source was discarded because the policy at one of the clinics required clerical staff to enter "Self" as the referral source instead of the actual referral source. Special housing was discarded because it proved too difficult to determine by street address all instances of special housing use. A standardized approach to combining raw data elements into classes for further analysis was developed and tested. Results from preliminary Study 2 indicated the discontinuation of the nested visit approach and the use of manual data gathering. Analysis of the larger (third) data sample resulted in a third determinant (appointment hour) being removed, since the actual appointment time of day was found to be sufficient and marginally more accurate. Please see Table 2 below for additional information about the specific determinants.

Table 2
Determinant Selection

Determinant	PS 1 n=2,000	PS 2 n=200	PS 3 N=13, 310	Determinant	PS 1 n=2,000	PS2 n=200	PS 3 N=13, 310
Gender[1]	0.002	♀	0.003	**Appt. Time**	p<0.001	p=0.643	p<0.001
Age[1]	†	p=0.004	p<0.001	**Appt. Hour**	*	*	‡
Marital Status[1]	p<0.001	p=0.579	p<0.001	**Appt. Date**	p=0.007	p=0.041	p=0.035
Employment[1]	p<0.001	p=0.093	p=0.019	**Appt Day**	*	*	p<0.001
Travel Distance[1]	p=0.003	p=0.699	p<0.001	**Use of Counseling**	*	*	p=0.001
Race/Ethnicity[1]	p<0.001	p=0.746	p<0.001	**Number of Appts**	†	*	p<0.001
Payer Type	p<0.001	p=0.868	p=0.001	**#of Non-adherent Appts**	†	*	p<0.001
Relationship of Contact Person	p<0.001	*	p=0.002	**#of Canceled Appts**	†	*	p<0.001
Primary Diagnosis	p=0.549	*	p<0.001	**Appt Maker**	p<0.001	*	p=0.271
Second Diagnosis	*	*	Δ	**Type of Provider**	p<0.001	*	p<0.001
Wait Days	p=0.025	p=0.772	p<0.001	**Referral Source**	p<0.001	*	†
Appt. Type	p=0.915	*	p=0.395	**Special Housing[1]**	p<0.001	*	†

1-Of patient
2-Admission, Discharge, and Transfer System (AKA billing and scheduling system)
3-Electronic Medical Record
4-Count including the proceeding 3 years in pilot studies 1&2 and proceeding 2 years in pilot study 3
5-Secondary diagnosis was modified the yes/no it exists rather than individual ICD codes
‡-Discarded because exact appointment time was a better determinant (didn't need both)
†- Discarded due to difficulties obtaining accurate data
♀-All female
*-Not used in this study
Δ-Modified from ICD code to yes/not that secondary diagnosis exists

Several determinants appear, from results of the three preliminary Studies, to be better determinants of visit non-adherence than others. Perhaps the most significant determinant of visit non-adherence is travel distance. Rather than a simple linear relationship (i.e., visit non-adherence increases steadily as travel distance increases), a tri-modal eye effect is observed. Proportionally greater numbers of non-adherent visits occur in the five to ten mile travel distance range than in the 11 to 30 mile range. Visit non-adherence then increases up to about 200 miles of travel distance, but is almost non-existent when patient travel distance to an appointment exceeds 200 miles. Access to transportation probably underlies this finding.

Patient age is a determinant that shows a consistent pattern across the first phases of this study. On average, non-adherent visits are associated with patient ages that are a little more than 6 years younger than the ages associated with adherent appointments.

Primary diagnosis may turn out to be especially useful when diagnosis codes captured for each visit are grouped into general diagnostic categories, based on ICD coding. In this circumstance, non-adherent visits tend to be disproportionately associated with depression, while visits associated with anxiety tend to be attended.

Because the literature suggests that it is a factor in visit non-adherence, race was included in all three Preliminary Studies. The data also suggest that race may be important in any useful prediction model. Further work needs to be done to isolate the possibility that race is confounded with another variable, such as payer. Literature also suggests that an individual's culture may play a role in acceptance of counseling care;

an idea that may further confound race, if culture is dictating the selection of provider type.

Analysis of provider type demonstrates a clear division between adherence and non-adherence to appointments when the provider of the appointment is an attending or resident physician. The visit adherence rate for resident staffed appointments is much less than that of attending physicians staffed appointments. There also appears to be a significant difference in the likelihood of visit non-adherence between new and return appointment types regardless of provider type. It is much more likely that a new visit will be non-adherent. New appointments are also affected by the general tendency for more data elements, especially diagnosis, to be missing than is normally the case in return appointments.

Patient marital status, payer, patient employment status, and the relationship of the patient's listed contact person are viewed as markers of the level of social support associated with the appointment. An example of the combined effect of social support was demonstrated by the analysis of the first data sample. Of the patients who were known to have a caseworker, the caseworker was listed as the contact person for less than one percent of the total visits where its use was indicated. Subsequently, clinic scheduling staff began using the contact person field to record case worker contact information, which allowed other healthcare workers to contact the case worker directly if communication with the patient was failing. This increased communication positively impacted visit adherence.

Generally, visits associated with a positive employment status (employed), with a commercial payer, and with the contact person listed as a close family member enjoy a higher adherence rate. Visits associated with contact person listed as “none” or where the payer is a non-Medicare government payer tended to be less adherent. To better understand the impact of family involvement, the rates of visit non-adherence were examined when the listed contact person was either a son or a daughter. Visits where the daughter was listed as the contact tended to have higher non-adherent rates than those where the son was listed as the contact person. The exact reason for this difference is not yet understood.

Patient gender also appears to be significant. Visits associated with female patients have a disproportionately higher rate of visit non-adherence. This appears to be especially true when patients are younger. This finding is at odds with the majority of the literature, suggesting it may be relevant only in a limited context and that further investigation is required.

The day of the week and the time of day at which a visit is scheduled appear to impact visit adherence. Appointments scheduled on Wednesdays and Friday afternoons appear to be especially vulnerable to visits non-adherence. Appointments scheduled late morning and late afternoon also suffer reduced visit adherence, although this effect is somewhat tempered by the patients' age.

Please see Table 3 below for a summary of the data collected and analyzed and the tools used in the process.

Table 3
Summary of Data and Tools

Determinant	Preliminary Study 1	Preliminary Study 2	Preliminary Study 3
Number of Visits	2,000	200	13,310
Clinic	Psychiatry/Continuity	Psychiatry/OB/GYN	Psychiatry
Data Format	Electronic/paper	Electronic/paper	Electronic
Basic Sample Demographics	Male, Female Age 18-94	Female, Ages 18-50	Male, Female Age 18-94
Date Collection Period	Jan-June 2005	FY 2004-2006	FY 2006-2007
Analysis Tool(s)	SAS	SAS	R, WEKA, ArcView
Analysis Method (s)	Binary logistic model Stepwise variable selection for frequency, percent, Chi-square, and odds ratio	Multinomial logistic regression Parameter significance tests, odds ratio, and analysis of deviance	Forward facing step-wise linear regression, logistic regression & general additive model Data clustering Geo-spatial mapping
Confidence Interval	95%	95%	95%
Trialed Items	2 test populations created by diving visit based on patient past adherence	Visits nested as a "patient" and "patients" analyzed rather than visits.	New determinant added: Appointment Hour

These three preliminary studies, while helping to form the methodology for subsequent investigation, suffered from several distinct limitations. One such limitation is the use of a sample population mostly from one type of clinical practice clinic. Another limitation is the use of a patient population receiving services at one academic medical center. A third, and important limitation, is the relatively small sample size, which may have caused bias in the results. Future work must expand the model to non-psychiatric visits and to various non-academic settings.

Another limitation is the current use of statistical methods that may have bias towards over-representation, as a function of the use of all visits (in the given timeframe) for each patient. Going forward, the use of statistical methods that better accommodate "within-subject" dependence is explored.

There are, however, some improvements to data collection that may positively impact future work. Because of efforts to improve the usability of the EMR in the research site, it may be possible, in successive work, to see if the use of referral source as a determinant will enhance the predictive model.

A sidebar discussion of results of preliminary Studies 1, 2, and 3 has focused on the effective use of data sources. Paper chart review has, in this study, been rendered unnecessary by the creative use of billing and scheduling data. In this particular practice, like many others in the US, the billing and scheduling systems in place are more stable and provide more structured data than the EMR or other clinical systems. This phenomenon is probably the result of systems designed to enable reimbursement for services rendered, an important aspect of health care delivery. Those same systems collect all the data elements needed for this study. However, some of the elements, such as contact person relationship, are more difficult to acquire, and some elements, such as zip codes (which requires an additional step to render into travel distance), may require additional manipulation before they are usable. Adjustments and enhancements to billing and scheduling systems could be a valuable tool in addressing visit non-adherence. Of course, this would depend on if their utility to practice management and provider productivity improvements are shown to justify the costs of modifications.

Current Study Methodology

Through this formative study, a context sensitive process has been developed that informed the present effort, and may be further utilized in future work by my collaborators and I, and, hopefully, by others. Two types of data analysis are used to establish the relative importance of the determinants as predictors of visit non-adherence. The first is the use of well-established statistical tools and methods and the second is the use of a data-mining tool. Statistical analyses include bi-variant analysis, histograms and linear and logistic regression, and well as such tests as outlier checking and “noise” evaluation. This model fitting and these tests were performed using the software package “SAS” under the guidance of an experienced bio-statistician.

Logistic regression with variable selection is a traditional strategy for deriving predictions on a binary outcome. Logistical regression is where the probability of a non-adherence visits is conditional on the set of predictors (X_1,,….X_k) and where B_1,…,B_k are regression coefficients which are to be estimated from the data. In addition to logistic regression, we used classification trees as an alternative technique. The resulting models were compared with respect to predictive accuracy and model parsimony [261].

This study also used Waikato Environment for Knowledge Analysis (WEKA) to classify and select features best suited to develop associations between the determinants. WEKA is a well-documented tool that is well suited to the type of data being used in this study. A coded data set was prepared for use in the WEKA part of the study. This is required to effectively leverage WEKA abilities for nominal and numeric data [270].

Once the statistical analysis and data mining results were available, the determinants were ranked according to utility. For this study, utility is a balance of ease of data element acquisition, accuracy of data element, statistical power, and association.

Once the set of determinants was fully identified and a useful model developed, the third step, comparison to other non-adherence or risk prediction algorithms, began. Several such tools are currently in use that hold potential for application to visit non-adherence. One such tool is the GAIL model, used in rapid calculation of a woman's breast cancer risk [204]. Advantages to using a (modified) GAIL include its acceptance by the health care community, its use of limited patient demographics as input, and its structure, which is simple enough to be used in a rapid and automated fashion [205].

Tools that are used by the airline industry to predict when a passenger may not show up for their scheduled flight (designed to enable over-booking) also hold promise. These models typically include a core set of client information, along with information regarding the attribute of a particular flight (such as origin point, cabin class, and flight "leg"). This set of mixed determinants perhaps can be considered corollaries for patient socio-demographic information and for logistical determinants, such as provider type, wait days, etc. Airline prediction algorithms typically do not include a corollary, however, for medical diagnosis.

Prediction models for box office success of films also offer clues as to what may be appropriate algorithms for health care use. Box offices prediction models include consideration for the attraction of client based on "star power," which may be similar in effect to that of provider type on health care visit adherence [71, 212]. The best fitting algorithm for health care visit adherence, however, may be in the models used for credit scoring. Much of the demographic information used in these tools is a one-to-one match with patient demographic determinants under consideration. Credit scoring tools also generally allow for inclusion of “past history” information that may line up with total visits, number of non-adherence visits, and number of canceled visits.

The following items were used as criteria for the selection or development of an algorithm for use in predictive determination of visit adherence. The successful comparison algorithm shall:

- Function with data commonly available from electronic sources.
- Incorporate all relevant determinants (per determinant testing).
- Be sufficiently robust to handle missing data elements.
- Be flexible enough to incorporate the additional (with adjustment) of any newly discovered determinants.
- Support the development of a risk calculator or other immediate use tools for active use in health care

In general, the data management process and the analysis process was, as planned, an iterative process. This process allowed an analysis cycle further focused on predictive accuracy and model parsimony. Both objectives are required for successful use of any decision tool developed from the model. Four such iterative cycles occurred prior to the validation cycle of this study. Please see Figure 3, below, for a general overview of the process.

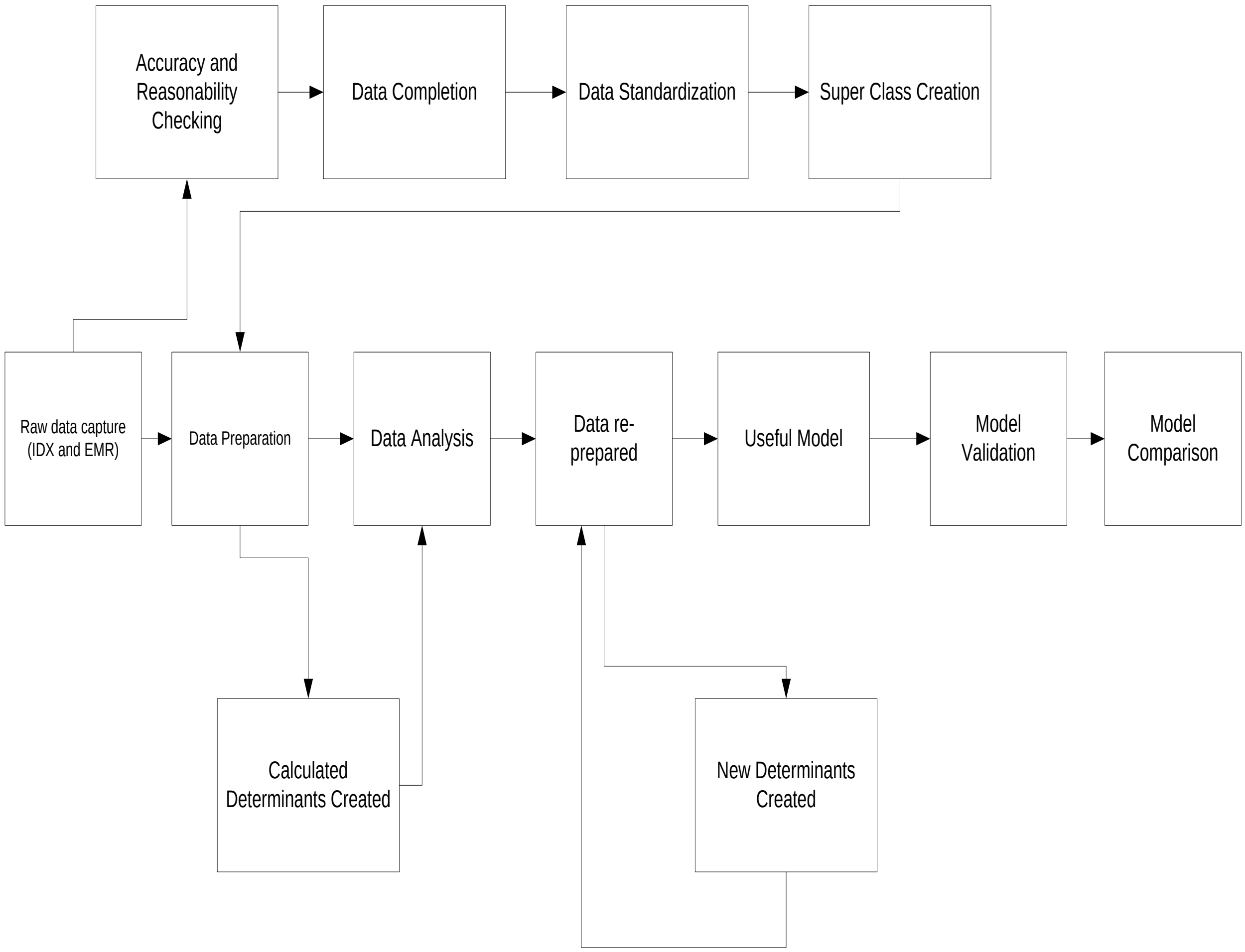

Figure 3. Data Management and Analysis Cycle.

Determinant selection

The selection of determinants used was informed by both the literature review and the results of the preliminary studies. Criteria for the selection of determinants for investigation include:

- Availability from electronic sources. All use of paper-based records to collect data elements has been discontinued

- The level of support use of the determinant found in the literature review
- The level of support for use of the determinant discovered in analysis carried out in the first three preliminary studies
- The potential for the effect of the determinant to be controlled for or managed by ambulatory care administration

Proposed determinants for this study include those selected on the basis of the literature review and preliminary studies, plus several developed specifically for this phase of the study. New candidate variables were developed to capture the needed historical perspective. Enhancements within the ADT systems and deepening understanding of the data also required that the working definition of the determinants be updated. Determinants included:

- Status of Appointment - defined as the adherent or non-adherent state of the visit being analysis
- Patient Gender - defined as the self-reported patient gender
- Patient Age - defined as patient's year of age at the service date
- Patient Travel Distance - defined as the distance between the patient's 5 digit zip code of resident and the five digit zip code associated with the care delivery site
- Visit Type - defined as either New (patient has not been seen in care delivery site in the previous three years) or Return (patient has been seen at care delivery site in the previous three years)
- Same Visit Type - defined as yes when visit under analysis is the same type (new or return) as immediate previous visit

- Service Date - defined as MM/DD/YYYY of the date of service. Service date, later in the analysis, was re-structured as season, whereby individual dates are combined into months and then months into season of the year. Season of the year is defined per US Naval Observatory Universal time, rounded to the nearest complete month [277]
- Appointment Wait Days - defined as the number of days elapsed between the date of request for a visit and the service delivery date
- Appointment Day of Week - defined as Monday, Tuesday, Wednesday, Thursday, or Friday
- Same Day of Week - defined as “yes” when the visit of analysis occurs on the same day of the week as the immediate previous patient visit
- Appointment Time - defined in standard military time; hour and minute
- Same Appointment Time - defined as “yes” when the visit of analysis occurs at the same time (within one-half hour) as the immediate previous visit
- Referring Provider - defined as actual referring provider name per IDX dictionary
- Payer - defined as a Financial Status Category (FSC) used by the clinic under investigation
- Same Payer - defined as “yes” when the payer of the visit of analysis is identical to the payer of the immediately previous visit
- Employment - defined as free text field in the IDX scheduling model that shows the name of the employer or other information regarding employment
- General (or primary) Diagnosis - defined as the four or five digit ICD-9 CM code, associated with the visit of analysis

- Secondary Diagnosis - defined as the four or five digit ICD-9 CM code or "None" associated with the visit of analysis
- Number of Previously Canceled Appointments - defined as the count of previous visits in the study period when the patient contacted the clinic of study at least 24 hours prior to a scheduled visit to indicate non-attendance
- Use of Non-MD Mental Health Services - defined as an appointment in the IDX scheduling system of a previously scheduled and arrived appointment with a counselor, case manager, or licensed clinical social worker outside of the clinic of study
- Patient Marital Status - defined as a free text filed in the IDX scheduling model indicating present marital status as reported by the patient.
- Percentage of Previous Visits non-Adherent - defined as a calculated field whereby the number of previous non-adherent visits (in the study timeframe) divided by the total number of visits in the study timeframe
- Relationship of Contact Person to patient - defined as a free text filed in the IDX scheduling model indicating patient self-reported present relationship of listed contact person to himself/herself.
- Number of Previous Non-Adherent Appointments - defined as the count of previous visits within the study timeframe that were non-adherent
- Patient Race - defined as patient's self identified race
- Total Number of Appointments - defined as the total count of appointments to the clinic of study in the study timeframe

- Immediate Previous Visit Non-Adherent - defined as “yes” when the visit immediately previous to the visit of study was non-adherent
- Maker of Appointment - defined as the initials of individual who made the appointment and his/her training status (trained or untrained)
- Provider Type - defined as the provider name which is tied to provider type based on licensure and training
- Same Provider Type - defined as “yes” when the provider type associated with the visit of study is identical to that of the immediately previous visit

Data capture

The data used in this study are all retrospective and collected under Institutional Review Board (IRB) approval. All data elements were collected at the individual visit level or were calculated from that level of data. This level of data granularity is required to allow the most expeditious use of the data sources and because any tool devised from this study would need to function at the point just prior to the scheduling of a visit, when visit level data are likely to be the only accessible data. Data were extracted from electronic sources and exported to Excel for review for accuracy and completion and were de-identified at the earliest possible point. Within the study “patients” are identified by a randomly generated code number that is not cross walked to the medical record number. Missing data were sometimes supplied by cross matching visits from the same patient within the complete data sample early in the process. Data values that are the result of keying errors into the ADT system (such a mis-keyed location) are removed by deletion of the individual record in entirety. Data elements electronically retrieved and automatically exported to Excel include patient identifier, appointment status, patient gender, patient age, patient home zip code, wait days until appointment, appointment type, service date, appointment day of week, appointment time, referring physician, payer, general (or primary) diagnosis, secondary diagnosis, patient employment, patient marital status, patient race, relationship of contact person, appointment maker, and physician/provider. Data elements that were manually extracted from the ADT system on a record by record basis include number of non-adherent visits, number of canceled visits, total number of appointments, and patient use of non-MD (i.e. counseling) appointments.

Because of recent efforts to improve the usability of the EMR in the research site, it was possible to reintroduce the use of referral source as a determinant and this field was back-filled. This study also debuts the use of several candidate determinants designed to ferret out temporal relationships. These include appointment time consistency (Same Time of Day), day of week consistency (Same Day of Week), appointment type consistency (Same Appointment Type), appointment status consistency (Previous Visit Non-adherent, Percent of Visits Non-adherent), payer consistency (Same Payer) and provider consistency (Same Provider). Seasonality of visit non-adherence is determined from month of service as determined by date of service.

Only data obtained from electronic sources (the ADT system or the Electronic Medical Record) are used. Data collected from the ADT system are drawn from both the Scheduling Module and the Charge Module. Please see Table 4 below for description of data completeness in the raw data sample.

Table 4
Data Completeness

Determinant Class	Determinant	Percent Complete
Social-Economic		
	Gender	98.8
	Age	98.7
	Marital Status	98.8
	Employment	94.7
	Travel Distance	99.8
	Race/Ethnicity	92.6
	Payer Type	98.7
	Relationship of Contact Person	95.6
Clinical Diagnosis		
	Primary (Gen) Diagnosis	92.3
	Second Diagnosis	100.0 (including "none" as legitimate)
Logistical		
	Wait Days	100.0
	Appt. Type	100.0
	Appt. Time	100.0
	Service Date (becomes "season")	100.0
	Appt Day	100.0
	Use of Non-MD Mental Health	78.3
	Number of Non-Adherent Appts	100.0
	Number of Canceled Appts	100.0
	Appt Maker	99.4
	Type of Provider	100.0
	Referral Source	98.4

Data Preparation Overview

A total of 19,428 patient visit records were processed for this study. This sample includes visits from the preliminary Studies 1 and 3, as well as a complete sample of all outpatient psychiatric visits (at the study site) for Fiscal Year 2009. Data samples from preliminary Studies 1 and 3 were re-drawn from the raw data and all data pre-processing from Studies 1 and 3 was removed. These three samples were then combined and re-processed as a single sample. It was also necessary to carry out 2 different pre-processing processes based on the criteria of the analysis tools (SAS and WEKA). All data elements were initially coded in accordance with the data dictionary developed as part of preliminary Study 3. Some data elements were combined to create super classes. For example, all types of commercial fee-for-service payers were combined into a single class. Sparse data also required the re-grouping of some data elements into more useful classes. The general data pre-processing was accomplished as follows.

Accuracy and Reasonableness Testing - The data sample was checked for accuracy and reasonableness of the values. As a result, five patients were identified (along with their visits) where confusion existed in their assigned medical record number. These patients and their respective visits were discarded from the sample. One patient, and visits, was removed from the sample because the listed zip code of residence indicated a more than 2,000 mile travel distance. Additionally, all non-adherent appointments and all canceled appointments were checked (and corrected if need be) to insure they had been properly categorized in the ADT system.

Data Completion - Data elements that were likely to have remained constant over the study time frame were backfilled from later appointments to earlier appointments, when the earlier appointments were missing data. Backfilling was done for patient race, patient gender, patient age (adjusted for the time elapsed between the appointments), and patient zip code of residence, referring provider (backfilled from additional system data rather than from subsequent visit data), payer, patient marital status, patient employment status, and patient primary and secondary diagnosis. Diagnoses, in psychiatry, are a special case. Psychiatric diagnoses tend not to vary over time once the initial diagnosis has been made. This is certainly borne out by the consistency of diagnosis exhibited by the data sample, where the type of successive diagnosis rarely varies significantly from the previous ones.

Standardization of data elements - All appointment times were converted to military times. All dates of services were converted to mmddyyyy format. Day of week was standardized to “Mon, Tues, Weds, Thu, Fri”.

Calculated determinants - Travel distance was derived by using MapQuest to calculate driving miles from zip code of patient residence to zip code of the study site. The total number of appointments, the number of canceled appointments, and the number of non-adherent appointments was calculated by count in the data collection time frame (2004-2009).

Super class creation - The granularity of values in some of the raw data left some determinants with an unacceptable level of complexity, which led to conclusions so minutely categorized as to be nearly un-interpretable. Grouping the values in the raw data allowed for exponential reduction in complexity and renders them more easily interpretable. Determinants that required grouping of raw data values into classes include appointment time of day, travel distances, visit type, referring provider, payer, employment status, general or primary diagnosis, secondary diagnosis, marital status, relationship of contact person, race, maker of appointment, and provider. Please see Table 5 below for the initial classification scheme.

Table 5
Initial Determinant Classification

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
Patient Gender	M,F,I	Expand to full word, Delete I	Male, Female
Patient Age	Range 1-100 years	Remove patients under age 18	Age range 18-96
Patient Marital Status	Single, Married, Separated, Divorced, Widowed, Other	Combine Other and Married	Single, Married, Separated, Divorced, Widowed
Patient Employment Status	Employed, Un-employed, Retired, Student, Home Maker, Disabled, Other	Combine Other and Employed	Employed, Unemployed, Disabled, Home Maker, Student
Patient Race	Caucasian, White, Black, African American, American, American Indian, Asian, Hispanic, Other	Combine Caucasian and White, Combine Black and African American, Combine American , American Indian, Asian, and Hispanic with Other	White, Black, Other
Payer Type	Financial Status Code	UNITED HC CHOICE UNIV EMPDIRECT CONTRACT-WC HEALTHLINK PPO UNITED HC CHOICE/PPO NON UNIV FIRST HEALTH PPO BLUE CROSS PPO-FSC 438 CORVEL WC PPO UNITED HC SELECT UNIV EMP CIGNA PPO WORKERS COMPENSATION BLUE CHOICE HMO POS ANTHEM BLUE ACCESS PPO/CHOICE P MERCY REFERRAL REQUIRED EMPLOYEE ASSISTANCE PROGRAM COVENTRY HEALTHCARE WORK COMP TEMPORARY	Managed Care

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
		BEECH STREET PPO ANTHEM BLUE PREFERRED HMO/PLUS HEALTHLINK HMO POS AETNA MANAGED CHOICE ETHIX PPO UNITED HEALTHCARE POS1 UNIV EMP GALAXY HEALTH NETWORK PPO MULTI PLAN PPOC OMMUNITY CARE NETWORK PPO FIRST HEALTH-WC PPO COX HEALTH NETWORK PPO BLUE CHOICE HMO1 NATIONAL PROVIDER NETWORK PPO CRMC EMPLOYEE HEALTH PLANMERCY HEALTH PLAN HMO1 UNIVERSITY OF MISSOURI RETIREES UNITED HEALTHCARE POS2 NON UNIV EMP/NON UP REF BLUE SHIELD ASSIGNED ST LOUIS NON CONTRACTED MANAGED CARE MERCY NON REFERRAL HEALTHLINK OPEN ACCESS COMMERCIAL ASSIGNED UNITED HC SELECT NON UNIV AETNA GREAT WEST HEALTHCARE COMMERCIAL THIRD UNITED HC SELECT NON UNIV EMP COMMERCIAL ASSIGNED SECOND	Commercial FFS

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
		MEDICAID BLUE CHOICE MISSOURI CARE MEDICAID NON CONTRACTED MC+ OUT OF STATE MEDICAID	Medicaid
		MEDICARE ASSIGNED MEDICARE PART A MEDICARE PRIVATE FEE FOR SERVICE MEDICARE PART A ONLY CRH MEDICARE PART A & B CRH ONLY RAILROAD MEDICARE ASSIGNED MEDICARE MANAGED CARE (HMO/PPO)	Medicare
		SELF PAY - NO INS PATIENT PAY/FEE FOR SERVICE	Self Pay

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
		GOVERNMENT AGENCIES MID MO MENTAL HEALTH CENTER DEPARTMENT OF MENTAL HEALTH TRICARE EXTRA KIDNEY ACQUISITION PREVENTION OF THE BLIND TRICARE PRIME REMOTE PRISONERS VA-FEE BASIS FSC CHAMPVA CRIME VICTIMS TRICARE ACTIVE DUTYADDRESS PHARMACY ONLY AUTOMOBILE MED PAY Unknown BAD ADDRESS PHARMACY ONLY MMMHC EMPLOYER SPONSORED EXAMS FACILITY ONLY BILLING DENTAL RESEARCH	Other
Relationship of listed contact person to the patient	Caseworker Daughter, Son, Boyfriend, Other, None, Nephew, Uncle, Father, Grandfather, Grandmother, Mother. Brother, Sister, Ex Husband, Ex Wife, Fiancée, Husband, Wife, Life Partner, Spouse, Partner	Combine Daughter, Son Aunt, Cousin, Niece, Nephew, Uncle, Father, Grandfather, Grandmother, Mother, Brother, Sister, Ex Husband, Ex Wife, Fiancée, Wife, Life partner, Spouse, Partner	Family
		Combine Caseworker, Boyfriend, Friend, Girlfriend, Other, None,	Non-Family
Primary (or first) diagnosis and Secondary diagnosis	314,314.01,314.9	Filter and Combine by CPT Diagnosis Sub-Category	Attention Deficit Disorder (ADD)

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
	309,309.1,309.24,309.28,309.3, 309.4,309.9,300, 300.01,300.02, 300.11,300.14 309.81	Filter and Combine by CPT Diagnosis Sub-Category	Anxiety
	296.4,296.41, 296.42,296.43, 296.45,296.5, 296.51,296.52, 296.53,296.54, 296.55,296.56, 296.63,296.64, 296.65,296.66	Filter and Combine by CPT Diagnosis Sub-Category	Bi-Polar Disorder
	296,296.02,296.1,296.11,296.15, 296.2,296.21, 296.22,296.23, 296.24,296.25, 296.26,296.3, 296.30,296.31, 296.33,296.34, 296.35,296.36, 300.4,311,296.7, 296.8,296.81, 296.82,296.89,2 96.9,296.99,297, 297.1,298.9	Filter and Combine by CPT Diagnosis Sub-Category	Depression
	299,299.01,299.8295.1,295.11, 295.12,295.295.23,295.3,295.31, 295.32,295.34, 295.35,295.32, 295.34,295.35 ,295.4,295.51, 295.6,295.62,295.7,295.71,295.72, 295.73,295.75, 295.8,295.9	Filter and Combine by CPT Diagnosis Sub-Category	Psychosis
	220.15,276.51,278.01,290,290.13,290.2, 290.21,290.42,290.43,291.89,292.84,29 2.89,293.1,293.81,293.82,293.83,293.84 ,293.89,294.1,294.11,294.8,294.9,300.2, 300.21,300.22,300.23,300.29,300.3,300. 7,300.81,301,301.13,301.22,301.6,301.7 ,301.83,301.84,301.9,302.6,302.85,303. 9,303.91,303.93,304,304.01,304.1,3042, 304.23,304.3,304.31,304.33,304.5,304.6 ,304.8,304.81,304.82,304.83,305,305.01 ,305.03,305.2305.21,305.5,305.6,305.61 ,305.8,306.51,307,307.1,307.23,307.3,3 07.42,307.44,307.45,307.5307.51,307.5 2,307.9,308.3,310.1,312,312.01,312.3,3 12.32,312.34,312.39	Filter and Combine by CPT Diagnosis Sub-Category	
Five digit zip code of patient origin	300+ zip codes	Convert to travel distance with MapQuest	Range 1-6 Miles= in town, 6-40 miles = local,41+=long distance
Wait days to appointment	Numerical Value from -24 to 1000+	n/a	Same

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
Type of appointment	New, Return, Therapy New, Therapy Return	Combine Therapy New with New, and Therapy Return with Return	New Return
Appointment time of day	From 8am to 7:30pm each day with appointments at 0:15,0:20,0:30,0:45	Change to military time	0800 to 1930 each day at 015,020,030,045
Appointment hour	From 8am to 7:30pm each day with appointments at 0:15,0:20,0:30,0:45	Round down to full hour	0800,0900,1000, 1100,1200,1300, 1400,1500,1600, 170,018,001,900
Appointment day of week	Mon,Tue,Wed,Thu,Fri	n/a	Mon,Tue,Wed, Thu,Fri
Appointment date	647 separate dates expressed as mm/dd/yyyy	Convert to months then to season	Winter=Jan. Feb,Mar Spring=Apr, May, Jun Summer=Jul, Aug, Sep, Fall=Oct, Nov,Dec
Patient use of non-MD mental health care providers	Scheduled Visits Record with LCSW or similar provider	Count as 0,1	No Yes
Total number of appointments	Scheduled Visit Record	Manual Count	1-1000
Total number of visit non-adherent appointments	Scheduled Visit Record	Manual Count	1-100
Total number of cancelled appointments	Scheduled Visit Record	Manual Count	1-100

Determinant	Raw Data (ADT,EMR) Format	Data Preparation Strategy	Final Data Format for Analysis
Maker of appointment	Maker Initials	Training records	Trained Un-trained
Provider Type	Provider Name	Provider Role	Attending Physician, Resident Physician, Counselors (includes PhD in Psychology)
Referral Source	Provider Name, or Institution Name, or "Self"	Combine all provider names and institution names to Provider	Provider Self

Determinants of consistency creation - Several new determinants were created from existing data as a means of injecting a measurement of previous patterns for several of the original determinants. These included, same visit type, same payer, same day of week, same time of day, same provider type, missing second diagnosis, and seasonality. These determinants are defined as follows:

Same Visit Type - defined as "yes" when the visit that immediately preceded the visit of study was the same visit type as the study visit

Same Payer - defined as "yes" when the visit that immediately preceded the visit of study utilized the same payer as the study visit

Same Day of Week - defined as "yes" when the visit that immediately preceded the visit of study occurred on the same day of the week as the study visit

Same Time of Day - defined as "yes" when the visit that immediately preceded the visit of study occurred at the same time of day as the study visit

Same Payer - defined as “yes” when the visit that immediately preceded the visit of study was scheduled with the same type of provider as the study visit

Missing Second Diagnosis - defined as “yes” when the visit of study has only a primary (general) diagnosis

Seasonality - defined as the time of year (winter= Jan, Feb, and Mar, spring= Apr, May, Jun, summer = Jul, Aug, Sep, and fall= Oct, Nov, Dec) of the service date.

Percent Non-adherent - defined as the total number of non-adherent visit divided by the total number of visits per patient. This is a cumulative determinant.

Previous Visit Non-adherent - defined as yes when the visit that immediately preceded the visit of study was non-adherent

Missing data management - most of the issues with missing or unknown data were dealt with during the sampling process. One notable exception to this occurred as a means to incorporate instances when no second diagnosis was assigned by the treating physician. It is not, in this data set, uncommon to see a single diagnosis listed. Therefore, a value of “none” was created and entered as a secondary diagnosis instead of a diagnosis based on the recorded ICD code when a visit had only one diagnosis code specified.

Data Formatting

Data formatting proved to be a non-trivial issue. All data elements were initially formatted according to a data dictionary developed in preliminary study 3 . Some data elements were combined to form super classes. The issues of sparse data were considered at several points in the testing process; some re-formatting was necessary to combine especially small samples of certain data elements, such a diagnosis or payer types, into more meaningful classes. Travel distance to care was determined by estimating (with geographical mapping tool) the number of miles from the patient's five digit zip code to the five digit zip code location of the clinic, by closest driving distance. The new candidate determinants (those that measure consistency) were completed by entering a yes/no response for each visit, based on information gathered from previous visits.

When there were no data available, each field was filled as "1000" (because it is a value not otherwise used) to allow for subsequent analysis.

As data analysis proceeded in this phase of the study, re-formatting occurred to maximize data analysis tool use and address sparse data issues. An additional determinant was added in the third round of analysis to allow for improved use of the secondary diagnosis determinant. A "presence of secondary diagnosis" determinant was developed and defined as simply "Yes" or "No."

Please see Table 6 below for information regarding the evolution of data formatting.

Table 6
Data Formatting

Determinant/ Source and Raw Data Format	Initial Format	Second Format	Third Format
Statue/ADT **ARR,NOS**	ARR=1 NOS=2	ARR=2 NOS=1	ARR=1 NOS=0
Gender/*ADT* **Male, Female** **Missing**	Female=1 Male=2 Missing=3	Female=2 Male=1	Female=2 Male=1
Age/*ADT* **Years of age**	No formatting needed	No formatting needed	No formatting needed
Marital Status/ ***ADT*** **Single, Widowed, Married, Divorced, Separated, Other**	Divorced=1 Married=2 Missing=3 Other=4 Separated=5 Single=6 Widowed=7	Divorced=4 Married=6 Other=1 Separated=3 Single=5 Widowed=2	Divorced=4 Married=6 Separated=3 Single=5 Widowed=2
Employment ***ADT*, None, Employed Student, Retired, Home-maker, Disabled**	Disabled=1 Employed=2 Home Maker=4 Missing =3 Retired=5 Student=6 Unemployed=7 Other=8 Non	Disabled=5 Employed=7 Home Maker=1 Retired=3 Student=2 Unemployed=6 Other=4	Disabled=5 Employed=7 Home Maker=1 Retired=3 Student=2 Unemployed=6
Travel Distance/*ADT* **5 digit zip code**	MapQuest travel Distance for each dyad	MapQuest travel Distance for each dyad	0-6 miles= In Town=1 7-40 miles=Local=2 41+ miles=Long Distance=3
Race/Ethnicity/ ***ADT*** **Census Categories in free text for race and ethnicity**	Black=1 Caucasian=2 Missing=3 Other=4	Black=2 Caucasian=3 Other=1	Black=2 Caucasian=3 Other=
Payer Type / ***ADT*** **FSC Codes**	Commercial FFS=1 Managed Care =2 Medicaid =4 Medicare=5 Missing=3 Other=6 Other Government=7 Self Pay=8	Commercial FFS=7 Managed Care =6 Medicaid =4 Medicare=5 Other=3 Other Government=2 Self Pay=1	Commercial FFS=7 Managed Care =6 Medicaid =4 Medicare=5 Other=3 Other Government=2 Self Pay=1

Determinant/ Source and Raw Data Format	Initial Format	Second Format	Third Format
Relationship of Contact Person / ***ADT*** **Free text, patient specified field**	Caseworker=1 Child=2 Friend=4 Missing=3 None=5 Other=6 Other Family=7 Parent=8 Sibling=9 Spouse=10	Caseworker=1 Child=3 Friend=4 None=5 Other=7 Other Family=6 Parent=8 Sibling=2 Spouse=9	Non-Family=1 Family=2
General (Primary) Diagnosis/ ***ADT*** **and** ***EMR*****, ICD -9 CM Code**	ADD=1 Anxiety=2 Behavior/Personality Disorder=4 Bi-polar=5 Dementia=6 Depression =7 Drug=8 Missing=3 Other=10 Psychosis=11	Anxiety=8 Behavior/Personality Disorder=10 Bi-polar=9 Dementia=4 Depression =11 Drug=2 Other=6 Psychosis=7 None=5	Anxiety=8 Behavior/Personality Disorder=10 Bi-polar=9 Depression =11 Other=6 Psychosis=7
Second Diagnosis / ***ADT*** **and** ***EMR*****, ICD-9 CM Code**	ADD=1 Anxiety=2 Behavior/Personality Disorder=4 Bi-polar=5 Dementia=6 Depression =7 Drug=8 Missing=3 Other=10 Psychosis=11	Anxiety=8 Behavior/Personality Disorder=10 Bi-polar=9 Dementia=4 Depression =11 Drug=2 Other=6 Psychosis=7	Anxiety=8 Behavior/Personality Disorder=10 Bi-polar=9 Depression =11 Other=6 Psychosis=7
Wait Days/ ***ADT*** **Count of days**	1-377	1-377	0 days=0 1-30 days= 1 31-90 days = 2 91-377 days=3
Appt. Type/ ***ADT*** **New Return, New Therapy, Ret. Therapy**	New=1 Return=2	New=1 Return=2	New=1 Return=2
Appt. Time/ ***ADT*** **Hour and minute**	8:15AM, 8:20 AM, 8:30 AM,8:45 AM 9:15 AM, 9:20 AM,9:30 AM,9:45 AM 10:15 AM, 10:20 AM,	Mid (day)=1 Afternoon=2 Morning=3	Mid (day)=1 Afternoon=2 Morning=3

Determinant/ Source and Raw Data Format	Initial Format	Second Format	Third Format
	10:30 AM,10:45 AM 11:15 AM, 11:20 AM, 11:30 AM,11:45 AM 12:15PM, 12:20 PM, 12:30 PM,12:45 PM, 1:15 PM, 1:20PM, 1:30 PM,1:45 PM 2:15 PM, 2:20PM, 2:30 PM,2:45 PM 3:15 PM, 3:20 PM ,3:30 PM,3:45 PM 4:15 PM, 4:20 PM 4:30 PM,4:45 PM 5:15 PM, 5:20 PM, 5:30 PM,5:45 PM 6:15 PM, 6:20 PM, 6:30 PM,6:45 PM 7:15 PM, 7:20 PM, 7:30 PM,7:45 PM		
Appointment Date/ *ADT* Calendar date	620 Calendar Dates	Jan.=1 Feb.=2 Mar.=3, Apr.=4, May=5, Jun.=6 Jul.=7, Aug.=8, Sep.=9, Oct.=10, Nov.=11, Dec.=12	Jan.,Feb.,Mar.= Winter=1 Apr., May, Jun.= Spring=2 Jul., Aug., Sep.=Summer=3 Oct., Nov., Dec.=Fall=4
Appt Day / *ADT,* Mon-Fri.	Mon=1 Tues=2 Weds=3 Thu=4 Fri=5	Mon=1 Tues=2 Weds=3 Thu=4 Fri=5	Mon=1 Tues=2 Weds=3 Thu=4 Fri=5
Non-MD Mental Health *ADT* Count by provider	None Known=1 Yes=2 Missing=3	No=2 Yes=1	No=2 Yes=1
Number of Appts/ *ADT* Appointment Count	Count (range1-59)	Count (range1-58)	Count (range1-58)

Determinant/ Source and Raw Data Format	Initial Format	Second Format	Third Format
Number of Non-adherent Appts/*ADT*, Appointment Count	Count (range 1-6)	Count (range 1-6)	Count (range 1-6)
Number of Canceled Appts/*ADT*, Appointment Count	Count (range 1-6)	Count (range 1-6)	0=None=0 1-6=1
Appt Maker/ *ADT* By maker initials and training status	Trained=1 Untrained=2	Trained=1 Untrained=2	Trained=1 Untrained=2
Type of Provider/ *ADT* Provider Name	Attending =1 Counselor =2 Missing=3 Resident=4	Attending =3 Counselor =1 Resident=2	Attending =3 Counselor =1 Resident=2
Referring Provider/ *ADT* Self Provider name	Counselor, Internal=1 External Health Center/Organization=2 Fellow, Internal=4 PCP, Internal=5 PCP External=6 PCP Undetermined=7 Resident, Internal=8 Self=9 Specialist, Internal=11 Missing=3 NP, External=12 NP, Internal=13	Any Provider=1 Self=2	Any Provider=1 Self=2
Same Appt Type	No=1 Yes=2 Missing=3 Unknown=4	*Removed*	*Removed*
Same Day of Week	No=1 Yes=2 Missing=3 Unknown=4	No=1 Yes=2	No=1 Yes=2
Same Time of Day	No=1 Yes=2 Missing=3 Unknown=4	No=1 Yes=2	No=1 Yes=2

Determinant/ Source and Raw Data Format	Initial Format	Second Format	Third Format
Same Payer	No=1 Yes=2 Missing=3 Unknown=4	No=1 Yes=2	No=1 Yes=2
Same Provider	No=1 Yes=2 Missing=3 Unknown=4	No=1 Yes=2	No=1 Yes=2
Percent Non-adherent	Percentage (range 0-77)	Percentage (range 0-77)	0-2% =1 3-77%=2
Previous Visit Non-adherent	Yes=1 No=2	Yes=1 No=2	Yes=1 No=2
Presence of Secondary Diagnosis	Not in use	Not in use	Yes=1 No=2

Data to be analyzed through data mining were re-formatted to a text format, replacing numerical representation with categorical variables when required.

Sampling

The sample used in this phase of study was primarily drawn from the ADT system (the EMR was only used to confirm suspect diagnoses) and represents all visits used in preliminary Studies 1 and 3 and all visits at the ambulatory psychiatric clinic of study for Fiscal year 2009. Total sample size was 19,428 visits. Both new (never seen in the psychiatric clinic before, or seen previously in the clinic with an elapsed period of at least three years time after the last date of service) and return patients were included in the initial sample. All visit types are included, except “lab only” visits were removed because the only provider seen at these visits is the lab technician. Although missing data elements were completed whenever possible, short of resorting to paper chart review, the data are considerably more complete for return patient appointments.

Diagnosis and visit history are potentially key sets of predictors that may not be available for some new patients. Therefore, the data sample was initially divided into two samples: one each for new and return visits. The sample of new patient visits was retained for potential model testing. Within the return visits sample, (12,000+ visits), seventy-five percent of visits were randomly selected to form the model development data set, with the remainder retained for model validation. To reduce the possibility of over-representation of some patients, the sample was further narrowed, so that it included only the last visit for each patient. Visits that were both the last visit and the only visit where removed at this point. This produced a sample of 2,170 visits, which comprised the initial development sample.

Two sample extraction processes were utilized. Initially, to improve sample stratification, the 2,170 individual psychiatric clinic visits were divided into ten test samples by randomized ten-fold selection. Random selection was used, but was restrained by allowing an individual patient to be represented in each subset only once. Selection with replacement was used. To improve on this sampling strategy, bootstrapping was used to create 100 replicates of 2,170 visits each. Bootstrapping is an automated subset search algorithm and is, as such, an application of re-sampling used in statistics. It is essentially a simulation method based on the data to be used in the analysis, whereby samples are drawn and re-drawn from the sample population. It is useful in the study because of the moderately sized data sample. Recent research shows that the effects of the correlation between predictor variables, the number of candidate predictor variables, the size of the sample, and the level of significance for entry and deletion of variables are of importance and may be differentially addressed by automated subset selection algorithms [248]. Bootstrapping can be computationally intensive, especially when large numbers of replicates are desired. To limit computational demands, while ensuring the quality of the sampling technique, a trial that compared the results of using 100 replicates and 1,000 replicates was carried out. No substantial differences were apparent, so 100 replicates were selected as a reasonable replicate level.

Sampling for data-mining

The sample for data mining was comprised of the sample of 2,170 return visits in the development sample. The retained validation sample was used to confirm the initial data analysis. Data were re-formatted to specifications required by the data mining tool, when required.

Testing

Beginning with the return visit sample, this study plan explored two distinct statistical models and one non-statistical method in the construction of a prediction rule for psychiatric non-adherent visits. There were a number of important considerations and influential decisions with regard to the appropriate use of analysis tools and strategies. In the absence of similar studies on which to base statistical model selection, this study substantiated the proposed analysis and model building techniques through trials on previous similar data sets. The three analytical methods utilized included statistical analysis, utilizing logistic regression with variable selection (which is a more traditional strategy for deriving predictions on a binary outcome), classification trees and regression via machine learning (as an alternative technique), and the comparison of the final model's determinants to those used in predictive models used in other health care situations and to those used in non-health care industries. The resulting models were compared based on predictive accuracy and model parsimony.

Statistical analysis

The construction of a prediction rule, for inclusion in a useful model, involves many decisions regarding the choice of statistical model, the selection of candidate determinants, the coding of specific determinants, and strategies for dealing with missing data. Each decision may reveal useful features of the sample population, but there is also the possibility of over-fitting or tailoring a model to the idiosyncrasies of a specific data set. The use of data captured from a clinical scheduling and billing system has its own limitations. Like all data used for research purposes rather than for the operational purposes (for which it was gathered), these data have several distinctive characteristics that must be addressed, methodologically, before they can be effectively used. One such issue is that the available data for existing patients are considerably richer than those available for new patients. Primary and secondary diagnosis and visit history (same provider, same payer, etc.) are potentially important determinants of visit non-adherence that are typically not available for new patients. Therefore, this study initially focused on the prediction of visit non-adherence for return visits and treated the issue of visit non-adherence for new patient visits as a separate problem.

This study focused on prediction based on the last visit for each patient when the last visit was at least the second visit for that patient. Therefore, each patient was represented in the sample only once. Approaching the sampling and analysis in this manner addressed the important consideration of how to best use multiple visits from the same patient. Regarding all visits as independent ignores the within-subject dependence between events and inflates the sample size, with the result that standard errors of the predictors can be downwardly biased, resulting in too many statistically significant associations. Treating all visits as independent undermines the potential for using prior visit history as a predictor. While it is possible to model the within-subject dependence with logistic regression, it is not possible with regression tree methods, nor is there a clear way to assess the consequences of ignoring the dependence with classification tree type methods (which are to be used as a second analysis strategy). This eliminates the need to accommodate within-subject dependencies and facilitates the use of both statistical and data mining methods, but leads to a potential increase in the number of determinants required to develop a usable model.

Missing data is a consideration in many studies and is also a concern here. Discarding observations with missing data can dramatically diminish the sample size, while imputing missing values is not practical for a prediction rule that will be used in real-time with data pulled from scheduling and billing systems. However, missing data elements can impede the initial development of a useful model. Therefore, this study began with the use of visits that represent a complete data record, and held those visit with missing data as a reserve validation population, depending on the results of the analysis and which determinants were included in the final model.

It has been established that regression-based variable selection methods, such as stepwise or backward selection, tend to capitalize on quirks of the developmental data, resulting in models that do not generalize and have poor predictive power when applied to a new data set [251, 278]. However, variable selection is often a necessary step to identify the important predictors while keeping models as parsimonious as possible. To mitigate the weaknesses of standard selection methods, this study randomly partitioned data by allocating 75% of patients to a developmental data set with the remaining 25% reserved for model validation. Within regression, partitioning the data into two samples (one for development and the other for testing the model) helps mitigate the weaknesses of standard variable selection methods. While such variable selection is needed to properly identify which determinants are important predictors, and to keep the models as parsimonious as possible, regression-based variable selection methods, such as stepwise or backward selection, tend to capitalize on quirks of the developmental data resulting in models that do not generalize and have poor predictive power when applied to a new data set [251, 278]. For the logistic regression modeling, the developmental data were further divided into 10 random subsets with forward variable selection preformed on each subset in an identical manner. The number of times a variable was selected, and the step at which the variable was selected, were recorded. Compelling predictors are those that are most frequently selected and that enter the model early. Once the “main-effects” were indentified and understood, substantively meaningful two-way interactions were considered for further exploration

The final logistic regression model yielded the predicted probability of a non-adherent visit for each observation in the data set. The ability of the model to discriminate between adherent and non-adherent appointments were summarized by the area under the ROC curve [260]. Cut-points in the estimated probability distribution were determined by examining the effects on sensitivity (the proportion of missed appointments predicted to be missed), specificity (the proportion of kept appointments predicted to be kept), and overall correct prediction rate.

Because it was expected that the estimates of the effects of the determinants derived from the developmental data were overly optimistic, they were also calculated for the validation data (those 25% of visits set aside) for confirmation purposes [251]. All statistical data manipulation and logistic regression analysis were done with SAS v9.2 Once the statistical analysis results were available, the determinants were ranked according to predictive utility. For this study, I defined utility as a balance of ease of data element acquisition, accuracy of data element, statistical power, and association.

Data Mining Process

The second analytical approach was the use of classification and regression tree methods in conjunction with random forests [262]. This study considered this strategy to be an alternative modeling strategy. This is because the data set used is large and [potentially] complex. The use of classification and regression trees may result in final trees that are too large to be easily interpreted and there may be a lack of a summary measure of the strength of association between each predictor and the outcome.

Therefore, the resulting models may be fragile and not generalizable to new data sets [279].

Classification trees are a recursive portioning method that considers each value on each candidate predictor as a split point for dividing subjects into two or more groups. Optimal cut-points are determined by maximizing a measure of node purity, such as the Gini Index for categorical predictors or minimizing squared error for continuous predictors [271]. The technique is recursive, in that the researcher has the option of considering each variable as a splitter at different levels of a tree. Splitting continues until nodes are homogeneous or until a minimum node size is reached, typically not less than ten subjects per terminal node. To avoid over fitting, a common strategy is to grow an initial large tree and then prune it back based on prediction error from a ten or twenty fold cross-validation step. Virtues of this method is that it can result in easily interpretable “trees”, works well with mixed data types, and may reveal important interaction effects that are less easy to identify with logistic regression methods.

The downside of recursive portioning methods are that final trees may be so large as to defy interpretation, there is no clear summary measure of the strength of association between each predictor and the outcome, and most significantly, the resulting tree-models may be fragile and not easily applied to new data sets [279]. A remedy for the shortcomings of these data mining methods is the use of random forests (RF) in model building. The idea behind RF is that growing many small trees, a forest, and pooling their predictions through voting or some other technique will result in better prediction than possible with any single tree. RF iteratively proceeds by drawing bootstrap samples of observations and growing trees from randomly selected subsets of the candidate predictors. While RF appears to work well as a predictive method, it is a “black box” technique that does not produce a final model that can be examined or incorporated into other applications. However, because the RF consists of a very large number of trees grown, each grown from a random subset of *variables,* it can produce a score for each predictor. For the prediction of visit non-adherence, we used RF methods as the first step in our tree-based modeling. By using the important variables identified in the RF step, we hope to obtain a relatively compact, stable, final tree that predicts well. Measures of predictive power from the RF step also provide a reference point for the maximum predictive power obtainable from any particular tree-based model.

General and the tree-based prediction used Waikato Environment for Knowledge Analysis (WEKA),version 3.2 from the University of Waikato [270]. WEKA is a collection of machine learning algorithms for data mining tasks that is free source and commonly used in academic settings, especially for educational purposes and research. The main strengths of WEKA include that is freely available under a general public license, runs on nearly every modern computing platform, contains a useful set of data pre-processing and modeling techniques, and is especially useful for the novice data miner because of its graphical user interface. WEKA supports clustering, classification, regression, visualization, and feature selection. This study employed WEKA’s main interface “Explorer” and initially used default settings and filters to transform data. It was not necessary to delete any data from the proposed 2,170 record set. Use of the classification capabilities of WEKA allowed the estimation of the accuracy of the prediction model produced and enabled the visualization of erroneous predictions, and Receiver Operating Curves (ROCs) and of the decision tree produced.

Model comparison

Once a predictive model is fully identified by statistical and/or machine learning methods, comparison of it to established prediction models discovered in the relevant literature can begin with an eye towards potential re-use of available algorithms associated with similar models. There are several such models currently in use that hold potential. One such tool is the GAIL model used in rapid calculation of a women's breast cancer risk [205]. Models used by the airline industry to predict when a passenger may not show up for their scheduled flight (designed to enable over-booking) were also compared. These models typically include a core set of client information along with information regarding the attributes of a particular flight (such as origin point, cabin class, and flight "leg") [54, 69]. Prediction models for box office success of films were also compared, particularly to include the capacity to allow consideration for the attraction of client based on "star power," which may be similar in effect to that of provider type on health care visit adherence [71, 212, 267, 280]. Credit scoring models were also compared to the predictive model developed in this study, as these may be the best fit for use in health care. Much of the demographic information used in these tools is a one-to-one match with patient demographic determinants under consideration in this study. Credit scoring tools also generally allow for inclusion of "past history" information that may line up with total visits, number of non-adherence visits and number of canceled visits [74, 220-225, 281-285]. Please see Table 7 below for the initial determinants with which the predictive models created in this study were compared. A successful comparison model shall:

- Function with data commonly available from electronic sources.
- Incorporate all relevant determinants (per determinant testing).
- Be sufficiently robust to handle missing data elements.
- Be flexible enough to incorporate the additional (with adjustment) of any newly discovered determinants.
- Support the development of a risk calculator or other immediate use tools for active use in health care.

Table 7

Comparison of Study Determinants to Determinants in Potentially Useful Predictive Models

Study Determinant	Fair Isaac (Credit Score)	Vantage (Credit Score)	Box Office Success	Airline	Hotel Yield Mgmt.	GAIL Model
Gender	If collected	n/a	n/a	Passenger gender	n/a	Females only
Age	If collected	n/a	n/a	n/a	n/a	Patient age
Marital Status	Others on accounts	n/a	n/a	n/a	n/a	n/a
Employment	Occupation	n/a	n/a	n/a	n/a	n/a
Travel Distance	n/a	n/a	n/a	Departure and arrival site	n/a	n/a
Race	n/a	n/a	n/a	n/a	n/a	Patient race/ ethnicity
Payer Types	Lender types	n/a	Box office revenue	Ticketed/ non ticketed	n/a	n/a
Relationship of Contact Person	n/a	n/a	n/a	n/a	n/a	Incidence of breast cancer in relatives
Referral Source	Bank reference	n/a	n/a	n/a	n/a	n/a

Study Determinant	Fair Isaac (Credit Score)	Vantage (Credit Score)	Box Office Success	Airline	Hotel Yield Mgmt.	GAIL Model
Primary Diagnosis	Type of account	n/a	n/a	n/a	n/a	Hx. of breast cancer
Second Diagnosis	Type of account	n/a	n/a	n/a	n/a	Previous biopsy for atypical hyperplasia
Wait Days for Appt.	Years at address	n/a	Binned days	Advance booking	Advance purchase days	n/a
Appt. Type	n/a	n/a	Genre/MPA A rating	Booking class	Room rate differences	n/a
Appt. Time	n/a	n/a	Sequel/ No sequel	Flight leg	n/a	n/a
Appointment Date	n/a	n/a	Year of release	Flight date	Special event days	n/a
Appt Day	n/a	n/a	n/a	Flight Number	n/a	n/a
Use of Counseling	New credit application	n/a	Competing films	Connecting flight	n/a	n/a
Number of Appts	Years in File/credit history	Credit avail-able	Opening screens	Number of bookings	Number of rooms	n/a
#of Non-Adherent Appts	Nonpayment of Bills	Paym't history	n/a	Number of no-shows	Daily no-show rates	n/a
#of Canceled Appts	Installment Credit	Kind of credit held	n/a	Number of previous cancels	Daily cancellation rates	n/a
Appt Maker	n/a	n/a	n/a	Booking agent	n/a	n/a
Type of Provider	n/a	n/a	Star actor/ director	n/a	n/a	n/a

Utilization of these three analysis techniques provides a number of benefits to this study. Logistic regression, which is most useful to study the relationship between independent and dependent variables, serves as a required underpinning for potential software development, while data mining, which focuses on the discovery of the relationships between all variables, serves as a vehicle by which additional discovery of interactions can be made. The model comparison exercise may potentially illuminate models in current use that could be adapted for use in the prediction of visit non-adherence. The three combine to form a more comprehensive picture of visit non-adherence prediction than any one, used singly, can provide.

CHAPTER 4- FINDINGS AND DISCUSSION

The analysis strategy employed in this study, much as the methodology, was driven by the need to create a usable model from which decision support tools could be developed. It was also important that the results of the analysis be explainable to a naive population of potential users, and that the entire process be easily replicable with other data samples. The end goal was to create a useful model that was as parsimonious as possible, while still retaining good predictive ability with applicability to as many types of ambulatory practice as possible. The strategy for analysis was carried out with three tools/strategies, as discussed in the methodology, and as shown in Figure 4 below.

Figure 4. Testing Flow

The initial development or discovery data sample included a total of 2,174 records. Each was a complete record with no missing data fields. Within this sample, 447 records were associated with non-adherent visits (20.56%) and 1,727 were associated with adherent visits (79.44%). This ratio is fairly consistent with reported visit non-adherence rates of 19 percent described in the literature.

The determinant set initially employed included:

- Status
- Gender
- Age
- Travel distance
- Visit type
- Same visit type
- Wait days
- Service date (Season)
- Appointment day of week
- Same day of week
- Appointment time
- Referring provider
- Payer
- Same payer
- General (primary) diagnosis
- Secondary diagnosis
- Marital status
- Employment status
- Number of previous non-adherent (NOS) appointments
- Number of previous cancelled (CNX) appointments
- Race
- Use of non-MD mental health appointments
- Relationship of contact person
- Total number of appointments
- Maker
- Provider
- Same provider

The statistical testing carried out in this study was cyclical in nature. Each regression cycle was followed by a reduction in the number of determinants to be used in the next cycle. Initially, a p-value of 0.20 was used to exclude determinants from further consideration. Determinants also need to enter the model as main effects both early in the model and frequently to be retained for the next analysis cycle. Figure 5, shown below, describes this process in elimination.

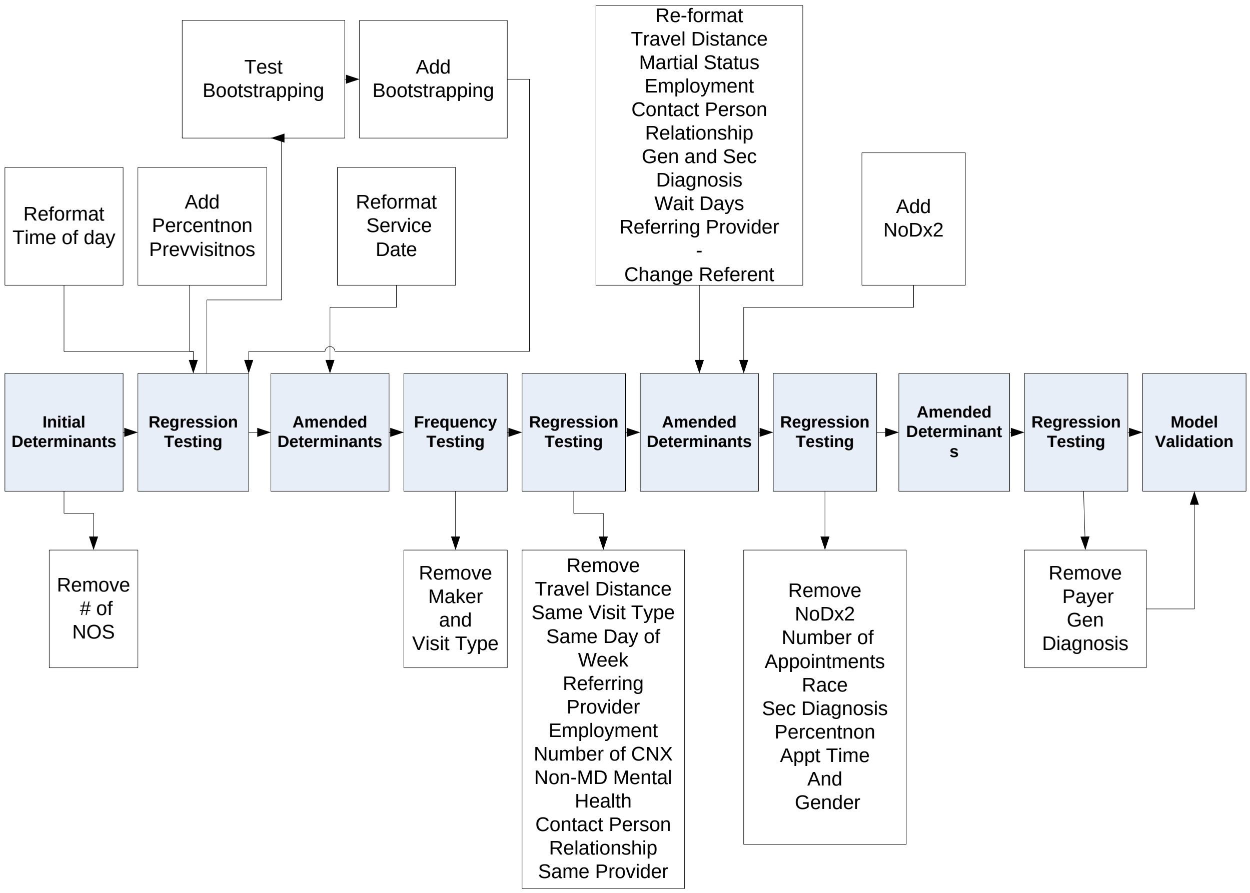

Figure 5. Statistical Analysis Process

<u>Statistical Analysis</u>

The first operations conducted in the analysis were a series of tests conducted to understand the characteristics of the data better. The list of determinants includes some that are categorical in nature and a number that are not. Those not categorical in nature include age, travel distance, number of previous non-adherent visits, number of previous cancelled visits, and total number of visits. One of the goals of the initial testing process was to see if these variables fell into natural ranges in preparation for converting them to categorical variables, if further analysis should require it. Another goal was to learn as much as possible about any unforeseen effects within the data.

Equally important in the initial analysis was the detection of any errors made in the formatting process. This initial analysis also permitted identification of any outliers in the data that could distort the rest of the analysis. And lastly, there was lingering concern about the use of the number of previous non-adherent visits determinant. Even before any additional analysis was carried out, it was abundantly apparent that the determinant was tremendously skewed and might, therefore, be unusable in its present form.

As a beginning investigative step, a simple regression was run using the entire set of determinants to see how the number of previous non-adherent visits determinant functioned. It was immediately apparent that the information represented in this determinant needed to be re-captured in another way. While the number of previous non-adherent visits constantly was the first step in a stepwise regression with forward selection, its use also meant that model validity (no maximum likelihood and no convergence) was in question in fifty percent of the regression runs (10 samples). Please see Tables 8 and 9 below for a summary of these ten trial regressions and their success or failure.

Table 8
Description of Regression Success Rate in Initial Trial

Trial	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Model Validity achieved
Sample 1	number NOS	maker	number CNX	same payer	sec diagnosis	0	0	0	0	yes
Sample 2	number NOS	number CNX	0	0	0	0	0	0	0	yes
Sample 3	number NOS	maker	number CNX	referring provider	sec diagnosis	age	number appts	0	0	yes
Sample 4	number NOS	maker	same payer	age	contact person	sec diagnosis	number CNX	0	0	yes
Sample 5	number NOS	maker	number CNX	same payer	gender	provider	0	0	0	no
Sample 6	number NOS	maker	number CNX	referring provider		wait days	employ ment	0	0	no
Sample 7	number NOS	number CNX	non-MD appoint ments	referring provider	0	0	0	0	0	no
Sample 8	number NOS	maker	0	0	0	0	0	0	0	yes
Sample 9	number NOS	maker	number CNX	0	0	0	0	0	0	yes
Sample 10	number NOS	maker	number CNX	employ ment	referring provider	gender	non-MD appoint ments	gen diagno sis	appt time	no

Table 9
Frequency of Occurrence of Determinants in Initial Trial

Frequency (All Determinants)	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Total
Number of NOS	10	0	0	0	0	0	0	0	0	10
Maker	0	8	0	0	0	0	0	0	0	8
Number of CNX	0	2	6	0	0	0	1	0	0	9
Same Payer	0	0	1	2	0	0	0	0	0	3
Non MD Appts	0	0	1	0	0	0	1	0	0	2
Referring Provider	0	0	0	3	1	0	0	0	0	4
Age	0	0	0	1	0	1	0	0	0	1
Employment	0	0	0	1	0	0	1	0	0	2
Sec Diagnosis	0	0	0	0	2	1	0	0	0	3
Gender	0	0	0	0	1	1	0	0	0	2
Provider	0	0	0	0	1	1	0	0	0	2
Wait Days	0	0	0	0	0	1	0	0	0	1
Number Appts	0	0	0	0	0	0	1	0	0	1
Gen Diagnosis	0	0	0	0	0	0	0	1	0	1
Appt Time	0	0	0	0	0	0	0	0	1	1

To identify if it was, indeed, the suspected determinant that was causing the failures, it was removed and a second set of regressions were performed. Removing this determinant resulted in successful regression with maximum likelihood and convergence achieved 100 percent of the time. Please see Tables 10 and 11 below for results of the regression run without the number of previous non-adherent visits determinant.

Table 10

Results of Regressions sans Number of Non-Adherent Visits

Trial	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Model Validity achieved
Sample 1	same payer	service date	payer	0	0	0	0	0	0	yes
Sample 2	age	service date	0	0	0	0	0	0	0	yes
Sample 3	age	Refer-ring provider		gen diagnos is	0	0	0	0	0	yes
Sample 4	gender	age	same payer	0	0	0	0	0	0	yes
Sample 5	service date	same payer	race	sec diagnos is	maker	number appts	appt time	contact person	0	yes
Sample 6	age	refer-ring provider	same payer	sec diagnos is	0	0	0	0	0	yes
Sample 7	service date	same payer	refer-ring provider	gen diagnos is	age	same visit type	sec diagnosis	appt day of week	gender	yes
Sample 8	refer-ring provi-der	age	same payer	0	0	0	0	0	0	yes
Sample 9	age	payer	gen diag nosis	0	0	0	0	0	0	yes
Sample 10	service date	age	gender	number appts	payer	0	0	0	0	yes

Table 11
Frequency of Occurrence of Determinants sans Number of Non-Adherent Visits

Frequency	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Total
Maker	0	0	0	0	1	0	0	0	0	1
Same Payer	1	2	3	0	0	0	0	0	0	6
Referring Provider	1	2	1	0	0	0	0	0	0	4
Age	4	3	0	0	1	0	0	0	0	8
Sec Diagnosis	0	0	0	2	0	0	1	0	0	3
Gender	1	0	1	0	0	0	0	0	1	3
Provider	0	0	1	0	0	0	0	0	0	1
Number Appts	0	0	0	1	0	1	0	0	0	2
Gen Diagnosis	0	0	1	2	0	0	0	0	0	3
Appt Time	0	0	0	0	0	0	1	0	0	1
Service Date	3	2	0	0	0	0	0	0	0	5
Payer	0	1	1	0	1	0	0	0	0	3
Race	0	0	1	0	0	0	0	0	0	1
Same Visit Type	0	0	0	0	0	1	0	0	0	1
Contact Person	0	0	0	0	0	0	0	1	0	1
Appt Day of Week	0	0	0	0	0	0	0	1	0	1

Simply removing the determinant, however, is not a viable solution, because it removes what might be a valuable determinant of visit non-adherence. Instead the number of previous non-adherent visits determinant was re-structured into two new determinants, which, together, encompassed the information contained in the number of previous non-adherent visits determinant. These two determinants, the percentage of previous visits that were non-adherent (Percentnon) and immediately previous visit non-adherent (Prevvisitnon) were created to encapsulate prior adherence history. Percentnon was calculated by dividing the number of non-adherent visits (at each given time in the stream of visits for each individual patient) by the total number of visits at the same time. Prevvistsnon was simply a yes/no recording of whether the visit in the sample had been directly preceded by an adherent (no) visit or a non-adherent visit (yes).

Further testing for increased knowledge about the determinants included the use of the Means Procedure for non-categorical determinants. Please see Tables 12, 13, 14, 15 and 16 below for results of each of those tests.

Table 12
Means Procedure Results for Age

Trial	Mean	Standard Deviation	Minimum	Maximum
Sample 1	44.6	14.5	3.0*	91.0
Sample 2	44.5	13.4	18.0	88.0
Sample 3	44.3	14.8	3.0*	85.0
Sample 4	43.5	13.8	19.0	81.0
Sample 5	42.6	14.4	3.0*	96.0
Sample 6	43.1	14.2	18.0	84.0
Sample 7	42.5	14.9	3.0*	86.0
Sample 8	45.9	15.3	3.0*	96.0
Sample 9	44.1	13.2	20.0	91.0
Sample 10	42.7	13.9	19.0	86.0

*Indicates visits missing ages. Five records were removed in a subsequent analysis cycle

The variance between the means among the 10 samples is reasonable; however, 50 percent of the samples show the presence of at least one record with a missing age (labeled as “3”). After further investigation, five records were removed from the data set, reducing the total number of records used. It may also be useful to conduct further investigation using the average age of patients. The median age of patients in this study is 44 years. This age (C44) will be used as the center point in further analysis.

Table 13
Means Procedure Results for Travel Distance

Trial	Mean	Standard Deviation	Minimum	Maximum
Sample 1	32.4	35.2	6.0	206.0
Sample 2	32.6	39.3	6.0	347.0
Sample 3	33.6	36.2	6.0	183.0
Sample 4	32.2	33.7	6.0	230.0
Sample 5	33.0	34.1	6.0	251.0
Sample 6	35.7	39.8	6.0	347.0
Sample 7	35.9	41.3	6.0	227.0
Sample 8	34.8	37.1	6.0	251.0
Sample 9	33.4	37.7	6.0	230.0
Sample 10	31.0	34.1	6.0	162.0

While the variance in the mean between samples is acceptable, the distribution in the data, as demonstrated by the large standard deviation (as compared to the mean), suggests that this determinant may benefit from re-formatting to increase its utility in a model. There is also one travel distance that seems to be an outlier (albeit a legitimate travel distance) that occurs in two of the samples. This value may have a significant effect on the regression. The large standard deviation in terms of the mean may also hint at decreased utility for this determinant in a final model.

Table 14
Means Procedure Results for Wait Days

Trial	Mean	Standard Deviation	Minimum	Maximum
Sample 1	52.5	41.8	0.0	189.0
Sample 2	53.4	42.1	0.0	182.0
Sample 3	50.3	39.9	0.0	182.0
Sample 4	54.2	43.2	0.0	188.0
Sample 5	52.1	39.4	0.0	182.0
Sample 6	48.3	37.6	0.0	183.0
Sample 7	56.5	45.6	0.0	189.0
Sample 8	56.8	46.6	0.0	337.0
Sample 9	49.7	40.7	0.0	182.0
Sample 10	50.3	39.1	0.0	182.0

Again, the variance in the means of the samples is reasonable, but a larger standard deviation and the presence of at least one potential outlier (maximum value in sample 8) warrants further investigation.

Table 15

Means Procedure Results for Number of Cancelled Appointments (CNX)

Trial	Mean	Standard Deviation	Minimum	Maximum
Sample 1	0.1	0.6	0.0	6.0
Sample 2	0.0	0.4	0.0	5.0
Sample 3	0.0	0.3	0.0	2.0
Sample 4	0.0	0.3	0.0	2.0
Sample 5	0.0	0.3	0.0	3.0
Sample 6	0.0	0.2	0.0	2.0
Sample 7	0.0	0.3	0.0	3.0
Sample 8	0.0	0.2	0.0	2.0
Sample 9	0.0	0.4	0.0	5.0
Sample 10	0.0	0.2	0.0	1.0

This determinant (Number of Cancelled Appointments) is obviously skewed. While a mean of 0 is understandable if the majority of patients are not cancelling appointments, the determinant itself may not be easily useable in a regression analysis because of lack of variability.

Table 17

Means Procedure Results forTotal Number of Appointments

Trial	Mean	Standard Deviation	Minimum	Maximum
Sample 1	5.9	5.2	1.0	44.0
Sample 2	5.4	3.8	2.0	26.0
Sample 3	6.0	4.7	2.0	36.0
Sample 4	5.2	3.7	2.0	24.0
Sample 5	5.8	4.4	1.0	30.0
Sample 6	6.1	5.4	2.0	38.0
Sample 7	5.9	4.8	2.0	31.0
Sample 8	5.7	5.6	2.0	48.0
Sample 9	5.8	4.8	2.0	32.0
Sample 10	6.0	5.1	2.0	40.0

There is slightly more variability in the means for these samples than in the previous four determinants. This may be driven by a small number of patients who have large visit counts.

Additional information regarding each determinant was obtained by frequency distribution analysis. Information of special concern included the differences found between the actual frequencies of values for each of the determinants and their respective expected frequencies and the presence of sparse data. Detailed frequency profiles (in table format) for each of the determinants may be found in Appendix 1 of this document. While these tables have great utility, not only in direct investigation of the determinants and in the development of data profiles needed for effective software development, they are large and difficult to visualize. To remedy this shortcoming, a visual representation of each of the determinants has been prepared (and follows) that details characteristics of each of the determinants.

Patient gender is considered an important factor in visit non-adherence by a number of researchers, but these studies yield mixed opinions on which gender is more likely to have non-adherent visits [19, 59, 105, 107]. The preliminary results obtained in this study show slightly fewer non-adherent males than expected and slightly more non-adherent females than expected. Please see Figure 6 for additional detail on the distribution of Gender by Status (adherent or non-adherent).

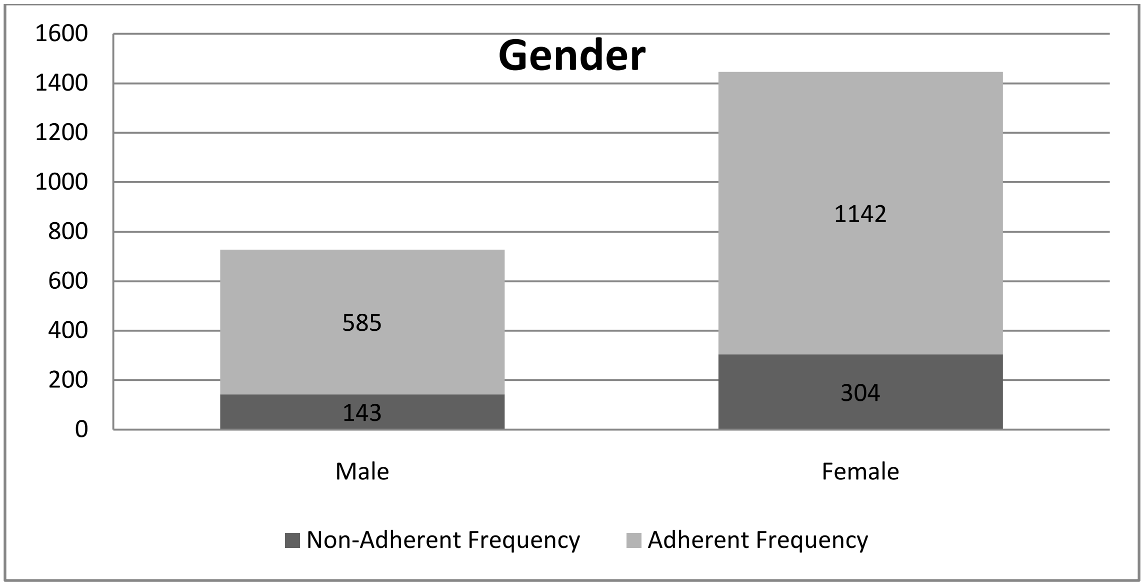

Figure 6. Distribution of Gender by Status

Age is often studied in healthcare. While important to the diagnostic process (and often included in studies for that very reason), it also is typically an easy determinant to collect and manage. The relevant literature supports the hypothesis that older patients tend to be more adherent to medical treatment, including visits, especially in females over the age of sixty years [110, 117-119, 283]. The distribution of age in the data used for this study suggests that older patients may, indeed, be more visit adherent than younger patients. The distribution also identified two additional concerns with age data. First, those visits with “3” designated in age represent visits where age was missing and which should have been removed from the sample earlier. Secondly, sparse data are an issue to be considered in further analysis. See Figure 7 below for a visual of the distribution.

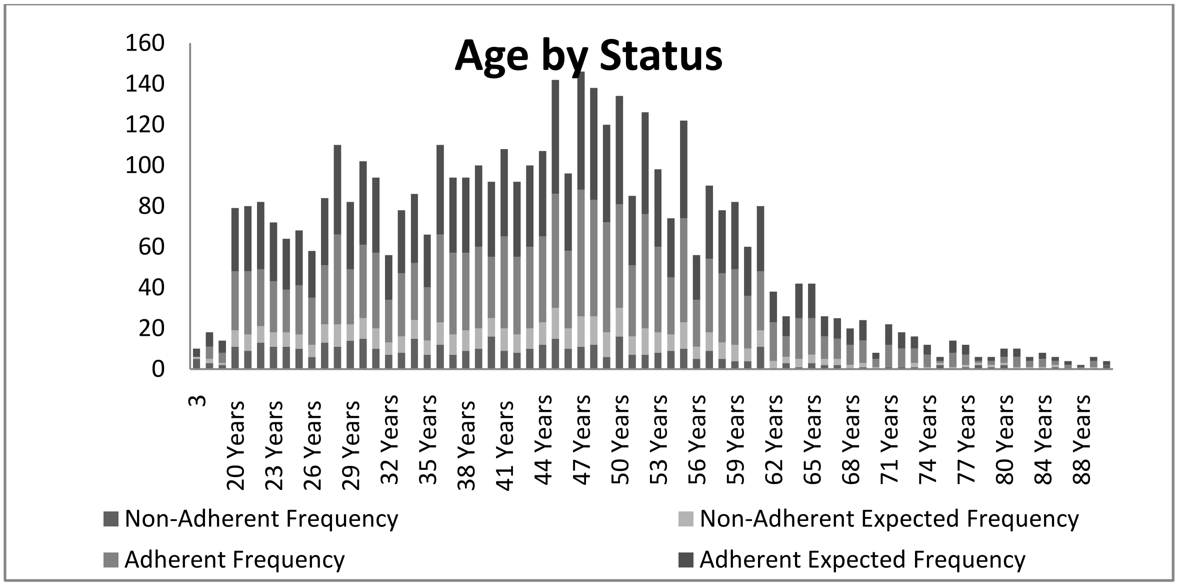

Figure 7. Distribution of Age by Status

Marital status is supported in the literature as a significant determinant of visit non-adherence [120]. In general, divorced marital status is positively associated with visit (treatment) non-adherence [121]. Divorced marital status has been identified as a determinant in the accurate prediction of visit non-adherence 75 percent of the time [121]. Our results show that the likelihood of an individual who is separated from their marital partner being associated with visit non-adherence is nearly fifty percent. A marital status of married, however, is significantly associated with visit adherence. Please see Figure 8 below for a visual of the distribution of marital states by status.

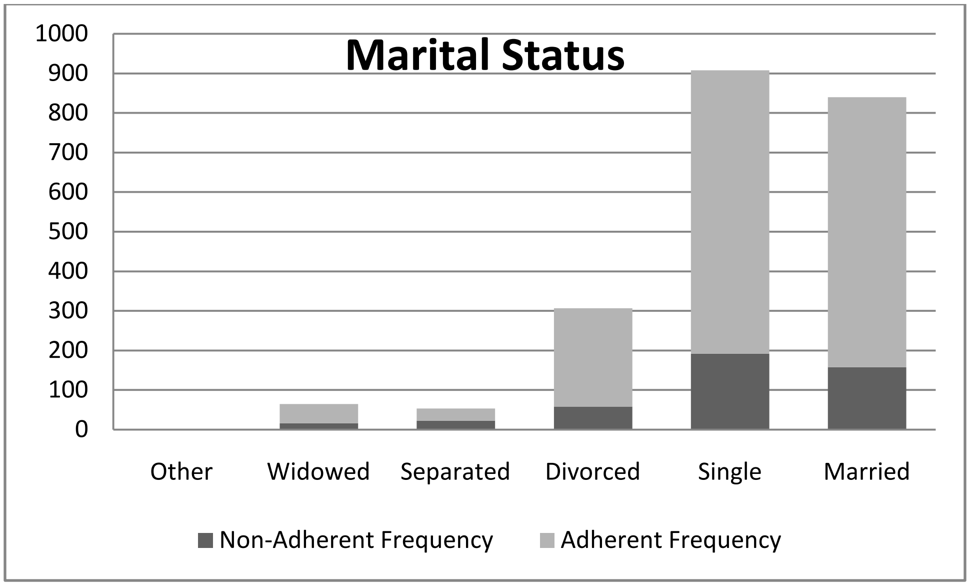

Figure 8. Distribution of Marital Status by Status

The role of patient employment status is substantiated in the literature, with evidence that an employed patient with a mental health condition is more likely to sustain treatment, including visits, and that patients who are unemployed tend to miss more appointments [123-124]. Results from this distribution analysis indicated that individuals who are retired are more likely to be visit adherent than either students or homemakers. The differences in adherences rates among those patients who are disabled, un-employed, or employed are less clear. The relationship of age and employment status will be further explored after a final model is selected. There is a distinct possibility that these may confound each other. The distribution is also marked by very sparse data in the employment category “Other.” Based on the low frequency, the decision was made to incorporate “Other” into the “employed” category for the purposes of further analysis. Please see Figure 9 below for a visual representation of the distribution of patient employment by status.

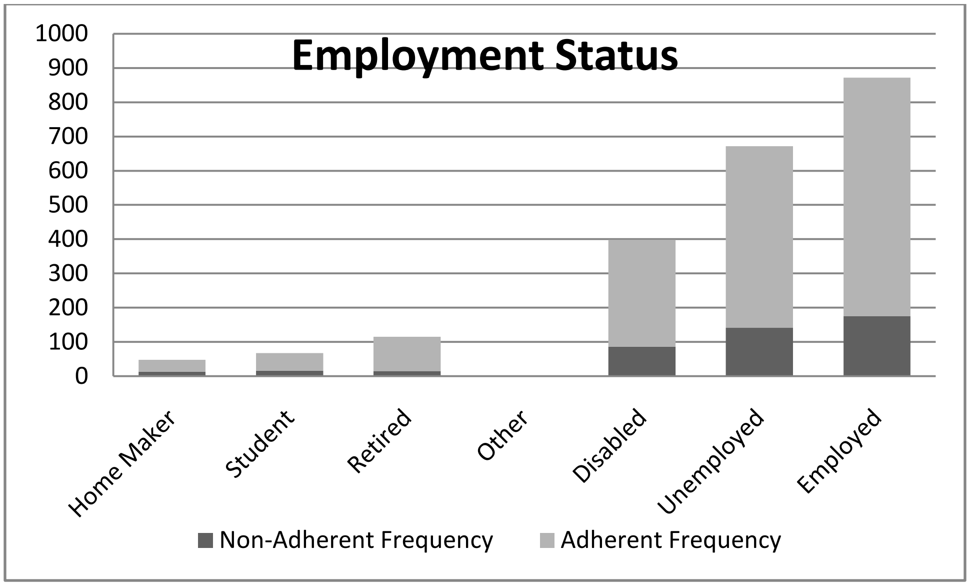

Figure 9. Distribution of Employment Status by Status

Per the relevant literature, patient race appears to be correlated to visit non-adherence [59, 107-108, 116, 119, 125-126]. The visit adherence rate is thought to be lower for patients with minority status, and especially low for those who self-identify are African American (Black) or Native American. Race is a difficult determinant to evaluate because of its potential correlation with payer (which could also be correlated with employment status) and with gender. The literature suggests that race, when combined with gender (specifically in African American females), is a marker of a tendency towards visit non-adherence; therefore, additional bi-variant analyses along these lines were conducted [132]. Asian or white patients tend to have lower rates of visit non-adherence [23, 116]. The low minority population in the sample meant that only three categories were viable (African American (Black), Caucasian (White) and Other (which contained Asian American, Hispanic American, Pacific Islanders, and Native Americans)). The results of the frequency distribution show that the sample is heavily skewed towards one race, but that there may not be substantial differences in adherence rates among the three categories. Please see Figure 10 below for a visual representation of the distribution of race by status.

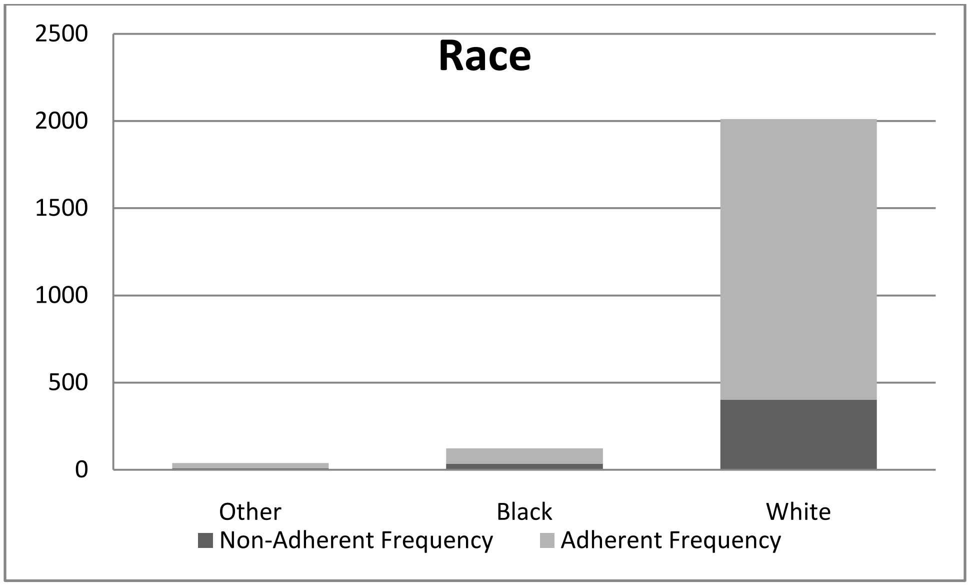

Figure 10. Distribution of Patient Race by Status

The literature considers payer type to be closely related to the cost of care, and the cost of care is significantly associated with visit non-adherence [120]. In many studies, uninsured patients have a significantly higher visit non-adherence rate, while patients with private or managed care insurance appear to have higher visit adherence rates [116, 119, 121, 135]. In our study, as in those described in the literature, patients with government-funded insurance based on income may also have a tendency towards visit non-adherence [136]. Patients with commercial fee-for-service insurance appear more likely to attend appointments. The rate of visit non-adherence in the self-pay population was lower than what other researchers had found, perhaps an effect of loss of Medicaid coverage on an already established patient population. Figure 11 below describes the payer distribution in this sample.

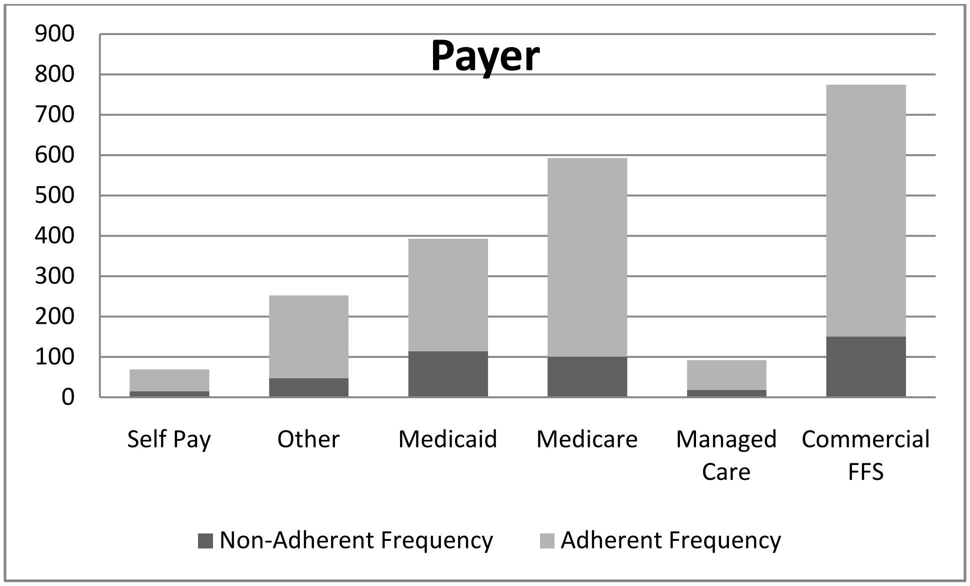

Figure 11. Distribution of Payer by Status

The relationship of listed contact person to the patient was defined, for the purposes of this study, as the relationship of the emergency contact person to the patient. Literature suggested that the lack of familial relationships can negatively affect visit adherence, while social support from family members and others decreases the likelihood of visit non-adherence [99, 125]. A striking number of visits within this study were associated with a contact person listed as “other,” indicating most frequently a non-familial relationship. Beyond that, it appears that patients who list spouses as contact person are more likely to be visit adherent than those that list parents. To minimize the effect of sparse data on the determinant, further analysis of the contact person relationship was carried out with the data re-partitioned as either “family” or “non-family”. The distribution of contact person relationship can be seen in Figure 12 below.

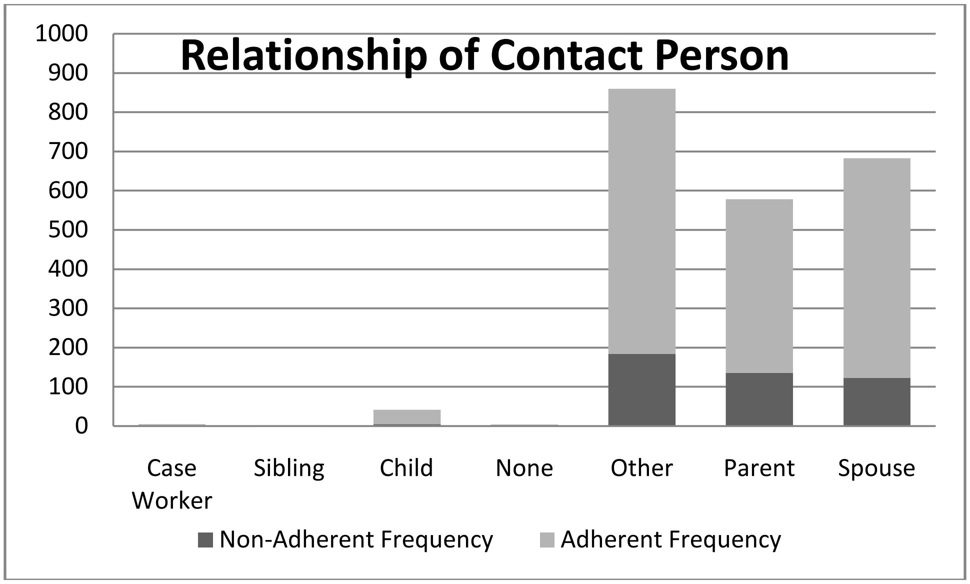

Figure12. Distribution of Relationship of Contact Person to Patient by Status

Primary (or first) diagnosis (defined as the first diagnosis listed by the physician at a visit as coded by ICD-9 CM codes) and secondary diagnosis (defined as the second diagnosis, if any, listed by the physician at a visit as coded by ICD-9 CM codes) are supported in the literature as determinants of visit non-adherence [36, 118, 138]. The results of this study seem to support the findings in the literature that patients with a diagnosis of schizophrenia or psychosis tend to have more difficulty with visit adherence [37, 121-123, 140]. Please see Figure 13 below for a visual of the distribution of general diagnosis by status of the appointment.

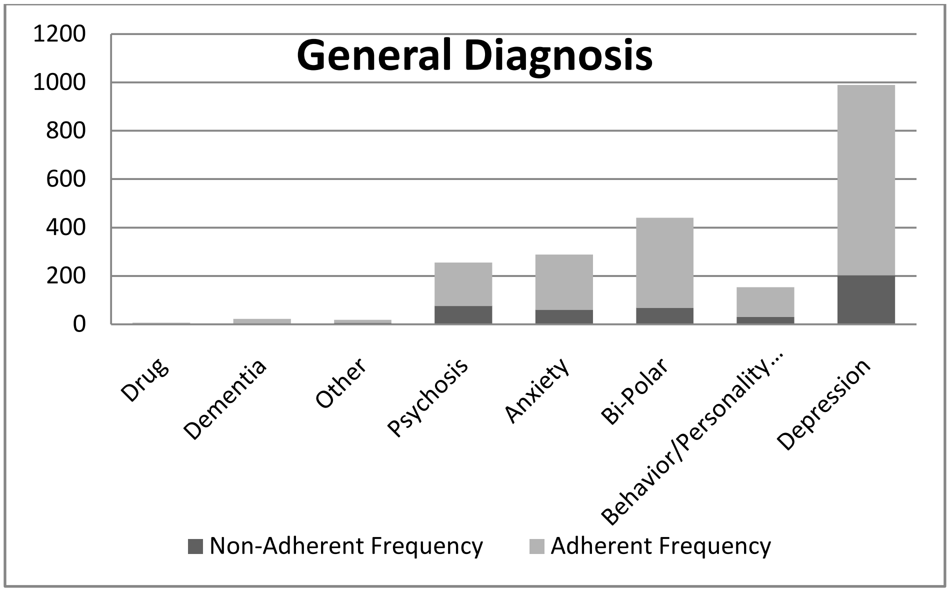

Figure 13. Distribution of General (Primary) Diagnosis by Status

The distribution of Secondary Diagnosis showed that a significant percentage of visits had no second diagnosis available. This lack is, perhaps, best accounted for by the billing practices of the providers rather than a reflection of actual disease, or lack thereof, although the absence of a second diagnosis certainly could be clinically correct. Because of the frequency of visits without a second diagnosis, a new determinant was created that simply recorded if the visit had or had not, a second diagnosis. Called NoDx2, the determinants replaced Sec Diagnoses in the next cycle of analysis. The same strategy had been employed (to no better effect) in preliminary Study 3. The change also proved ineffective in this analysis. Please see Figure 14 below for the distribution of secondary diagnoses.

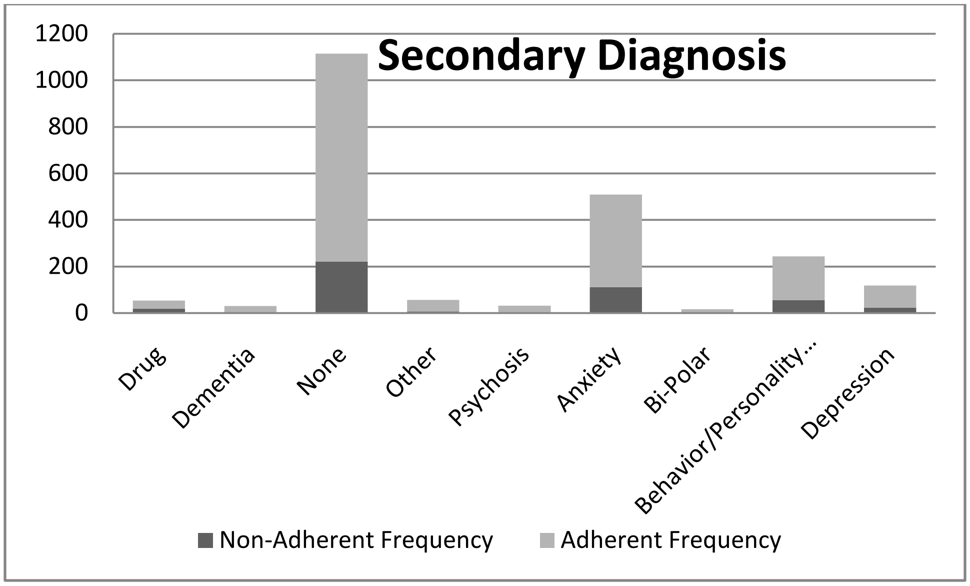

Figure 14. Distribution of Secondary Diagnosis by Status

For the purposes of this study, travel distance is defined as the distance from the five digit zip code of patient origin (as defined by US Postal Service codes) to the five digit zip code of the Psychiatry Clinic. Travel distance and mode of transport are considered important determinants of visit non-adherence and are well studied [103, 120]. Logically, and in previous studies, it seems that patients with longer travel distances would have a higher risk of visit non-adherence [116, 152]. Some research supports that living less than 15 miles from the care delivery site is also significantly associated with visit non-adherence in the form of treatment termination [96, 121]. In this study, travel distance is hampered by sparse data. To help offset this effect, a set of ranges was created, based on what appeared to be natural cut off points in the data. These ranges (0 - 6 miles = “in town”, 7 - 40 miles =”local”, and 41+ miles = long distance) permit improved analysis. It is interesting that while the distribution of visits shows that expected drop in the number of visits decreasing by increasing travel miles, that certain peaks exist in the data where increased visit non-adherence seems to occur. One of these peaks occurs between 35 and 40 miles distant from the clinic site, and represents an area of the state from which a disproportionate number of referrals originates. A second such peak occurs at approximately the 70 mile range. This tri-modal distribution of visit non-adherence had been documented in preliminary Studies 1 and 3. Please see Figure 15 below for a visual of the distribution of non-adherent and adherent visits.

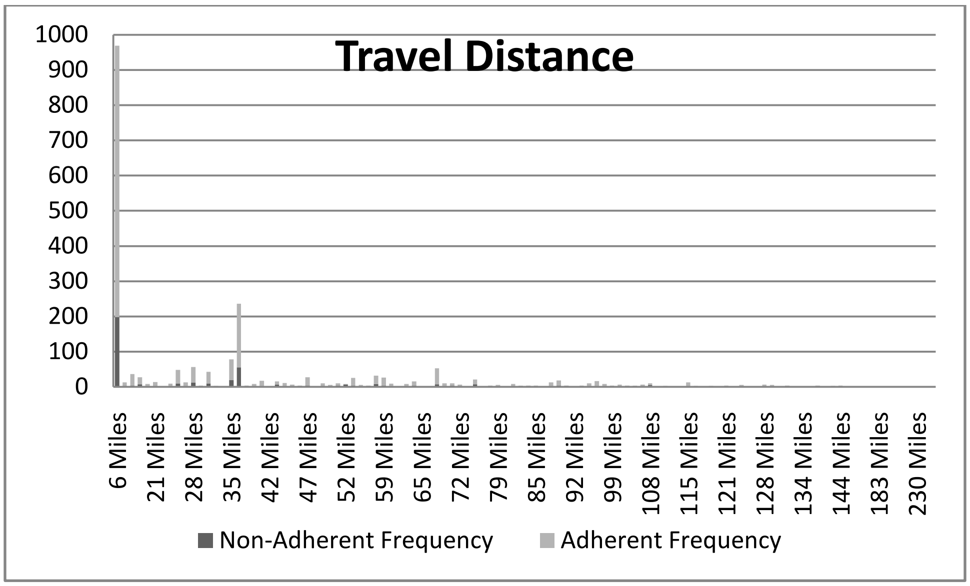

Figure 15. Distribution of Travel Distance by Status

In this study, wait days to appointment is defined as the number of days that elapse from the time of a request for an appointment to the date service is available. The number of wait days as an important determinant of visit non-adherence is supported in a number of studies, with most studies suggesting that longer wait time leads to more non-adherent visits, and that a wait time of more than two weeks also increases the risk of visit non-adherence [1, 7, 19, 23, 26, 28, 52, 95, 118, 148, 153-159]. Several studies, disputed that, and indicated that there was no demonstrable relationship between wait time and missed appointments [37, 133, 160]. The results of this study appear to support the latter finding. No large effect of wait time on visits adherence was observed, perhaps because this study focused on return patient visits. A visual of the distribution of wait days by visits status appears below in Figure 16.

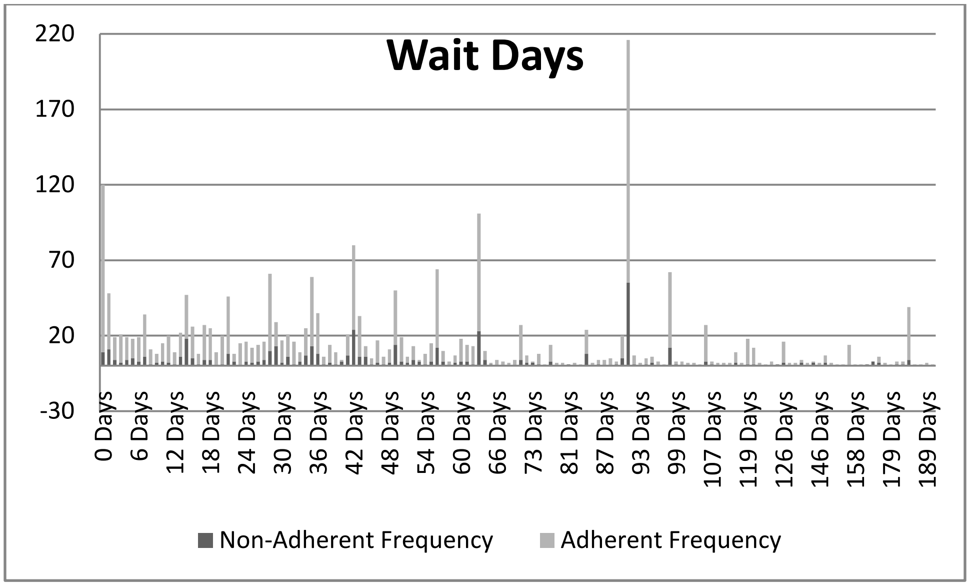

Figure 16. Distribution of Wait Days by Status

In this study, as in most others, appointment time of day is defined as standard military time of the scheduled appointment. The time of the appointment has been significantly associated with visit non-adherence [28, 162]. The main effect of time of day on visit adherence in this study appears to be that morning visits are slightly less adherent that those in the afternoon. Please see below (Figure 17).

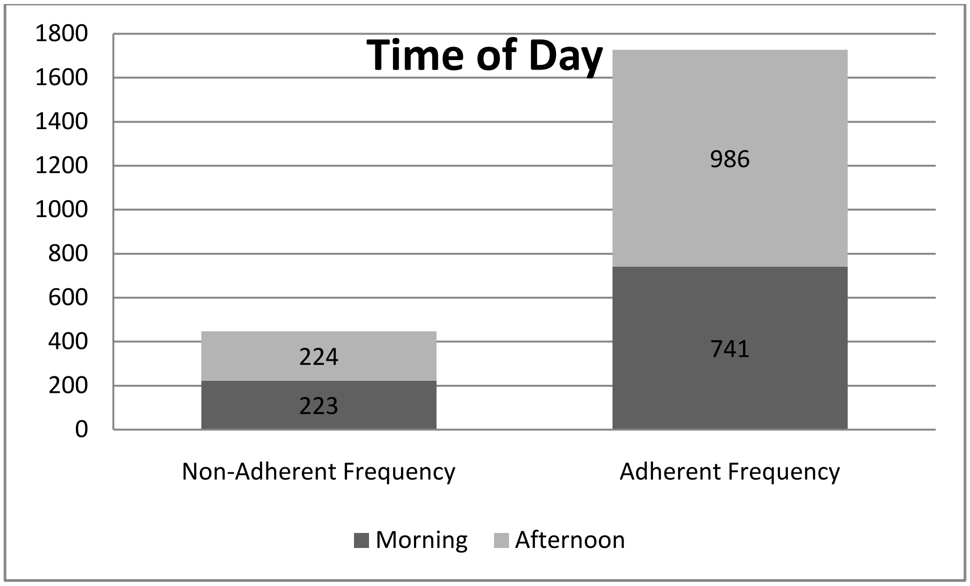

Figure 17. Distribution of Time of Day by Status

This study organically included a determinant for the appointment hour as well as the time of the appointments. Appointment Hour was rejected for further analysis when it proved to be less discriminatory that the exact visits time

Appointment day of week was defined as Monday, Tuesday, Wednesday, Thursday, or Friday. As suggested by one study conducted in an outpatient psychiatric clinic, Mondays appear to have a higher non-adherence rate [286]. Please see Figure 18 below for the distribution of non-adherent and adherent visits by the day of week in which they occurred.

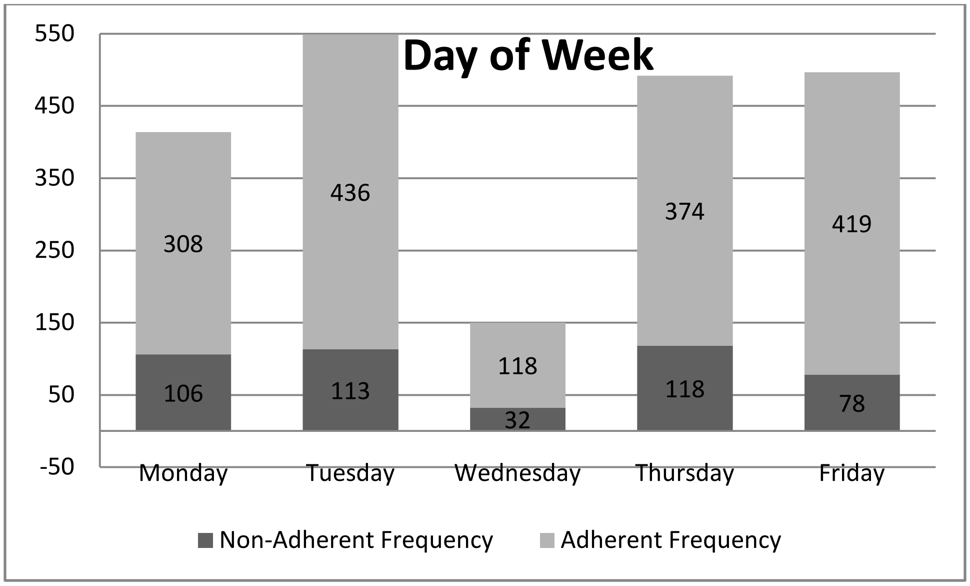

Figure 18. Distribution of Day of Week by Status

Appointment date was defined as the date of service. Because of the large distribution of appointment dates, and the relatively low visit counts for some days, we looked to the literature for ways to group services dates into a more usable form. Research suggested that there are seasonal patterns in visit non-adherence and in the causes for which care is needed [105]. Subsequent to the discovery of the Purdue study, conducted at a Veteran's Administration facility, seasonality of visits entered into our discussion of potential determinants [80, 113]. Therefore, we modified service or appointment date to a measure of seasonality. It appears that fall is the season of the year that experiences the largest proportion of non-adherent visits. Please see a visual of this phenomenon directly below.

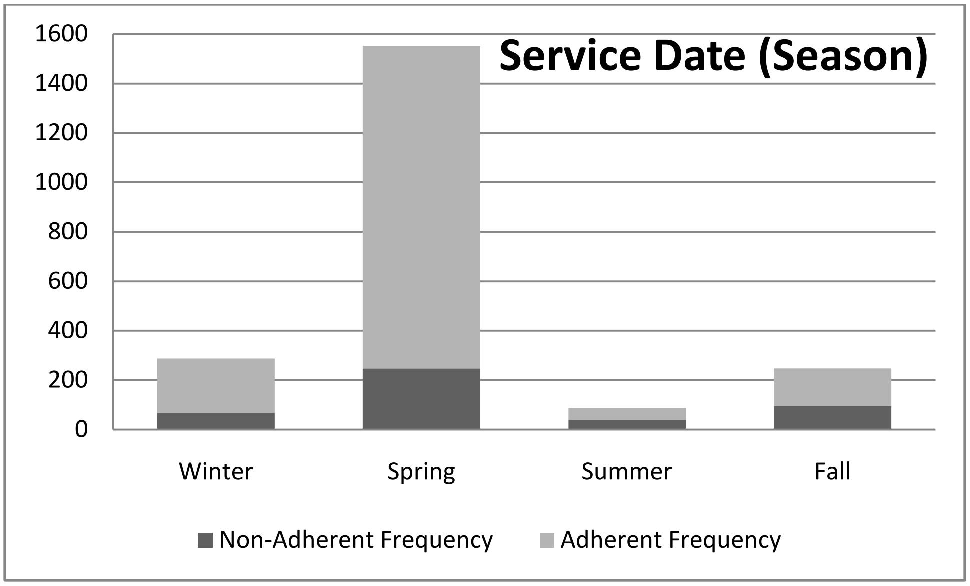

Figure 19. Distribution of Service Date (Season) by Status

Patient use of non-MD mental health care providers is defined as a documentable use, by the patient, of mental health services provided by counselors, psychologists, social workers, nurses, and lay mental health workers outside on the ambulatory psychiatric clinic environment. Previously completed research reported that previous instances of any mental health visit non-adherence are considered to an accurate measure of the likelihood of future psychiatric visit adherence [118]. This study shows little effect of patient use of Non-MD mental health services on the likelihood of visit non-adherence. Please see Figure 20 below.

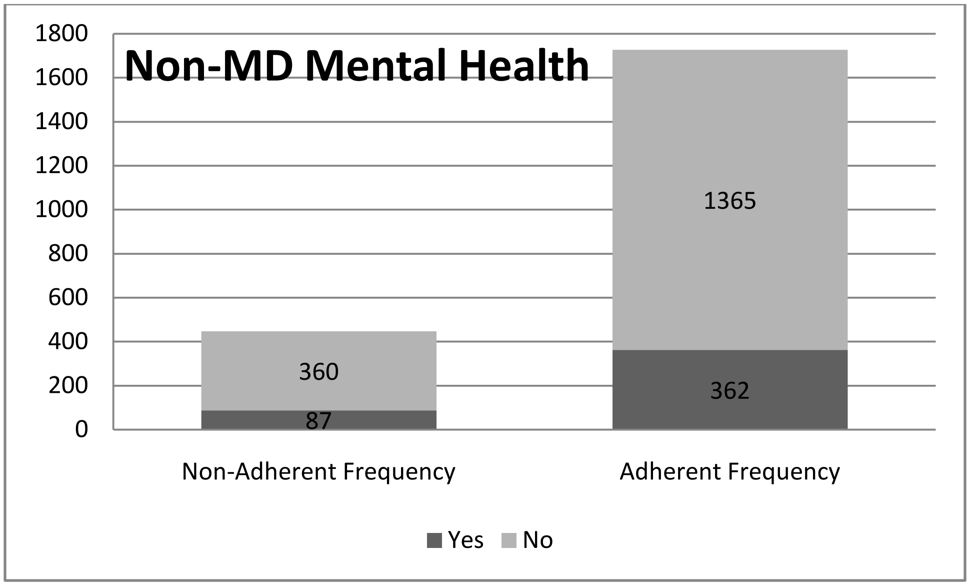

Figure 20. Distribution of Use of Non-MD Mental Health Services

Given the characteristics of the billing and schedule system that formed the basis of data collection for this study, total number of appointments, total number of visit non-adherent appointments, and total number of cancelled appointments are all interrelated. Conclusions from relevant literature about number of appointments include the conclusion that larger numbers of appointments may lead to decreased visit adherence, and that the percent of non-canceled appointments to kept appointments was independently significant [20, 108]. Also, the likelihood of visit non-adherence has been effectively measured by calculating and analyzing the percentage of adherent visits to scheduled appointments, hence the utilization of a determinant (Percentnon) to replace number of non-adherent visits when that determinant proved difficult to analyze [36, 108, 163]. However, it appears that patients in the sample used for this study are not repeatedly cancelling appointments. Please see Figure 21 below for more detail. Likewise, the total number of appointments appears to be less contributory to visits non-adherence that previously thought. Please see Figure 22 below for detail on the distribution of the number of visits a patient has to the status of the appointment.

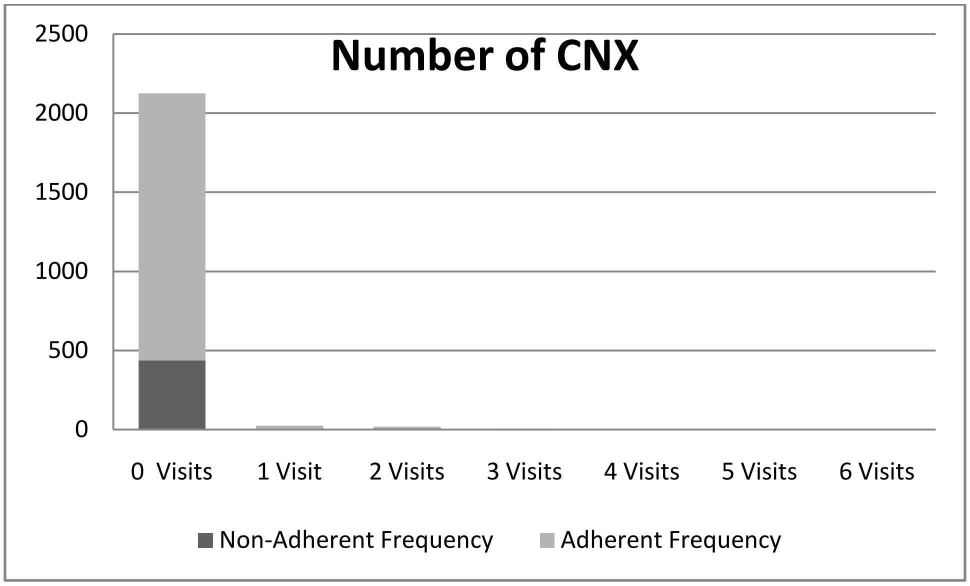

Figure 21. Distribution of Number of Cancelled Appointments (numbercnx) by Status

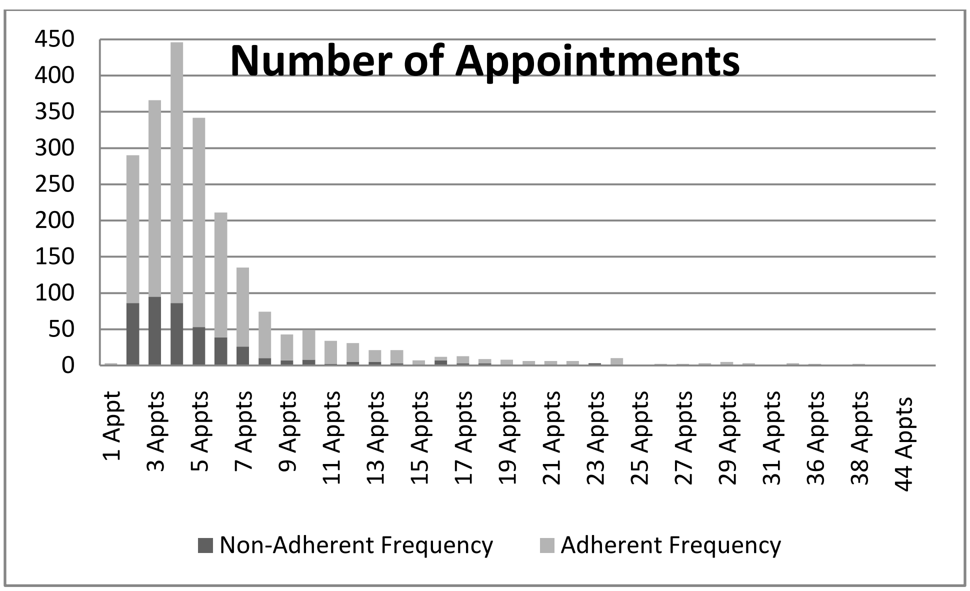

Figure 22. Distribution of Number of Appointment by Status

For the purposes of this study, maker of appointment was defined as the individual who entered the visit as an appointment in the Admission, Discharge, and Transfer system used by the clinic of study. In spite of the potential value of the determinant “maker” to the prediction of visit non-adherence, as suggested by the literature, this determinant demonstrated poor variability in this study, and was discontinued for that reason [33, 94, 123, 136, 164-165]. Please see Figure 23 below for additional detail on the distribution of maker by status.

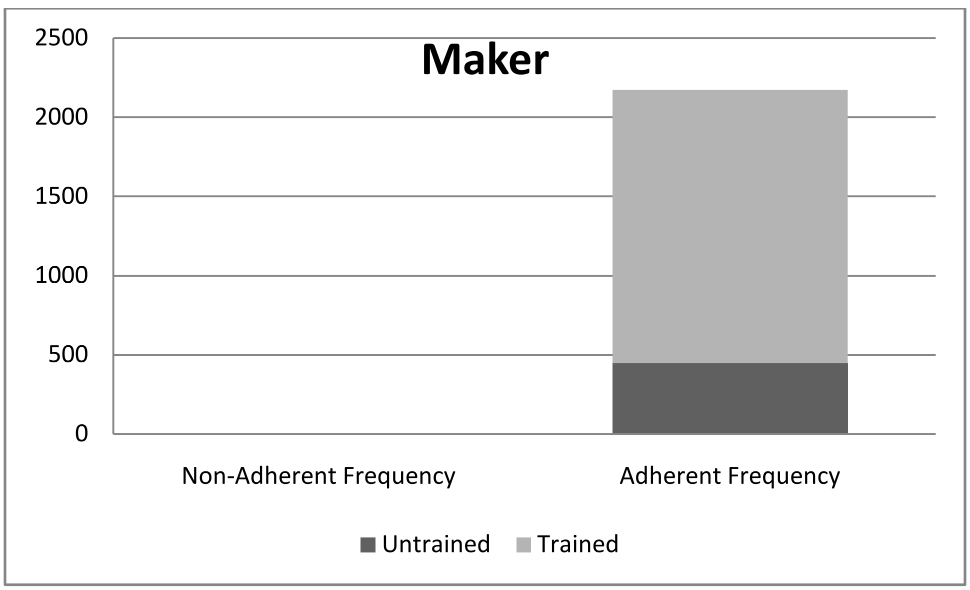

Figure 23. Distribution of Maker by Status

Provider type was defined as the general class of physician/provider (attending physician, resident physician, or counselor) for the visit. The specific treating physician is associated with visit adherence, and visit non-adherence rates appear to be higher is resident clinics [79, 116, 125, 166, 168]. It is, therefore, not surprising that this study found that visits scheduled with a resident have a higher non-adherence rate. Please see Figure 24 below for more detail regarding the distribution of provider and appointment status.

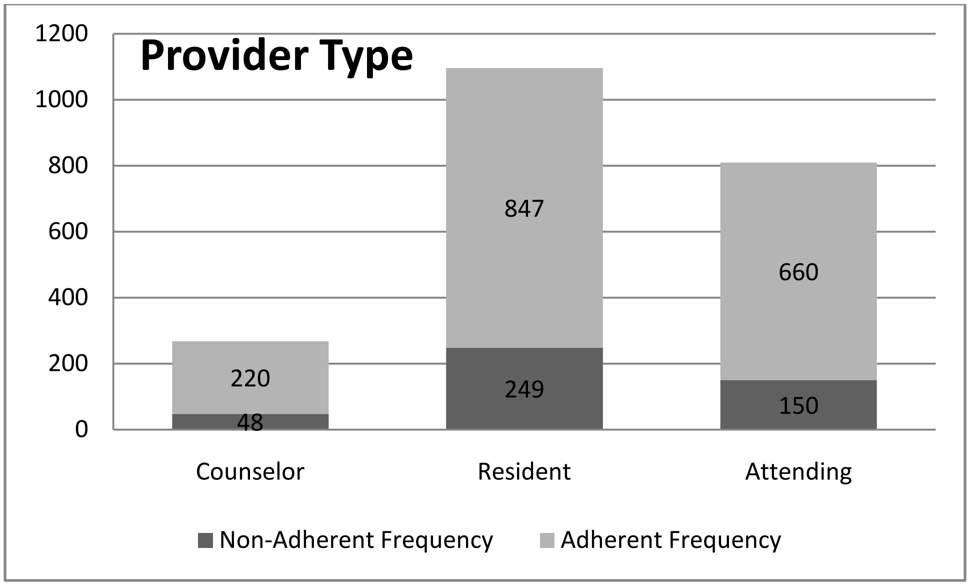

Figure 24. Distribution of Provider by Status

Referral source is defined as the type of person or entity that sent the patient to the clinic of study for care. Literature suggests that referral source impacts the rate of visit non-adherence, and that patients who are self-referred may have better visit adherence [28, 117-118, 133]. In this initial analysis of referring provider, its impact on the model may be somewhat diluted by the presences of sparse data in some classes. To offset this effect, or lack thereof, the values were reformatted as “self” and “provider” before any further analysis. Please see Figure 25 below for additional detail.

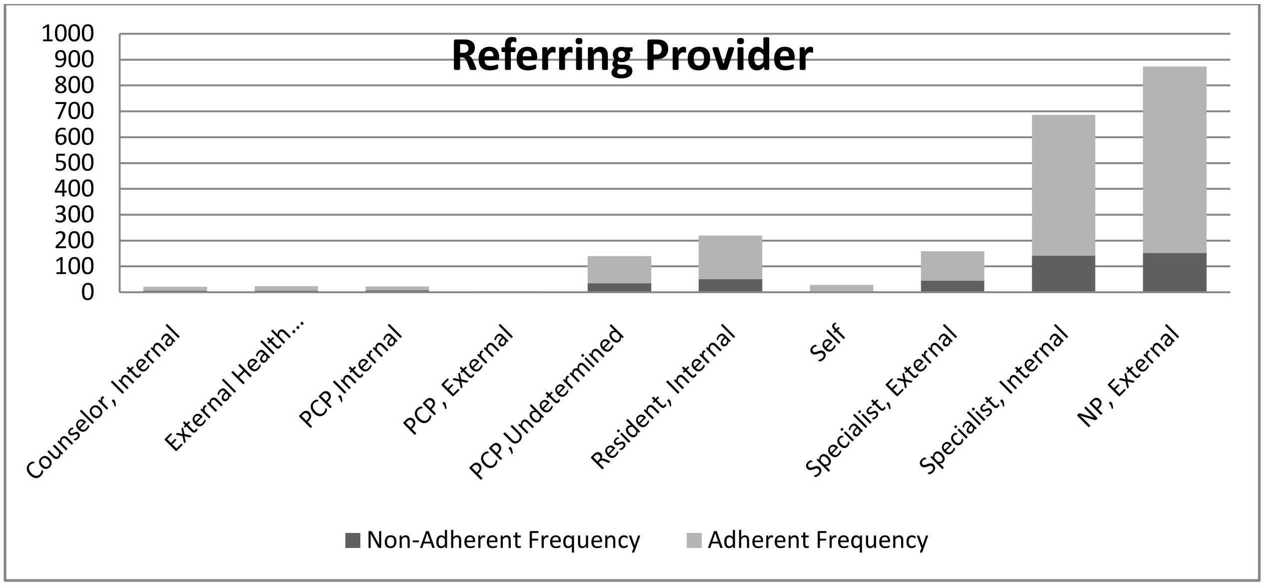

Figure 25. Distribution of Referring Provider by Status

Based on findings in the literature, a number of new candidate determinants were introduced in this study. Some proved to be more successful than other. They are briefly discussed in the following few paragraphs.

The determinant “Same Visit Type” was retired because of lack of variability. Please see below (Figure 26)

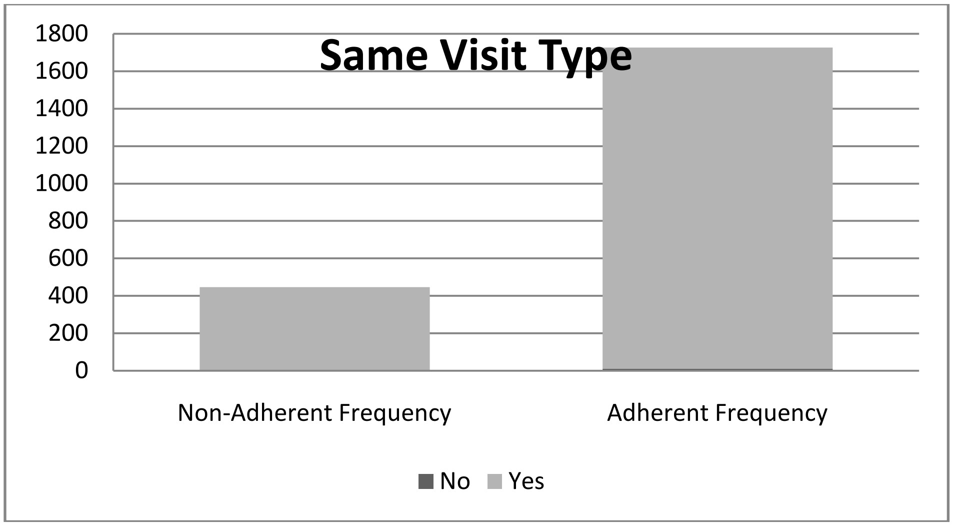

Figure 26. Distribution of Same Visit Type by Status

Same Day of Week was also not especially informative.

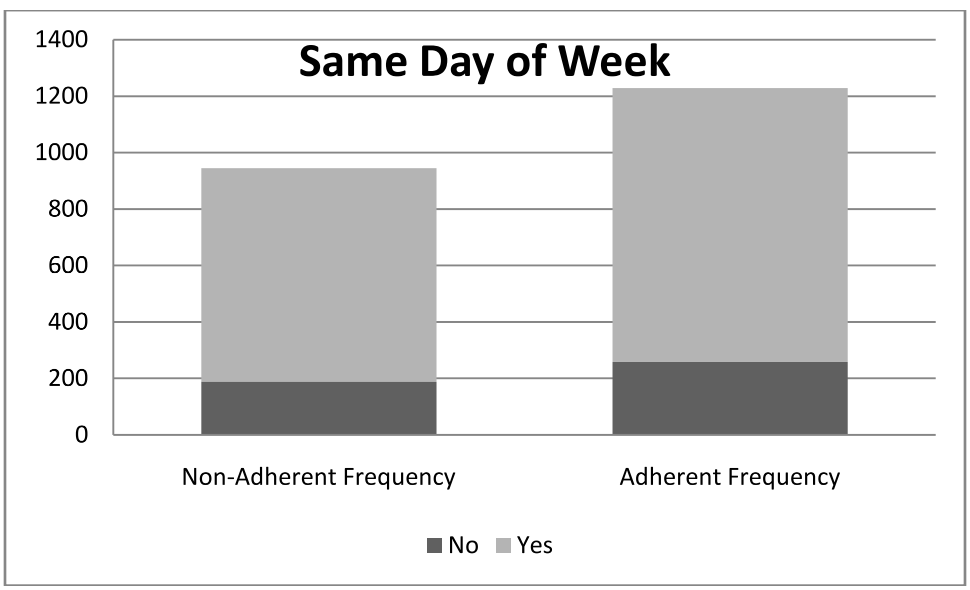

Figure 27. Distribution of Same Day of Week by Status

Same Time of Day was also marginally of interest.

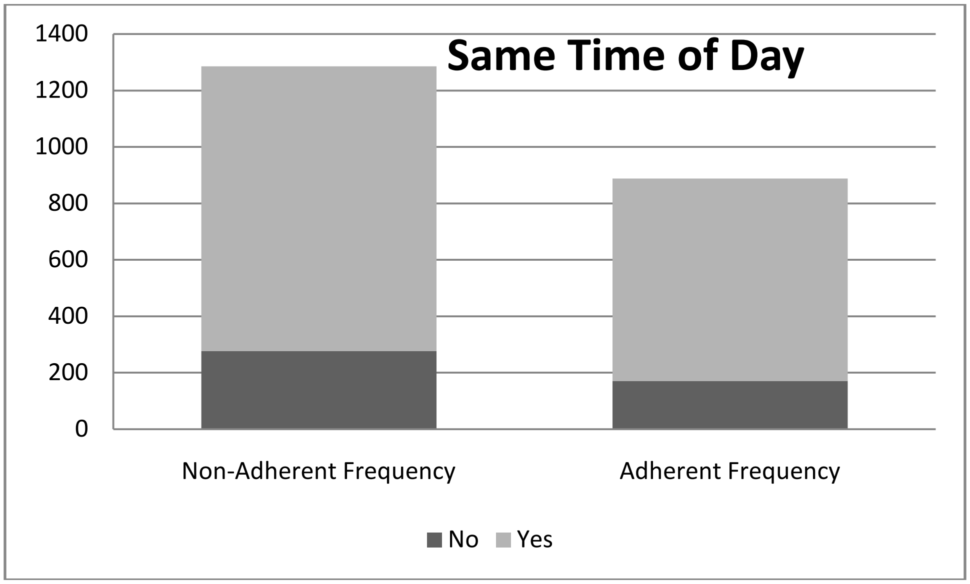

*Figure 28.*Distribution of Same Time of Day by Status

However, Same Payer appears to hold sufficient variability in its distribution with regards to adherence status to warrant further investigation. Please See Figure 29 below.

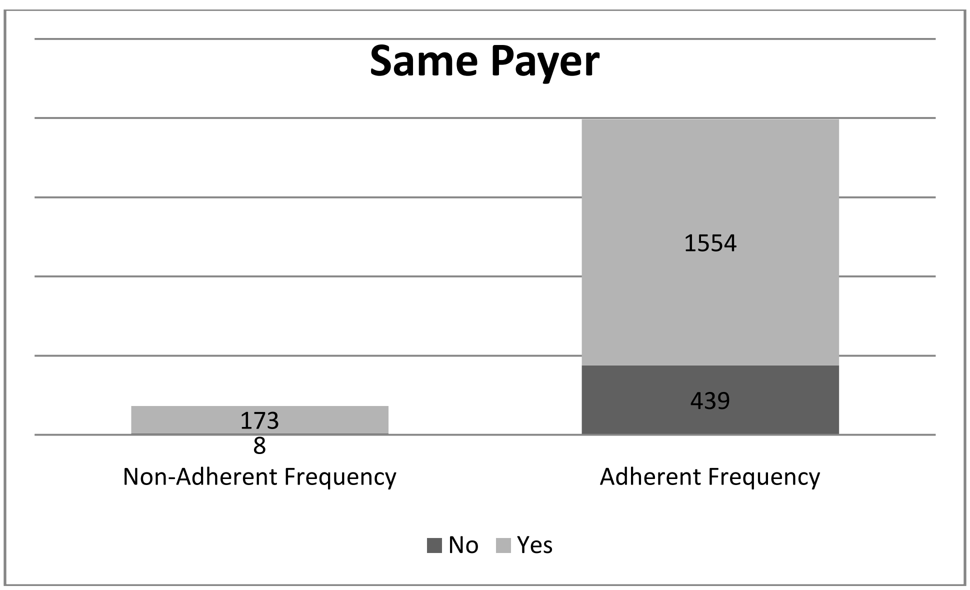

Figure 29. Distribution of Same Payer by Status

Same Provider was also retained for further analysis, although it is somewhat suspect.

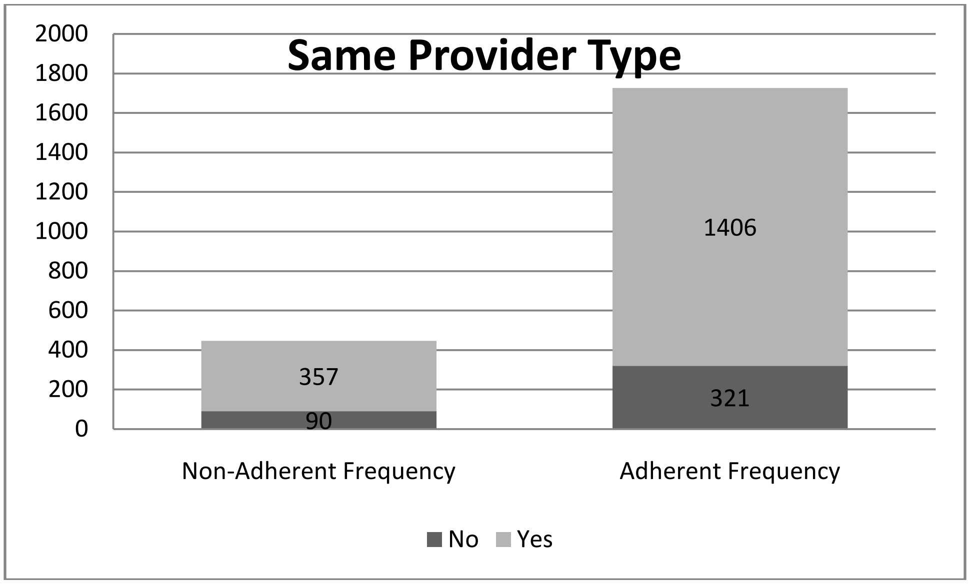

Figure 30. Distribution of Same Provider Type by Status

The two determinants created to replace the number of non-adherent visits showed very different distributions. Percentnon is very diffuse, with areas of very sparse data and no clear ranging except for the value of 0. Prevvisitnos, however, exhibited a very interesting distribution. Given that distribution, it holds promise as a strong determinant of potential visit non-adherence. Of the two, prevvistnos and percentnon, prevvisitnos is also the easiest to extract from electronic memory or from human understanding, as no calculation is needed beyond locating the previous visit. Please see Figures 31 and 32 below for further detail on these two determinants.

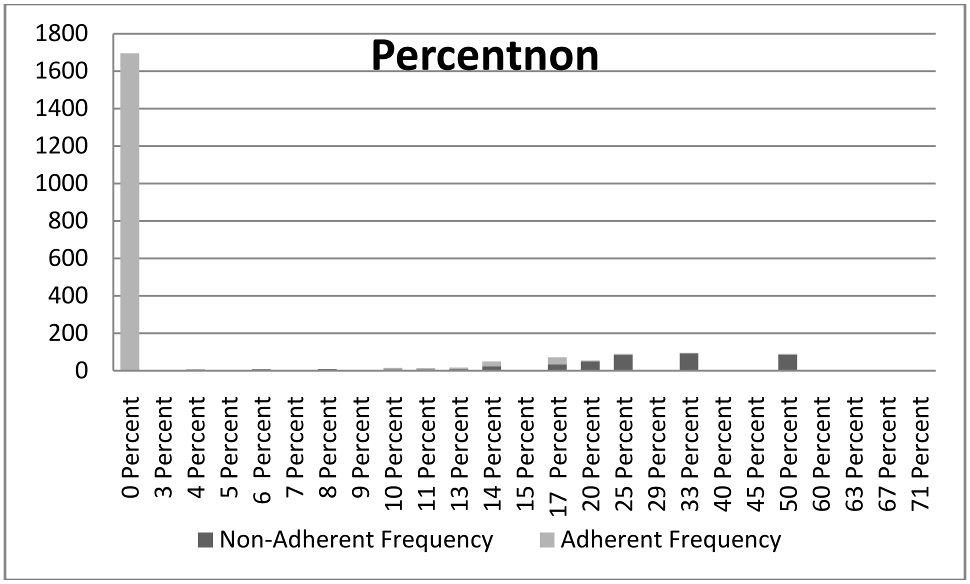

*Figure 31.*Distribution of Percent Non-Adherent (Percentnon) by Status

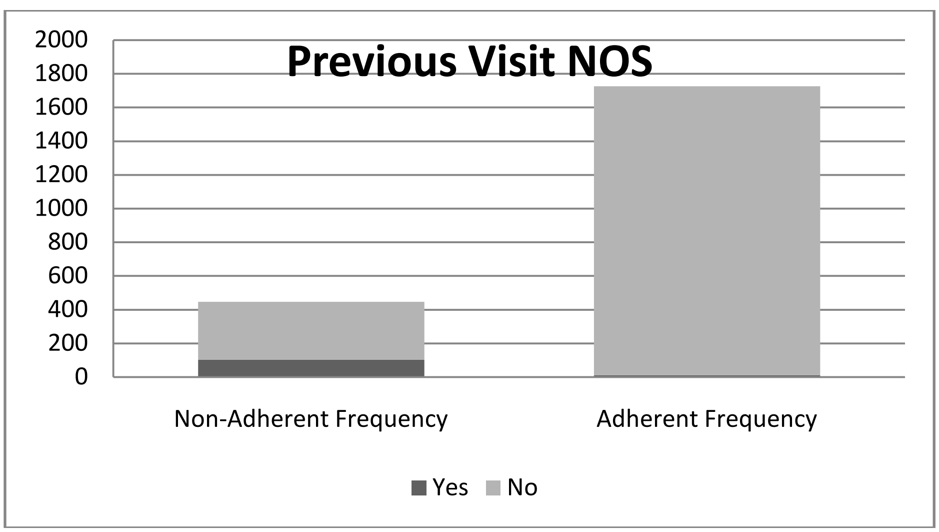

Figure 32. Distribution of Previous Visit Non-Adherent by Status

Several other potentially problematic distributions were observed to ensure that some major, unrealized, effect was not unduly affecting the analyses. Below is a series of visualization of these distributions, none of which exhibited sufficient cause for concern. The distribution of payer by patient race does not appear to be skewed, nor does the distribution of patient race by type of provider. The distribution of gender by age appears reflective of the sample (more females than males in general) and of the distribution of the genders in the general population (more women living longer than men). The distribution of type of payer by the type of employment appears to be consistent and appropriate with the general population. For examples, more retired person have Medicare, as do those that are disabled, while those who are employed tend to have commercial fee-for-service type insurance. Please see the following four figures (Figures 33, 34, 35, and 36) for further detail.

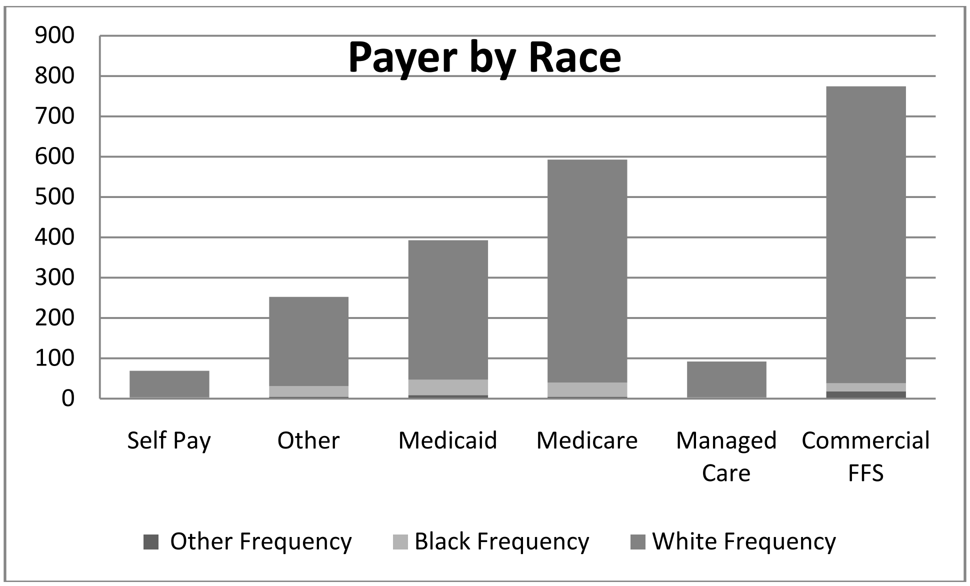

Figure 33. Distribution of Payer by Race

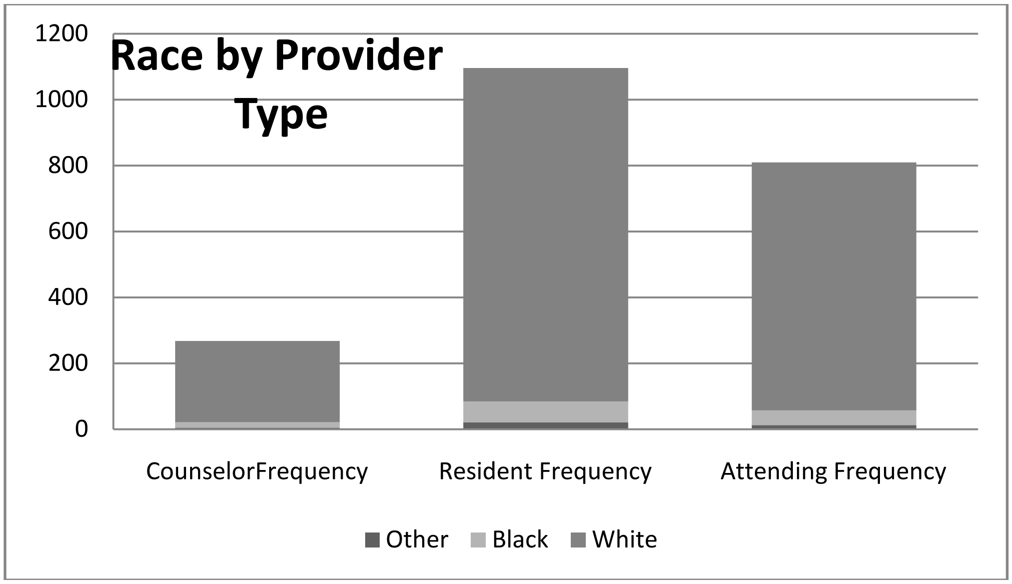

Figure 34. Distribution of Provider by Race

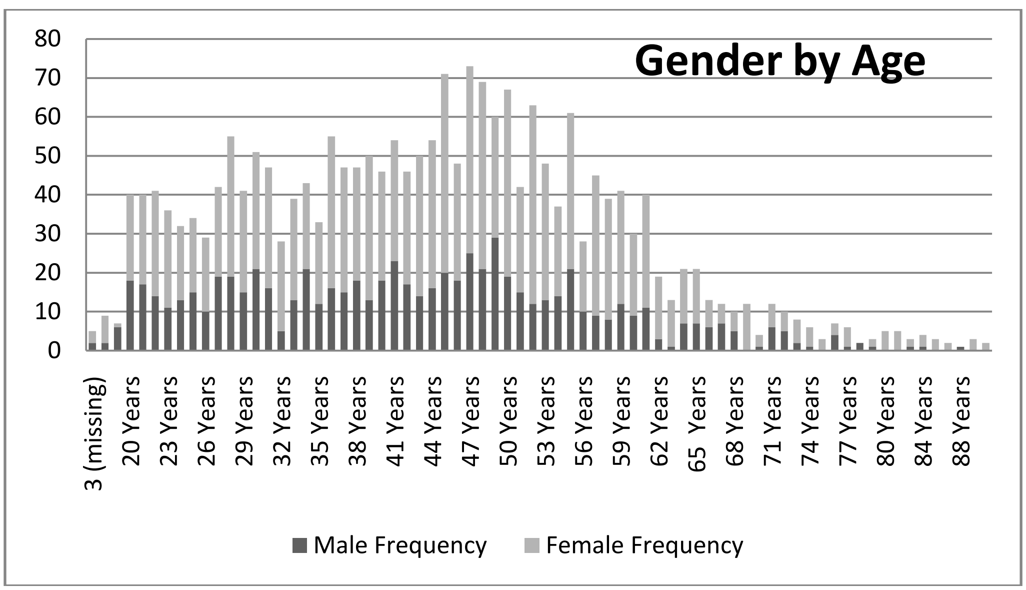

Figure 35. Distribution of Gender by Age

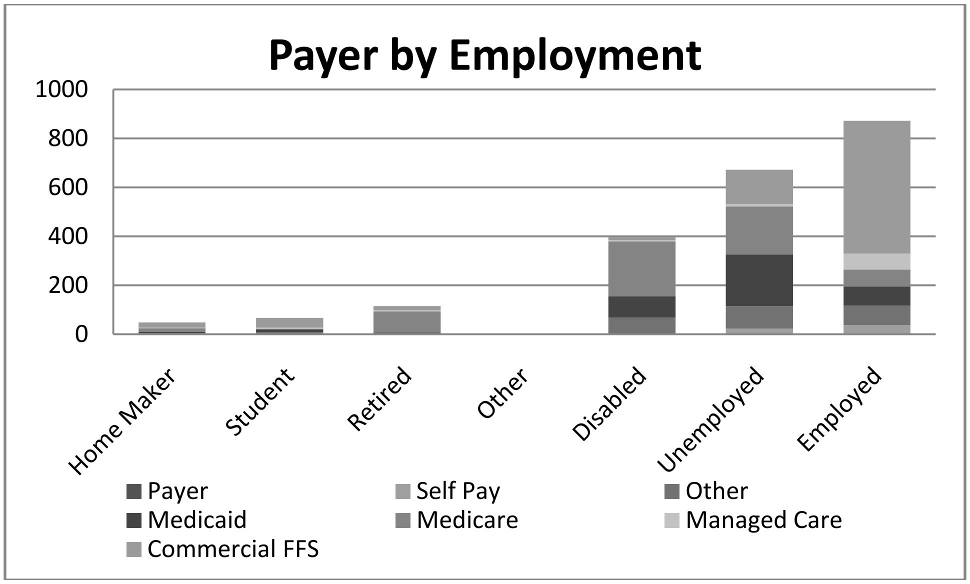

Figure 36. Distribution of Payer by Employment

Table 17 (see below) shows the overall frequency and p values for each of the determinants at this point in the analysis. Based on these results, several determinants were removed from further consideration. These included:

- Travel Distance
- Same Visit Type
- Same Day of Week
- Same Time of Day
- Referring Provider
- Employment
- Number of Cancelled Appointments
- Non-MD Mental Health
- Relationship of Contact Person
- Same Provider

Table17

Intermediate Frequency Procedure Results for Status by Each Categorical Determinant

Independent Variable		Population Size (n)	Percent Non-Adherent	Percent Adherent	P-value
Gender	Female	1446	66.13	68.01	0.4522
	Male	728	31.99	33.87	n/a
Marital Status	Widowed	65	2.84	3.58	0.0005
	Separated	53	1.74	5.15	n/a
	Divorced	307	14.42	12.98	n/a
	Single	908	41.46	42.95	n/a
	Married	841	39.55	35.35	n/a
Employment	Home Maker	48	2.03	2.91	0.2950
	Student	67	2.95	3.58	n/a
	Retired	115	5.79	3.36	n/a
	Disabled	398	18.07	19.24	n/a
	Unemployed	672	30.75	31.54	n/a
	Employed/Other	874	40.42	39.37	n/a
Travel Distance	In Town	969	44.64	44.30	0.8145
	Local	603	27.45	28.86	n/a
	Distant	602	27.91	26.85	n/a
Race	Other	39	1.74	2.01	0.0746
	Black	123	5.10	7.83	n/a
	White	2012	93.17	90.16	n/a
Payer Type	Self pay	69	3.13	3.36	0.0004
	Other Payer	252	11.81	10.47	n/a
	Medicaid	393	16.16	25.50	n/a
	Medicare	593	28.49	22.60	n/a
	Managed Care	92	4.28	4.03	n/a
	Commercial FFS	775	36.16	33.78	n/a
Relationship of Contact Person	Non-Family	869	39.66	41.16	0.5641
	Family	1305	60.34	58.84	n/a
Gen. Diagnosis	Drugs	6	0.12	0.89	<0.0001
	Dementia	23	1.27	0.22	n/a
	Other Diagnoses	18	0.75	1.12	n/a
	Psychosis	255	10.36	17.00	n/a
	Anxiety	289	13.26	13.42	n/a
	Bi-polar	440	21.54	15.21	n/a
	Behavior/Personality Disorder	153	7.12	6.71	n/a
	Depression	990	45.57	45.41	n/a
Sec. Diagnosis	Drugs	54	2.03	4.25	0.0443

Independent Variable		Population Size (n)	Percent Non-Adherent	Percent Adherent	P-value
	Dementia	30	1.62	0.45	n/a
	None	1114	51.71	49.44	n/a
	Other Diagnoses	57	2.90	1.57	n/a
	Psychosis	31	1.56	0.89	n/a
	Anxiety	509	22.99	25.06	n/a
	Bi-polar	17	0.81	0.67	n/a
	Behavior/Personality Disorder	243	10.89	12.30	n/a
	Depression	119	5.50	2.37	n/a
Wait Days	Zero	120	6.43	2.01	0.0010
	1-30 Days	659	30.57	29.31	n/a
	31-90 Days	884	39.37	45.64	n/a
	91+ Days	511	23.62	23.04	n/a
Appt Type	All return	All return	All return	All return	n/a
Appt Time	AM	964	42.81	49.89	0.0081
	PM	1210	57.09	42.91	n/a
Appt Date (Season)	Winter	288	12.80	14.99	<.0001
	Spring	1552	75.56	55.26	n/a
	Summer	87	2.78	8.72	n/a
	Fall	247	8.86	21.03	n/a
Appt Day of Week	Monday	486	22.00	23.71	0.0238
	Tuesday	549	25.25	25.28	n/a
	Wednesday	150	6.83	7.16	n/a
	Thursday	492	21.66	26.40	n/a
	Friday	497	24.26	17.45	n/a
MD Mental Health	Yes	449	20.96	19.46	0.4856
	No	1725	79.04	80.54	n/a
Number of Adherent Appts	Modified to Prevvisitnos and Prevapptnos	n/a	n/a	n/a	
Number of Canceled Appts	None	2125	97.74	97.76	0.9786
	1-6	49	2.26	2.24	n/a
Appt Maker	Removed/No variation	Removed/ No variation	Removed/ No variation	Removed/No variation	n/a
Type of Provider	Counselor	268	12.74	10.74	0.0419
	Resident	1096	49.04	55.70	n/a

Independent Variable		Population Size (n)	Percent Non-Adherent	Percent Adherent	P-value
	Attending	810	38.22	33.56	n/a
Referring Provider	Any Provider	1955	90.27	88.59	0.2924
	Self	219	9.73	11.41	n/a
Same Visit Type	No	14	99.33	0.67	0.9358
	Yes	2160	99.36	0.64	n/a
Same Day of Week	No	945	57.72	42.28	0.5702
	Yes	1229	56.22	43.78	n/a
Same Time of Day	No	1286	58.48	61.74	0.2111
	Yes	888	41.52	58.48	n/a
Same Payer	No	181	10.02	1.79	<.0001
	Yes	1993	89.98	98.21	n/a
Same Provider	No	411	18.59	20.13	0.4565
	Yes	1763	81.41	79.87	n/a
Previous Visit adherent*	Yes	115	0.69	23.04	<.0001
	No	2059	99.31	76.96	n/a
Presence of Secondary Diagnosis*	Added	Added	Added	Added	n/a

Before continuing with the analysis, a test was made to ensure that bootstrapping with 100 replicated was sufficient. The use of 100 replicates conserved computational costs, but might cause concern if the number of samples was seen as insufficient. The results of the testing showed that, for the purposes of this study, 100 replicates yielded results similar to those from 1000 replicates. Please see Tables 18 and 19 below for regression main effects summaries of each trial.

Table 18

Effect Entered for Bootstrap 1000 Replicates

Frequency	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11	Step 12
prevvistnos	1000	0	0	0	0	0	0	0	0	0	0	0
samepayer	0	14	473	317	139	49	6	2	0	0	0	0
season	0	969	30	1	0	0	0	0	0	0	0	0
maritalstatus	0	13	200	202	235	200	81	37	12	5	0	0
waitdays	0	2	199	245	203	176	63	51	27	8	2	0
ageC44	0	1	45	92	257	306	127	60	43	11	2	2
apptdayof week	0	0	7	29	49	75	224	215	106	40	13	0
gendiagnosis	0	0	7	5	8	44	190	172	137	49	20	0
provider	0	0	14	46	32	47	104	155	114	73	10	0
payer	0	1	24	60	57	58	69	73	74	58	13	1

Table 19

Effect Entered for Bootstrap 100 Replicates

Frequency	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11	Step 12
prevvistnos	100	0	0	0	0	0	0	0	0	0	0	0
samepayer	0	2	50	32	14	2	0	0	0	0	0	0
season	0	97	3	0	0	0	0	0	0	0	0	0
waitdays	0	0	15	27	25	21	4	5	2	0	1	0
maritalstat us	0	1	19	20	25	18	10	3	2	0	0	0
ageC44	0	0	5	8	16	33	19	5	3	2	0	1
apptdayof week	0	0	0	3	4	10	28	24	7	4	2	0
gendiagnos is	0	0	2	0	3	3	18	18	13	5	2	0
provider	0	0	2	2	6	5	6	17	14	8	0	0
payer	0	0	4	8	5	8	5	10	4	4	1	2

Utilizing bootstrapping (with 100 replicates), another regression was carried out utilizing the remaining determinants. The results can be seen in Table 20 below.

Table 20

Summary of Effects Entered

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11	Step 12
Gender	0	0	0	0	1	0	2	1	1	2	0	0
Age	0	1	3	7	29	30	9	10	3	2	0	0
Wait Days	0	2	22	24	23	18	5	2	2	1	0	0
Season	0	93	7	0	0	0	0	0	0	0	0	0
Day of Week	0	0	0	2	2	8	18	24	11	8	0	0
Appointment Time	0	0	0	0	0	1	2	1	2	3	1	0
Payer	0	0	1	4	10	5	10	2	9	7	1	0
Same Payer	0	2	41	34	14	5	2	0	0	0	0	0
Gen Diagnosis	0	0	0	0	0	5	29	21	14	1	0	0
Sec Diagnosis	0	0	0	1	0	1	2	10	11	2	0	1
Marital Status	0	2	23	23	20	22	5	0	2	0	0	0
Number Non-Adherent	0	0	0	0	0	0	0	0	0	0	0	0
Race	0	0	0	0	0	0	0	0	0	0	0	0
Number of Appointments	0	0	0	0	0	0	0	0	0	0	0	0
Previous Visit Non-Adherent	100	0	0	0	0	0	0	0	0	0	0	0
Provider	0	0	1	4	1	4	15	13	5	5	0	0

Based on the findings from the regression above, a third regression analysis was conducted to further refine the model. This test began with a set of determinants that included the following;

- Prevvistnos
- Same payer
- Season
- Waitdays
- Marital Status
- Age C44
- Appointment Day of Week
- General Diagnosis
- Provider
- Payer

The summary of main effects for this regression (See Table 21 and Table 22 below) showed that payer no longer entered the model quickly enough or frequently enough to continue its use and that its p value was > 0.05.

Table 21
Effects Entered

Frequency	Step 1	Step 2	Step 3	Step 4	Step 5
Prevvisitnos	100	0	0	0	0
Same Payer	0	2	50	32	14
Season	0	97	3	0	0
Wait days	0	0	15	27	25
Marital status	0	1	19	20	26
Age C44	0	0	5	8	16
apptdayofweek	0	0	0	3	4
gendiagnosis	0	0	2	0	3
provider	0	0	2	2	6
payer	0	0	4	8	6

Table 22
Analysis of Findings

Effect	DF	Wald	pr>ChiSq
Age C44	1	13.237	0.0003
Waitdays	3	17.5195	0.0006
Season	3	73.0693	<.0001
Appt day of week	4	12.0633	0.0169
Payer	5	6.4694	0.2632
Same payer	1	23.8638	<.0001
Gen diagnosis	5	8.7683	0.1187
Marital status	4	27.6755	<.0001
Prevvisitnos	1	128.1593	<.0001
Provider	2	6.0635	0.0482

With payer removed, the regression was rerun to determine the effects on the model. Tables 23 and 24 below describe the effects of the removal of payer from the model, showing that it appeared that the model would be sufficiently robust and perhaps more usable, with general diagnosis removed.

Table 23
Effects Entered with Payer Removed

Frequency	Step 1	Step 2	Step 3	Step 4	Step 5
Age C44	0	0	5	9	23
Prevvisitnos	100	0	0	0	0
Same payer	0	2	52	31	13
Season	0	97	3	0	0
Waitdays	0	0	15	31	25
Marital status	0	1	20	20	25
Appt day of week	0	0	0	4	5
Provider	0	0	2	5	6
Gen diagnosis	0	0	3	0	3

Table 24
Analysis of Findings with Payer Removed

Effect	DF	Wald	pr>ChiSq
Age C44	1	18.9242	<.0001
Waitdays	3	18.2861	0.0004
Season	3	73.4885	<.0001
Appt day of week	4	12.0027	0.0173
Same payer	1	23.3094	<.0001
Gen diagnosis	5	8.8702	0.1114
Marital status	4	29.1609	<.0001
Prevvisitnos	1	130.3961	<.0001
Provider	2	8.2840	0.0159

Removing general diagnosis has an additional positive influence on the usability of the final model. General diagnosis was the final remaining determinant in the model that could potentially be highly specific to a particular clinical practice. Its removal may mean that the final model can be applied to a wider variety of clinical practices. Removing general diagnosis achieved the following model, which appears to be the best model for the data. Please see Tables 25, 26, 27, 28, 29 and 30 below for details.

Table 25

Effects Entered with Payer and Gen Diagnosis Removed

Frequency	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8
Prevvisitnos	100	0	0	0	0	0	0	0
Same payer	0	2	54	30	13	1	0	0
Season	0	97	3	0	0	0	0	0
Waitdays	0	0	15	32	24	18	6	5
Age C44	0	0	6	8	25	42	13	5
Marital status	0	1	20	20	26	21	7	3
Appt day of week	0	0	0	4	6	11	49	14
Provider	0	0	2	6	6	7	17	36

Table 26

Analysis of Findings with Payer and Gen Diagnosis Removed

Effect	DF	Wald	pr>ChiSq
Age C44	1	17.1449	<.0001
Waitdays	3	18.9794	0.0003
Season	3	73.9354	<.0001
Appt day of week	4	12.3405	0.015
Same payer	1	24.5064	<.0001
Marital status	4	28.2287	<.0001
Prevvisitnos	1	131.7266	<.0001
Provider	2	8.3460	0.0154

Table 27

Analysis of Maximum Likelihood Estimates for Final Model

Parameter		DF	Estimate	Standard Error	Wald Chi-Sq	Pr>ChiSq
Intercept		1	-0.0652	0.2827	0.0531	0.8177
Age C44		1	-0.0208	0.0050	17.1449	<.0001
Wait Days	None	1	-0.6999	0.2722	6.6094	0.0101
Wait Days	1-30days	1	-0.1260	0.1400	0.8108	0.3679
Wait Days	31-90 days	1	0.3763	0.1219	9.5294	0.0020
Season	Winter	1	-0.2861	0.1461	3.8349	0.0502
Season	Spring	1	-0.7774	0.1081	51.7405	<.0001
Season	Summer	1	0.4983	0.2204	5.1122	0.0238
Appt day of week	Mon	1	0.0764	0.1218	0.3940	0.5302
Appt day of week	Tue	1	0.00955	0.1170	0.0066	0.9353
Appt day of week	Wed	1	-0.0775	0.1998	0.1503	0.6382
Appt day of week	Thu	1	0.3210	0.1178	7.4193	0.0065

Table 27
Analysis of Maximum Likelihood Estimates for Final Model

Parameter		DF	Estimate	Standard Error	Wald Chi-Sq	Pr>ChiSq
Same payer	No	1	-0.9736	0.1967	24.5064	<.0001
Marital Status	Widowed	1	0.3132	0.2689	1.3568	0.2441
Marital Status	Separated	1	1.0560	0.2629	16.1379	<.0001
Marital Status	Divorced	1	-0.3588	0.1610	4.9666	0.0258
Marital Status	Single	1	-0.5662	0.1362	17.2763	<.0001
Prev visit nos	Yes	1	1.8929	0.1649	131.726	<.0001
Provider	Counselor	1	-0.1339	0.1402	0.9126	0.3394
Provider	Resident	1	0.2528	0.0974	6.7378	0.0094

Table 28
Odds Ratio Estimates for Final Model

Effect		Point Estimate	95% Wald CI (Lower)	95% Wald CI (Upper)
Age C44		0.979	0.97	0.989
Wait Days	None vs 91+days	0.317	0.151	0.663
Wait Days	1-30days vs 91+days	0.562	0.393	0.805
Wait Days	31-90 days vs 91+days	0.929	0.691	1.250
Season	Winter vs Fall	0.427	0.278	0.656
Season	Spring vs Fall	0.261	0.187	0.364
Season	Summer vs Fall	0.935	0.501	1.747
Appt day of week	Mon vs Fri	1.050	1.027	2.192
Appt day of week	Tue vs Fri	1.404	0.970	2.030
Appt day of week	Wed vs Fri	1.287	0.741	2.235
Appt day of week	Thu vs Fri	1.916	1.326	2.770
Samepayer	No vs Yes	0.143	0.066	0.308
Marital Status	Widowed vs Married	2.133	1.106	4.111
Marital Status	Separated vs Married	4.483	2.344	8.571
Marital Status	Divorced vs Married	1.089	0.744	1.595
Marital Status	Single vs Married	0.885	0.662	1.194
Prev visit nos	Yes vs No	44.074	23.089	84.131

Provider	Counselor vs Attending	0.985	0.638	1.522
Provider	Resident vs Attending	1.450	1.109	1.897

Table 29
Association of Predicted Probabilities and Observed Responses in Final Model

Percent Concordant	76.5
Percent Discordant	23.1
Percent Tied	0.4
c (ROC)	0.767

Table 30
Classification Table for Final Model

Probability Level	Correct Percentage	Sensitivity Percentage	Specificity Percentage	False Positive Percentage	False Negative Percentage
0	20.4	100	0	79.6	.
0.1	43.2	89.1	31.4	75.0	8.1
0.2	73.4	60.9	76.7	60.0	11.6
0.3	81.9	43.7	91.7	72.7	13.6
0.4	83.9	36.4	96.0	30.0	14.5
0.5	83.9	28.3	98.1	20.9	15.8
0.6	84.0	24.2	99.3	10.1	16.3
0.7	83.7	22.9	99.3	10.6	16.6
0.8	83.3	20.6	99.4	10.8	17.0
0.9	82.2	13.1	99.9	3.3	18.2
1	79.6	0	100	.	20.4

The Hosmer and Lemeshow Goodness of Fit test (chisq) returned a value of 0.0314, which indicates that there is a lack of fit in the model.

The final model, as derived from the development data set, can be shortly represented

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{Age}) + \underline{\beta}_2(\text{Wait Days}) + \underline{\beta}_3(\text{Season}) + \underline{\beta}_4(\text{Appt. Day of Week}) + \beta_5(\text{Same Payer}) + \underline{\beta}_6(\text{Marital Status}) + \beta_7(\text{Prevvisits NOS}) + \underline{\beta}_8(\text{Provider})$$

where underscores represent vectors of regression coefficients that correspond to categorical variables with more than two categories.

Before checking the model with the validation data set (25% withheld from the original data set), bi-variant analyses were performed on the determinants that contributed to the model as a main effect. For brevity's sake, a visualization of each of these analyses is shown below, along with a short description. Frequency tables, from which these visuals were generated, may be found in Appendix 2.

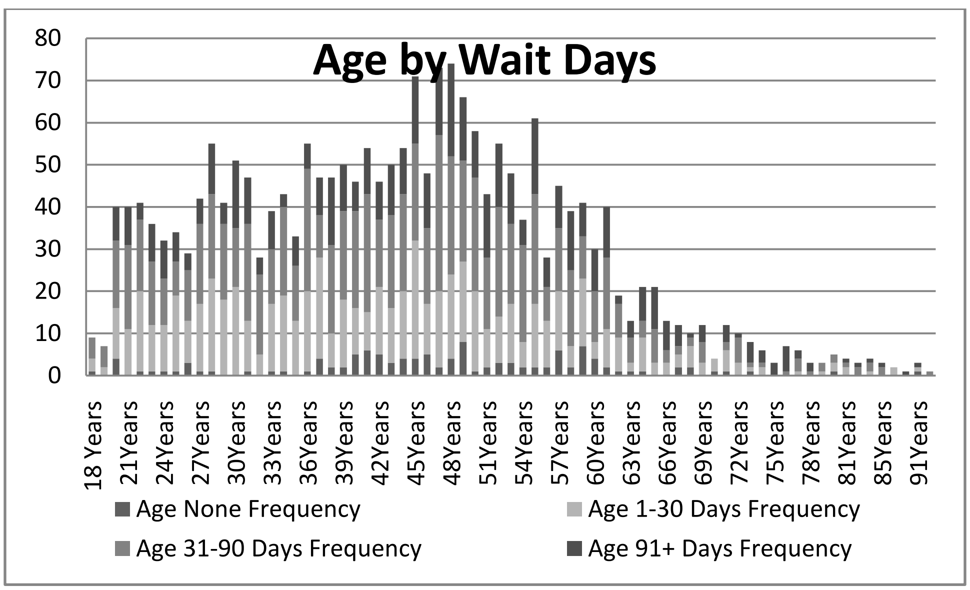

Figure 37. Frequency Distribution of Age by Wait Days

The number of days waiting for an appointment does not appear to be influenced by the age of the patient unduly.

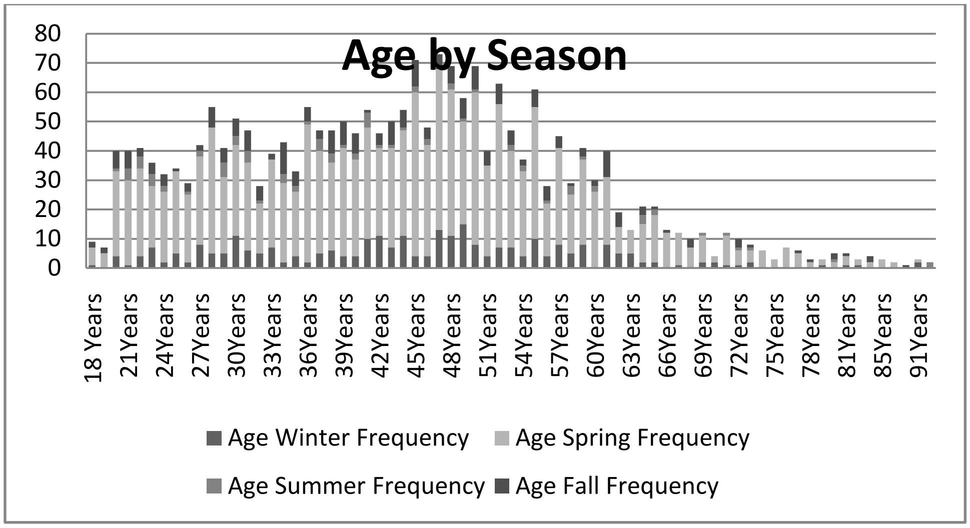

Figure 38. Frequency Distribution of Age by Season

The distribution of patients by age doesn't appear to be affected by the season of the year.

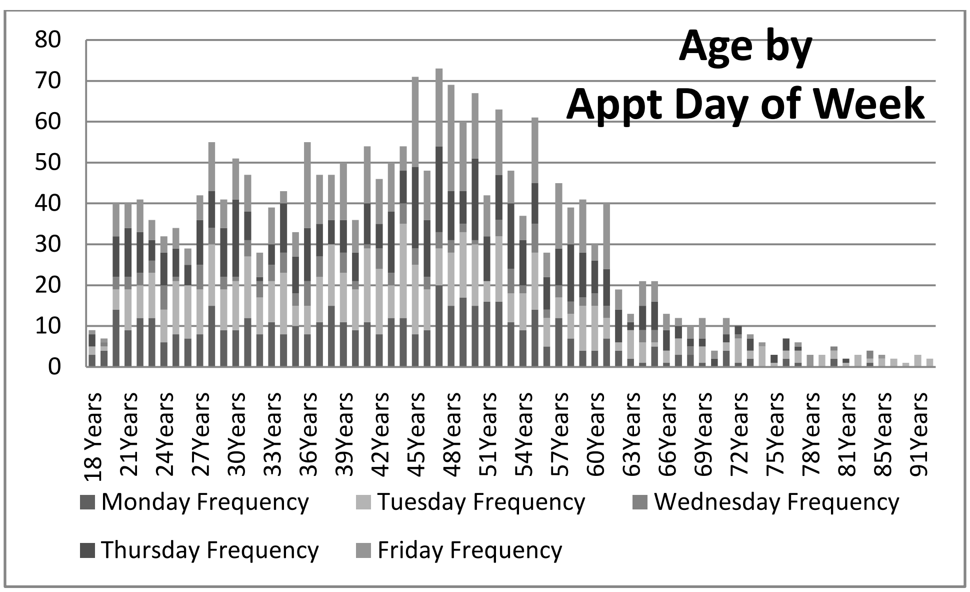

Figure 39. Frequency Distribution of Age by Day of Week

Patient age appears to be evenly distribution over the week.

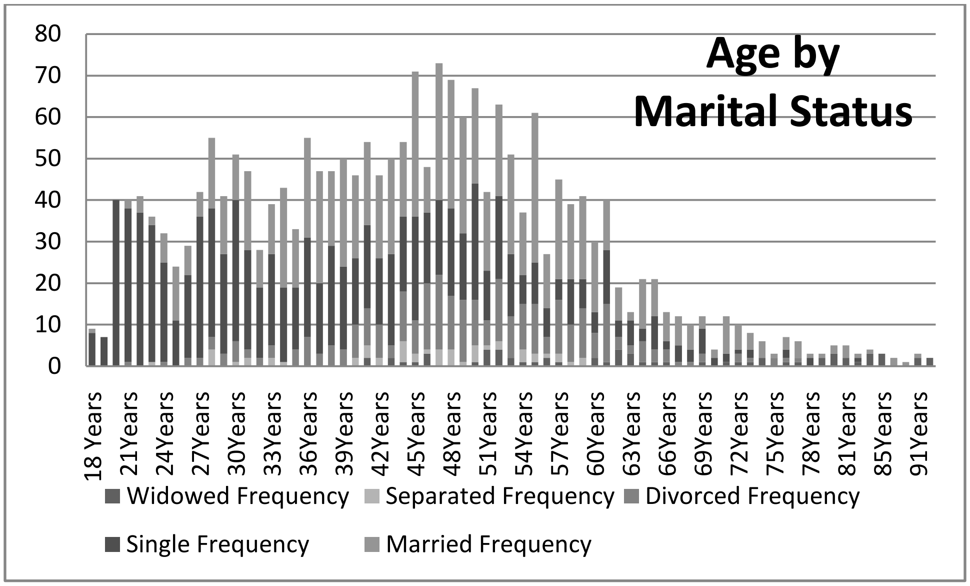

Figure 40.Frequency Distribution of Age by Marital Status

This distribution is consistent with what is known of the general population. Younger patients tend to be single, while older patients tend to be widowed more frequency. It's interesting that some individuals in their 40s are widowed.

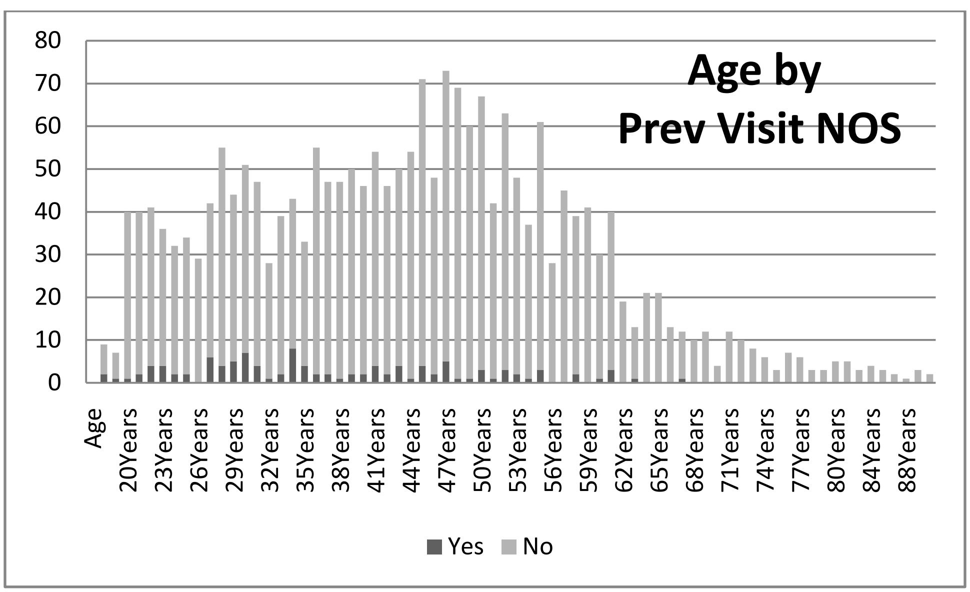

Figure 41. Frequency Distribution of Age by Prev Visit NOS

Younger patients tend to be more visit non-adherent.

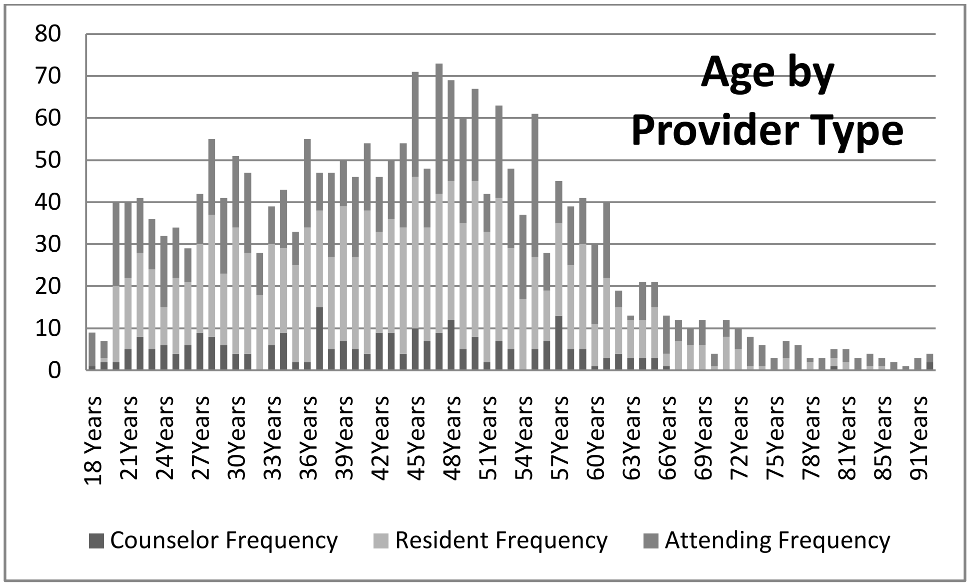

Figure 42. Frequency Distribution of Age by Prev Visit NOS

There is a relatively even distribution of patients (by age) across provider types, except that the use of counselors tends to drop off towards the upper years.

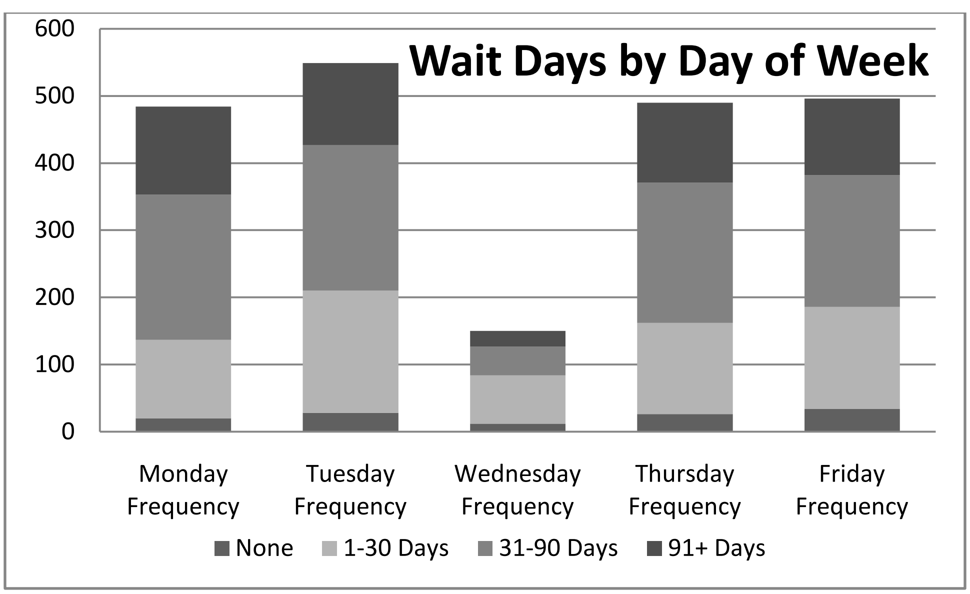

Figure 43. Frequency Distribution of Wait Days by Appt Day of Week

There appears to be a small tendency for patients to wait fewer days to be seen if their appointment occurs on Tuesday or Friday. There are more walk-in visits on Fridays.

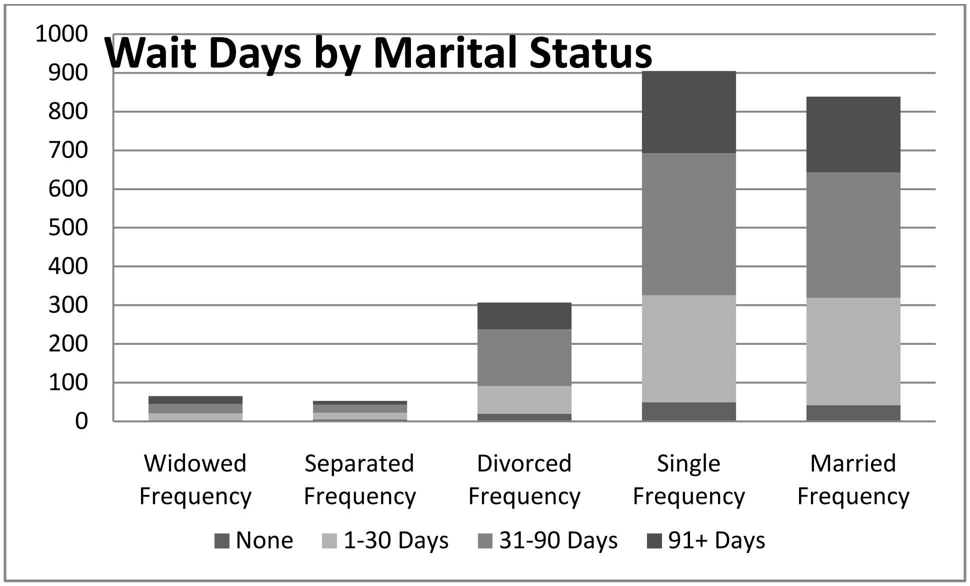

Figure 44. Frequency Distribution of Wait Days by Marital Status

The wait for an appointment appears to be about the same regardless of marital status.

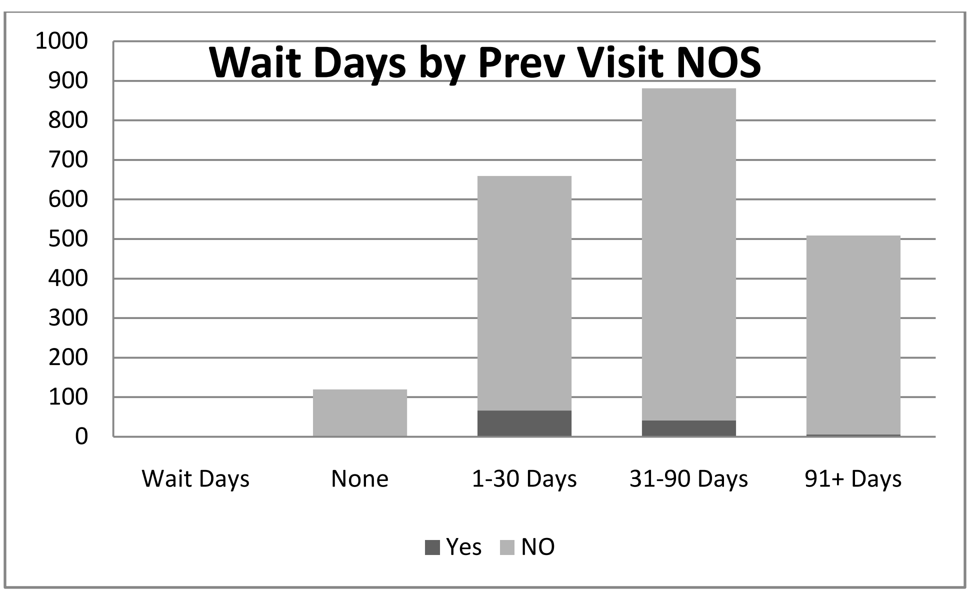

Figure 45. Frequency Distribution of Wait Days by Prev Visits NOS

Comparably, there or more non-adherent visits than would be expected when the wait for an appointment ranges from one to thirty days.

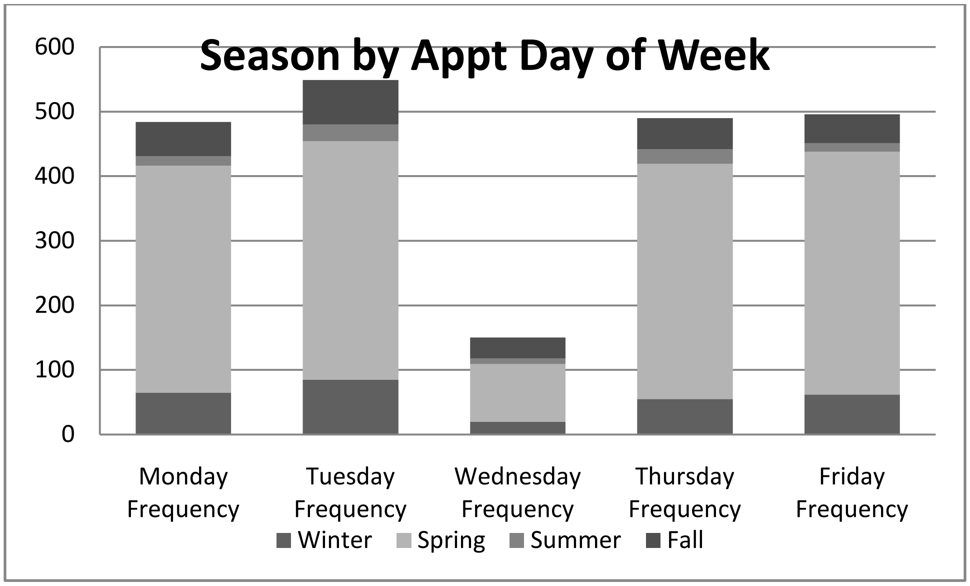

Figure 46. Frequency Distribution of Season by Appt Day of Week

Appointment day of week does not appear to vary by season.

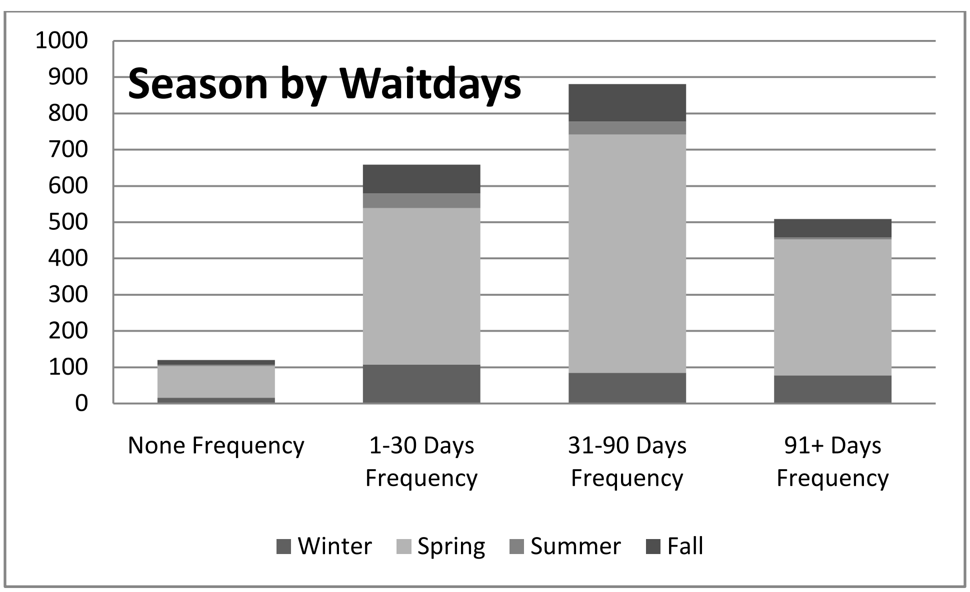

Figure 47. Frequency Distribution of Season by Wait Days

The number of days of waiting to be seen does not appear to vary much by season of the year.

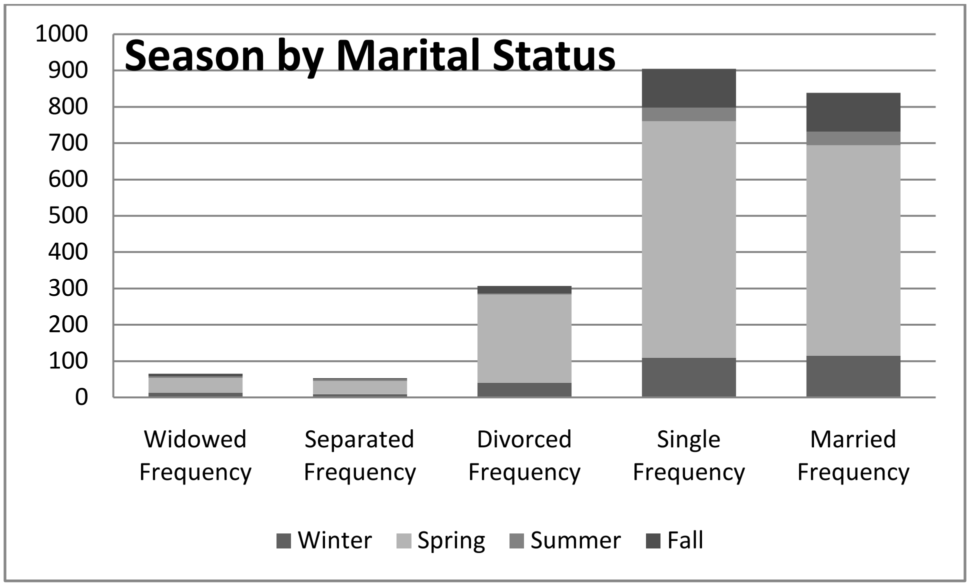

Figure 48. Frequency Distribution of Season by Marital Status

The distribution of individual's marital status does not appear to vary much by season.

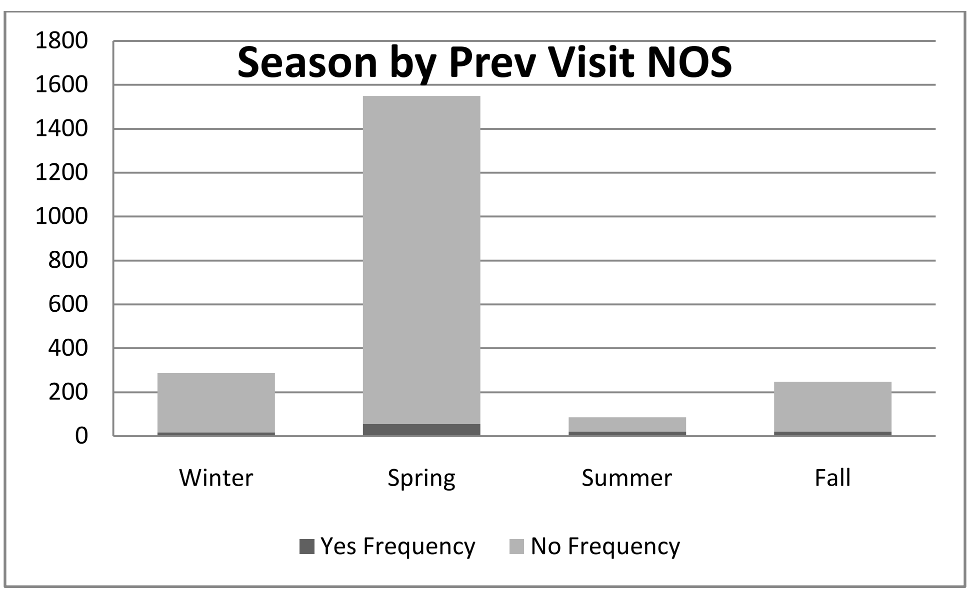

Figure 49. Frequency Distribution of Season by Prev Visit NOS

The rate of visit adherence is pretty constant, regardless of the season.

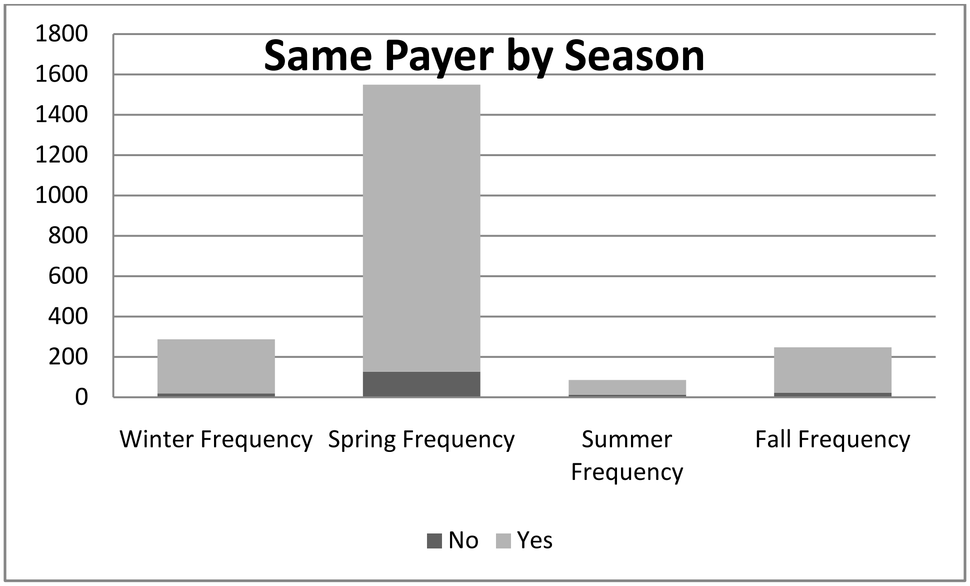

Figure 50. Frequency Distribution of Same Payer by Season

Payer doesn't vary by season.

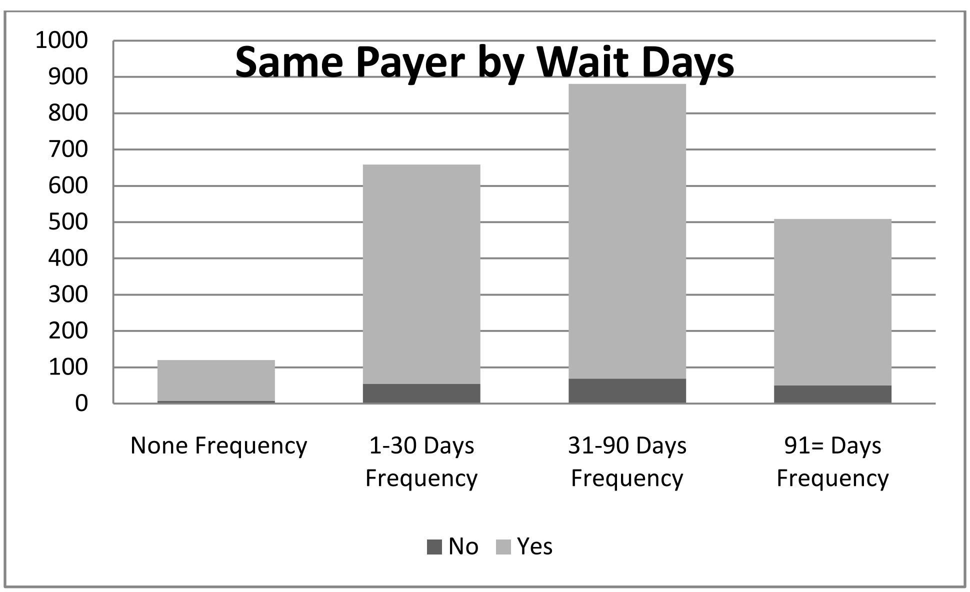

Figure 51. Frequency Distribution of Same Payer by Wait Days

A change in payer may be slightly associated with a longer wait time.

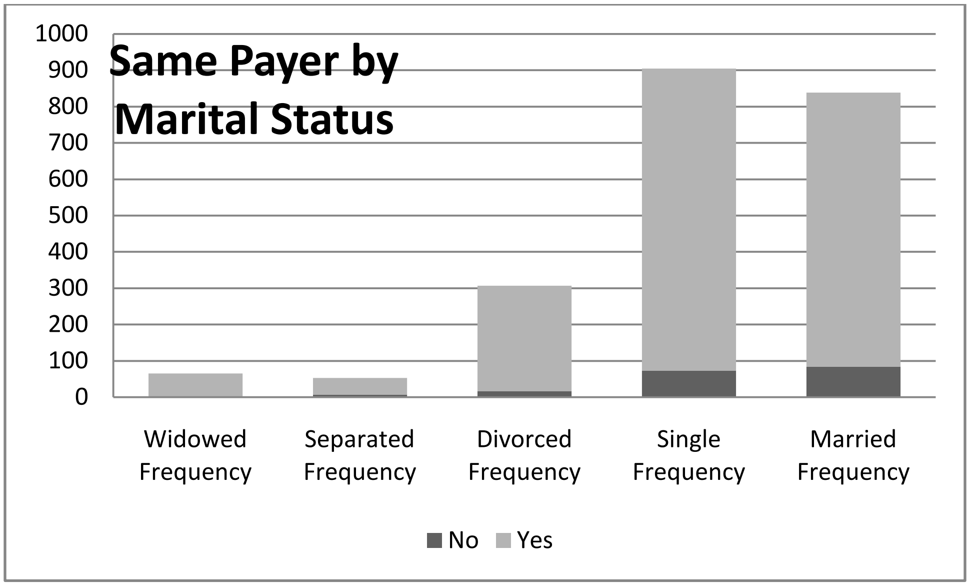

Figure 52. Frequency Distribution of Same Payer by Marital Status

Those who are widowed tend to have the same payer. This is consistent with the fact that older individuals are both more likely to have been widowed and more likely to have Medicare as a payer. It is interesting that married individuals have a greater tendency to have a different payer than those that are single. This may be a reflection of a psychiatric population whose illness makes it less likely that they will marry.

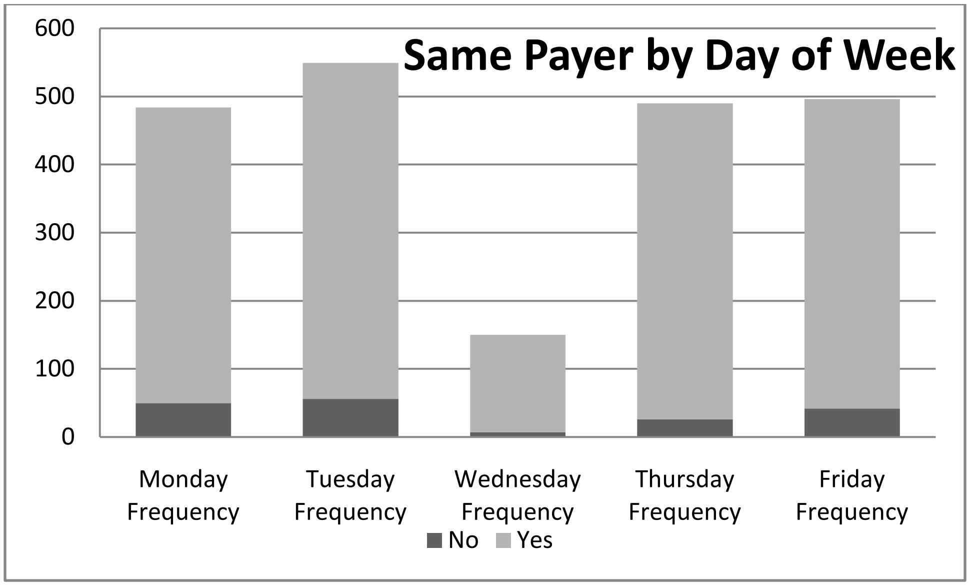

Figure 53. Frequency Distribution of Same Payer by Day of the Week

The day of the week on which the patient is scheduled does not appear to be influenced by the type of payer involved.

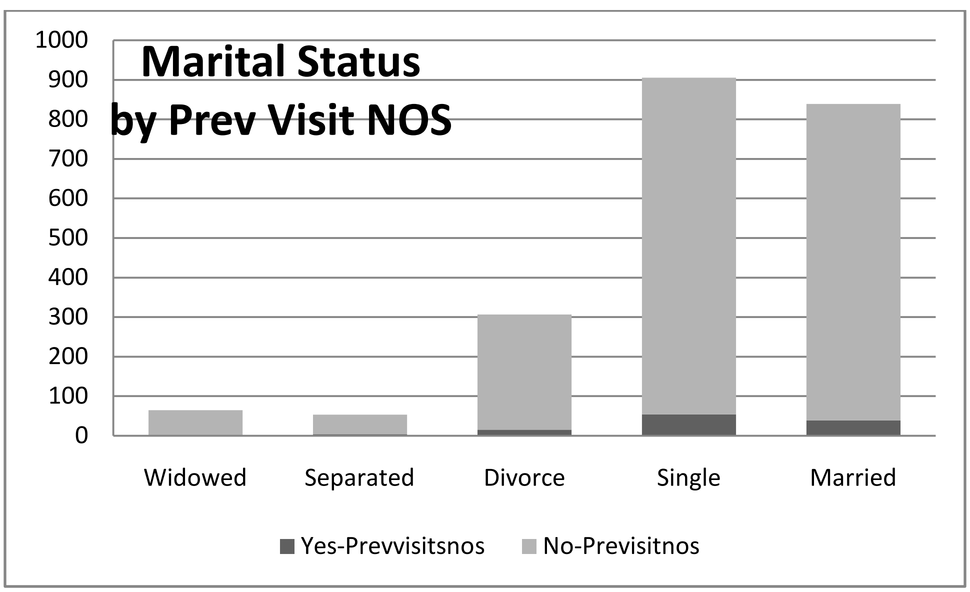

Figure 54. Frequency Distribution of Marital Status by Previous Visit NOS

Patients that are married are less likely to have had their previous visit non-adherent than those that are single. The samples for widowed and separated may be so small that little can be determined from them.

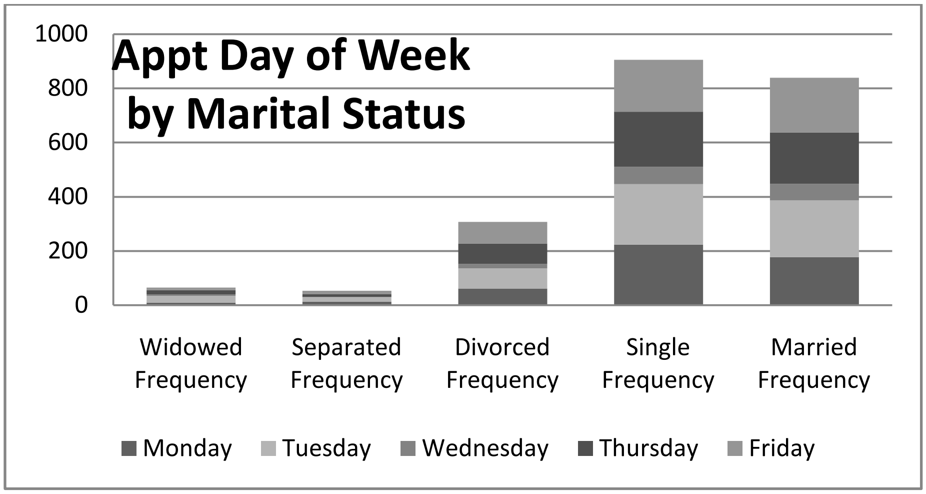

Figure 55. Frequency Distribution of Day of Week by Marital Status

Patients of all marital statuses are evenly scheduled over the days of the week.

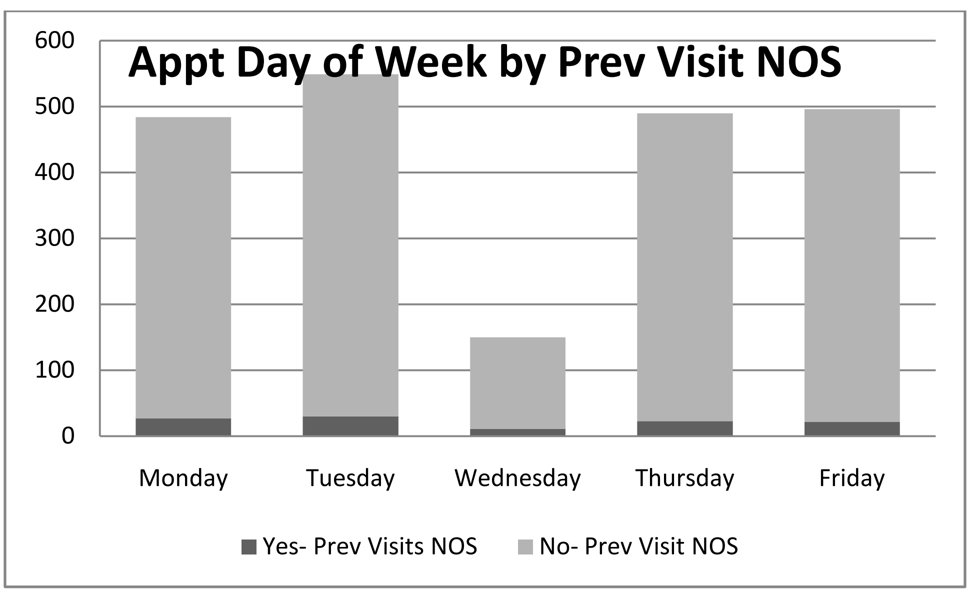

Figure 56. Frequency Distribution of Appt. Day of Week by Prev Visit NOS

The day of the week that an appointment is scheduled does not appear to be influenced by the occurrence of a previous non-adherent appointment.

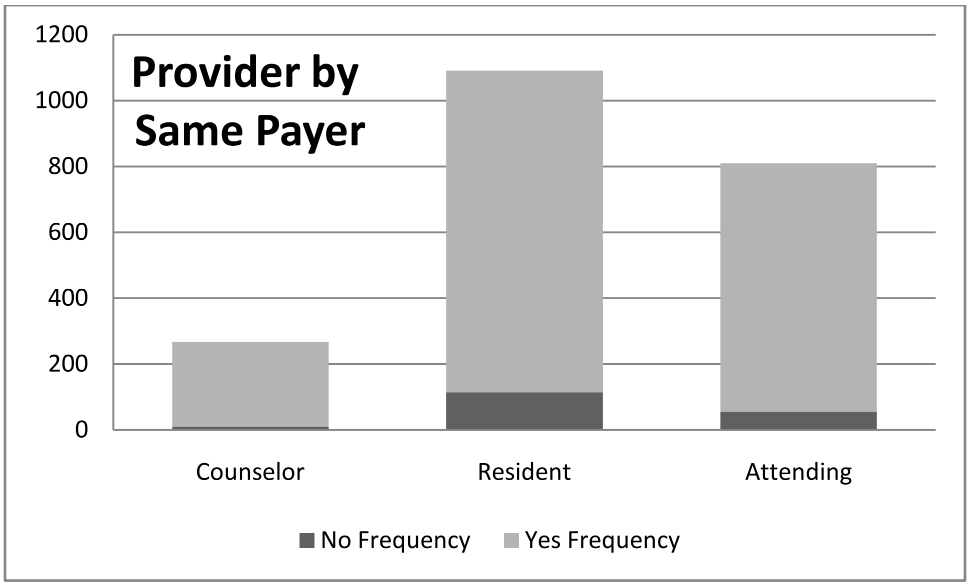

Figure 57. Frequency Distribution of Provider by Same Payer

A change in payer does not appear to be associated with a particular provider type.

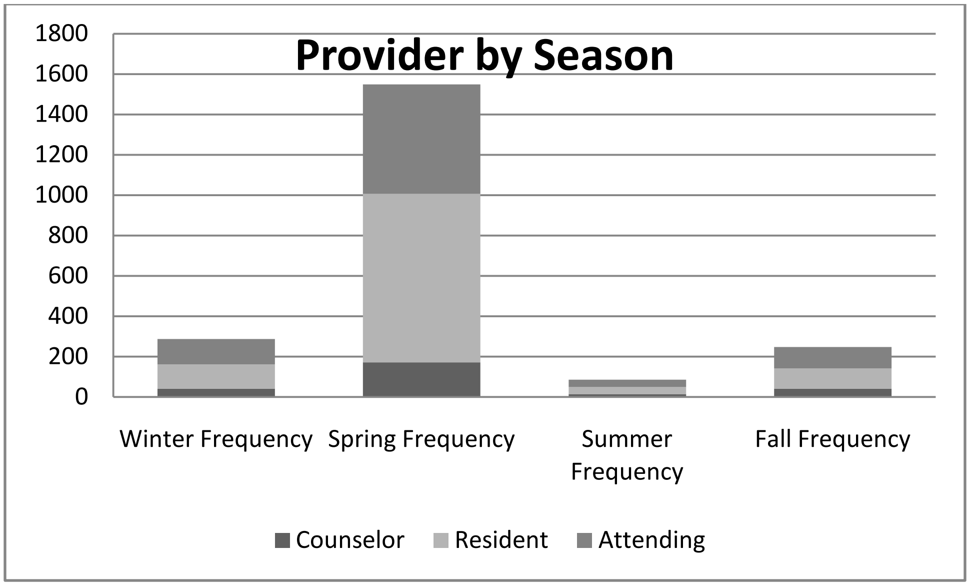

Figure 58. Frequency Distribution of Provider by Season

Providers appear to be scheduled equally across the seasons.

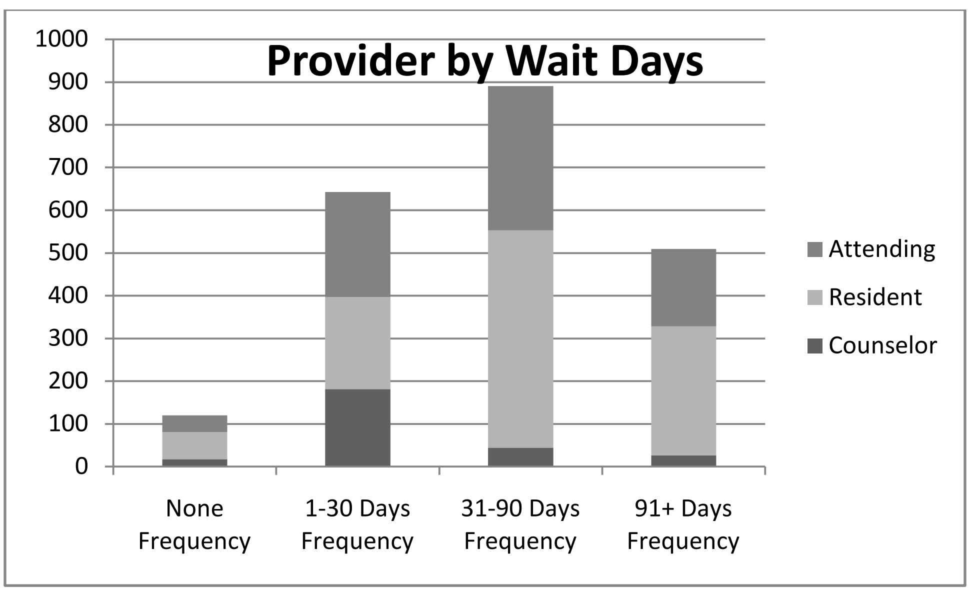

Figure 59. Frequency Distribution of Provider by Wait Days

The wait days associated with seeing a counselor tend to fall more frequently in the 1-30 days range, while those scheduled with resident tend to fall in the 31-90 days range.

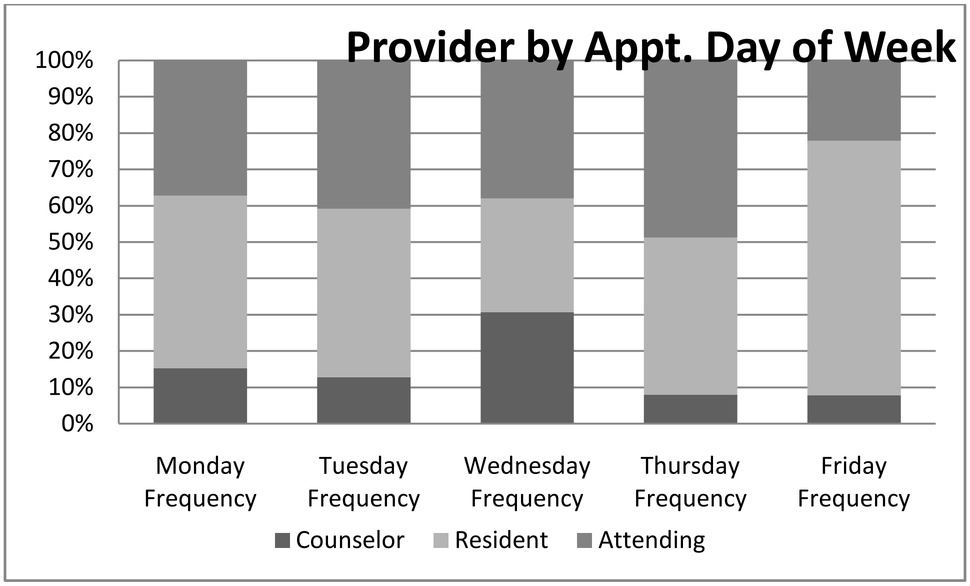

Figure 60. Frequency Distribution of Provider by Day of Week

Residents see more patients on Fridays and counselors and attending physicians see fewer.

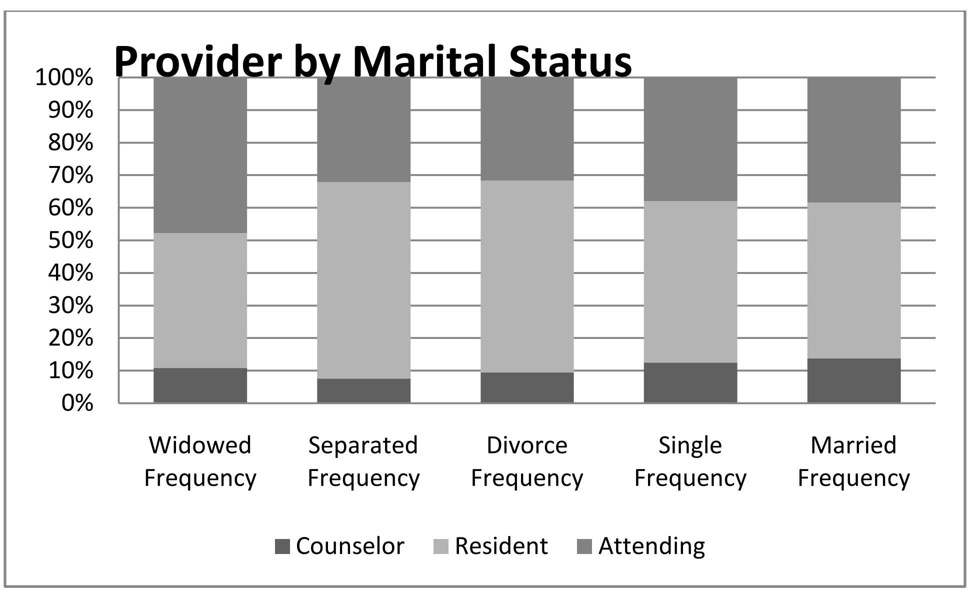

Figure 61. Frequency Distribution of Provider by Marital Status

Marital status does seem to affect the type of provider seen.

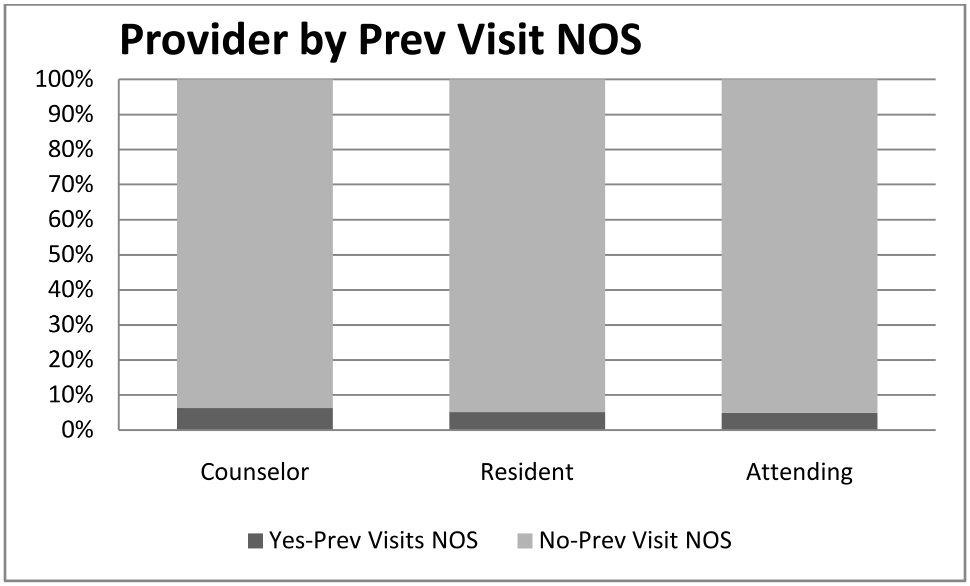

Figure 62. Frequency Distribution of Provider by Prev Visit NOS

The type of provider seen does not appear to be impacted by a previous non-adherent visit.

For model validation purposes, twenty-five percent of the total visit sample was retained during the model development phase of this study. In the validation process, the parameter estimates were treated as known (from the analysis done of the development data set) and final and then regression was used to produce the predictors for the validation data. The following results were obtained.

Table 31
Analysis of Findings in Validation Model

Effect	DF	Wald	pr>ChiSq
Age C44	1	9.7351	0.0018
Waitdays	3	7.8388	0.0495
Season	3	27.7206	<.0001
Appt day of week	4	3.2564	0.5159

Effect	DF	Wald	pr>ChiSq
Same payer	1	5.3158	0.2110
Marital status	4	2.7233	0.6051
Prevvisitnos	1	.9727	0.3240
Provider	2	13.7317	0.0010

Table 32
Analysis of Maximum Likelihood Estimates for Validation Model

Parameter		DF	Estimate	Standard Error	Wald Chi-Sq	Pr>ChiSq
Intercept		1	-2.1141	0.4756	19.7557	<.0001
Age C44		1	-0.0334	0.0107	9.7351	0.0018
Wait Days	None	1	-1.4510	0.7818	3.4442	0.0635
Wait Days	1-30days	1	0.2771	0.3190	0.7544	0.3851
Wait Days	31-90 days	1	0.8124	0.3045	7.1202	0.0076
Season	Winter	1	-0.3666	0.2603	0.9831	0.1591
Season	Spring	1	-0.9791	0.2040	23.0448	<.0001
Season	Summer	1	0.3139	0.3507	0.8012	0.3707
Appt day of week	Mon	1	0.2656	0.2494	1.0580	0.3037
Appt day of week	Tue	1	0.2280	0.2410	0.9015	0.3424
Appt day of week	Wed	1	-0.6697	0.5142	1.6963	0.1928
Appt day of week	Thu	1	0.2890	0.2419	1.4269	0.2323
Samepayer	No	1	-0.5592	0.2425	5.3158	0.0211
Marital Status	Widowed	1	-0.1412	0.7095	0.0396	0.8422
Marital Status	Separated	1	-0.4338	0.9403	0.2129	0.6445
Marital Status	Divorced	1	0.2505	0.3880	0.4169	0.5185
Marital Status	Single	1	-0.0454	0.3503	0.0168	0.8969
Prev visit nos	Yes	1	0.1440	0.1460	0.9727	0.3240
Provider	Counselor	1	0.1085	0.2283	0.2260	0.6345
Provider	Resident	1	0.4758	0.1767	7.2487	0.0071

Table 33
Odds Ratio Estimates for Validation Model

Effect		Point Estimate	95% Wald CI (Lower)	95% Wald CI (Upper)
Age C44		0.967	0.947	0.988
Wait Days	None vs 91+days	0.163	0.020	1.341

Table 33
Odds Ratio Estimates for Validation Model

Effect		Point Estimate	95% Wald CI (Lower)	95% Wald CI (Upper)
Wait Days	1-30daysvs 91+days	0.919	0.443	1.906
Wait Days	31-90 days vs 91+days	1.570	0.820	3.006
Season	Winter vs Fall	0.247	0.098	0.625
Season	Spring vs Fall	0.134	0.059	0.303
Season	Summer vs Fall	0.488	0.158	1.503
Appt day of week	Mon vs Fri	1.435	0.701	2.938
Appt day of week	Tue vs Fri	1.396	0.696	2.800
Appt day of week	Wed vs Fri	0.568	0.1.48	2.189
Appt day of week	Thu vs Fri	1.482	0.732	3.003
Samepayer	No vs Yes	0.327	0.126	0.846
Marital Status	Widowed vs Married	0.600	0.113	3.196
Marital Status	Separated vs Married	0.446	0.045	4.412
Marital Status	Divorced vs Married	0.888	0.425	1.852
Marital Status	Single vs Married	0.660	0.380	1.147
Prev visit nos	Yes vs No	1.334	0.753	2.363
Provider	Counselor vs Attending	1.999	0.948	4.215
Provider	Resident vs Attending	2.887	1.646	5.062

Table 34
Association of Predicted Probabilities and Observed Responses in Validation Model

Percent Concordant	77.4
Percent Discordant	22.3
Percent Tied	0.3
c (ROC)	0.775

Table 35
Classification Table for Validation Model

Probability Level	Correct Percentage	Sensitivity Percentage	Specificity Percentage	False Positive Percentage	False Negative Percentage
0	20.4	100	0.0	79.6	.
0.1	43.2	89.1	31.4	75.0	8.1
0.2	73.4	60.9	76.7	60.0	11.6
0.3	81.9	43.7	91.7	72.7	13.6
0.4	83.9	36.4	96	30.0	14.5
0.5	83.9	28.3	98.1	20.9	15.8
0.6	84.0	24.2	99.3	10.1	16.3
0.7	83.7	22.9	99.3	10.6	16.6
0.8	83.3	20.6	99.4	10.8	17.0
0.9	82.2	13.1	99.9	3.3	18.2
1	79.6	0	100	.	20.4

There was some variation in the frequencies of some determinants that may have influenced the results. Please see Table 36 below.

Table 36
Frequency Comparison Between Development Data and Validation Data

Determinant		Development Frequency	Validation Frequency
Wait days	None	5.53	7.00
	1-30 Days	30.38	33.39
	31-90 Days	40.62	37.70
	91+ Days	23.47	21.90
Season	Winter	13.23	14.72
	Spring	71.42	73.61
	Summer	3.96	5.03
	Fall	11.39	6.64
Appt Day of Week	Mon	22.31	22.80
	Tues	25.31	25.31
	Wed	6.92	7.00
	Thu	22.59	23.88
	Fri	22.87	21.01
Same Payer	No	8.34	9.87

Determinant		Development Frequency	Validation Frequency
	Yes	91.66	90.13
Marital Status	Widowed	3.00	4.13
	Separated	2.44	1.08
	Divorced	14.15	15.44
	Single	41.72	43.27
	Married	38.68	36.09
Prev Visit NOS	Yes	5.21	94.79
	No	19.57	80.43
Provider	Counselor	12.36	15.08
	Resident	50.30	46.14
	Attending	37.34	37.78

Data Mining Results

As a second method of analysis, this study used classification and classification trees as a means to further explore the data. The use of data mining can be effective as a means to explore the relationships determinants have with each other, and may, in the process, uncover previously unconsidered relationships. Decision trees are perhaps most widely used as practical forms of machine learning and data mining. Of all data mining methods, they have been most widely researched and have been applied to a large variety of data mining problems, although not often applied to health care data.

The classification software used in this study was the random tree function contained in the WEKA software packages. The random forests strategy essentially builds a number of classification trees, which increases reliability but decreases the capacity for visualization of the resulting model. These trees are built by a process that is known as partitioning. The partitioning done in this study was based on one of the independent variables, which is also known as the splitting attribute. Branches were created for different values of this splitting attribute. The choice of splitting attribute is done by picking the attribute that will partition the original sample into sub-samples that are as homogenous as possible [270]. This process creates the root nodes and leaf nodes that make up the tree.

The ultimate goal of building a tree model is to end up with the smallest tree possible that retains the purest leaf nodes. The purer a leaf node is, the more precise its classification.

This data mining exercise was carried out in three phases. The first phase involved the large, original set of determinants and the random forests method, the second phase used the reduced set of determinants found in the statistical analysis along with random forests, and the third phase used a single tree classification that results in a visual classification tree. Please see Figure 63 below for detail on the process flow used for data mining in this study.

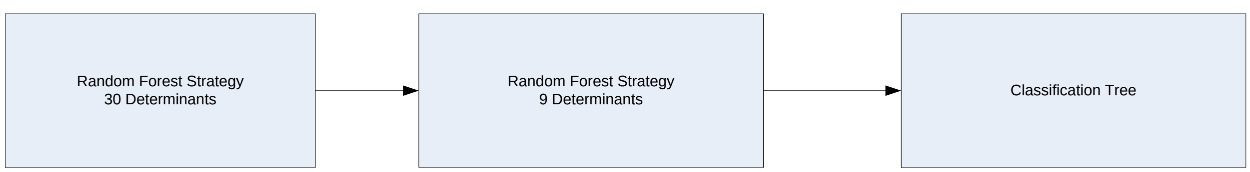

Figure 63. Data Mining Process Flow

In the first phase, random forest was used with 50/50 split between non-overlapping sub-samples. The full sample, without any splitting, is considered the root node and each of the sub-samples is considered a node. The data sample was identical to the original used for the more traditional statistical testing and contained thirty attributes. These attributes (determinants) included:

- status
- gender
- age
- traveldistance
- visittype
- samevisittype
- waitdays
- servicedate
- apptdayofweek
- samedayofweek
- appttime
- sametimeofday
- referringprovider
- payer
- samepayer
- gendiagnosis
- secdiagnosis
- employment
- numbernos
- percentnos
- numbercnx
- race
- nonmdappointments
- contactpersonrelationship
- numberappointments
- prevvistnos
- maker
- provider
- sameprovider

The test mode was set up where 50 percent of the data were used to train (or develop) the initial set of trees and the remaining fifty percent were used to conduct testing. A random forest of ten trees was created.

The example tree (prepared as a single tree) appears as follows:

```
percentnos < 3.5 : no (1389/4) [716/0]
percentnos >= 3.5
|   gendiagnosis = Anxiety
|   |   apptdayofweek = WEDS : yes (2/0) [2/1]
|   |   apptdayofweek = MON
```

```
|   |   |   maritalstatus = Widowed : yes (0/0) [0/0]
|   |   |   maritalstatus = Divorced : no (1/0) [0/0]
|   |   |   maritalstatus = Single : no (4/0) [4/0]
|   |   |   maritalstatus = Married : yes (6/0) [3/2]
|   |   |   maritalstatus = Separated : yes (0/0) [0/0]
|   |   |   maritalstatus = Missing : yes (0/0) [0/0]
|   |   |   maritalstatus = Other : yes (0/0) [0/0]
|   |   apptdayofweek = THU : no (10/0) [2/0]
|   |   apptdayofweek = FRI : no (14/1) [7/2]
|   |   apptdayofweek = TUES
|   |   |   waitdays < 34 : yes (7/1) [4/0]
|   |   |   waitdays >= 34 : no (7/0) [10/2]
|   gendiagnosis = Depression : no (188/33) [74/16]
|   gendiagnosis = Bi-Polar : no (66/16) [25/6]
|   gendiagnosis = Psychosis
|   |   servicedate = Dec : no (5/0) [2/0]
|   |   servicedate = Feb : no (3/1) [2/1]
|   |   servicedate = May : no (8/1) [3/0]
|   |   servicedate = Jun
|   |   |   referringprovider = PCP Internal : no (6/2) [0/0]
|   |   |   referringprovider = PCP External : no (3/0) [4/0]
|   |   |   referringprovider = Resident Internal : no (0/0) [0/0]
|   |   |   referringprovider = Specialist external : no (0/0) [0/0]
|   |   |   referringprovider = Self : no (6/0) [0/0]
|   |   |   referringprovider = Missing : no (0/0) [0/0]
|   |   |   referringprovider = Specialist internal : yes (1/0) [2/1]
|   |   |   referringprovider = Fellow,Internall : no (0/0) [0/0]
|   |   |   referringprovider = External Health Center : no (1/0) [0/0]
|   |   |   referringprovider = PCPUndertermined : yes (1/0) [1/0]
|   |   |   referringprovider = Counselor InternalExternal Health Center : no (0/0) [0/0]
|   |   |   referringprovider = Counselor InternalMissing : no (0/0) [0/0]
|   |   servicedate = Nov : yes (1/0) [0/0]
|   |   servicedate = Apr : no (6/3) [4/0]
|   |   servicedate = Mar : no (6/1) [1/0]
|   |   servicedate = Aug
|   |   |   waitdays < 14 : yes (2/0) [1/0]
|   |   |   waitdays >= 14 : no (2/0) [2/0]
|   |   servicedate = Sep : yes (5/0) [2/0]
|   |   servicedate = Jul : no (1/0) [3/3]
|   |   servicedate = Oct : no (5/1) [5/0]
|   |   servicedate = Jan : no (0/0) [0/0]
|   gendiagnosis = Behavior/Personaility : no (35/6) [14/2]
|   gendiagnosis = Dementia : no (1/0) [1/0]
|   gendiagnosis = Drug : no (2/0) [2/1]
|   gendiagnosis = Other : no (6/0) [3/1]
|   gendiagnosis = Missing : yes (30/3) [17/2]
```

The model summary is as follows:

Table 37
Model Summary for Complete Determinant Set

Correctly Classified Instances	1341	97.6693 %
Incorrectly Classified Instances	32	2.3307 %
Kappa statistic	0.9341	
Mean absolute error	0.0677	
Root mean squared error	0.1496	
Relative absolute error	19.1535%	
Root relative squared error	35.6286%	
Total Number of Instances	1373	

The detailed analysis of accuracy included the following.

Table 38
Analysis of Accuracy with Complete Determinant Set

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.984	0.048	0.986	0.984	0.985	0.993	ARR
	0.952	0.016	0.946	0.952	0.949	0.993	NOS
Weighted Avg.	0.977	0.041	0.977	0.977	0.977	0.993	

Confusion was minimal with 1,347 visits correctly classified and 32 misclassified.

The second phase was run with the same data set, and the same 50/50 split, but with the variables reduced to the following:

- status
- age
- waitdays
- season
- apptdayofweek
- samepayer
- maritalstatus
- prevvisitnos
- provider

The test mode was retained at a split, with 50%for training and the remainder for testing. A random forest of 10 trees was constructed

The model summary of the training data is a follows:

Table 39

Model Summary for Reduced Determinant Set

Correctly Classified Instances	868	80.0738 %
Incorrectly Classified Instances	216	19.9262 %
Kappa statistic	0.3448	
Mean absolute error	0.2295	
Root mean squared error	0.3835	
Relative absolute error	70.6364 %	
Root relative squared error	94.5564 %	
Total Number of Instances	1084	

The detailed analysis of accuracy included the following:

Table 40

Analysis of Accuracy with Reduced Determinant Set

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.418	0.099	0.525	0.418	0.465	0.735	ARR
	0.901	0.582	0.855	0.901	0.878	0.735	NOS
Weighted Avg.	0.801	0.482	0.787	0.801	0.792	0.735	

Confusion was larger in this set, with 875 visits correctly classified and 179 misclassified.

The test data, however, were quite a different story. Utilizing what was "learned" from the training data set, the testing data set showed greatly improved results.

The model summary of the testing data is a follows.

Table 41
Model Summary for Testing Determinant Set

Correctly Classified Instances	2124	97.9253%
Incorrectly Classified Instances	45	2.0747%
Kappa statistic	0.9358	
Mean absolute error	0.0567	
Root mean squared error	0.1406	
Relative absolute error	17.472%	
Root relative squared error	34.9073%	
Total Number of Instances	2169	

The detailed analysis of accuracy included the following.

Table 42
Analysis of Accuracy with Testing Determinant Set

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.943	0.012	0.954	0.943	0.949	0.997	ARR
	0.988	0.057	0.986	0.988	0.987	0.997	NOS
Weighted Avg.	0.979	0.047	0.979	0.979	0.979	0.997	

Confusion was diminished in this set, with 2,124 visits correctly classified and 45 misclassified.

Because of the difficulty using the random forest utility in WEKA to generate visual representation of trees, a second classification strategy was used. The tree formation protocol used to continue this study was the J48 Classifier Tree with 0.25 Confidence Interval (which is the default setting). This particular classification tree was chosen because it results in a visually interpretable tree. The determinants used were:

- age
- waitdays

- season
- apptdayofweek
- samepayer
- maritalstatus
- prevvisitnos

The test mode used was 10-fold cross-validation

The resulting model appeared as follows:

```
prevvisitnos = Yes
|   waitdays = one month return
|   |   age <= 51: ARR (60.0/6.0)
|   |   age > 51: NOS (6.0)
|   waitdays = ninety day return: ARR (41.0)
|   waitdays = more than ninety day return: ARR (6.0)
|   waitdays = none: ARR (0.0)
prevvisitnos = No: NOS (2056.0/341.0)
```

The model summary of the testing data is a follows.

Table 43
Model Summary for J48 Classifier Determinant Set

Correctly Classified Instances	1817	83.7713 %
Incorrectly Classified Instances	352	16.2287 %
Kappa statistic	0.3076	
Mean absolute error	0.2695	
Root mean squared error	0.3683	
Relative absolute error	83.0052 %	
Root relative squared error	91.4311 %	
Total Number of Instances	2169	

The detailed analysis of accuracy included the following:

Table 44
Analysis of Accuracy with J48 Classifier Determinant Set

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class

	0.229	0.006	0.902	0.229	0.365	0.595	ARR
	0.994	0.771	0.834	0.994	0.907	0.595	NOS
Weighted Avg.	.0.838	0.616	0.848	0.838	0.796	0.595	

Please see Figure 64 below for a representation of the classification tree.

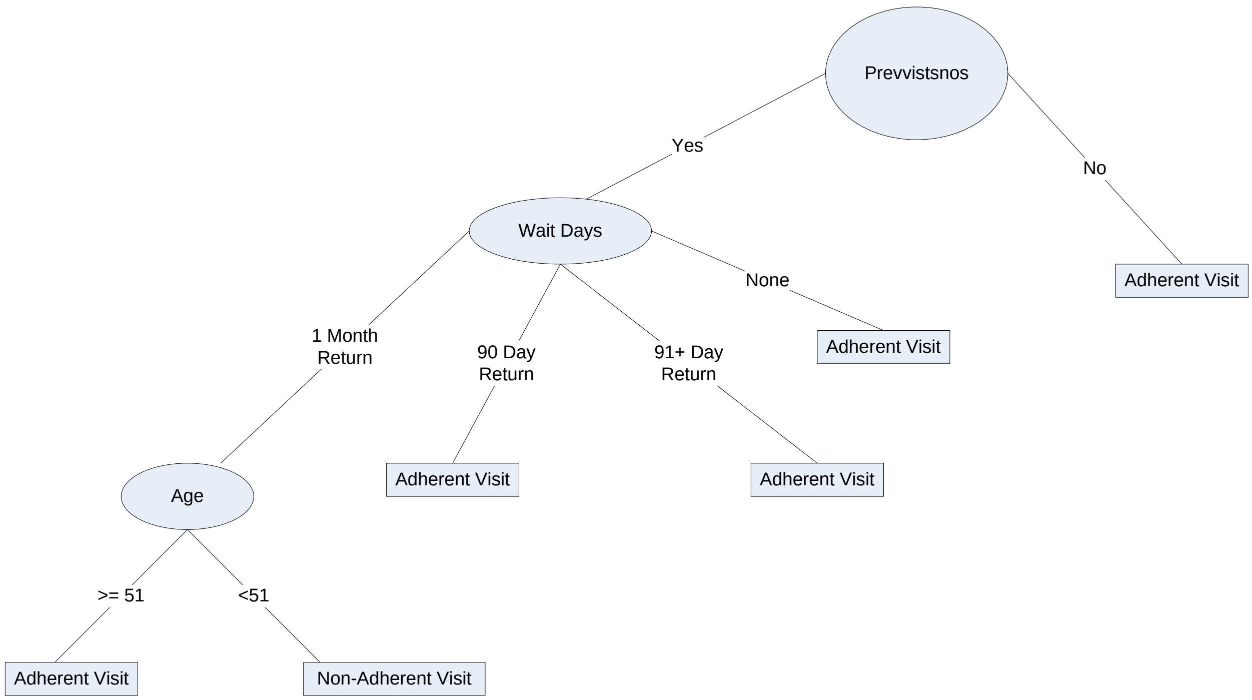

Figure 64. Classification Tree

Predictive ability appears to be better in the entire classification tree methods used. This may be a function of the training and testing sets or of the number of samples available and the degree of colinearity found in the determinants, especially between relationship of contact person and marital status [248]. Additional analysis with a larger data set should be carried out to confirm results found in this study.

Comparison of Useful Models

The third component of the analysis in this study was the comparison of the determinants in the model formed from the statistical analysis with those found in other models used by health care and non-health care industries. Table 45 below shows how the models found in the literature compare, on the basis of determinants used, with the visit non-adherence model developed in this study.

Table 45
Comparison of Study Determinants to Determinants in Potentially Useful Predictive Models

Study Determinant	Fair Isaac (Credit Score)	Vantage (Credit Score)	Box Office Success	Airline	Hotel Yield Mgmt.	GAIL Model
Age	If collected	n/a	n/a	n/a	n/a	Patient age
Marital Status	Others on accounts	n/a	n/a	n/a	n/a	n/a
Same Payer	Lender types	n/a	Box office revenue	Ticketed/ non ticketed	n/a	n/a
Wait Days for Appt.	Years at address	n/a	Binned days	Advance booking	Advance purchase days	n/a
Season	n/a	n/a	Year of release	Flight date	Special event days	n/a
Appt Day	n/a	n/a	n/a	Flight Number	n/a	n/a
Type of Provider	n/a	n/a	Star actor/ director	n/a	n/a	n/a

Limitations

This study is bound by a number of limitations. The most significant of the limitations may be that the sample size is smaller than might be desired and that the sample came from a single health care system and mostly from a single clinic setting. Another limitation lies in the reality of care provision in a rural area. The travel distances used in this study may be very specific to the locality and may not generalize to another setting for that reason. Given the strong support for travel distance as a determinant of visits non-adherence found in the literature, the results of this study may have underutilized travel distance as a predictor. Considering the originating location of the patient rather than travel distance may be a more useful way of incorporating the travel factor into the model at some point. Work done in preliminary Study 3 shows that at least one of the travel distance ranges may have been strongly affected by visits/patients originating in a specific location in that range. This location, composed of one city and several smaller co-located towns, forms a referral cell from which a disproportionate number of patients seek psychiatric services with a much larger spread of diagnoses than would be expected. Please see Figures 65-71 below for additional information.

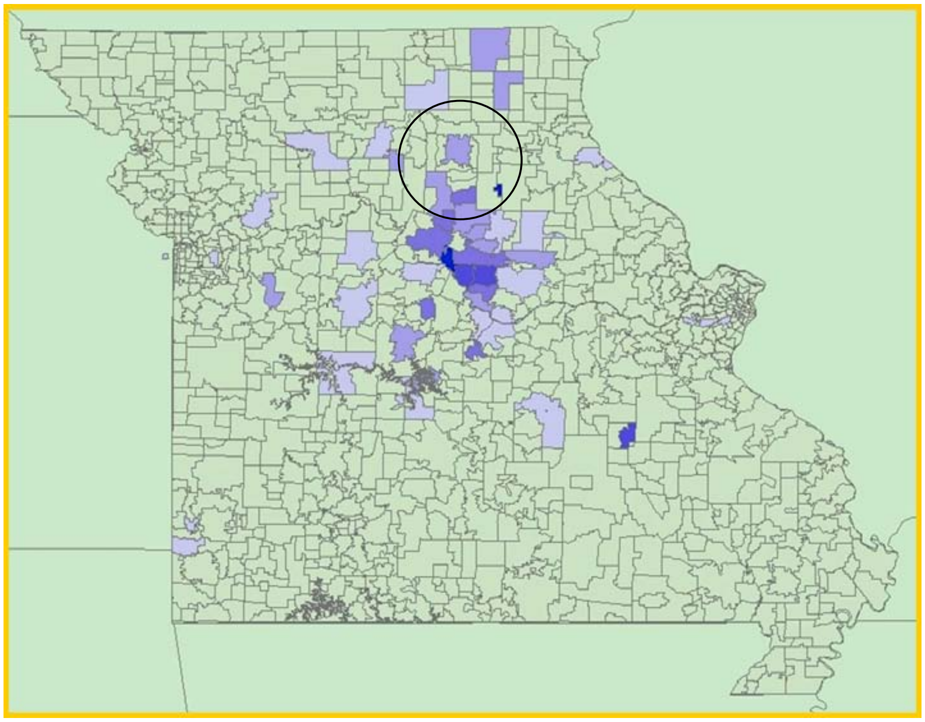

Figure 65. Distribution of Referrals for Attention Deficit Disease. Darker color indicates higher density of Attention Deficit Disorder than was expected for population density

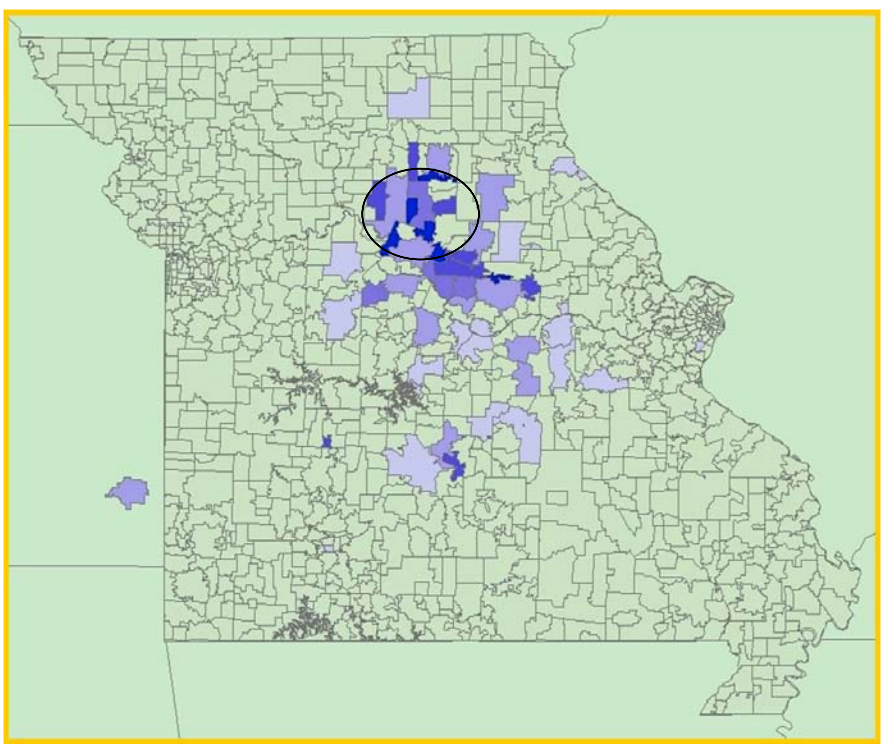

Figure 66. Distribution of Referrals for Anxiety. Darker color indicates higher density of Anxiety/Panic Disorder/PTSD than was expected for population density

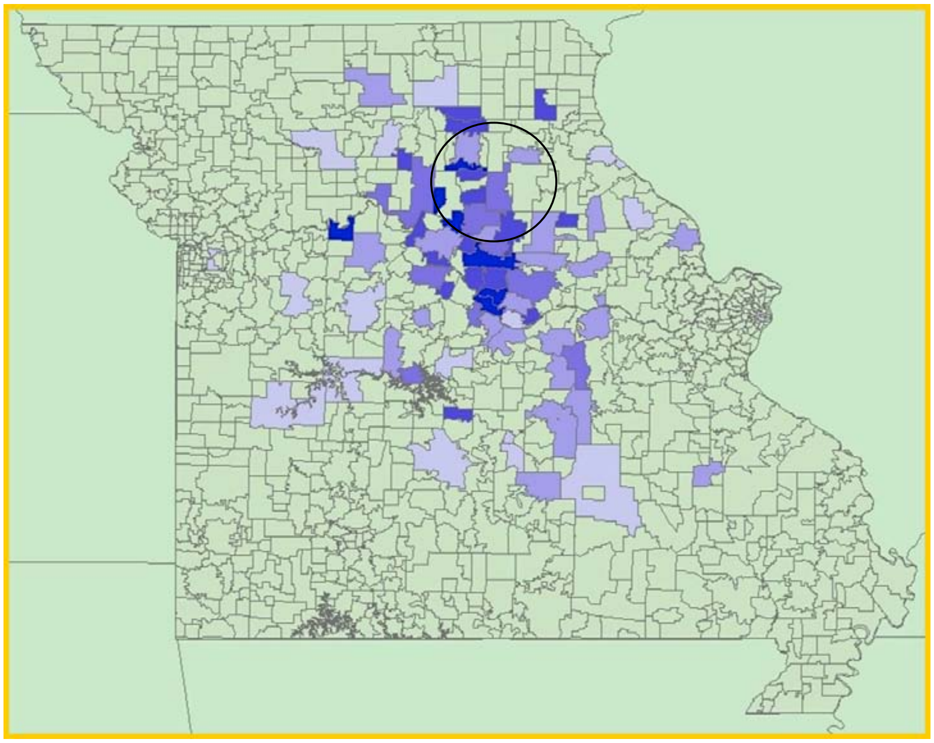

Figure 67. Distribution of Referrals for Bi-Polar Disorder.   Darker color indicates higher density of Bi-Polar Disorder than was expected for population density

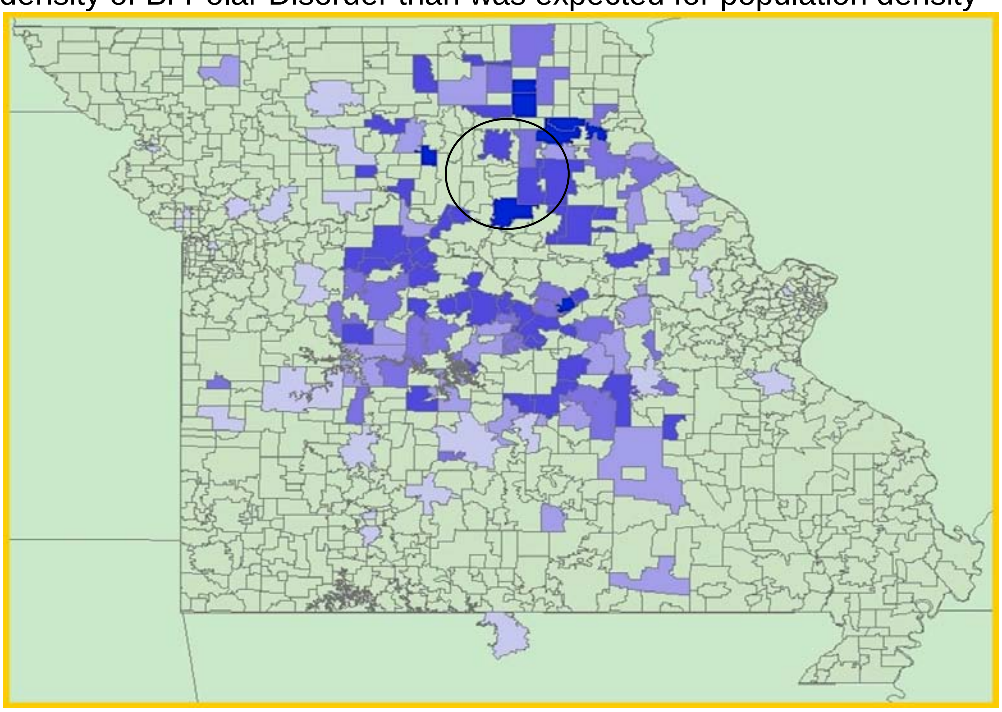

Figure 68. Distribution of Referrals for Depression. Darker color indicates higher density of Depression than was expected for population density

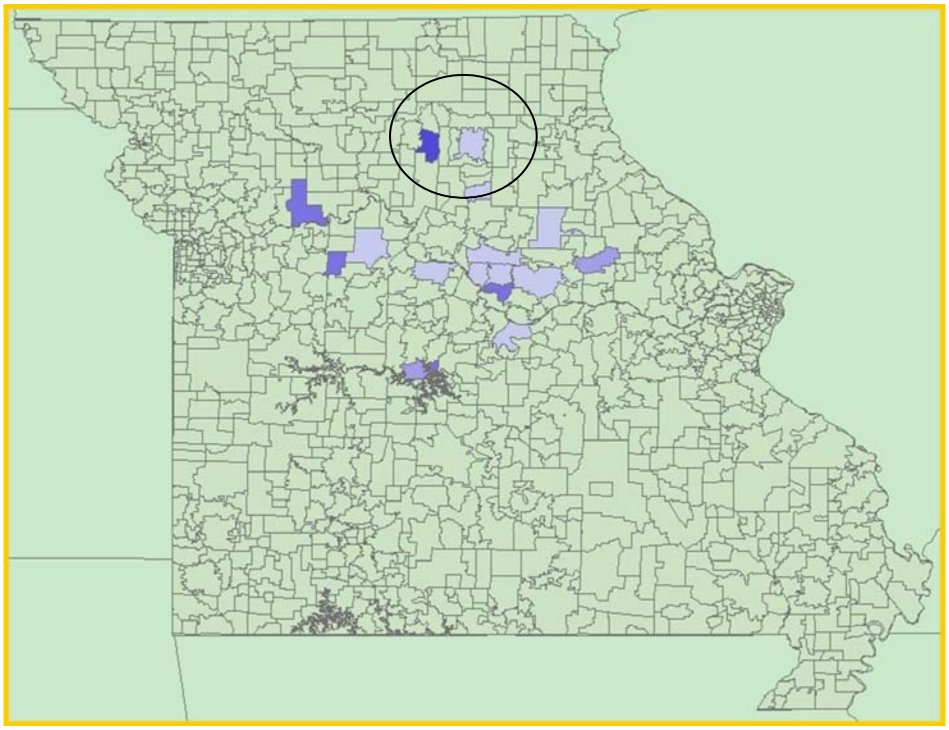

Figure 69. Distribution of Referrals for Dementia. Darker color indicates higher density of Dementia than was expected for population density.

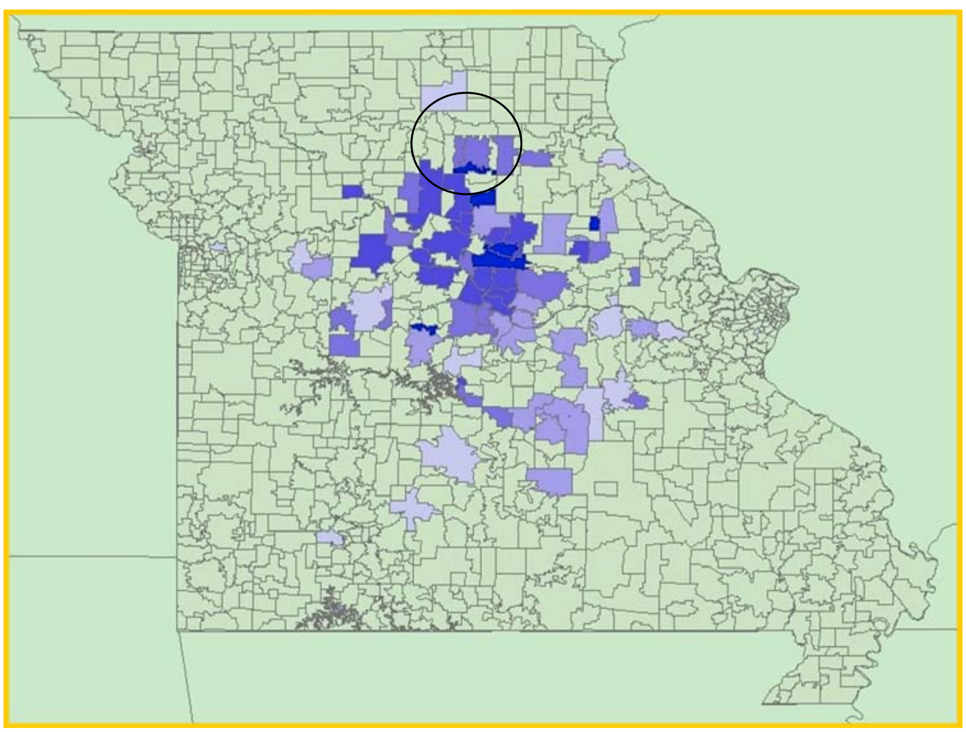

Figure 70. Distribution of Referrals for Psychosis. Darker color indicates higher density

of Psychosis than was expected for population density.

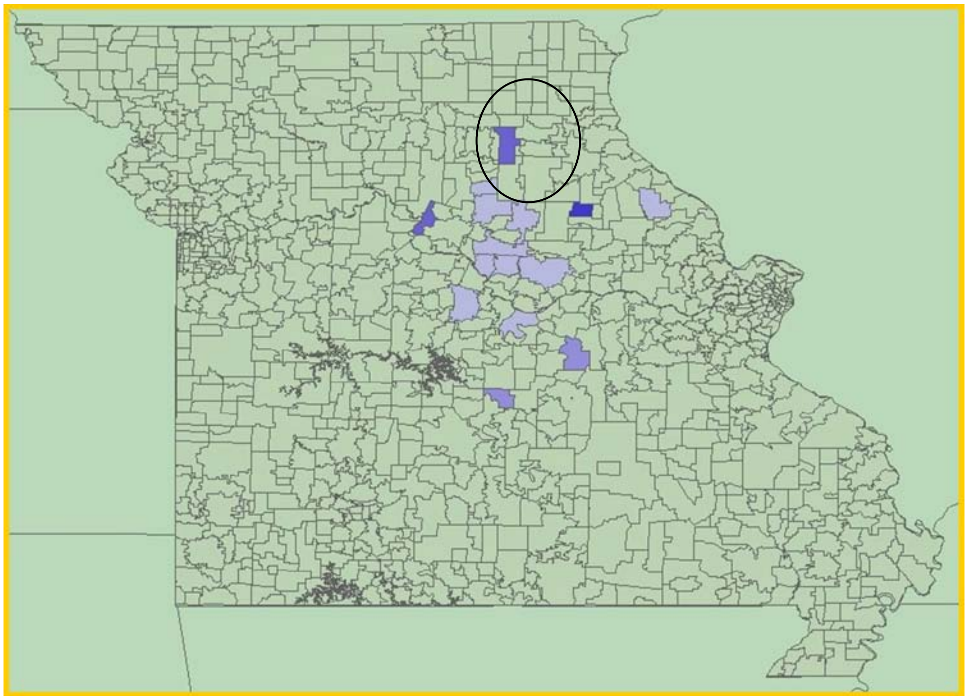

Figure 71. Distribution of Referrals for Drug/Alcohol Abuse. Darker color indicates higher density of Drug/Alcohol Abuse than was expected for population density.

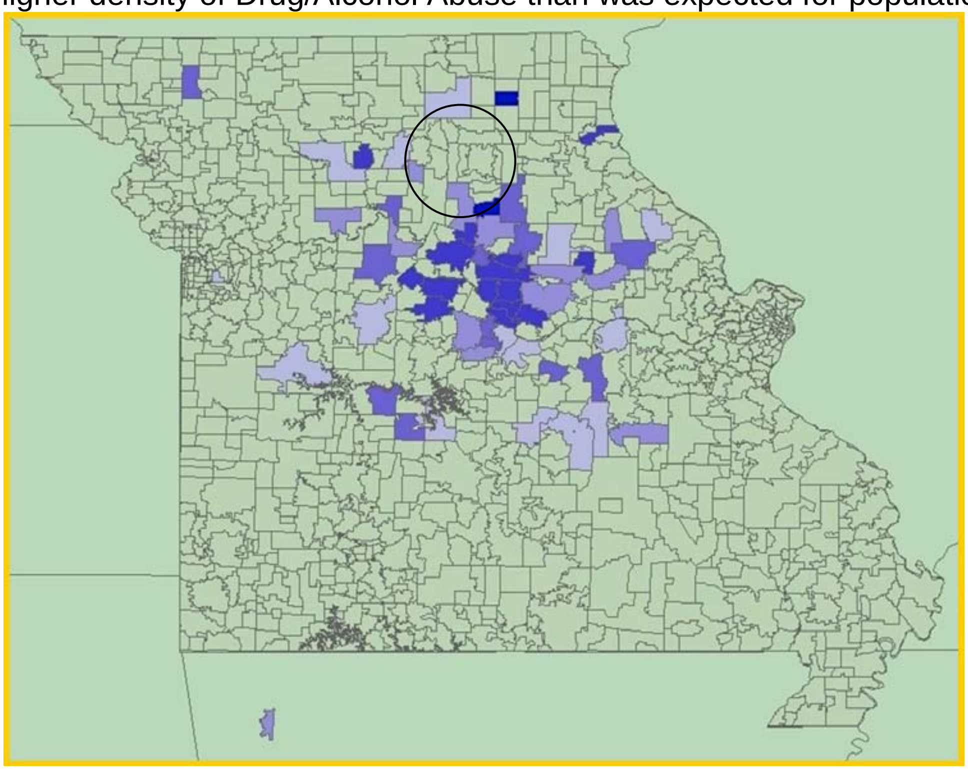

Figure 72. Distribution of Referrals for Other Diagnoses. Darker color indicates higher

density of Other Mental Health Diagnoses than was expected for population densit**y.**

This study is also limited because it used records for return visits that also contained a complete data set. Because incomplete data tend to be found in records for new or visit non-adherent patients, bias in the results is a distinct possibility.

Furthermore, while regression is a well-proven analysis tool, there is a possibility, however slight, that since this study employed a single data set, that one or more of the determinants in the final model may, in fact, not be a truly independent predictor of visit non-adherence.

Although general and secondary diagnosis were not included in the final regression model, it is possible that their use might be re-instated in future work. Because the grouping of diagnoses into super classes was accomplished through the use of ICD-9 CM codes, it will be necessary to carefully crosswalk ICD-10 codes to ICD 9 codes before proceeding with this avenue of research.

Discussion

The results of this study demonstrate that it may be possible to determine the likelihood of a non-adherent visit, prior to scheduling, using a limited set of determinants that can be collected from electronic sources. While much further work is warranted, the prospect that a limited set of determinants, usually readily available to scheduling staff, can, when properly used, predict visit adherence constitutes an improvement over the results of prior studies and opens the door for improved management of visit non-adherence.

The overall objectives of this study included the introduction of an evidence-based model of visit non-adherence, improved prediction of ambulatory psychiatric visit non-adherence, and support for the development of scheduling tools that could be used to decrease ambulatory visit non-adherence.

In addressing the three specific aims of the study (to further refine a set of determinants of visit non-adherence, to fit a *useful* model for the prediction of visit non-adherence for ambulatory psychiatry, including a replicable procedure, and compare the resulting model to other health and non-health related prediction models), three investigative tools were used. The use of more traditional statistical methods, including logistic regression, paired with a machine learning method and the comparison of previously established models successfully demonstrated the applicability of a number an informatics tools to a health care problem and leveraged the characteristics of each tool to improve the model and inform future work. There were a number of reasons to utilize regression as a means to create a model. Because the primary research interest lay in the relationship of the determinants to the outcome of interest, regression was an appropriate tool. Secondly, regression is a tool that is well proven as a technique for addressing problems of prediction, as was the case in this study. Although a primary hypothesis guided investigation, the use of regression analysis also allowed the investigation to be an iterative process that may well have uncovered an alternative hypothesis. In fact, the final model differs from the original in that it does not incorporate general or secondary diagnosis.

Removing diagnosis may have potential to allow the model to be generalized to other medical practices more easily. While the type of diagnosis a patient has may well help determine the likelihood of visit non-adherence, given the symptoms associated with certain diagnoses, the identification of a diagnosis via ICD codes is problematic. Diagnosis coding is variable and is often dependent on contractually-based reimbursement or on individual physician idiosyncrasies. It may be difficult to generalize diagnosis or acuity of symptoms across providers and types of practices.

The use of classification trees serves as a means from which additional information may be learned about the relationship of the determinants to each other. While regression is focused on the independent-dependent variable relationship, classification trees can be used to explore the relationships among all variables without establishing independence. This capability is likely one of the reasons that the model developed from the full determinant set differs significantly from that developed though logistic regression. A useful future step in this study may be further investigation of the roles of determinants found differentially important to prediction.

The order in which regression and data mining were used in this study could have been reversed. The decision to proceed with regression first was made on the basis of its relative acceptability to the probable end users of the study. Regression is well recognized as a respectable analysis tool in the health care delivery field, whereas data mining and machine learning are viewed as somewhat experimental and, perhaps, less credible. The use of data mining might well have led to the retention of “relationship of contact person” as a determinant. “Contact person” dropped out of the regression analysis fairly early, but is retained as a strong measure in the classification tree done using WEKA. The differences in the results of the two analyses and the importance of determinants found in each supports the general philosophy of the research in the use of multiple tools to provide multiple perspectives to the question of visit non-adherence. Re-inclusion of this determent may very well be a consideration in future studies.

The comparison of the final model of visit non-adherence to that used in other healthcare predictions and to those used in non-health care fields (such as airline management) offers the possibility of re-use of previous tools. Although this study found that none of the comparison models was adequate for visit non-adherence prediction, it demonstrated another area where the science of informatics might be successfully applied to health care problems. Prediction models, such as models used in the increasing sophisticated credit scoring industry, have the potential to both inform investigation and reduce the need for the development of new tools.

CHAPTER 5-CONCLUSIONS AND RECOMMENDATIONS

This study reflects an effort to address the need for a predictive model that can be used to identify visit non-adherence. Visit non-adherence is a costly issue that has been seen as an intractable problem in the delivery of health care services. Visit non-adherence also is a barrier to success in preventative health care provision. Research done prior to this study tends to focus on either the development of redundant scheduling practices or on the isolation of patient characteristics that constitute a “bad patient”. This study focused on leveraging past work done in the area of visit non-adherence prediction by using a combination of investigative tools to create a solution that could be effectively applied to ambulatory clinic operations.

This study offers new information that may be useful to both providers of health care and researchers. It makes a contribution to science and to the field of informatics in particular, as well as offers the potential of commercialization in the form of a software application. The results of this study, conditioned on further study in a variety of practise environments, hold potential to substantively reduces costs to the mental health care system as well as reduce human suffering. The model as developed could constitute a major improvement to the care process. Properly applied, without its use as a means to profile patients, it can introduce changes that allow clinic operations staff to manage and control important aspects of the visit. Its use may provide significant benefit to patients (who should have increased access to care), to the health care system (through improved management of resources), and to the general populations (in terms of reduced costs for health care).

Understanding the role various determinants play in visit non-adherence and how they combine to lead to visit non-adherence may also prove useful in the evaluation of care quality. Since much of the present strategy of clinical care quality is outcomes based, and retrospective, the effect of visit non-adherence may be underestimated or simply unaccounted for in present measurement. Use of an evidence-based model of visit non-adherence may lead to improved process measurements of visit non-adherence, which, in turn, can be used to modify and monitor operational processes in ambulatory clinic practice.

The study also provides opportunities for a number of improvements. To the scientific community, including those that practice medicine, this study addresses a seemingly intractable problem with a process and a model that could be replicated in a number of environments. It directly addresses a health care delivery issue that creates both higher health care costs and decreased health care quality. If the vicious cycle of visit non-adherence can be curtailed, in such a manner that utilization of health care resources can be confined, when appropriate, to outpatient appointments, some of the health care costs associated with increases in patient morbidity can be reduced or eliminated. Visit non-adherence reduction also holds out hope that individual patients may benefit, especially when that benefit includes reduced levels of suffering. The health care system, as a whole, also may benefit given that visit non-adherence is well supported in the relevant literature as an area of improvement that could lead to improved quality of care, increased access to services, and improved utilization of physicians.

To the health informatician, this study provides a relatively rare example of the application of informatics tools and knowledge to health care. Informatics tools, especially machine learning techniques, have not been consistently applied to health care operational or care delivery issues. Much of the machine learning used in health care has been confined to the realm of genomics. If the utility of machine learning to health care operations issues can successfully be demonstrated, it is hopeful that they will be increasingly recognized as valuable tools in the health care arena.

This research can create a significant paradigm shift if it replaces the current “bad patient” or “bad schedule” mentality with an evidence-based solution that helps ensure maximum return on health dollars with a significant reduction in human suffering. Creating this much needed paradigm shift is largely dependent on the publication of results in appropriate venues and to audiences that can place tools created from the model built in this study into clinical practice.

Results from the three preliminary studies have already been presented a number of times and have received national interest [287]. The researchers have also used the presentations to network with others interested in this work and have several promises of assistance with dissemination once final results are available. Additional publication work needs to be accomplished; especially publication that can bring potential solutions to health care administrators and clinicians that modifies their present concept of visit non-adherence and demonstrates that informatics tools can be effectively applied to long-standing health care issues. Additionally, the concept of preventable visit non-adherence needs to be brought to the attention of policy makers, so that new, evidence-based polices can better support clinical practice and patient well-being.

To achieve these objectives, publication of results will need to take place in at least two specific publication streams. First, the results should be published in academic journals as a means to encourage other researchers to continue to investigate the visit non-adherence phenomena and to encourage academicians to consider adding machine learning methods to the arsenal of investigate techniques they use in their research. Publication also needs to take place in health care trade journals. The concepts of “bad patients” and “bad schedules” are so accepted among health care administrator and clinic managers that repeated exposure to another model will be required before it will be accepted. Trade journals and conferences are the most effective venues to reach this audience.

The data used and results gleaned from this study also lend themselves to publication where the focus is on the methodology used. The use of iterative regression and statistical discovery methods, as they were used in this study, support the idea that more traditional statistical methods may be effectively used as “data mining” tools. Although statistics form the basis on which machine learning rests, the use of machine learning methods has, in recent years, overshadowed the use of traditional statistical methods as data mining tools.

Given that the ultimate objective of this research was to develop a model of visit non-adherence to support the development of tools that can be predictably used by health care providers to reduce the rate of visit non-adherence, the development of commercial software tools seems required. Such software could be created as “lay-over” applications designed to work in tandem with scheduling software, or could be created as standalone applications, or perhaps, even as calculators. Software applications have great potential for improving quality of care, through such applications as Computerized Physician Order Entry (CPOE), e-prescribing, and medication interaction alerts. To date, however, there has been no significant innovation within the scheduling system itself to improve quality of care, despite the fact that there are few differences between medical scheduling applications and those used for scheduling other appointments, such as those for haircuts, photography sessions, or automobile maintenance, where quality improvement, or at least service improvement, are often a primary concern.

The development of a software product from this project would be relatively straightforward. The determinants used in this model are already available within the scheduling system. Building a software application would essentially consist of reaping these determinants, in real time, from the scheduling system and loading them into a parsing and analysis module to automatically calculate the probability of a given visit being non-adherent. This probability could then be efficiently presented to clinical staff at the time of booking for assistance in choosing the appointment with minimal probability of visit non-adherence. The application could also track visit adherence histories for the life of the patient-provider relationship and utilize patient specific histories to further improve visit non-adherence prediction. Such an application could also allow reoccurring scheduling to take place based on past visit adherence. Provider and patient scheduling incompatibilities could be recognized early in the scheduling process and resolved by employing a preferential scheduling pattern for certain patients or by expanding the scheduling of their appointment to include multiple providers. The application could also provide patient and provider specific flagging.

Potentially, and relatively expediently, visit scheduling could be transformed from what is now a mostly "guesswork" operations to a well-defined, evidence based process. Rather than booking a patient into a time slot without regard for reducing the likelihood of a non-adherent visits, schedulers will be able to make informed decisions. As they schedule a visit (or increasingly as the patient self schedules a visit) decision support and warnings incorporated into to scheduling software could inform users of the risk of non-adherence for a particular visit and suggest ways to diminish the risk.

Given the present lack of comparable products in the market, such a tool could be quite profitable if barriers, such as administrator and provider attitudes, to use were effectively addressed.

A sidebar discussion of the results of this study continues to be focused on the effective use of data sources. In this particular study practice, like many others in the US, the billing and scheduling systems in place are more stable and provide more structured data than the EMR or other clinical systems. While this phenomenon is probably the result of systems designed to enable reimbursement for services rendered, an important aspect of health care delivery, these same systems collect all the data elements needed to predict visit non-adherence. Adjustments and enhancements to billing and scheduling systems could be a valuable tool in addressing visit adherence.

As with any good initial study, there are multiple opportunities for future work. To address limitations imposed by the sample, my collaborators and I plan to enlarge the sample population, both by increasing the number of visits studied within the present patient population and by working to create data sharing partnerships with several other health care delivery systems to expand the patient base.

Several potential determinants remain to be investigated when the availability of information from electronic sources increases. These include the length of time between psychiatry appointments and its impact on visit non-adherence, as well as a better way of measuring the relationship between the first and second diagnosis and a strategy to take into account the effects of a somatic health diagnosis on visit non-adherence.

Future work should also include the introduction of other computational tools to the problem. The application of such tools as neural networks to the issue of visit non-adherence, while not in the scope of this small study, should be reconsidered as such tools become increasingly applicable to health care data.

Appendix 1

Table 46
Status by Gender

Gender	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Male	143	150	585	578
Female	304	297	1142	1149

Table 47
Status by Age

Age	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
3	5	1	0*	4
18 Years	3*	2	6	7
19 Years	2*	1	5	6
20 Years	11	8	29	31
21 Years	9	8	31	32
22 Years	13	8	28	33
23 Years	11	7	25	29
24 Years	11	7	21	25
25 Years	10	7	24	27
26 Years	6	6	23	23
27 Years	13	9	29	33
28 Years	11	11	44	44
29 Years	14	8	27	33
30 Years	15	10	36	41
31 Years	10	10	37	37
32 Years	7	6	21	22
33 Years	8	8	31	31
34 Years	15	9	28	34
35 Years	7	7	26	26
36 Years	12	11	43	44
37 Years	7	10	40	37
38 Years	9	10	38	37
39 Years	10	10	40	40
40 Years	16	9	30	37
41 Years	9	11	45	43
42 Years	8	9	38	37
43 Years	10	10	40	40
44 Years	12	11	42	42
45 Years	15	15	56	56

Age	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
46 Years	10	10	38	38
47 Years	11	15	62	58
48 Years	12	14	57	55
49 Years	6	12	54	48
50 Years	16	14	51	53
51 Years	7	9	35	34
52 Years	7	13	56	50
53 Years	8	10	42	38
54 Years	9	8	28	29
55 Years	10	13	51	48
56 Years	5	6	23	22
57 Years	9	9	36	36
58 Years	5	8	34	31
59 Years	4*	8	37	33
60 Years	4*	6	26	24
61 Years	11	8	29	32
62 Years	0*	4	19	15
63 Years	3*	3	10	10
64 Years	1*	4	20	17
65 Years	3*	4	18	17
66 Years	2*	3	11	10
67 Years	2*	3	10	10
68 Years	0*	2	10	8
69 Years	1*	2	11	10
70 Years	0*	1	4*	3
71 Years	1*	0	11	10
72 Years	1*	0	9	8
73 Years	1*	2	7	6
74 Years	0*	1	6	5
75 Years	2*	1	1*	2
76 Years	0*	1	7	6
77 Years	1*	1	5	5
78 Years	2*	1	1*	2
79 Years	1*	1	2*	2
80 Years	2*	1	3*	4
81 Years	0*	1	5	4
82 Years	0*	1	3*	2
84 Years	0*	1	4*	3
85 Years	1*	1	2*	2
86 Years	0*	0	2*	2

Age	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
88 Years	0*	0	1*	1
91 Years	0*	1	3*	2
96 Years	2*	0	0*	2

*=Sparse Data

Table 48
Distribution of Marital Status by Status

Marital Status	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Other	0*	0	1*	1
Widowed	16	13	49	52
Separated	23	11	30	42
Divorced	58	63	249	244
Single	192	187	716	721
Married	158	173	682	667

*=Sparse data

Table 49
Distribution of Employment Status by Status

Employment Status	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Home Maker	13	10	35	38
Student	16	14	51	53
Retired	15	24	100	91
Other	1	1	1	2
Disabled	86	82	312	316
Unemployed	141	138	531	534
Employed	175	179	697	693

Table 50
Distribution of Patient Race by Status

Race	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Other	9	8	30	31
Black	35	25	88	98
White	403	414	1609	1598

Table 51
Distribution of Payer by Status

Payer	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
Self Pay	15	14	54	55
Other	48	52	204	200
Medicare	114	81	279	312
Medicaid	101	122	492	471
Managed Care	18	19	74	73
Commercial FFS	151	159	624	616

Table 52
Distribution of Relationship of Contact Person to Patient by Status

Relationship of Contact Person	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
Case Worker	0*	1	5	4
Sibling	08	0	2*	2
Child	5	9	37	33
None	08	1	4*	3
Other	184	177	676	683
Parent	135	119	443	459
Spouse	123	140	560	543

*=Sparse data

Table 53
Distribution of General (Primary) Diagnosis by Status

General Diagnosis	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
Anxiety	4*	1	2*	5
Dementia	1*	5	22	18
Other	5	4	13	14
Psychosis	76	52	179	203
Anxiety	60	59	229	230
Bi-Polar	68	90	372	350
Behavior/Personality Disorder	30	31	123	122
Depression	203	204	787	786

*=Sparse Data

Table 54
Distribution of Secondary Diagnosis by Status

Secondary Diagnosis	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
Drug	19	11	35	43
Dementia	2*	6	28	24
None	221	229	893	885
Other	7	12	50	45
Psychosis	4	6	27	25
Anxiety	112	105	397	404
Bi-Polar	3*	3	14	14
Behavior/Personality Disorder	55	50	188	193
Depression	24	24	95	95

*=Sparse Data

Table 55
Distribution of Travel Distance by Status

Travel Distance (Miles)	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
6	198	199	771	770
13	3*	3	10	10
15	4*	7	32	29
16	7	0	20	21
20	2*	0	6	6
21	2*	0	12	11
23	0*	0	2*	2
24	2*	0	7	7
25	9	10	39	38
26	2*	3	11	10
28	12	12	44	44
30	0*	1	4*	3
31	9	9	34	34
32	1*	1	2*	2
34	0*	0	1*	1
35	19	16	59	62
36	55	49	181	187
37	0*	1	4*	3
40	2*	2	6	6
41	2*	3	15	13
42	0*	1	4*	3
43	6*	3	9	12
44	0*	2	11	9

Travel Distance (Miles)	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
45	**2***	1	4*	5
46	2*	1	2*	3
47	4*	6	23	22
48	1*	0	0*	1
49	3*	2	7	8
50	0*	1	5	4
51	4*	2	6	8
52	6*	2	2*	6
53	4*	5	21	20
56	0*	1	5	4
57	0*	1	4*	3
58	8	7	24	25
59	4*	5	22	21
60	2*	2	7	7
61	2*	0	0*	2
62	2*	2	6	6
64	4*	3	11	12
65	0*	0	2*	2
67	0*	0	2*	2
68	7	11	46	42
69	0*	2	10	8
71	4*	2	6	8
72	0*	1	6	5
74	0*	0	1*	1
75	7	4	14	15
76	0*	0	2*	2
77	0*	1	4*	3
79	0*	1	5	4
80	0*	0	2*	2
81	0*	2	8	6
82	3*	1	0*	2
83	0*	1	4*	3
85	2*	1	2*	3
86	0*	0	2*	2
89	0*	3	13	10
90	4*	4	14	14
91	3*	1	1*	3
92	0*	2	1*	8
93	0*	1	4*	3
94	2*	2	8	8

Travel Distance (Miles)	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
95	2*	3	14	13
96	0*	2	8	6
99	0*	1	4*	3
101	0*	1	6	5
103	2*	1	2*	3
104	3*	1	0*	2
105	2*	1	4*	5
108	5	2	5	8
109	0*	0	2*	2
111	0*	1	3*	2
112	0*	0	1*	1
114	2*	0	0*	2
115	0*	6	13	10
116	1*	1	2*	2
117	0*	0	1*	1
118	2*	1	1*	2
120	0*	0	2*	2
121	0*	1	4*	3
122	0*	0	2*	2
123	2*	1	3*	4
125	0*	0	2*	2
127	0*	0	2*	2
128	2*	1	4*	5
129	1*	1	4*	4
131	0*	0	2*	2
132	2*	1	2*	3
133	2*	0	0*	2
134	2*	0	0*	2
137	0*	0	2*	2
138	2*	1	2*	3
141	0*	0	2*	2
142	0*	1	3*	2
144	0*	1	4*	3
151	0*	0	2*	2
162	0*	0	2*	2
164	0*	0	2*	2
169	0*	0	2*	2
183	0*	0	2*	2
192	0*	0	2*	2

Travel Distance (Miles)	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
203	0*	0	2*	2
206	0*	0	1*	1
227	0*	0	1*	1
230	0*	0	2*	2
251	0*	0	2*	2
347	0*	0	2*	2

*= Sparse Data

Table 56
Distribution of Wait Days by Status

Wait Days	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
0	9	25	111	2
1	11	10	37	2
2	4	4	15	3
3	2	4	19	4
4	4	4	15	2
5	5	4	13	5
6	3	4	16	10
7	6	7	28	5
8	1	2	10	9
9	2	2	6	6
10	3	3	12	12
11	2	4	18	16
12	1	2	8	7
13	6	5	16	17
14	18	10	29	37
15	5	5	21	21
16	1	2	7	6
17	4	6	23	21
18	4	5	21	20
19	0	2	9	7
20	1	4	19	16
21	8	9	38	37
22	3	2	5	6
23	0	3	15	12
24	3	3	13	13
25	2	2	10	10
26	3	3	11	11
27	4	3	12	13

Wait Days	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
28	10	13	51	48
29	13	6	16	23
30	2	3	15	14
31	6	4	15	17
32	1	3	15	13
33	3	2	6	7
34	7	5	18	20
35	13	12	46	47
36	8	7	27	28
37	1	1	5	5
38	2	3	12	11
39	1	2	8	7
40	3	1	1	3
41	7	4	14	17
42	24	16	56	64
43	6	7	27	26
44	6	3	7	10
45	1	1	4	4
46	2	3	15	14
47	1	1	5	5
48	2	2	9	9
49	14	10	36	40
50	3	4	16	15
51	2	1	4	5
52	4	3	9	10
53	3	1	1	3
54	1	2	7	6
55	3	3	12	12
56	12	13	52	51
57	3	2	7	8
58	0	1	3	3
59	2	1	5	5
60	3	4	15	15
61	3	3	11	11
62	1	3	12	10
63	23	21	78	80
64	4	2	6	8
65	1	0	1	2
66	0	1	4	3
67	0	1	3	2
68	0	0	2	2

Wait Days	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
69	0	1	4	3
70	4	6	23	21
71	2	1	5	6
73	2	1	1	2
74	1	2	7	6
75	0	0	1	1
77	3	3	11	11
78	0	0	2	2
80	0	0	2	2
81	1	0	0	1
82	0	0	2	2
83	0	0	1	1
84	8	5	16	19
85	0	0	2	2
86	1	1	3	3
87	0	1	4	3
88	0	1	5	4
89	1	1	2	2
90	5	4	15	16
91	55	44	161	172
92	1	1	6	6
93	0	1	2	2
94	1	1	4	4
95	2	1	4	5
96	0	1	3	3
97	0	0	1	1
98	12	13	50	49
99	1	1	2	2
100	0	1	3	2
102	0	0	2	2
103	0	0	2	2
104	0	0	1	1
105	3	6	24	21
107	1	1	2	2
108	0	0	2	2
109	0	0	2	2
111	0	0	2	2
112	2	2	7	7
115	0	0	2	2
119	1	4	17	15
120	1	3	11	10

Wait Days	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
121	0	0	2	2
122	0	0	1	1
124	0	1	3	2
125	1	0	0	1
126	2	3	14	13
127	0	0	2	2
132	0	0	2	2
133	2	1	2	3
139	0	0	2	2
140	2	1	1	2
146	0	0	2	2
147	2	1	5	6
150	0	0	2	2
152	0	0	1	1
153	0	0	1	1
154	0	1	14	3
158	0	0	1	1
160	0	0	1	1
165	1	0	0	1
168	3	1	0	2
175	2	1	4	5
176	0	0	2	2
179	0	0	1	1
180	1	1	2	2
181	1	1	2	2
182	4	8	35	31
183	0	0	1	1
188	0	0	1	1
189	0	0	2	2
337	0	0	1	1

Table 57
Distribution of Appointment Time by Status

Appt Time	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Morning	223	198	741	766
Afternoon	224	249	986	961

Table 58
Distribution of Day of Week by Status

Day of Week	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Monday	106	100	308	386
Tuesday	113	113	436	436
Wednesday	32	31	118	119
Thursday	118	101	374	391
Friday	78	102	419	395

Table 59
Distribution of Service Date (Season) by Status

Service Date (Season)	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Winter	67	59	221	229
Spring	247	319	1305	1233
Summer	39	18	48	69
Fall	94	51	153	196

Table 60
Distribution of Use of Non-MD Mental Health Services

Non MD Mental Health Appointments	Non-Adherent Frequency	Expected Frequency	Adherent Frequency	Expected Frequency
Yes	87	92	362	366
No	360	355	1365	1370

Table 61

Distribution of Number of Cancelled Appointments (numbercnx) by Status

Numbercnx	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
0 Visits	437	435	1688	1688
1 Visit	2	5	22	19
2 Visits	7	3	10	14
3 Visits	0	1	4	3
4 Visits	0	0	1	1
5 Visits	0	0	2	2
6 Visits	1	0	0	1

Table 62

Distribution of Number of Appointment by Status

Number of Appointments	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
1	0*	1	3*	2
2	86	60	204	230
3	95	75	271	291
4	86	92	360	354
5	53	70	289	272
6	39	43	172	168
7	26	28	109	107
8	10	15	64	59
9	7	9	36	34
10	8	10	41	39
11	2*	7	32	27
12	5*	6	26	25
13	5*	4	16	17
14	3*	4	18	17
15	1*	1	6	6
16	7	2	5	10
17	3*	3	10	10
18	3*	2	6	7
19	0*	2	8	6
20	1*	1	5	5
21	0*	1	6	5
22	0*	1	6	5
23	3*	1	0*	2
24	0*	2	10	8

Number of Appointments	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
25	0*	0	1*	1
26	0*	0	2*	2
27	0*	0	2*	2
28	1*	1	2*	2
29	1*	1	4*	4
30	1*	1	2*	2
31	0*	0	1*	1
32	1*	1	2*	2
36	0*	0	2*	2
37	0*	0	1*	1
38	0*	0	2*	2
40	0*	0	1*	1
44	0*	0	1*	1
48	0*	0	1*	1

Table 63
Distribution of Maker by Status

Maker	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Untrained	0	0	447	447
Trained	1	1	1726	1726

Table 64
Distribution of Provider Type by Status

Provider Type	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	48	55	220	213
Resident	249	225	847	871
Attending	150	167	660	643

Table 65
Distribution of Referring Provider by Status

Referring Provider	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
Counselor, Internal	6	4	15	17
External Health Center/Organization	6	5	17	18
PCP, Internal	7	5	15	17
PCP, External	0*	1	3*	2
PCP, Undetermined	35	29	105	111
Resident, Internal	51	45	168	174
Self	3*	6	25	22
Specialist, External	45	32	113	126
Specialist, Internal	141	141	545	545
NP, External	153	180	721	694

*=Sparse data

Table 66
Same Visit Distribution of Same Visit Type by Status

Same Visit Type	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
No	3*	3	11	11
Yes	444	444	1716	1716

*=Sparse Data

Table 67
Distribution of Same Day of Week by Status

Same Day of Week	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
No	189	194	258	253
Yes	756	750	971	976

Table 68
Distribution of Same Time of Day by Status

Same Time of Day	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
No	276	264	171	183
Yes	1010	1022	717	705

Table 69

Distribution of Same Payer by Status

Same Payer	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
No	8	37	439	110
Yes	173	144	1554	1583

Table 70-

Distribution of Same Provider Type by Status

Same Provider Type	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
No	90	85	321	326
Yes	357	362	1406	1401

Table 71

Distribution of Percent Non-Adherent (Percentnon) by Status

Percent Non	Non-Adherent Frequency	Non-Adherent Expected Frequency	Adherent Frequency	Adherent Expected Frequency
0 Percent	0*	349	1696	1347
3 Percent	3*	1	0*	2
4 Percent	8	2	0*	6
5 Percent	1*	0	1*	2
6 Percent	9	2	1*	8
7 Percent	4*	1	0*	3
8 Percent	10	2	0*	8
9 Percent	2*	1	2*	2
10 Percent	8	2	8	7
11 Percent	7	2	7	6
13 Percent	9	2	9	7
14 Percent	25	6	25	24
15 Percent	0*	1	0*	1
17 Percent	36	8	36	29
20 Percent	50	11	5	44
25 Percent	85	19	6	72
29 Percent	0*	0	1*	1
33 Percent	94	20	3*	77
40 Percent	2*	0	0*	2
45 Percent	0*	0	1*	1
50 Percent	87	19	4*	72
60 Percent	2*	0	0*	2

Percent Non	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
63 Percent	1*	0	0*	1
67 Percent	2*	0	0*	2
71 Percent	2*	0	0*	2

*= Sparse data

Table 72
Distribution of Previous Visit Non-Adherent by Status

Previous Visit NOS	Non-Adherent		Adherent	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Yes	103	24	12	91
No	344	423	1715	1635

Table 73
Distribution of Payer by Race

Payer	Other		Black		White	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Self Pay	1	1	2	4	66	64
Other	5	5	26	14	221	233
Medicaid	9	7	38	22	346	364
Medicare	5	11	35	34	553	549
Managed Care	1	2	2	5	89	85
Commercial FFS	18	14	20	44	737	717

Table 74
Distribution of Race by Provider Type

Race	Counselor		Resident		Attending	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Other	5	5	21	20	13	15
Black	16	15	63	62	44	46
White	247	248	1012	1014	753	750

Table 75
Distribution of Gender by Age

Age in Years	Male		Female	
	Frequency	Expected Frequency	Frequency	Expected Frequency
3 (missing)	2	2	3	3
18 Years	2	3	7	6
19 Years	6	2	1	5
20 Years	18	13	22	27
21 Years	17	13	23	27
22 Years	14	14	27	27
23 Years	11	12	25	24
24 Years	13	11	19	21
25 Years	15	11	19	23
26 Years	10	10	19	19
27 Years	19	14	23	28
28 Years	19	18	36	37
29 Years	15	14	26	27
30 Years	21	17	30	34
31 Years	16	16	31	31
32 Years	5	10	23	19
33 Years	13	13	26	26
34 Years	21	14	22	29
35 Years	12	11	21	22
36 Years	16	18	39	37
37 Years	15	16	32	31
38 Years	18	16	29	31
39 Years	13	17	37	33
40 Years	18	15	28	31
41 Years	23	18	31	36
42 Years	17	15	29	31
43 Years	14	17	36	33
44 Years	16	18	38	36
45 Years	20	24	51	47
46 Years	18	16	30	32
47 Years	25	24	48	49
48 Years	21	23	48	46
49 Years	29	20	31	40
50 Years	19	22	48	45
51 Years	15	14	27	28
52 Years	12	21	51	42
53 Years	13	16	35	32
54 Years	14	12	23	25

Age in Years	Male		Female	
	Frequency	Expected Frequency	Frequency	Expected Frequency
55 Years	21	20	40	41
56 Years	10	9	18	19
57 Years	9	15	36	30
58 Years	8	13	31	26
59 Years	12	14	29	27
60 Years	9	10	21	20
61 Years	11	13	29	27
62 Years	3	6	16	13
63 Years	1	4	12	9
64 Years	7	7	14	14
65 Years	7	7	14	14
66 Years	6	4	7	9
67 Years	7	4	5	8
68 Years	5	3	5	7
69 Years	0	4	12	8
70 Years	1	1	3	3
71 Years	6	4	6	8
72 Years	5	3	5	7
73 Years	2	3	6	5
74 Years	1	2	5	4
75 Years	0	1	3	2
76 Years	4	2	3	5
77 Years	1	2	5	4
78 Years	2	1	0	2
79 Years	1	1	2	2
80 Years	0	2	5	3
81 Years	0	2	5	3
82 Years	1	1	2	2
84 Years	1	1	3	3
85 Years	0	1	3	2
86 Years	0	1	2	1
88 Years	1	0	0	1
91 Years	0	1	3	2
96 Years	0	1	2	1

Table 76
Distribution of Payer by Employment

Payer	Home Maker		Student		Retired		Other	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Self Pay	0	1	3	2	3	4	0	0
Other	4	6	6	8	6	13	0	0
Medicaid	6	9	10	12	0	21	0	0
Medicare	15	13	3	18	84	31	2	1
Managed Care	2	2	5	3	5	5	0	0
Commercial FFS	21	17	40	24	17	41	0	0

Table 77
Distribution of Payer by Employment Continued

Payer	Disabled		Unemployed		Employed	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Self Pay	3	13	24	21	36	28
Other	65	46	91	78	80	101
Medicaid	88	72	210	121	79	158
Medicare	223	109	197	193	69	238
Managed Care	5	17	10	28	65	37
Commercial FFS	14	142	140	240	543	311

Appendix 2-SAS Commands

Regression 1

```
DATA nonadherence;
INPUT status gender age waitdays season apptdayofweek appttime payer samepayer gendiagnosis
secdiagnosis maritalstatus percentnon race numberappointments  prevvisitnos provider @@;
CARDS;
;
run;

data NONADHERENCE; set NONADHERENCE; ROW= _n_;
  AgeC44= Age-44; * center age at the grand mean;
run;

 proc surveyselect noprint data=NONADHERENCE OUT= BootSamp method=URS
     sampsize= 2169 rep= 100 outhits;
     id row status gender ageC44 waitdays season apptdayofweek appttime payer samepayer
gendiagnosis
secdiagnosis maritalstatus percentnon race numberappointments  prevvisitnos provider ;
run;

proc logistic descending data=BootSamp; by replicate;
 class gender waitdays season apptdayofweek appttime
          payer samepayer gendiagnosis secdiagnosis maritalstatus
           prevvisitnos provider   ;
model status= gender agec44 waitdays season apptdayofweek appttime
          payer samepayer gendiagnosis secdiagnosis maritalstatus
           prevvisitnos provider  /
                    selection = forward sle= 0.05;
ods output ModelBuildingSummary= MBS;
run;
```

Regression 2

```
DATA nonadherence;
INPUT status gender age waitdays season apptdayofweek appttime payer samepayer gendiagnosis
secdiagnosis maritalstatus percentnon race numberappointments  prevvisitnos provider @@;
CARDS;
;
run;

data NONADHERENCE; set NONADHERENCE; ROW= _n_;
  AgeC44= Age-44; * center age at the grand mean;
run;

 proc surveyselect noprint data=NONADHERENCE OUT= BootSamp method=URS
     sampsize= 2169 rep= 100 outhits;
     id row status gender ageC44 waitdays season apptdayofweek appttime payer samepayer
gendiagnosis secdiagnosis maritalstatus percentnon race numberappointments prevvisitnos provider;
run;

proc logistic descending data=BootSamp; by replicate;
 class gender waitdays season apptdayofweek appttime
          payer samepayer gendiagnosis secdiagnosis maritalstatus race prevvisitnos provider;
model status= gender agec44 waitdays season apptdayofweek appttime
```

```
        payer samepayer gendiagnosis secdiagnosis maritalstatus percentnon race numberappointments
prevvisitnos provider/
                 selection = forward sle= 0.05;
ods output ModelBuildingSummary= MBS;RUN;
```

Regression 3

```
DATA nonadherence;
INPUT status age waitdays season apptdayofweek samepayer gendiagnosis maritalstatus prevvisitnos
provider @@;
CARDS;
;
run;

data NONADHERENCE; set NONADHERENCE; ROW= _n_;
  AgeC44= Age-44; * center age at the grand mean;
        * check coding;
        run;

proc surveyselect noprint data=NONADHERENCE OUT= BootSamp method=URS
    sampsize= 2169 rep= 100 outhits seed= 121212;
    id row status ageC44 waitdays season apptdayofweek
                 samepayer gendiagnosis maritalstatus prevvisitnos provider;
run;


proc logistic descending data=BootSamp;by replicate;
 class waitdays season apptdayofweek samepayer gendiagnosis maritalstatus prevvisitnos provider;
model status= agec44 waitdays season apptdayofweek samepayer gendiagnosis maritalstatus
prevvisitnos provider/
                 selection = forward sle= 0.05;
ods output ModelBuildingSummary= MBS;RUN;

proc freq data= MBS order=freq;
        table effectentered*step/ norow nocol nopercent; run;

proc logistic descending data=NONADHERENCE;
 class  waitdays season apptdayofweek samepayer gendiagnosis maritalstatus prevvisitnos provider;
        model status= agec44 waitdays season apptdayofweek samepayer gendiagnosis maritalstatus
prevvisitnos provider/
                 lackfit ctable ;
run;
```

Validation

```
DATA nonadherence;
INPUT status age waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider @@;
CARDS;
;
run;

data NONADHERENCE; set NONADHERENCE; ROW= _n_;
  AgeC44= Age-44; * center age at the grand mean;
        * check coding;
        run;
```

```
proc surveyselect noprint data=NONADHERENCE OUT= BootSamp method=URS
     sampsize= 2169 rep= 100 outhits seed= 121212;
     id row status ageC44 waitdays season apptdayofweek
                    samepayer maritalstatus prevvisitnos provider;
run;


proc logistic descending data=BootSamp;by replicate;
 class waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider;
model status= agec44 waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider/
                    selection = forward sle= 0.05;
ods output ModelBuildingSummary= MBS;RUN;

proc freq data= MBS order=freq;
        table effectentered*step/ norow nocol nopercent; run;

proc logistic descending data=NONADHERENCE outest=OENONADHERENCE;
 class  waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider;
        model status= agec44 waitdays season apptdayofweek samepayer
                        maritalstatus prevvisitnos provider/ lackfit ctable pprob=(0 to 1 by .1);
        output out=LRO_DevDat p= phat;
run;

proc freq data=nonadherence; table status; run;

goptions reset=all;
proc univariate data=LRO_DevDat; class status;
    histogram phat / nrows=2 ;*nmidpoints = 40 HOFFSET=2 nrows=4;
    inset n = "N" (4.0)  mean = "Mean" (5.2) median= "Median" (5.2)
                        std = "Std Dev" (5.2) Min= "Min" (5.2) Max= "Max" (5.2) /
     pos = ne height = 2.5;
          *format Lap_n_Conv lapf.;
          *
          title "Estimated Probability of Nonadherence:  Dev Data";
  run;

        proc sort data=LRO_DevDat; by status;
proc boxplot data=LRO_DevDat;
         plot phat*status/boxstyle = schematic ; * nohlabel schematicidfar  boxwidthscale = 1 bwslegend;
   *   inset nobs mean(5.1) stddev(5.1) min max / header = 'Overall Summary Statistics' pos= tm; *
format=f6.0;
     insetgroup n mean(5.2) stddev(5.2) min (5.2) Q1 (5.2) Q2 (5.2) Q3 (5.2) max (5.2) / header =
'Summary Stats by Status';
 title h=1.2 'Estimated Probability of Nonadherence:  Dev Data';
 run;



DATA validation;
INPUT status age waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider @@;
CARDS;
;
run;
```

```
data Validation; set Validation;
  AgeC44= Age-44; * center age at the grand mean;
          * check coding;
          run;



proc logistic descending data=validation inest=OENONADHERENCE;
 class  waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider;
          model status= agec44 waitdays season apptdayofweek samepayer maritalstatus prevvisitnos
provider
                              / lackfit ctable pprob=(0 to 1 by .1) maxiter=0;
run;


proc logistic descending data=nonadherence ;
 class  waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider;
          model status= agec44 waitdays season apptdayofweek samepayer maritalstatus prevvisitnos
provider
                              / lackfit ctable pprob=(0 to 1 by .1) ; title 'Final Logistic Model Fit to Development
Data';
run;



proc logistic descending data=validation ;
 class  waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider;
          model status= agec44 waitdays season apptdayofweek samepayer maritalstatus prevvisitnos
provider
                              / lackfit ctable pprob=(0 to 1 by .1) ; title 'Final Logistic Model Fit to Validation
Data';
run;

proc freq data=validation; table status; run;

data V; set validation; dsn= 'V';
data D; set nonadherence; dsn= 'D';
data DV; set D V;

proc means maxdec=2 data=DV; class dsn;
 var agec44 waitdays season apptdayofweek samepayer maritalstatus prevvisitnos provider;
 run;

 proc freq data=DV; table dsn*(waitdays season apptdayofweek samepayer maritalstatus prevvisitnos
provider)/chisq; run;
```

Appendix 3 Bi-Variant Analysis Tables

Table 78
Frequency Distribution of Age by Waitdays

	None		1-30 Days		31-90 Days		91+ days	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
18	1*	0	3*	3	5	4	0*	2
19	0*	0	2*	2	5	3	0*	2
20	4*	2	12	12	16	16	8	9
21	0*	2	11	12	20	16	9	9
22	1*	2	19	13	17	17	4*	10
23	1*	2	11	11	15	15	9	9
24	1*	2	11	10	11	13	9	8
25	1*	2	18	10	8	14	7	8
26	3*	2	10	9	12	12	4*	7
27	1*	2	16	13	19	17	6	10
28	1*	3	22	17	20	22	12	13
29	0*	2	18	12	18	17	5	10
30	0*	3	21	15	14	21	16	12
31	1*	2	12	14	23	19	11	11
32	0*	2	5	9	19	11	4*	7
33	1*	2	16	12	13	16	9	9
34	1*	2	18	13	21	17	3*	10
35	0*	2	13	10	13	13	7	8
36	1*	3	19	17	29	22	6	13
37	4*	3	24	14	10	19	9	11
38	2*	3	8	14	21	19	16	11
39	2*	3	16	16	21	20	11	12
40	5	3	11	14	23	19	7	11
41	6	3	9	16	28	22	11	13
42	5	3	16	14	16	19	9	11
43	3*	3	13	15	22	20	12	12
44	4*	3	16	16	23	22	11	13
45	4*	4	28	22	23	29	16	17
46	5	3	12	15	18	19	13	11
47	2*	4	18	22	37	30	16	17
48	4*	4	20	23	28	28	22	16
49	8	3	19	18	24	24	15	14
50	1*	4	19	36	27	27	11	16
51	2*	2	9	16	17	17	15	10
52	3*	4	11	34	26	26	15	15
53	3*	3	14	19	19	19	12	11

Table 78

Frequency Distribution of Age by Waitdays

	None		1-30 Days		31-90 Days		91+ days	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
54	2*	2	6	11	23	15	6	7
55	2*	3	15	19	26	25	18	14
56	2*	2	11	9	8	11	7	7
57	6	2	14	14	15	18	10	11
58	2*	2	5	12	18	16	14	9
59	7	2	16	12	10	17	8	10
60	4*	2	4*	9	12	12	10	7
61	2*	2	9	12	17	16	12	9
62	1*	1	8	6	8	8	2*	4
63	1*	1	2*	4	6	5	4*	3
64	1*	1	8	6	4*	9	8	5
65	0*	1	3*	6	8	9	10	5
66	0*	1	3*	4	3*	5	7	3
67	2*	1	3*	4	2*	5	5	3
68	2*	1	5	0	2*	4	1*	2
69	0*	1	3*	4	5	5	4*	3
70	1*	0	3*	1	0*	2	0*	1
71	1*	1	5	4	2*	5	4*	2
72	0*	0	3*	3	6	4	1*	2
73	1*	0	1*	2	1*	3	5	2
74	0*	0	2*	1	1*	2	3*	1
75	0*	0	0*	1	0*	1	3*	1
76	0*	0	0*	2	1*	3	6	2
77	0*	0	1*	2	3*	2	2*	1
78	0*	0	0*	1	1*	1	2*	1
79	0*	0	1*	1	2*	1	0*	1
80	1*	0	2*	2	2*	2	0*	1
81	0*	0	2*	2	1*	2	1*	1
82	0*	0	0*	1	2*	1	1*	1
84	0*	0	1*	1	2*	2	1*	1
85	0*	0	0*	1	2*	1	1*	1
86	0*	0	2*	1	0*	1	0*	0
88	0*	0	0*	0	0*	0	1*	0
91	1*	0	1*	1	0*	1	1*	1
96	0*	0	0*	1	1*	1	0*	0

* = Sparse Data

Table 79
Frequency Distribution of Age by Season

	Winter		Spring		Summer		Fall	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
18	1*	1	6	6	0*	0	2*	1
19	0*	1	5	5	0*	0	2*	1
20	4*	5	29	29	1*	2	6	5
21	1*	5	29	29	4*	2	6	5
22	4*	5	30	30	4*	2	3*	5
23	7	5	21	26	4*	1	4*	4
24	2*	4	24	23	2*	1	4*	4
25	5	4	28	24	0*	1	1*	4
26	2*	4	23	21	1*	1	3*	3
27	8	6	30	30	2*	2	2*	5
28	5	7	43	39	0*	2	7	6
29	5	5	26	29	5	2	5	5
30	11	7	31	36	3*	2	6	6
31	6	6	30	34	4*	2	7	5
32	5	4	17	20	1*	1	5	3
33	7	5	30	28	0*	2	2*	4
34	2*	6	27	31	3*	2	11	5
35	4	4	22	24	2*	1	5	4
36	2*	7	47	40	1*	2	5	6
37	5	6	35	34	4*	2	3*	5
38	6	6	30	34	3*	2	8	5
39	4*	7	37	36	1*	2	8	6
40	4*	6	33	33	2*	2	7	5
41	10	7	38	39	5	2	1*	6
42	11	6	30	33	1*	2	4*	5
43	7	7	34	36	1*	2	8	6
44	11	7	36	39	1*	2	6	6
45	4*	9	56	51	2*	3	9	8
46	4*	6	38	34	2*	2	4*	5
47	13	10	55	52	0*	3	5	8
48	11	9	50	49	2*	3	6	8
49	15	8	35	43	1*	2	7	7
50	8	9	52	48	1*	3	8	8
51	4*	6	31	30	0*	2	5	5
52	7	8	49	45	0*	2	7	7
53	7	6	33	34	2*	2	5	5
54	4*	5	29	26	2*	1	2*	4

Table 79

Frequency Distribution of Age by Season

	Winter		Spring		Summer		Fall	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
55	10	8	45	44	0*	2	6	7
56	4*	4	18	20	1*	1	5	3
57	8	6	33	32	0*	2	4*	5
58	5	5	20	28	3*	2	1*	4
59	8	5	29	29	1*	2	3*	5
60	0*	4	26	21	2*	1	2*	3
61	8	5	23	29	0*	2	9	5
62	5	3	9	14	0*	1	5	2
63	5	2	8	9	0*	1	0*	1
64	2*	3	13	15	3*	1	3*	2
65	2*	3	16	15	2*	1	1*	2
66	0*	2	12	9	0*	1	1*	1
67	1*	2	11	9	0*	0	0*	1
68	0*	1	7	7	0*	0	3*	1
69	2*	2	9	9	1*	0	0*	1
70	2*	1	2*	3	0*	0	0*	0
71	1*	2	10	9	1*	0	0*	1
72	1*	1	5	7	1*	0	3*	1
73	2*	1	4*	6	1*	0	1*	1
74	0*	1	6	4	0*	0	0*	1
75	0*	0	3*	1	0*	0	0*	0
76	0*	1	7	5	0*	0	0*	1
77	0*	1	5	4	0*	0	1*	1
78	0*	0	2*	2	0*	0	1*	0
79	1*	0	2*	2	0*	0	0*	0
80	0*	1	2*	4	1*	0	2*	1
81	1*	1	3*	4	0*	0	1*	1
82	1*	0	2*	2	0*	0	0*	0
84	0*	1	2*	3	0*	0	2*	0
85	0*	0	3*	2	0*	0	0*	0
86	0*	0	2*	1	0*	0	0*	0
88	0*	0	0*	1	0*	0	1*	0
91	2*	0	1*	2	0*	0	0*	0
96	0*	0	0*	1	2*	0	0*	0

* = Sparse Data

Table 80

Frequency Distribution of Age by Day of Week

	Monday		Tuesday		Wednesday		Thursday		Friday	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
18	3*	2	2*	2	0*	1	3*	2	1*	2
19	4*	2	1*	2	1*	0	0*	2	1*	2
20	14	9	5	10	3*	3	10	9	8	9
21	9	9	10	10	3*	3	12	9	6	9
22	12	9	8	10	3*	3	10	9	8	9
23	12	8	11	9	3*	2	5	8	5	8
24	6	7	8	8	6	2	8	7	4*	7
25	8	8	13	9	1*	2	7	8	5	8
26	7	6	13	7	0*	2	5	7	4*	7
27	8	9	11	11	6	3	11	9	6	10
28	15	0	15	14	4*	4	9	12	12	13
29	9	9	10	10	3*	3	12	9	7	9
30	9	11	12	13	1*	4	19	12	10	12
31	12	10	15	12	4*	3	7	11	9	11
32	8	6	9	7	4*	2	1*	6	6	6
33	11	15	10	10	4*	3	5	9	9	10
34	8	10	15	11	5	3	12	10	3*	10
35	10	7	5	8	3*	3	9	7	6	8
36	8	12	7	14	6	4	13	12	21	13
37	11	10	11	12	5	3	8	11	12	11
38	15	10	15	12	0*	3	6	11	11	11
39	11	11	12	13	5	3	8	11	14	11
40	9	10	10	12	2*	3	7	10	8	11
41	11	12	18	14	1*	4	10	12	14	12
42	8	10	16	12	5	3	6	10	11	11
43	12	11	8	13	3*	3	15	11	12	11
44	12	12	23	14	5	4	8	12	6	12
45	8	16	17	18	4*	5	20	16	22	16
46	9	11	10	12	3*	3	14	11	12	11
47	20	16	9	18	4*	5	21	16	19	17
48	15	15	13	17	3*	5	12	16	26	16
49	17	13	16	15	2*	4	8	14	17	14
50	15	15	15	17	1*	5	20	15	16	15
51	16	9	5	11	0*	3	11	9	10	10
52	16	14	16	16	4*	4	11	14	16	14
53	11	11	7	12	6	3	16	11	8	11
54	9	8	9	9	2*	3	11	8	6	8

Table 80
Frequency Distribution of Age by Day of Week

	Monday		Tuesday		Wednesday		Thursday		Friday	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
55	14	14	14	15	7	4	10	14	16	14
56	5	6	7	4	2*	2	8	6	6	6
57	12	10	5	11	3*	3	9	10	16	10
58	7	9	6	10	3*	3	14	9	9	9
59	4*	9	11	10	2*	3	11	9	13	9
60	4*	7	11	8	3*	2	8	7	4*	7
61	7	9	5	10	3*	3	9	9	16	9
62	4	4	2*	5	0*	1	8	4	5	4
63	2*	3	7	3	0*	1	2*	3	2*	3
64	1*	5	5	5	3*	1	6	5	6	5
65	5	5	1*	5	3*	1	7	5	5	5
66	1*	3	3*	3	0*	1	5	3	4*	3
67	3*	3	4*	3	0*	1	3*	3	2*	3
68	3*	2	0*	3	2*	1	2*	2	3*	2
69	1*	3	4*	3	0*	1	2*	3	5	3
70	0*	1	0*	1	0*	0	2*	1	2*	1
71	4*	3	2*	3	0*	1	2*	3	4*	3
72	1*	2	6	3	1*	1	2*	2	0*	2
73	2*	2	2*	2	0*	0	3*	2	1*	2
74	0*	1	5	2	0*	1	0*	1	1*	1
75	0*	1	1*	1	0*	0	2*	1	0*	1
76	2*	2	2*	2	0*	0	3*	2	0*	2
77	1*	1	3*	2	0*	0	1*	1	1*	1
78	0*	1	0*	1	0*	0	0*	1	3*	1
79	0*	1	3*	1	0*	0	0*	1	0*	1
80	2*	1	2*	1	1*	0	0*	1	0*	1
81	0*	1	1*	1	0*	0	1*	1	0*	1
82	0*	1	3*	1	0*	0	0*	1	0*	1
84	1*	1	1*	1	2*	0	0*	1	0*	1
85	0*	1	2*	1	0*	0	0*	1	1*	1
86	0*	0	2*	1	0*	0	0*	0	0*	0
88	0*	0	1*	0	0*	0	0*	0	0*	0
91	0*	1	3*	1	0*	0	0*	1	0*	0
96	0*	0	2*	1	0*	0	0*	0	0*	0

* = Sparse Data

Table 81
Frequency Distribution of Age by Marital Status

	Widowed		Separated		Divorced		Single		Married	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
18	0*	0	0*	0	0*	1	8	4	1*	3
19	0*	0	0*	0	0*	1	7	3	0*	3
20	0*	1	0*	1	0*	6	40	17	0*	15
21	0*	1	0*	1	1*	6	37	17	2*	15
22	0*	1	0*	1	0*	6	37	17	4*	16
23	0*	1	1*	1	0*	5	33	15	2*	14
24	0*	0	0*	1	1*	5	24	13	7	12
25	0*	1	0*	1	0*	5	11	14	13	13
26	0*	1	0*	1	2*	4	20	12	7	11
27	0*	1	0*	1	2*	6	34	18	6	16
28	0*	2	4*	1	3*	8	31	23	17	21
29	0*	1	0*	1	3*	6	24	17	14	16
30	0*	2	1*	1	5	7	34	21	11	11
31	0*	1	2*	1	2*	7	24	20	19	19
32	0*	1	0*	1	2*	4	17	12	9	9
33	0*	1	2*	1	3*	6	22	16	12	12
34	0*	1	1*	1	0*	6	18	18	24	24
35	0*	1	0*	1	4*	5	15	14	14	14
36	0*	2	0*	1	7	8	24	23	24	21
37	0*	1	0*	1	3*	7	17	20	27	18
38	0*	1	0*	1	5	7	24	20	18	18
39	0*	1	0*	1	4*	7	20	21	26	19
40	0*	1	2*	1	8	7	16	20	20	18
41	2*	2	3*	1	9	8	20	23	20	21
42	0*	1	2*	1	8	7	16	19	20	18
43	2*	1	0*	1	3*	7	22	21	23	19
44	1*	2	5*	1	12	8	18	23	18	21
45	1*	2	2*	2	8	10	25	30	35	28
46	3*	1	1*	1	16	7	17	20	11	19
47	0*	2	4*	2	18	10	18	30	33	28
48	0*	2	4*	2	13	10	21	29	31	27
49	0*	2	1*	2	15	8	16	25	28	23
50	1*	2	4*	2	11	9	28	28	23	26
51	4*	1	1*	1	6	66	12	18	19	16
52	4*	2	2*	2	15	9	20	26	22	24
53	2*	1	0*	1	10	7	15	20	24	19
54	1*	1	3*	1	11	5	7	15	15	14

Table 81

Frequency Distribution of Age by Marital Status

	Widowed		Separated		Divorced		Single		Married	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
55	1*	2	2*	1	12	9	10	26	36	24
56	2*	1	1*	1	4*	4	7	12	13	11
57	1*	1	2*	1	13	6	5	19	24	17
58	0*	1	1*	1	9	6	11	16	18	15
59	0*	1	2*	1	12	6	7	17	20	16
60	2*	1	0*	1	6	4	5	13	17	12
61	1*	1	0*	1	14	6	13	17	12	15
62	4*	1	0*	0	3*	3	4*	8	8	7
63	3*	0	0*	0	2*	2	6	5	2*	5
64	1*	1	0*	1	5	3	3*	8	12	8
65	1*	1	0*	1	3*	3	8	8	9	8
66	1*	0	0*	0	3*	2	2*	5	7	5
67	0*	0	0*	0	1*	2	4*	5	7	5
68	0*	0	0*	0	1*	1	3*	4	6	4
69	1*	0	0*	0	2*	2	6	5	3*	5
70	0*	0	0*	0	0*	1	2*	2	2*	2
71	1*	0	0*	0	0*	1	2*	5	9	5
72	1*	0	0*	0	2*	1	1*	4	6	4
73	1*	0	0*	0	1*	1	2*	3	4*	3
74	2*	0	0*	0	0*	1	0*	3	4*	2
75	0*	0	0*	0	2*	0	0*	1	1*	1
76	1*	0	0*	0	1*	1	2*	3	3*	3
77	1*	0	0*	0	1*	1	0*	3	4*	2
78	0*	0	0*	0	0*	0	2*	1	1*	1
79	2*	0	0*	0	0*	0	0*	1	1*	1
80	3*	0	0*	0	0*	1	0*	2	2*	2
81	2*	0	0*	0	0*	1	0*	2	3*	2
82	1*	0	0*	0	0*	0	1*	1	1*	1
84	3*	0	0*	0	0*	1	0*	2	1*	2
85	3*	0	0*	0	0*	0	0*	1	0*	1
86	0*	0	0*	0	0*	0	0*	1	2*	1
88	0*	0	0*	0	0*	0	0*	0	1*	0
91	2*	0	0*	0	0*	0	0*	1	1*	1
96	2*	0	0*	0	0*	0	0v	1	0*	1

* = Sparse Data

Table 82

Frequency Distribution of Age by Prev Visit NOS

	Yes		No	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency
18	2*	0	7	9
19	1*	0	6	7
20	1*	2	39	38
21	2*	2	38	38
22	4*	2	37	39
23	4*	2	32	34
24	2*	2	30	30
25	2*	2	32	32
26	0*	2	29	27
27	6	2	36	40
28	4*	3	51	52
29	5	2	39	39
30	7	3	44	48
31	4*	2	43	45
32	1*	1	27	27
33	2*	2	37	37
34	8	2	35	41
35	4*	2	29	31
36	2*	3	53	52
37	2*	2	45	45
38	1*	2	46	45
39	2*	3	48	47
40	2*	2	44	43
41	4*	3	50	51
42	2*	2	44	44
43	4*	3	46	47
44	1*	3	53	51
45	4*	4	67	67
46	2*	3	46	45
47	5	4	68	69
48	1*	4	68	65
49	1*	3	59	57
50	3*	3	64	64
51	1*	2	41	40
52	3*	3	60	60
53	2*	3	46	45
54	1*	2	36	35
55	3*	3	58	58
56	0*	1	28	27

Table 82

Frequency Distribution of Age by Prev Visit NOS

	Yes		No	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency
57	0*	2	45	43
58	2*	2	37	37
59	0*	2	41	39
60	1*	2	29	28
61	3*	2	37	38
62	0*	1	19	18
63	1*	1	12	12
64	0*	1	21	20
65	0*	1	21	20
66	0*	1	13	12
67	1*	1	11	11
68	0*	1	10	9
69	0*	1	12	11
70	0*	0	4*	4
71	0*	1	12	11
72	0*	0	10	9
73	0*	0	8	8
74	0*	0	6	6
75	0*	0	3*	3
76	0*	0	7	7
77	0*	0	6	6
78	0*	0	3*	3
79	0*	0	3*	3
80	0*	0	5	5
81	0*	0	5	5
82	0*	0	3*	3
84	0*	0	4*	4
85	0*	0	3*	3
86	0*	0	2*	2
88	0*	0	1*	1
91	0*	0	3*	3
96	0*	0	2*	2

* = Sparse Data

Table 83

Frequency Distribution of Age by Provider Type NOS

	Counselor		Resident		Attending	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
18	1*	1	0*	5	8	4
19	2*	1	1*	4	4*	3
20	2*	5	18	20	20	15
21	5	5	17	20	18	15
22	8	5	20	21	13	15
23	5	4	19	18	12	13
24	6	4	9	16	17	12
25	4*	4	18	17	12	13
26	6	4	15	15	8	11
27	9	5	21	21	12	16
28	8	7	29	28	18	21
29	6	5	17	21	18	15
30	4*	6	30	26	17	19
31	4*	6	24	24	19	18
32	0*	4	18	14	10	10
33	6	5	24	20	9	15
34	9	5	20	22	14	16
35	2*	4	23	17	8	12
36	2*	7	32	28	21	21
37	15	6	23	24	9	18
38	5	6	22	24	20	18
39	7	6	32	25	11	19
40	5	6	22	23	19	17
41	4*	7	34	27	16	20
42	9	6	24	23	13	17
43	9	6	27	25	14	19
44	4*	7	30	27	20	20
45	10	9	36	36	25	27
46	7	6	27	24	14	18
47	9	9	33	37	31	27
48	12	9	33	35	24	26
49	5	7	30	30	25	22
50	8	8	37	34	22	25
51	2*	5	31	21	9	16
52	7	8	34	32	22	24
53	5	6	24	24	19	18
54	0*	5	17	19	20	14
55	5	8	22	31	34	23

Table 83
Frequency Distribution of Age by Provider Type NOS

	Counselor		Resident		Attending	
Age	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
56	7	4	12	14	9	10
57	13	6	22	23	10	17
58	5	5	20	20	14	15
59	5	5	25	21	11	15
60	1*	4	10	15	19	11
61	3*	5	19	20	18	15
62	4*	3	11	10	4*	7
63	3*	2	9	7	1*	5
64	3*	3	9	11	9	9
65	3*	3	12	11	6	9
66	1*	2	3	6	9	5
67	0*	1	7	6	5	4
68	0*	1	6	5	4*	4
69	0*	1	6	6	6	4
70	0*	0	1*	2	3*	1
71	0*	1	8	6	4*	4
72	0*	1	5	5	5	4
73	0*	1	1*	4	7	3
74	0*	1	1*	3	5	2
75	0*	0	0*	2	3*	1
76	0*	1	3*	4	4*	3
77	0*	1	0*	3	6	2
78	0*	0	2*	2	1*	1
79	0*	0	0*	2	3*	1
80	1*	1	2*	3	2*	2
81	0*	1	2*	3	3*	2
82	0*	0	0*	2	3*	1
84	0*	0	1*	2	3*	1
85	0*	0	1*	2	2*	1
86	0*	0	0*	1	2*	1
88	0*	0	0*	0	1*	0
91	0*	0	0*	1	3*	1
96	2*	0	0*	1	2*	1

* = Sparse Data

Table 84

Frequency Distribution of Wait Days by Appt Day of Week

Wait Days	Monday		Tuesday		Wednesday		Thursday		Friday	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
None	20	27	28	30	12	8	26	27	34	27
1-30 Days	117	147	182	167	72	46	136	149	152	150
31-90 Days	216	197	217	223	43	61	209	199	196	201
91+ Days	131	114	122	129	23	35	119	115	114	116

Table 85

Frequency Distribution of Wait Days by Marital Status

Wait Days	Widowed		Separated		Divorced		Single		Married	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
None	3*	4*	5*	3*	20	17	50	50	42	46
1-30 Days	18	20	17	16	71	93	276	275	277	255
31-90 Days	24	26	21	22	146	124	366	368	324	341
91+ Days	20	15	10	12	70	72	213	121	196	197

* = Sparse Data

Table 86

Frequency Distribution of Wait Days by Prev Visits NOS

	Yes		NO	
Wait Days	Frequency	Expected Frequency	Frequency	Expected Frequency
None	0*	6	120	114
1-30 Days	66	34	593	625
31-90 Days	41	46	840	835
91+ Days	6	27	503	482

* = Sparse Data

Table 87

Frequency Distribution of Season by Appt Day of Week

Season	Monday		Tuesday		Wednesday		Thursday		Friday	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Winter	65	64	85	73	20	20	55	65	62	66
Spring	351	346	369	392	89	107	364	350	376	354
Summer	15	19	26	22	9	6	23	19	13	20
Fall	53	55	69	63	32	17	48	56	45	56

Table 88

Frequency Distribution of Season by Wait Days

Season	None		1-30 Days		31-90 Days		91+ Days	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Winter	16	16	108	87	85	117	78	67
Spring	87	86	431	471	657	629	374	364
Summer	4*	5	40	26	36	35	6	20
Fall	13	14	80	75	103	100	51	58

* = Sparse Data

Table 89

Frequency Distribution of Season by Marital Status

Season	Widowed		Separated		Divorced		Single		Married	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Winter	13	9	9	7	41	41	109	120	115	111
Spring	41	46	36	38	242	219	651	646	579	599
Summer	3*	3*	4*	2*	3*	12	38	36	38	34
Fall	8	7	4	6	21	35	107	103	107	96

* = Sparse Data

Table 90

Frequency Distribution of Season by Prev Visit NOS

Season	Yes		No	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Winter	18	15	269	272
Spring	55	81	1494	1468
Summer	20	4*	66	82
Fall	20	13	227	234

* = Sparse Data

Table 91

Frequency Distribution of Same Payer by Season

Same Payer	Winter		Spring		Summer		Fall	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
No	19	34	127	129	12	7	23	21
Yes	268	263	1422	1420	74	79	224	226

Table 92

Frequency Distribution of Same Payer by Wait Days

Same Payer	None		1-30 Days		31-90 Days		91= Days	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
No	7	10	55	55	69	74	50	42
Yes	113	110	604	604	812	807	459	467

Table 93

Frequency Distribution of Same Payer by Marital Status

Same Payer	Widowed		Separated		Divorced		Single		Married	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
No	0*	5	7	4*	17	26	73	76	84	70
Yes	65	60	46	49	290	281	832	829	755	769

* = Sparse Data

Table 94

Frequency Distribution of Same Payer by Day of the Week

Same Payer	Monday		Tuesday		Wednesday		Thursday		Friday	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
No	50	40	56	46	7	13	26	41	42	41
Yes	434	444	493	503	143	137	464	449	454	455

Table 95

Frequency Distribution of Same Payer by Previous Visit NOS

Same Payer	Yes-Prevvisitsnos		No-Previsitnos	
	Frequency	Expected Frequency	Frequency	Expected Frequency
No	2*	9	179	172
Yes	11	104	1877	1884

* = Sparse Data

Table 96

Frequency Distribution of Marital Status by Previous Visit NOS

Marital Status	Yes-Prevvisitsnos		No-Previsitnos	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Widowed	1*	3	64	62
Separated	4*	3	49	50
Divorce	15	16	292	291
Single	54	47	851	858
Married	39	44	800	795

* = Sparse Data

Table 97

Frequency Distribution of Day of Week by Marital Status

Day of Week	Widowed		Separated		Divorced		Single		Married	
	Freq	Exp Freq	Freq	Exp Freq	Freq	Exp Freq	Freq	Exp Freq	Freq	Exp Freq
Mon	9	15	13	12	62	69	223	202	177	187
Tues	25	16	16	13	74	78	224	229	210	212
Wed	6	4*	2*	4*	17	21	64	63	61	58
Thurs	15	15	9	12	74	69	203	204	189	190
Fri	10	15	13	12	80	70	191	207	202	192

* = Sparse Data; Freq = Frequency; Exp Freq = Expected Frequency

Table 98

Frequency Distribution of Provider by Same Payer

Provider	No		Yes	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	11	22	257	246
Resident	115	91	976	1000
Attending	55	68	755	742

Table 100

Frequency Distribution of Provider by Season

Provider	Winter		Spring		Summer		Fall	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	41	35	171	191	15	11	41	30
Resident	121	144	836	779	34	43	100	124
Attending	125	107	542	578	37	32	106	92

Table 99

Frequency Distribution of Appt. Day of Week by Prev Visit NOS

Day of Week	Yes		No	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Monday	27	25	457	459
Tuesday	30	29	519	520
Wednesday	11	8	139	142
Thursday	23	26	467	464
Friday	22	26	474	470

Table 101

Frequency Distribution of Provider by Wait Days

Provider	None		1-30 Days		31-90 Days		91+ Days	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	17	15	181	81	44	109	26	63
Resident	64	60	216	331	509	443	302	256
Attending	39	45	246	246	338	329	181	190

Table 102

Frequency Distribution of Provider by Day of Week

Provider	Monday		Tuesday		Wednesday		Thursday		Friday	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	74	60	70	68	46	19	39	60	39	61
Resident	230	243	255	276	47	75	212	246	347	249
Attending	180	181	224	205	57	56	239	183	110	185

Table 103

Frequency Distribution of Provider by Marital Status

Provider	Widowed		Separated		Divorce		Single		Married	
	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	7	8	4*	6	29	38	113	112	115	104
Resident	27	33	32	27	181	154	449	445	402	422
Attending	31	24	17	20	97	115	343	338	322	313

* = Sparse Data

Table 104

Frequency Distribution of Provider by Prev Visit NOS

Provider	Yes		No	
	Frequency	Expected Frequency	Frequency	Expected Frequency
Counselor	17	14	251	254
Resident	56	57	1035	1034
Attending	40	42	770	768

Bibliography:

1. Centorrino, F., et al., *Factors associated with noncompliance with psychiatric outpatient visits.* Psychiatric Services, 2001. **52**(3): p. 378-80.

2. McDonnell, P.J. and M.R. Jacobs, *Hospital admissions resulting from preventable adverse drug reactions.[see comment].* Annals of Pharmacotherapy, 2002. **36**(9): p. 1331-6.

3. Schiff, G.D., et al., *Decompensated heart failure: symptoms, patterns of onset, and contributing factors.[see comment].* American Journal of Medicine, 2003. **114**(8): p. 625-30.

4. Senst, B.L., et al., *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): p. 1126-32.

5. Rodgers, P.T. and D.M. Ruffin, *Medication nonadherence: Part II--A pilot study in patients with congestive heart failure.* Managed Care Interface, 1998. **11**(9): p. 67-9.

6. Misdrahi, D., et al., *Compliance in schizophrenia: predictive factors, therapeutical considerations and research implications.* Encephale, 2002. **28**(3 Pt 1): p. 266-72.

7. Compton, M.T., et al., *Predictors of missed first appointments at community mental health centers after psychiatric hospitalization.* Psychiatric Services, 2006. **57**(4): p. 531-7.

8. Katon, W.J., *Clinical and health services relationships between major depression, depressive symptoms, and general medical illness.* Biological Psychiatry, 2003. **54**(3): p. 216-26.

9. Burkhart, P.V. and E. Sabate, *Adherence to long-term therapies: evidence for action.* Journal of Nursing Scholarship, 2003. **35**(3): p. 207.

10. Rae, H. *No-show hospital patients cost £20m in 12 months*. 2008 [cited 2008 09/10]; Available from: http://www.chroniclelive.co.uk/north-east-news/todays-evening-chronicle/2008/05/30/no-show-hospital-patients-cost-20m-in-12-months-72703-20997799/.

11. Puccia, S.V., *Can we schedule smarter? Is there a pattern of time-dependent variables in missed psychotherapy appointments?* 2004, Puccia, Sarah V : Chicago School Of Professional Psychology, US.

12. Rodolfa, E.R., Rapaport, R., & Lee, V. E. , *Variables related to premature terminations in a university counseling service.* Journal of Counseling Psychology, 1983. **30**: p. 87–90.

13. Shonick, W. and B.W. Klein, *An approach to reducing the adverse effects of broken appointments in primary care systems: development of a decision rule based on estimated conditional probabilities.* Medical Care, 1977. **15**(5): p. 419-29.

14. Ciechanowski, P., et al., *Where is the patient? The association of psychosocial factors and missed primary care appointments in patients with diabetes.* General Hospital Psychiatry, 2006. **28**(1): p. 9-17.

15. Luppa, M., et al., *Cost-of-illness studies of depression: a systematic review.* Journal of Affective Disorders, 2007. **98**(1-2): p. 29-43.

16. Moore, C.G., P. Wilson-Witherspoon, and J.C. Probst, *Time and money: effects of no-shows at a family practice residency clinic.* Family Medicine, 2001. **33**(7): p. 522-7.

17. Simon, G.E., M. VonKorff, and W. Barlow, *Health care costs of primary care patients with recognized depression.* Archives of General Psychiatry, 1995. **52**(10): p. 850-6.

18. World Health Organization, *The World Health Report: 2001- Mental Health: New Undrstanding, New hope.* 2001, World Health Organization: Geneva Switzerland.

19. Bean, A.G. and J. Talaga, *Predicting appointment breaking.* Journal of Health Care Marketing, 1995. **15**(1): p. 29-34.

20. Cashman, S.B., et al., *Patient health status and appointment keeping in an urban community health center.* Journal of Health Care for the Poor & Underserved, 2004. **15**(3): p. 474-88.

21. Dove, H.G. and K.C. Schneider, *The usefulness of patients' individual characteristics in predicting no-shows in outpatient clinics.* Medical Care, 1981. **19**(7): p. 734-40.

22. Izard, T., *Managing the habitual no-show patient.* Family Practice Management, 2005. **12**(2): p. 65-6.

23. Kruse, G.R. and B.M. Rohland, *Factors associated with attendance at a first appointment after discharge from a psychiatric hospital.* Psychiatric Services, 2002. **53**(4): p. 473-6.

24. Mallard, S.D., et al., *Same-day scheduling in a public health clinic: a pilot study.* Journal of Public Health Management & Practice, 2004. **10**(2): p. 148-55.

25. Yale Medical Group. *YMG Practice Standards*. Available from: www.med.yale.edu/yfp/standards/standards.html.

26. Oppenheim, G.L., J.J. Bergman, and E.C. English, *Failed appointments: a review.* Journal of Family Practice, 1979. **8**(4): p. 789-96.

27. Swenson, T.R. and G. Pekarik, *Interventions for reducing missed initial appointments at a community mental health center.* Community Mental Health Journal, 1988. **24**(3): p. 205-18.

28. Vikander, T., et al., *New-patient no-shows in an urban family practice center: analysis and intervention.* Journal of Family Practice, 1986. **22**(3): p. 263-8.

29. Belardi, F.W., S. Craig, FW, *A controlled trial of an advanced access appointment system in a residency family medicine clinic.* Family Medicine, 2004. **36**(5): p. 341-345.

30. Hashim, M.J., P. Franks, and K. Fiscella, *Effectiveness of telephone reminders in improving rate of appointments kept at an outpatient clinic: a randomized controlled trial.* Journal of the American Board of Family Practice, 2001. **14**(3): p. 193-6.

31. Johnson, B.J., J.W. Mold, and J.M. Pontious, *Reduction and management of no-shows by family medicine residency practice exemplars.* Annals of Family Medicine, 2007. **5**(6): p. 534-9.

32. Kennedy, J.G. and J.T. Hsu, *Implementation of an open access scheduling system in a residency training program.* Family Medicine, 2003. **35**(9): p. 666-70.

33. Martin, C., T. Perfect, and G. Mantle, *Non-attendance in primary care: the views of patients and practices on its causes, impact and solutions.* Family Practice, 2005. **22**(6): p. 638-43.

34. Schnitzer, P.K., *"They don't come in!" Stories told, lessons taught about poor families in therapy.* American Journal of Orthopsychiatry, 1996. **66**(4): p. 572-82.

35. Turner, B.J. and F.M. Hecht, *Improving on a coin toss to predict patient adherence to medications.[see comment][comment].* Annals of Internal Medicine, 2001. **134**(10): p. 1004-6.

36. Kimmel, P.L., et al., *Behavioral compliance with dialysis prescription in hemodialysis patients.* Journal of the American Society of Nephrology, 1995. **5**(10): p. 1826-34.

37. Glyngdal, P., P. Sorensen, and K. Kistrup, *Non-compliance in community psychiatry: failed appointments in the referral system to psychiatric outpatient treatment.* Nordic Journal of Psychiatry, 2002. **56**(2): p. 151-6.

38. Scholle, S.H., et al., *Addressing depression in obstetrics/gynecology practice.* General Hospital Psychiatry, 2003. **25**(2): p. 83-90.

39. DiMatteo, M., *Adherence to treatment*, in *Behavioral medicine in primary care: A practical guide*, C.J. Feldman MD, Editor. 1997, Appleton and Lange: Stamford. p. 136-140.

40. Smith, K.J., L.M. Subich, and C. Kalodner, *The transtheoretical model's stages and processes of change and their relation to premature termination.* Journal of Counseling Psychology, 1995. **42**(1): p. 34-39.

41. George, A. and G. Rubin, *Non-attendance in general practice: a systematic review and its implications for access to primary health care.* Fam. Pract., 2003. **20**(2): p. 178-184.

42. Deyo, R.A. and T.S. Inui, *Dropouts and broken appointments. A literature review and agenda for future research.* Medical Care, 1980. **18**(11): p. 1146-57.

43. Deane, F.P., *Improving attendance at intake in children's outpatient services of a community mental health centre.* Child: Care, Health & Development, 1991. **17**(2): p. 115-21.

44. El-Mallakh, R.S., et al., *Follow-up after inpatient psychiatric hospitalization with partial control of the system responsiveness variable.* Psychiatry, 2004. **67**(3): p. 294-8.

45. Eytan, A., Gex-Fabry, M., Ferrero, F., Bertschy, G. , *Missed appointments at outpatient psychiatric clinics in Geneva: A pilot study.* Schweizer Archive fur Neurologie und Psychiatrie 2004. **155**(3): p. 125-28.

46. Sisco, M. *How to calculate and convey the true cost of downtime*. 2002 [cited 2008 09/10]; Available from: http://articles.techrepublic.com.com/5100-10878_11-1038783.html.

47. Keller, M.B. and R.J. Boland, *Implications of failing to achieve successful long-term maintenance treatment of recurrent unipolar major depression.* Biological Psychiatry, 1998. **44**(5): p. 348-60.

48. Keller, M.B. and R.W. Shapiro, *"Double depression": superimposition of acute depressive episodes on chronic depressive disorders.* American Journal of Psychiatry, 1982. **139**(4): p. 438-42.

49. Ulmer, T., Troxler, C., *The Economic Cost of Missed Appointments and the Open Access System.* Community Health Scholars, 2004.

50. Gelb, B.D., *Telephoned appointment-scheduling by a physician's office: does it work?* Journal of Health Care Marketing, 1989. **9**(4): p. 61-3.

51. Morse, D.L., et al., *Waning effectiveness of mailed reminders on reducing broken appointments.* Pediatrics, 1981. **68**(6): p. 846-9.

52. Phillips, L., et al., *Preventive mental health care: Accessing the target population.* Australian and New Zealand Journal of Psychiatry, 1999. **33**(6): p. 912-917.

53. Nicholson, I.R., *Factors involved in failure to keep initial appointments with mental health professionals.* Hospital & Community Psychiatry, 1994. **45**(3): p. 276-8.

54. Mueller, T.I. and A.C. Leon, *Recovery, chronicity, and levels of psychopathology in major depression.* Psychiatric Clinics of North America, 1996. **19**(1): p. 85-102.

55. Levy, G., M.K. Zamacona, and W.J. Jusko, *Developing compliance instructions for drug labeling.* Clinical Pharmacology & Therapeutics, 2000. **68**(6): p. 586-91.

56. McEvoy, B., R. Nydegger, and G. Williams, *Factors related to patient compliance in the treatment of acne vulgaris.* International Journal of Dermatology, 2003. **42**(4): p. 274-80.

57. Morisky, D.E., L.W. Green, and D.M. Levine, *Concurrent and predictive validity of a self-reported measure of medication adherence.* Medical Care, 1986. **24**(1): p. 67-74.

58. Noble, L., *Doctor–patient communication and adherence to treatment*, in *Adherence to treatment in medical conditions.* , M.K. Myers LB, Editor. 1998, Harwood Academic Publishers. p. 51-82.

59. Rhee, M.K., et al., *Patient adherence improves glycemic control.* Diabetes Educator, 2005. **31**(2): p. 240-50.

60. Berg, J.S., et al., *Medication compliance: a healthcare problem.* Annals of Pharmacotherapy, 1993. **27**(9 Suppl): p. S1-24.

61. Burnier, M., *Long-term compliance with antihypertensive therapy: another facet of chronotherapeutics in hypertension.* Blood Pressure Monitoring, 2000. **5 Suppl 1**: p. S31-4.

62. Clarke, A.R., Schnieden, V. Hamilton, B. A., Dudley, A., M., Beard, J, Einfeld, S, L., Buss, R, Tobin, Mt, Knowles, M, Stevens, G, Gibbs, N.,, *Factors associated with treatment compliance in young people following an Emergency Department Presentation for delibrate self-harm.* Archives of Suicide Research, 2004 **8**(2): p. 147-152.

63. Horne, R., *Adherence to medication: A review of existing research.* , in *Adherence to treatment in medical conditions*, M.K. Meyers LB, Editor. 1998, Harwood Academic. p. 285-310.

64. Turner, B.J., et al., *Predicting adherence to colonoscopy or flexible sigmoidoscopy on the basis of physician appointment-keeping behavior.* Annals of Internal Medicine, 2004. **140**(7): p. 528-32.

65. Sewitch, M.J., et al., *Patient nonadherence to medication in inflammatory bowel disease.[see comment].* American Journal of Gastroenterology, 2003. **98**(7): p. 1535-44.

66. Hays RD, D.M., *Key issues and suggestions for patient compliance assessment: sources of information, focus of measures, and nature of response options.* Journal of Compliance in Health Care, 1987. **2**: p. 37-53.

67. Subramanian, J., S. Stidham, Jr., and C.J. Lautenbacher, *Airline yield management with overbooking, cancellations, and no-shows.* Transportation Science, 1999. **33**(2): p. 147-167.

68. Garrow, L.A.a.K., F.S., *Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information.* Journal of Air Transport Management, 2004. **10**(6): p. 401-411.

69. Gorin, T.B., W. White, M., *No-show forecasting: A blended cost-based PNR-adjusted approach.* Journal of Revenue and Pricing Management, 2006. **5**(3): p. 188-206.

70. Neuling, R.R., S. and Kalka, K., *New Approaches to origin and destination and no-show forecasting: excavating the passinger name treasure.* Journal of Revenue and Pricing Management, 2004. **3**(1): p. 62-73.

71. Sharda, R. and D. Delen, *Predicting box-office success of motion pictures with neural networks.* Expert Systems with Applications, 2006. **30**(2): p. 243-254.

72. Walls, W.D., *Modeling Movie Success When 'Nobody Knows Anything': Conditional Stable-Distribution Analysis Of Film Returns* Journal of Cultural Economics, 2005. **29**: p. 177-190.

73. Ekstrom, M.A., H.C. Bjornsson, and C.I. Nass, *A Reputation Mechanism for Business-to-Business Electronic Commerce That Accounts for Rater Credibility.* Journal of Organizational Computing and Electronic Commerce, 2005. **15**(1): p. 1 - 18.

74. Lifang, P., C. Zhong, and L. Qi. *Model and method for evaluating creditability of C2C electronic trade*. in *Proceedings of the ACM Conference on Electronic Commerce Proc. of the 8th Int. Conf. on Electronic Commerce 2006 - The New E-Commerce: Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet*. 2006.

75. Lawrence, R.H., SJ. Cherrier, J. *Passenger-Based Predictive Modleing of Airline No-show Rates*. in *Conference on Knowlwdge Discovery in Data*. 2003. Washington DC: ACM.

76. Hayden, K. (2006) *Aviacsa Becomes First Airline in Mexinco to Select Sabre AirMax Revenue Manager*. PHX-Coroporate.

77. Netessine, S.S., R., *Introduction to the Theory and Practice of Yield Management.* INFORMS Transactions on Education, 2002. **3**(1): p. 34-44.

78. Dennis, C.L. and L. Chung-Lee, *Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review.* Birth, 2006. **33**(4): p. 323-31.

79. McIvor, R., E. Ek, and J. Carson, *Non-attendance rates among patients attending different grades of psychiatrist and a clinical psychologist within a community mental health clinic.* Psychiatr Bull, 2004. **28**(1): p. 5-7.

80. Daggy, J., Sands, L., Lawley, M., Willis, D. *Modeling No-Show Behavior in a Midwestern VA Medical Center*. [cited 2008 09/10].

81. Robinson, J.W. and D.L. Roter, *Psychosocial problem disclosure by primary care patients.* Social Science & Medicine, 1999. **48**(10): p. 1353-62.

82. Roter, D.L., et al., *Effectiveness of interventions to improve patient compliance: a meta-analysis.* Medical Care, 1998. **36**(8): p. 1138-61.

83. Carrion, P.G., et al., *Compliance with clinic attendance by outpatients with schizophrenia.* Hospital & Community Psychiatry, 1993. **44**(8): p. 764-7.

84. Fava, M., et al., *Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study.* Psychiatric Clinics of North America. **26**(2): p. 457-94.

85. Mojtabai, R., M. Olfson, and D. Mechanic, *Perceived need and help-seeking in adults with mood, anxiety, or substance use disorders.* Archives of General Psychiatry, 2002. **59**(1): p. 77-84.

86. Amankwaa, L.C., *Postpartum depression among African-American women.* Issues in Mental Health Nursing, 2003. **24**(3): p. 297-316.

87. Appel, H.B., *Depression among minority women during pregnancy and postpartum.* The Sciences and Engineering 2005. **66**((3-B)): p. 1420.

88. Freeman, G. and P. Hjortdahl, *What future for continuity of care in general practice?* BMJ, 1997. **314**(7098): p. 1870-3.

89. Harpole Jr, D.H., *Prognostic Modeling in Early Stage Lung Cancer: An Evolving Process from Histopathology to Genomics.* Thoracic Surgery Clinics, 2007. **17**(2): p. 167-173.

90. Wyatt, J., Altman, D. (1995) *Commentary: Prognostic models: clinically useful or quickly forgotten?* , 1539-1541.

91. Benway, C.B., V. Hamrin, and T.J. McMahon, *Initial appointment nonattendance in child and family mental health clinics.* American Journal of Orthopsychiatry, 2003. **73**(4): p. 419-28.

92. Hartmann, A., Schulgen, G., Olschewski, M., & Herzog, T. , *Modeling psychotherapy outcome as an event in time: An application of multistage analysis. .* Journal of Consulting and Clinical Psychology, 1997. **65**: p. 262–268.

93. Horsley, B.P., et al., *Appointment keeping behavior of Medicaid vs non-Medicaid orthodontic patients.* American Journal of Orthodontics & Dentofacial Orthopedics, 2007. **132**(1): p. 49-53.

94. Lacy, N.L., et al., *Why we don't come: patient perceptions on no-shows.* Annals of Family Medicine, 2004. **2**(6): p. 541-5.

95. Grunebaum, M.L., P. Callahan, M. Leon,A. Olfson, M and Portera,L., *Predictors of missed appointment for psychiatric consultations in a primary care clinic.* Psychiatric Services, 1996. **47**(8): p. 848-852.

96. Otero, J., et al., *Factors associated with missed initial psychiatric visits.* Actas Espanolas de Psiquiatria, 2001. **29**(3): p. 153-158.

97. Meyer, W.S., *Why They Don't Come Back: A Clinical Perspective on the No-Show Client.* Clinical Social Work Journal, 2001. **29**(4): p. 325-339.

98. Bigby, J., et al., *Appointment reminders to reduce no-show rates. A stratified analysis of their cost-effectiveness.* JAMA, 1983. **250**(13): p. 1742-5.

99. Agras, W.S., *Understanding compliance with the medical regimen: the scope of the problem and a theoretical perspective.* Arthritis Care & Research, 1989. **2**(3): p. S2-7.

100. Richmond, R., *Discriminating variables among psychotherapy dropouts from a psychological training clinic.* Professional Psychology: Research and Practice, 1992. **23**(2): p. 123-130.

101. Olkin, R. and R. Lemle, *Increasing attendance in an outpatient alcoholism clinic: a comparison of two intake procedures.* Journal of Studies on Alcohol, 1984. **45**(5): p. 465-8.

102. Gans, J.S., & Counselman, E. F. , *The missed session: A neglected aspect of psychodynamic psychotherapy.* Psychotherapy: Theory, Research, Practice, Training, , 1996. **33**(1): p. 43–50.

103. Jhanjee, I., et al., *Parents' health and demographic characteristics predict noncompliance with well-child visits.* Journal of the American Board of Family Practice, 2004. **17**(5): p. 324-31.

104. Klein, E.B., Stone, W. N., Hicks, M. H., & Pritchard, I. L, *Understanding dropouts.* Journal of Mental Health Counseling, , 2003. **25**: p. 89–100.

105. Livianos-Aldana, L., et al., *Patients who miss initial appointments in community psychiatry? A Spanish community analysis.* International Journal of Social Psychiatry, 1999. **45**(3): p. 198-206.

106. Weighill, V.E., J. Hodge, and D.F. Peck, *Keeping appointments with clinical psychologists.* British Journal of Clinical Psychology, 1983. **22**(Pt 2): p. 143-4.

107. Hertz, P. and P.L. Stamps, *Appointment-keeping behavior re-evaluated.* American Journal of Public Health, 1977. **67**(11): p. 1033-6.

108. Goldman, L., et al., *A multivariate approach to the prediction of no-show behavior in a primary care center.* Archives of Internal Medicine, 1982. **142**(3): p. 563-7.

109. Bean, A.G. and J. Talaga, *Appointment breaking: causes and solutions.* Journal of Health Care Marketing, 1992. **12**(4): p. 14-25.

110. Catz, S.L., et al., *Predictors of outpatient medical appointment attendance among persons with HIV.* AIDS Care, 1999. **11**(3): p. 361-73.

111. Acri Cavaleri, M., *Understanding barriers to mental health care for poor urban women.* Humanities and Social Sciences 2005. **65**((11-A)): p. 43-46.

112. Cawley, M.E. and F.M. Stevens, *Non-attendance at outpatient clinics at the Regional Hospital, Galway, Ireland.* Social Science & Medicine, 1987. **25**(11): p. 1189-96.

113. Albaladejo Monreal E, A.M.A., Jiménez Martínez JM, López-Picazo JJ, Martínez López J, del Olmo Fernández A., *Scheduled consultation in primary care. Analysis of several relevant factors]. .* Aten Primaria, 1990. **7(4)**(Apr): p. 283-8.

114. Logsdon, M.C., et al., *Raising the awareness of primary care providers about postpartum depression.* Issues in Mental Health Nursing, 2006. **27**(1): p. 59-73.

115. Murray, M. and C. Tantau, *Same-day appointments: exploding the access paradigm.* Family Practice Management, 2000. **7**(8): p. 45-50.

116. Smith, C.M. and B.P. Yawn, *Factors associated with appointment keeping in a family practice residency clinic.* Journal of Family Practice, 1994. **38**(1): p. 25-9.

117. Baruch, G., A. Gerber, and P. Fearon, *Adolescents who drop out of psychotherapy at a community-based psychotherapy centre: a preliminary investigation of the characteristics of early drop-outs, late drop-outs and those who continue treatment.* British Journal of Medical Psychology, 1998. **71**(Pt 3): p. 233-45.

118. Carpenter, P.J., et al., *Who keeps the first outpatient appointment?* American Journal of Psychiatry, 1981. **138**(1): p. 102-5.

119. Margolis, K.L., et al., *Predictors of failure to attend scheduled mammography appointments at a public teaching hospital.* Journal of General Internal Medicine, 1993. **8**(11): p. 602-5.

120. Gruzd, D.C., C.L. Shear, and W.M. Rodney, *Determinants of no-show appointment behavior: the utility of multivariate analysis.* Family Medicine, 1986. **18**(4): p. 217-20.

121. Swett, C., Jr. and J. Noones, *Factors associated with premature termination from outpatient treatment.[see comment].* Hospital & Community Psychiatry, 1989. **40**(9): p. 947-51.

122. Marcus, S.M., et al., *Depressive symptoms among pregnant women screened in obstetrics settings.[see comment].* Journal of Women's Health, 2003. **12**(4): p. 373-80.

123. Berghofer, G., et al., *Predictors of treatment discontinuity in outpatient mental health care.* Social Psychiatry & Psychiatric Epidemiology, 2002. **37**(6): p. 276-82.

124. Brown, K.A., et al., *Correlates of missed appointments in orofacial injury patients.* Oral Surgery Oral Medicine Oral Pathology Oral Radiology & Endodontics, 1999. **87**(4): p. 405-10.

125. Abe-Kim, J., D. Takeuchi, and W.C. Hwang, *Predictors of help seeking for emotional distress among Chinese Americans: family matters.* Journal of Consulting & Clinical Psychology, 2002. **70**(5): p. 1186-90.

126. Leong, F.T.L., Wagner, N. S., & Tata, S. P. , *Racial and ethnic variations in help-seeking attitudes*. Handbook of multicultural counseling ed. J.M.C. J. G. Ponterotto, L. A. Suzuki, & C. M. Alexander 1995, Thousand Oaks: Sage. 415–438.

127. Sue, S., N. Zane, and K. Young, eds. *Research on psychotherapy with culturally diverse populations*. 1994, John Wiley & Sons: Oxford, England. (1994). Handbook of psychotherapy and behavior change (4th ed.). (pp. 783-817). xvi, 864.

128. McMiller, W.P. and J.R. Weisz, *Help-seeking preceding mental health clinic intake among African-American, Latino, and Caucasian youths.[see comment].* Journal of the American Academy of Child & Adolescent Psychiatry, 1996. **35**(8): p. 1086-94.

129. Palacios, M., & Franco, J. N. , *Counseling Mexican-American women.* Journal of Multicultural Counseling and Development, , 1986. **14**: p. 124–131.

130. Paolillo, J.G. and T.W. Moore, *Appointment compliance behavior of community mental health patients: a discriminant analysis.* Community Mental Health Journal, 1984. **20**(2): p. 103-8.

131. Wierzbicki, M. and G. Pekarik, *A meta-analysis of psychotherapy dropout.* Professional Psychology: Research and Practice, 1993. **24**(2): p. 190-195.

132. Tidwell, R., *The "No-Show" Phenomenon and the Issue of Resistance Among African American Female Patients at an Urban Health Care Center.* Journal of Mental Health Counseling, 2004. **26**(1): p. 1-12.

133. Sparks, W.A., J.A. Daniels, and E. Johnson, *Relationship of referral source, race, and wait time on preintake attrition.* Professional Psychology: Research and Practice, 2003. **34**(5): p. 514-518.

134. Yonkers, K.A., et al., *Onset and persistence of postpartum depression in an inner-city maternal health clinic system.* American Journal of Psychiatry, 2001. **158**(11): p. 1856-63.

135. Sobey, W.S., *Barriers to postpartum depression prevention and treatment: a policy analysis.* Journal of Midwifery & Women's Health, 2002. **47**(5): p. 331-6.

136. Forrest, C.B., et al., *Specialty referral completion among primary care patients: results from the ASPN Referral Study.* Annals of Family Medicine, 2007. **5**(4): p. 361-7.

137. Mellins, C.A., et al., *Longitudinal study of mental health and psychosocial predictors of medical treatment adherence in mothers living with HIV disease.* AIDS Patient Care & Stds, 2003. **17**(8): p. 407-16.

138. Davies, B.R., S. Howells, and M. Jenkins, *Early detection and treatment of postnatal depression in primary care.* Journal of Advanced Nursing, 2003. **44**(3): p. 248-55.

139. Coleman, V., Morgan, MA., Zinberg, S., Schulkin J. , *Rates and predictors of depression treatment among pregnant women in hospital-affiliated obstetrics practices. .* American Jouranl of Obstetrics and Gynecology, 2003. **189**(1): p. 267-73.

140. Granboulan, V., et al., *Predictive factors of post-discharge follow-up care among adolescent suicide attempters.* Acta Psychiatrica Scandinavica, 2001. **104**(1): p. 31-6.

141. Petrou, S., et al., *Economic costs of post-natal depression in a high-risk British cohort.* British Journal of Psychiatry, 2002. **181**: p. 505-12.

142. Flynn, H.A., F.C. Blow, and S.M. Marcus, *Rates and predictors of depression treatment among pregnant women in hospital-affiliated obstetrics practices.* General Hospital Psychiatry, 2006. **28**(4): p. 289-95.

143. Cardone, I.A., et al., *Psychosocial assessment by phone for high-scoring patients taking the Edinburgh Postnatal Depression Scale: communication pathways and strategies.* Archives of Women's Mental Health, 2006. **9**(2): p. 87-94.

144. Sword, W.A., P.D. Krueger, and M.S. Watt, *Predictors of acceptance of a postpartum public health nurse home visit: findings from an Ontario survey.* Canadian Journal of Public Health. Revue Canadienne de Sante Publique, 2006. **97**(3): p. 191-6.

145. Jones, P.K., S.L. Jones, and J. Katz, *Improving compliance for asthmatic patients visiting the emergency department using a health belief model intervention.* Journal of Asthma, 1987. **24**(4): p. 199-206.

146. Sherbourne, C.D., et al., *Antecedents of adherence to medical recommendations: results from the Medical Outcomes Study.* Journal of Behavioral Medicine, 1992. **15**(5): p. 447-68.

147. Killaspy, H., et al., *Prospective controlled study of psychiatric out-patient non-attendance. Characteristics and outcome.[see comment].* British Journal of Psychiatry, 2000. **176**: p. 160-5.

148. Peeters, F.P. and H. Bayer, *'No-show' for initial screening at a community mental health centre: rate, reasons and further help-seeking.* Social Psychiatry & Psychiatric Epidemiology, 1999. **34**(6): p. 323-7.

149. Okano, T., et al., *Effectiveness of antenatal education about postnatal depression: A comparison of two groups of Japanese mothers.* Journal of Mental Health, 1998. **7**(2): p. 191-198.

150. Virji, A., *A study of patients attending without appointments in an urban general practice.* BMJ, 1990. **301**(6742): p. 22-6.

151. Little, B., et al., *The failed appointment.* Journal - Oklahoma State Medical Association, 1991. **84**(9): p. 455-8.

152. Campbell, B., D. Staley, and M. Matas, *Who misses appointments? An empirical analysis.* Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie, 1991. **36**(3): p. 223-5.

153. Benoit C., W.R., Bonfonti A., Nuernberger K. , *Social determinants of mental health disparities among new mothers.* Centres Excellence Women's Health Res Bull, 2006. **5**(1): p. 9-11.

154. Gallucci, G., W. Swartz, and F. Hackerman, *Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center.[see comment].* Psychiatric Services, 2005. **56**(3): p. 344-6.

155. Chen, A., *Noncompliance in community psychiatry: a review of clinical interventions.* Hospital & Community Psychiatry, 1991. **42**(3): p. 282-7.

156. Folkins, C., P. Hersch, and D. Dahlen, *Waiting time and no-show rate in a community mental health center.* American Journal of Community Psychology, 1980. **8**(1): p. 121-123.

157. Cavanaugh, R.M., Jr., *Utilizing the phone appointment for adolescent follow-up.* Clinical Pediatrics, 1990. **29**(6): p. 302-4.

158. Benjamin-Bauman, J., M.L. Reiss, and J.S. Bailey, *Increasing appointment keeping by reducing the call-appointment interval.* Journal of Applied Behavior Analysis, 1984. **17**(3): p. 295-301.

159. Quattlebaum, T.G., P.M. Darden, and J.B. Sperry, *Effectiveness of computer-generated appointment reminders.* Pediatrics, 1991. **88**(4): p. 801-5.

160. Freund, R.D., Russell, T. T., & Schweitzer, S. , *Influence of length of delay between intake session and initial counseling session on client perceptions of counselors and counseling outcomes.* Journal of Counseling Psychology, , 1991. **38**: p. 3-8.

161. Neal, R.D., et al., *Missed appointments in general practice: retrospective data analysis from four practices.[see comment].* British Journal of General Practice, 2001. **51**(471): p. 830-2.

162. Neal, R.D., et al., *Reasons for and consequences of missed appointments in general practice in the UK: questionnaire survey and prospective review of medical records.* BMC Family Practice, 2005. **6**: p. 47.

163. Pekarik, G. and K. Finney-Owen, *Outpatient clinic therapist attitudes and beliefs relevant to client dropout.* Community Mental Health Journal, 1987. **23**(2): p. 120-30.

164. Chisholm, D., et al., *Health services research into postnatal depression: results from a preliminary cross-cultural study.* British Journal of Psychiatry - Supplementum, 2004. **46**: p. s45-52.

165. Gunter-Hunt, G., K.J. Ferguson, and G.G. Bole, *Appointment-keeping behavior and patient satisfaction: implications for health professionals.* Patient Counselling & Health Education, 1982. **3**(4): p. 156-60.

166. Cohen, A.D., et al., *Nonattendance of adult otolaryngology patients for scheduled appointments.* Journal of Laryngology & Otology, 2007. **121**(3): p. 258-61.

167. McGuff, R., D. Gitlin, and M. Enderlin, *Clients and therapists confidence and attendance at planned individual therapy sessions.* Psychological Reports, 1996. **79**(2): p. 537-8.

168. Hixon, A.L., R.W. Chapman, and J. Nuovo, *Failure to keep clinic appointments: implications for residency education and productivity.* Family Medicine, 1999. **31**(9): p. 627-30.

169. Miller LS., B.B.C.A., *Improving the identification and treatment of postpartum depression in a managed care organization.* Journal of Clinical Outcomes Management, 2004. **11**(3): p. 157-61.

170. Feierabend, R.H., *Physician ability to predict appointment-keeping behavior of prenatal patients.* Journal of Family Practice, 1996. **42**(5): p. 482-6.

171. Brown, J.T., C.C. Fulkerson, and E.R. Delong, *The resident leaves the clinic: the effects of changing physicians on appointment-keeping behavior.* Journal of General Internal Medicine, 1986. **1**(2): p. 98-100.

172. Lanska, M.J., et al., *Effect of resident turnover on patients' appointment-keeping behavior in a primary care medical clinic.* Journal of General Internal Medicine, 1986. **1**(2): p. 101-3.

173. Weingarten, N., D.L. Meyer, and J.A. Schneid, *Failed appointments in residency practices: who misses them and what providers are most affected?* Journal of the American Board of Family Practice, 1997. **10**(6): p. 407-11.

174. Dietrich, A.J., et al., *Depression care attitudes and practices of newer obstetrician-gynecologists: a national survey.* American Journal of Obstetrics & Gynecology, 2003. **189**(1): p. 267-73.

175. Guse, C.E., et al., *The effect of exit-interview patient education on no-show rates at a family practice residency clinic.* Journal of the American Board of Family Practice, 2003. **16**(5): p. 399-404.

176. Rigatelli, M., G.M. Galeazzi, and G. Palmieri, *Consultation-liaison psychiatry in obstetrics and gynecology.* Journal of Psychosomatic Obstetrics & Gynecology, 2002. **23**(3): p. 165-72.

177. Bishop, G. and A.C. Brodkey, *Personal responsibility and physician responsibility--West Virginia's Medicaid plan.* New England Journal of Medicine, 2006. **355**(8): p. 756-8.

178. Sewitch, M.J., et al., *Measuring differences between patients' and physicians' health perceptions: the patient-physician discordance scale.* Journal of Behavioral Medicine, 2003. **26**(3): p. 245-64.

179. Taylor, R.B., et al., *Purpose of the medical encounter: identification and influence on process and outcome in 200 encounters in a model family practice center.* Journal of Family Practice, 1980. **10**(3): p. 495-500.

180. Solberg, V., et al., *Asian-American students' severity of problems and willingness to seek help from university counseling centers: Role of previous counseling experience, gender, and ethnicity.* Journal of Counseling Psychology, 1994. **41**(3): p. 275-279.

181. Miller, L.G., et al., *How well do clinicians estimate patients' adherence to combination antiretroviral therapy?* Journal of General Internal Medicine, 2002. **17**(1): p. 1-11.

182. Rothenbacher, D., G. Ruter, and H. Brenner, *Prognostic value of physicians' assessment of compliance regarding all-cause mortality in patients with type 2 diabetes: primary care follow-up study.* BMC Family Practice, 2006. **7**: p. 42.

183. Murri, R., et al., *Patient-reported and physician-estimated adherence to HAART: social and clinic center-related factors are associated with discordance.* Journal of General Internal Medicine, 2004. **19**(11): p. 1104-10.

184. Lasser, K.E., et al., *Missed appointment rates in primary care: the importance of site of care.* Journal of Health Care for the Poor & Underserved, 2005. **16**(3): p. 475-86.

185. Becker, M.H., et al., *The Health Belief Model and prediction of dietary compliance: a field experiment.* Journal of Health & Social Behavior, 1977. **18**(4): p. 348-66.

186. Nickerson, K.J., J.E. Helms, and F. Terrell, *Cultural mistrust, opinions about mental illness, and Black students' attitudes toward seeking psychological help from White counselors.* Journal of Counseling Psychology, 1994. **41**(3): p. 378-385.

187. Chesney, A.P., et al., *Physician-patient agreement on symptoms as a predictor of retention in outpatient care.* Hospital & Community Psychiatry, 1983. **34**(8): p. 737-9.

188. Koch, A. and L.S. Gillis, *Non-attendance of psychiatric outpatients.* South African Medical Journal. Suid-Afrikaanse Tydskrif Vir Geneeskunde, 1991. **80**(6): p. 289-91.

189. Dias-Vieira, C., *An analogue study of stigma, help-seeking attitudes, and symptom severity in postpartum depression.* The Sciences and Engineering, 2006. **67**(2-B): p. 1145.

190. Edwards E. , T.S., *A qualitative study of stigma among women suffering postnatal illness.* Journal of Mental Health Counseling,, 2005. **14**(5): p. 471-81.

191. Murray, L., et al., *Self-exclusion from health care in women at high risk for postpartum depression.* Journal of Public Health Medicine, 2003. **25**(2): p. 131-7.

192. Small, R., et al., *Missing voices: What women say and do about depression after childbirth.* Journal of Reproductive and Infant Psychology, 1994. **12**(2): p. 89-103.

193. Grote, N.K., et al., *Culturally Relevant Psychotherapy for Perinatal Depression in Low-Income Ob/Gyn Patients.* Clinical Social Work Journal, 2004. **32**(3): p. 327-347.

194. Sparr, L.F., M.C. Moffitt, and M.F. Ward, *Missed psychiatric appointments: who returns and who stays away.[see comment].* American Journal of Psychiatry, 1993. **150**(5): p. 801-5.

195. Kravitz, R.L., et al., *Recall of recommendations and adherence to advice among patients with chronic medical conditions.* Archives of Internal Medicine, 1993. **153**(16): p. 1869-78.

196. Nurmi, H., *Monotonicity and its cognates in the theory of choice.* Public Choice, 2004. **121**(1-2): p. 25-49.

197. Donovan, J.L. and D.R. Blake, *Patient non-compliance: deviance or reasoned decision-making?* Social Science & Medicine, 1992. **34**(5): p. 507-13.

198. Golin, C.E., M.R. DiMatteo, and L. Gelberg, *The role of patient participation in the doctor visit. Implications for adherence to diabetes care.* Diabetes Care, 1996. **19**(10): p. 1153-64.

199. Rice, J.M. and J.R. Lutzker, *Reducing noncompliance to follow-up appointment keeping at a family practice center.* Journal of Applied Behavior Analysis, 1984. **17**(3): p. 303-11.

200. Sledge, W.H., et al., *Effect of time-limited psychotherapy on patient dropout rates.[see comment].* American Journal of Psychiatry, 1990. **147**(10): p. 1341-7.

201. Snell, M.N., et al., *Predicting counseling center clients' response to counseling: A 1-year follow-up.* Journal of Counseling Psychology, 2001. **48**(4): p. 463-473.

202. Gerson, L.W., G. McCord, and S.L. Wiggins, *A strategy to increase appointment keeping in a pediatric clinic.* Journal of Community Health, 1986. **11**(2): p. 111-21.

203. Ferber, R., Piskie, R.. *Subjective Probabilities and Buying Intentions.* Review of Economics and Statistics,, 1965. **47**(August): p. 322-325.

204. Armstrong, K., A. Eisen, and B. Weber, *Assessing the risk of breast cancer.[see comment].* New England Journal of Medicine, 2000. **342**(8): p. 564-71.

205. Tice, J.A., et al., *Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population.* Breast Cancer Research & Treatment, 2005. **94**(2): p. 115-22.

206. Belobaba, P., *Application of a Probabilistic Decision Model to Airline Seat Inventory Control.* Operations Research, 1989. **37**(2).

207. Allison, P.D., *Event history analysis.* SAGE, 1984(Beverly Hills, CA).

208. Hueglin, C.V., F. *Data mining techniques to improve forecast accuracy in airline business*. in *The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2001.

209. McGill, G.V.R., G., *Revenue management research overview and prospect.* Transportation Science, 1999. **33**(2): p. 233-256.

210. Boyd, A., *Airline Alliance Revenue Management.* OR/MS Today, 1998. **25**(October).

211. Sawhney, M., Eliashberg,J, *A Parsimonious Model for Forecasting Gross Box office Revenues of Motion Pictures.* Marketing Science, 1996. **15**(2): p. 113-131.

212. Zufryden, F.S., *Linking advertising to box office performance of new film releases: A marketing planning model.* Journal of Advertising Research, 1996. **36**(4): p. 29-41.

213. Litman, B., Ahn,H.,, ed. *Predicting financial success of motion pictures*. The motion picture mega-industry, ed. B. Litman. 1998, Allyn & Bacon Publishing, Inc: Boston, MA.

214. Sochay, S., *Predicting the performance of motion pictures.* The Journal of Media Economics, 1994. **7**(4): p. 1-20.

215. Puig, C., *Has Warner Bros. Lost Its Way? Scripts, Budgets Blamed For Bad Film Year,.* USA Today, 1998. **January 26**: p. 1D.

216. Radas, S. and S.M. Shugan, *Seasonal marketing and timing new product introductions.* Journal of Marketing Research, 1998. **35**(3): p. 296-315.

217. Ramya, N. and C. Pradeep, *A Bayesian Model to Forecast New Product Performance in Domestic and International Markets.* Marketing Science, 1999. **18**(2): p. 115-136.

218. Ravid, S., *Information, blockbusters, and stars: A study of the film industry,.* Journal of Business, 1999. **72**(4): p. 463-492.

219. Eliashberg, J., Shugan,SM., *Film Critics: Influencers or Predictors?* Journal of Marketing Research, 1997. **61**(April): p. 68-78.

220. CreditInfocenter. *Credit scores and what influences them, according to Fair Isaac*. 1999 [cited 200808/19];Availablefrom: http://www.creditinfocenter.com/creditreports/scoring/scoringinfluences.shtml.

221. Jiang, Y., *Research on Credit Management about C to C Model in Domestic Auction Website.* University of Electronic Science and Technology of China, 2005(June).

222. Komegay, A., Schwab,MR., de Almeida,MC. *Credit Score Simulation*. 2006 [cited 2008 September 2008]; Patent Application].

223. Piao, C.A., J. Fang, M., *Study on credit evaluation method and algorithm for C2C E-commerce* IEEE DOI, 2007. **10.1109/ICEBE**: p. 77.

224. Shao, J., S., Li,U, , *Credit Evaluation System Research of Auction Website.* Statistics and Decision, 2006. **Feb**.

225. Chunhui, P., A. Jing, and F. Meiqi, *Study on Credit Evaluation Model and Algorithm for C2C E-Commerce*, in *Proceedings of the IEEE International Conference on e-Business Engineering*. 2007, IEEE Computer Society.

226. Page, R.C., et al., *Longitudinal validation of a risk calculator for periodontal disease.* Journal of Clinical Periodontology, 2003. **30**(9): p. 819-27.

227. Boyle, M. and M. Green, *Pressure sores in intensive care: defining their incidence and associated factors and assessing the utility of two pressure sore risk assessment tools.* Australian Critical Care, 2001. **14**(1): p. 24-30.

228. Bemmaor, A., *Predicting Behavior from Intention-to-Buy Measures: The Parametric Case.* Journal of Marketing Research, 1996. **32**(May): p. 176-191.

229. Granbois, D., Summers,JO . *Primary and Secondary Validity of Consumer Purchase Probabilities.* Journal of Consumer Research, 1975. **1**(March): p. 31-38.

230. Morrison, D., *Purchase Intentions and Purchase Behavior.* Journal of Marketing Research, 1979. **43**(Spring): p. 65-74.

231. Belk, R.W., ed. *Issues in the Intention-Behavior Discrepancy*. Research in Consumer Behavior, ed. J. Sheth. Vol. 1. 1985, JAI Press: Greenwich, CT. 1-34.

232. Marti, S., *Trust and Reputation in Peer-to-Peer Networks*. 2005, Stanford University.

233. Traupman, J.W., R. , *EM-trust: A Robust Reputation Algorithm for Peer-to-Peer Marketplaces.* Computer Science Division (EECS) University of California Berkeley, California, 2005. **July**: p. 395.

234. Jamieson, L., Bass,FM. , *Adjusting Stated Purchase Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods,.* Journal of Marketing Research, 1989. **26**(August): p. 336-45.

235. Soe-Tsyr, Y. and S. Hao, *A learning-enabled integrative trust model for e-markets.* Applied Artificial Intelligence, 2004. **18**(1): p. 69-95.

236. Durad, M., Cao,Y., Liehuang,Z., *A Novel Evidential Trust Evaluation Algorithm.* WSEAS Transactions on Computers,, 2006. **5**(June).

237. Manski, C.F., *The Use of Intentions Data to Predict Behavior: A Best Case Analysis,.* Journal of the American Statistical Association,, 1990. **85**(December): p. 934-40.

238. Morwitz, V.G. and D. Schmittlein, *Using segmentation to improve sales forecasts based on purchase intent: Which "intenders" actually buy?* Journal of Marketing Research, 1992. **29**(4): p. 391-405.

239. Tauber, E.M., *Predictive validity in consumer research.* Journal of Advertising Research, 1975. **15**(5): p. 59-64.

240. Urban, G.L., J.R. Jauser, and J.H. Roberts, *Prelaunch forecasting of new automobiles.* Management Science, 1990. **36**(4): p. 401-21.

241. Urban, G., Weinberg,BD, Hauser,JR., *Premarket Forecasting of Really-New Products.* Journal of Marketing Research, 1996. **60**(January): p. 47-60.

242. Infosino, W., *Forecasting New Product Sales From Likelihood of Purchase Ratings,.* Marketing Science, 1986. **5**(Fall): p. 372-384.

243. Juster, F., *Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design.* Journal of the American Statistical Association,, 1966. **61**(September): p. 658-96.

244. Kalwani, M.U. and A.J. Silk, *On the Reliability and Predictive Validity of Purchase Intention Measures.* Marketing Science, 1982. **1**(3): p. 243-286.

245. Krider, R., and Weinberg, CB., *Competitive dynamics and the introduction of new products: The motion picture timing game.* Journal of Marketing Research, 1998. **1**: p. 1-15.

246. White, A.A., et al., *Cause and effect analysis of closed claims in obstetrics and gynecology.* Obstetrics & Gynecology, 2005. **105**(5 Pt 1): p. 1031-8.

247. Steyerberg, E.W., et al., *Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets.* Medical Decision Making, 2001. **21**(1): p. 45-56.

248. Derksen, S. and H.J. Keselman, *Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables* British Journal of Mathematical and Statistical Psychology, 1992. **45**: p. 265-282.

249. Dougherty, J., Kohavi, R. & Sahami, M. *Supervised and Unsupervised Discretization of Continuous Features, In Proc*. in *Twelfth International Conference on Machine Learning*. 1995. Los Altos, CA.: Morgan Kaufmann.

250. Fisher, R., *The use of multiple measurements in taxonomic problems.* Eugen, 1936. **7**: p. 179-188.

251. Harrell, F.E., *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis* 2001: Springer-Verlag Inc (Berlin; New York)

252. Long, J.S., & Freese, J. , *Regression models for categorical dependent variables using Stata.* . 2001 College Station: Stata Press.

253. Song, D., R.G. Sands, and Y.L. Wong, *Utilization of mental health services by low-income pregnant and postpartum women on medical assistance.* Women & Health, 2004. **39**(1): p. 1-24.

254. Ferrone, C.R., et al., *Multivariate prognostic model for patients with thick cutaneous melanoma: importance of sentinel lymph node status.* Annals of Surgical Oncology, 2002. **9**(7): p. 637-45.

255. Barr, R.G., et al., *Impaired flow-mediated dilation is associated with low pulmonary function and emphysema in ex-smokers: the Emphysema and Cancer Action Project (EMCAP) Study.[see comment].* American Journal of Respiratory & Critical Care Medicine, 2007. **176**(12): p. 1200-7.

256. Bedeschi, E., et al., *Urban air pollution and respiratory emergency visits at pediatric unit, Reggio Emilia, Italy.* Journal of Toxicology & Environmental Health Part A, 2007. **70**(3-4): p. 261-5.

257. Wagner, B.D., et al., *Relationship of body mass index with outcomes after coronary artery bypass graft surgery.* Annals of Thoracic Surgery, 2007. **84**(1): p. 10-6.

258. Dominici, F., et al., *On the use of generalized additive models in time-series studies of air pollution and health.[see comment].* American Journal of Epidemiology, 2002. **156**(3): p. 193-203.

259. Lopez-Caudana, A.E., et al., *Predictors of bone mineral density in female workers in Morelos State, Mexico.* Archives of Medical Research, 2004. **35**(2): p. 172-80.

260. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology, 1982. **143**(1): p. 29-36.

261. Breiman, L., et al., *Classification and regression trees*. Cole Advanced Books & Software. 1984, Monterey, CA: Wadsworth & Brooks.

262. Breiman, L., *Random forests.* Machine Learning, 2001. **45**(1): p. 5-32.

263. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. , *Classification and regression trees*. 1984: Wadsworth Co.

264. Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection.* in *The fourteenth international joint conference on artificial intelligence (IJCAI)*. 1995. Montreal, Quebec, Canada: Morgan Kaufman Publishing.

265. Nguyen ND, F.S., Center JR, Eisman JA, Nguyen TV., *Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks.* Osteoporosis International, 2008. **19**(10): p. 1431-1444.

266. Principe, C., Euliano,NR., Lefebre, WC., *Neural and adaptive systems: Fundamentals through simulations,*. 2000, New York: John Wiley and Sons.

267. Sharda, R., *Neural Networks for the MS/OR Analyst: An Application Bibliography.* INTERFACES, 1994. **24**(2): p. 116-130.

268. White, B., *Starting a revolution in office-based care.* Family Practice Management, 2001. **8**(9): p. 29-35.

269. White, H., *Connectionist nonparametric regression. Multilayer feedforward networks can learn arbitrary mappings.* Neural Networks, 1990. **3**: p. 535-549.

270. University of Wailkato. *Weka 3: Data Mining Software in Java*. Available from: www.cs.waikato.ac.nz/ml/weka/.

271. Zhang, H. and B. Singer, *Recursive partitioning in the health sciences* 1999: Springer-Verlag Inc (Berlin; New York).

272. Se June, H., H. Jonathan, and N. Ramesh, *Ensemble Modeling Through Multiplicative Adjustment of Class Probability*, in *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02)*. 2002, IEEE Computer Society.

273. Liu, H., Hussain,F., Chew,TL, Dash,M. . , *Discretization An Enabling Technique,*. Data Mining and Knowledge Discovery, 2002. **6**: p. 393-423.

274. Bianca, Z. and E. Charles, *Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers*, in *Proceedings of the Eighteenth International Conference on Machine Learning*. 2001, Morgan Kaufmann Publishers Inc.

275. Apte, C.e.a., *A probabilistic estimation framework for predictive modeling analytics.* IBM Systems Journal, 2002. **41**(3): p. 438-448.

276. Morwitz, V.G., E. Johnson, and D. Schmittlein, *Does measuring intent change behavior?* Journal of Consumer Research, 1993. **20**(1): p. 46-61.

277. US Naval Observatory. *Universal Time-Seasons*. 2010; Available from: http://aa.usno.navy.mil/data/docs/EarthSeasons.php.

278. Shelley, D. and H.J. Keselman, *Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables* British Journal of Mathematical and Statistical Psychology, 1992. **45**: p. 265-282.

279. Beck, R.A., *Statistical learning from a regression perspective*. 1 ed. 2008: Springer.

280. Sharda, R., A. R., and M. H., E. *Forecasting Gate Receipts Using Neural Networks and Rough Sets,*. in *The proceedings of the International DSI Conference*. 2000. Athens, Greece.

281. Anthony, S. *YouNoodle: Better Innovation Through Algorithms?* 2008 [cited 2008 08/19]; Availablefrom: http://discussionleader.hbsp.com/anthony/2008/08/younoodle_innovation_through_a.html.

282. Chun, C., *E-commerce Creditability Research Based on B2B Model.* XiDian University, 2005. January: p. 48-52.

283. Logsdon, M.C., *Depression in adolescent girls: screening and treatment strategies for primary care providers.* Journal of the American Medical Womens Association, 2004. 59(2): p. 101-6.

284. Minnesota Home Buyer Assistance Program. *How Credit Scores Work, How a Score is Calculated*. [cited 2008 08/19].

285. Staff, B. *Factors that Affect Your Credit Score*. 2006 [cited 2008 September]; Factors in a credit score].

286. Eytan, A., M. Gex-Fabry, and G. Bertschy, *Missed Monday appointments.* Psychiatric Services, 2005. **56**(5): p. 613.

287. Alafaireet, P. Houghton, H., Petroski, G Gong, Y & Savage. G. *Determining the structure of psychiatric visit non-adherence*. Journal of Ambulatory Care Management, 2010, *33(2) 108-116*

288. Alafaireet, P., Houghton, H., Savage, G. Gong, Y. *Using Electronic Data Sources to Understand the Determinants of Psychiatric Visit Non-Adherence*. AMIA National Conference. San Francisco CA. Nov., 2009

289. Alafaireet, P., Houghton, H., Anderson, B., Savage, G. Gong, Y. *Going Postal- A Practical Use of Geo-Spatial Maps in Psychiatry*. AMIA National Conference. San Francisco CA . Nov., 2009

290. Alafaireet, P., Houghton, H., Anderson, B., Gong, Y. "*Going Postal- A Practical Use of Geo-spatial Mapping in Psychiatry.* 25th Annual Missouri Life Sciences Research Poster Session. April 15, 2009

291. Alafaireet, P., Houghton, H. *Get Your Share of 100 Billion Dollars in Lost Revenue- Leverage the Practice Tools You Have and Reduce No-Show Visits by 83%+.*" MGMA Annual Conference. Oct.2008 (Butterfield S (Feb 2009). Research reveals reasons underlying patient no-shows. ACP Internist, http://www.acpinternist.org/archives/2009/02/no-shows.htm.)

292. Alafaireet, P. Wiles, C. World Health Congress-Innovation and Technology. *Beyond Outcomes and Process Indicators: Benchmarking the Transformationa*l. December 2007

293. Alafaireet, P. Information Processing in the Service of Mankind and Health ICICT- *GUI Design Characteristics Needed for Physicians Use* IEEE CCC Code: 0-7803-9770-3/06 December 2006

VITA

Patricia E. Alafaireet is the Director of Applied Health Informatics in the Health Management and Informatics Department at University of Missouri's School of Medicine. She holds a clinical faculty appointment and teaches Enterprise Information Architecture. Her research interests include visit non-adherence prevention, GUI design aesthetics for physician use and the application of informatics tools to day-to-day clinic operations [289-293]. Her latest publications focus on the development of tools and processes needed to support the prediction of visit non-adherence [287,288].

Dr. Alafaireet has expertise the development and application of information and knowledge systems. This expertise has been honed by 27 years of management experience, much of it in health care and human services and international health care delivery. Dr. Alafaireet has designed a number of software applications, including one for medical practice reporting and has extensive research in the area of graphic user interface development. Other projects completed include evaluation of RFID technology for use in health care and the development of risk calculators. She also holds a Masters of Health Administration