Using Electronic Data Sources to Understand the Determinants of Psychiatric Visit Non-Adherence

Patricia E. Alafaireet, MHA1,3, Howard L. Houghton, MD1,2, Grant T. Savage, PhD1,3, Yang Gong, MD, PhD1, 3

1Department of Health Management & Informatics
2Department of Psychiatry
3MU Informatics Institute
University of Missouri, Columbia, MO

Abstract

Technology cannot be effectively used to drive improvements in health care quality and health care cost reduction until significantly more healthcare visits are attended. Visit attendance is often perceived as an intractable problem. This formative study identified and analyzed a set of visit adherence determinants to delineate a structure of adherence. The study is distinguished from previous work because it employs three types of determinants (socio-economic, diagnosis, and logistical), captured from an ADT system and an EMR, to predict visit non-adherence.

1. Introduction

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 adherence is insufficiently understood. Previous research into medical adherence (also called no-shows) focuses either on a specific set of demographic characteristics of a class of patients (race, social economic status, gender, age, etc.) or on supporting the utilization of a specific visit scheduling strategy (open access or overbooking, for example) to meet care standards. Little is known about the multiple determinants of adherence for a health care visit and less is known about how these determinants interact. This lacunae has fostered the mindset that visit non-adherence is an intractable problem.

Selecting visit non-adherence determinants is an essential step toward developing a prognostic algorithm to maximize visit adherence. Our focus on establishing the structure of visit adherence is the product of four years of data collection and analysis by a multi-disciplinary team of investigators, with expertise in ambulatory care management, health care informatics, psychiatry practice, human factors, and biostatistics within the context of adherence in a psychiatry practice. Psychiatry was selected because it typically functions with an eleven to nineteen percent non-adherence rate, and individuals with mental health conditions frequently consume two or more times the normal amounts of other medical services.

This study identified and analyzed the determinants best suited for inclusion in the proposed visit adherence structure. One significant difference in this effort that distinguishes it from its predecessors is its focus on three contributory classes of determinants (patient socio-economic factors, disease state [clinical diagnosis] and clinic/care provider logistical factors) as they are discussed in the relevant literatures. This multi-factor structure is hypothesized to provide better prognostic value than the use of any of the factors can singularly provide. Another significant difference is that this study focuses on determinants that can be successfully used at the visit level rather than the patient level.

This study draws upon work done not only in health care but also on similar types of work done in the airline6, 8, 15 (seat abandonment), hospitality19 (hotel room reservation abandonment), cinema 17, 20 (box office success prediction) and banking5, 11 (credit scores) industries. The peer-reviewed literature provided support for twenty five potentially useful determinants (see Table 1).

This research can create significant paradigm shift if it replaces the current “bad patient” or “bad schedule” mentality with an evidence based solution.
for minimizing non-adherence visits. The ultimate objective of this research is to develop a structure of adherence which, in turn, would allow for a decision support tool that health care providers would use prognostically to reduce the rate of visit non-adherence.

2. Methods

This formative study employed both quantitative and qualitative methods, and used data obtained retrospectively (under IRB approval) from the Admission, Discharge, and Transfer (ADT) system, the Electronic Medical Record (EMR), and from paper charts used by ambulatory care clinics at an academic medical practice. Three data sets were collected and 3 pilot studies were carried out. Biostatistical staff and clinical data administrators supported all of the studies. These three pilot 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.

Pilot 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. The visits were a complete data sample for that time frame. Data in Pilot 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. Data for Pilot Study 3 were obtained as a complete sample (every visit) from the Psychiatry clinic in FY 2006-2007. In each of the three studies, data regarding potential determinants were extracted from the ADT 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 either the electronic system or from paper charts.

Data collection for the first two pilot studies proved to be an intensely manual process, requiring counts of the number of visits and the review of more than 1000 paper charts. Several electronically obtained data elements had to be extensively checked for accuracy. Pilot Study 3 used only data obtained from electronic sources. Data completeness was above 90% for each of the determinants. When there was no data available, each field was filled as “NA” to allow for subsequent row-by-row omission during statistical analyses. 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 statistical analysis of Pilot Study 1 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.

Visits for patient in Pilot Study 2 were nested within a “patient” and were analyzed as a patient rather than as a visit. Statistical analysis for Pilot Study 2 (also using SAS) was a multinomial logistic regression using parameter significance tests, odds ratio, and analysis of deviance. The confidence interval was also set at 95%.

Statistical analysis of Pilot Study 3 was accomplished as follows. A reserved data sample of 2,000 visits (for future testing) and 10 randomly selected data samples (for a total of 11,310 visits) were created. The marginal distribution of each of the determinants was analyzed with histogram and bi-variant analysis for each of the ten data samples and forward facing step-wise linear regression (maintaining a 95% confidence interval).

The determinant selection criteria also included the statistical relevance of each and the results of an analysis of data quality and availability.

3. Results

As a result of the Pilot Studies 1 and 2 (2,200 visits from 2004-2005) 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 Pilot Study 2 indicated the discontinuation of the nested visit approach and the use of manual data gathering.

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.

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. Based on the results of the regression analysis, the proposed determinants appeared to fall naturally into three categories (high, medium and low) based on their apparent contribution to the occurrence of visit non-adherence and thus, their utility to the development of the adherence structure. Results are shown in Table 1.

4. Discussion and Conclusions

Results from the three pilot studies, suggest that the determinants of visit non-adherence are multifactorial in nature and that a degree of correlation exists among the drivers.

Statistical analysis carried out in these three pilot studies supported the use of patient travel distance, age, primary diagnosis, employment, type of payer, appointment time, patient marital status, provider type, and race in the development of a structure of visit non-adherence. The results also support using three classes of determinants (socio-economic, clinical diagnosis, and logistical). Through this formative study we have also developed a context sensitive process that can be followed in future work by this research team, and, hopefully, by others.

One distinct limitation of the pilot study is the relatively small sample size, which may have caused bias in the results

A primary limitation to this research is its use of data from a single provider organization’s patient population.

5. Future Work

This work is yet incomplete. The research team is enlarging 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. The study is continuing and investigating additional determinates such as the season in which the appointment occurs (fall, winter, etc.) and the length of time between psychiatry appointments. Additional future avenues of investigation include developing a determinant measuring the relationship between the first and second diagnosis and a strategy to take into account the effects of a somatic health diagnosis.

Other work underway includes the testing of several algorithms (GAIL, Fair Isaac and others) for their potential to be a part of the decision tools for reducing visit non-adherence.

The aim of this research was to develop a structure of adherence that would serve as the underpinning for decision support tools to be used by health providers and their staff to prospectively decrease the likelihood of a non-adherent visit. The development of those decision tools should improve both the quality of care and reduce its costs, while reducing human suffering.

6. Acknowledgement

We gratefully acknowledge the contribution to this work by Greg Petroski, PhD, bio-statistician at the University of Missouri.
## Table 1 - Determinate Selection

<table>
<thead>
<tr>
<th>Determinant</th>
<th>Source and Form</th>
<th>Literature support (# of Articles)</th>
<th>Pilot Study 1 (2,000 Visits)</th>
<th>Pilot Study 2 (200 Visits)</th>
<th>Pilot Study 3 (11, 310 Visits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>ADT*, Male, Female</td>
<td>139</td>
<td>0.0015</td>
<td>0.003161</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>ADT, Years of age</td>
<td>198</td>
<td>† 0.00356</td>
<td>3.35e-38</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td>ADT, Single, Widowed Married, Divorced, Separated</td>
<td>11</td>
<td>&lt;0.0001</td>
<td>0.57864</td>
<td>0.000826</td>
</tr>
<tr>
<td>Employment</td>
<td>ADT, None, Employed Student, Retired, Home-maker, Disabled</td>
<td>11</td>
<td>&lt;0.0001</td>
<td>0.092743</td>
<td>0.019332</td>
</tr>
<tr>
<td>Travel Distance</td>
<td>ADT,5 digit zip code</td>
<td>2</td>
<td>0.0027</td>
<td>0.69987</td>
<td>7.97e-06</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>ADT, Census Codes</td>
<td>126</td>
<td>&lt;0.0001</td>
<td>0.74626</td>
<td>2.95e-12</td>
</tr>
<tr>
<td>Payer Types</td>
<td>ADT, FSC Codes</td>
<td>36</td>
<td>&lt;0.0001</td>
<td>0.86776</td>
<td>0.001037</td>
</tr>
<tr>
<td>Relationship of Contact Person to Patient</td>
<td>ADT, Free text, patient specified field</td>
<td>6</td>
<td>&lt;0.0001</td>
<td>*</td>
<td>0.0021137</td>
</tr>
<tr>
<td>Primary Diagnosis</td>
<td>ADT,EMR*.ICD Code</td>
<td>88</td>
<td>0.5490</td>
<td>*</td>
<td>6.62e-17</td>
</tr>
<tr>
<td>Second Diagnosis</td>
<td>ADT,EMR ICD code</td>
<td>88</td>
<td>*</td>
<td>*</td>
<td>Δ</td>
</tr>
<tr>
<td>Wait Days for Appt.</td>
<td>ADT, Count of days until appointment</td>
<td>18</td>
<td>0.0248</td>
<td>0.77214</td>
<td>0.000256</td>
</tr>
<tr>
<td>Appt. Type</td>
<td>New, Return, New Therapy, Return Therapy</td>
<td>14</td>
<td>0.9147</td>
<td>*</td>
<td>0.394747</td>
</tr>
<tr>
<td>Appt. Time</td>
<td>Hour / minute</td>
<td>110</td>
<td>&lt;0.0001</td>
<td>0.64271</td>
<td>5.53e-185</td>
</tr>
<tr>
<td>Appt. Hour</td>
<td>Hour only</td>
<td>1 10</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Appointment Date</td>
<td>Calendar date</td>
<td>110</td>
<td>0.0073</td>
<td>0.04123</td>
<td>0.035281</td>
</tr>
<tr>
<td>Appt Day</td>
<td>Mon-Fri.</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>0.000232</td>
</tr>
<tr>
<td>Use of Counseling</td>
<td>Count by provider</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>0.001137</td>
</tr>
<tr>
<td>Number of Appts</td>
<td>Appointment Count</td>
<td>5</td>
<td>†</td>
<td>*</td>
<td>3.6e-06</td>
</tr>
<tr>
<td>#of Non-adherent Appts</td>
<td>Appointment Count</td>
<td>5</td>
<td>†</td>
<td>*</td>
<td>2.1e-188</td>
</tr>
<tr>
<td>#of Canceled Appts</td>
<td>Appointment Count</td>
<td>5</td>
<td>†</td>
<td>*</td>
<td>0.000366</td>
</tr>
<tr>
<td>Appt Maker</td>
<td>By maker initials</td>
<td>4</td>
<td>&lt;0.0001</td>
<td>*</td>
<td>0.271289</td>
</tr>
<tr>
<td>Type of Provider</td>
<td>Provider Name</td>
<td>24</td>
<td>&lt;0.0001</td>
<td>*</td>
<td>0.000151</td>
</tr>
<tr>
<td>Referral Source</td>
<td>Self, provider name</td>
<td>3</td>
<td>&lt;0.0001</td>
<td>*</td>
<td>†</td>
</tr>
<tr>
<td>Special Housing</td>
<td>Street Address</td>
<td>11</td>
<td>&lt;0.0001</td>
<td>*</td>
<td>†</td>
</tr>
</tbody>
</table>

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
References


