HVMA: Mitigating the Effects of Patient No-shows

*Project Team*
Kelli Crosby, Bryan Gillan, Artem Golyshkin, John Griffith, Amanda Price

*Project Advisor*
Prof. James Benneyan

*Abstract*
The objective of this project is to alleviate the negative impact of patient no-shows as measured by provider idle time and revenue loss at the department of Obstetrics and Gynecology (OB/GYN) in the Harvard Vanguard Medical Associates (HVMA) Quincy medical center. A logistic regression and digital simulation are employed to systematically optimize the scheduling system. The solution is an integration of appointment sequencing and strategic double-booking based on appointment type. This method is proven to effectively maximize revenue and provider utilization while minimizing patient wait time and doctor overtime.
**Need for Project**

This project is necessary in order to reduce wasted time and money caused by patient no-shows at the Quincy HVMA facility.

The need for the project is established by patient no-shows that occur on a regular basis (18.2% of appointments). A no-show appointment is defined as one in which a patient fails to show up for a scheduled appointment or cancels their appointment within 24 hours. No-shows result in undesired idle time which affects both the providers and patients. Idle time equates to missed opportunities to generate revenue for HVMA and patients lose the opportunity of being seen during this unoccupied appointment slot. Based on data from 2009, no-shows resulted in a loss of approximately $446,000 in revenue. Below is the calculation used to approximate revenue loss:

\[
2,230 \text{ (No Shows in 2009)} \times \$200 \text{ (approximate revenue loss per appointment)} = \$446,000
\]

**Design Project Objectives and Requirements**

Develop scheduling and overbooking approaches that will mitigate the negative effects of no-shows by improving provider utilization.

**Design Objectives**

The project objective is to develop strategic overbooking practices aimed to mitigate the negative effects of no shows. Considered overbooking solutions were tested with simulation models to determine the compliance with constraints listed in Table 1 of “Design Requirements”.

**Design Requirements**

Table 1 shows the customer’s desired design requirements/metrics of measurement for improving the effects of no-shows. Provider overtime and patient wait time should not exceed 30 minutes, while the primary factor of concern is increasing the overall utilization rate of the provider to greater than 90%.

<table>
<thead>
<tr>
<th>Design Requirements</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider Utilization</td>
<td>≥ 90%</td>
</tr>
<tr>
<td>Patient Wait Time</td>
<td>≤ 30 mins</td>
</tr>
<tr>
<td>Overtime</td>
<td>≤ 30 mins</td>
</tr>
</tbody>
</table>
Design Concepts considered

Three primary design concepts were considered: Wave scheduling, strategic double booking, and strategic compression.

Table 2: Considered Strategies

<table>
<thead>
<tr>
<th>Considered Strategies</th>
<th>Cons</th>
<th>Pros</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wave Scheduling</strong></td>
<td>- High patient wait time</td>
<td>- Accomodates variability in patient arrival time</td>
</tr>
<tr>
<td></td>
<td>- High provider over time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Low Utilization</td>
<td></td>
</tr>
<tr>
<td><strong>Strategic Double Booking</strong></td>
<td>- Increased patient wait time</td>
<td>- Improved provider utilization</td>
</tr>
<tr>
<td></td>
<td>- Increased provider over time</td>
<td></td>
</tr>
<tr>
<td><strong>Compression</strong></td>
<td>- High patient wait time</td>
<td>- Increases number of available appointment block per session</td>
</tr>
<tr>
<td></td>
<td>- High provider over time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Difficult to implement</td>
<td></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
<td>- Does not affect utilization rate</td>
<td>- Absorbs patient wait time caused by double booking</td>
</tr>
</tbody>
</table>

Research was performed to study all current clinical overbooking practices. Based on practicality of implementation and their potential to achieve desired results, methods were narrowed down to the three most effective ones – wave scheduling, compression, and double-booking. Sequencing appointments within a day based on show rates was also experimented with and combined with the above methods.

The primary overbooking approaches considered were:

1. **Wave Scheduling** - Wave scheduling offers scheduling of several patients for a fixed time slot and serving those patients in the order in which they arrive. This method allows for a ready supply of patients on hand and reduces provider underutilization caused by no-shows and late arrivals.

2. **Strategic Double Booking** - Employs scheduling two patients in one appointment slot. Obstetrics appointments are significantly more likely to show up and insurance companies pay a flat rate for the entire pregnancy, making a less significant impact on the revenue stream. Therefore, it was important to focus specifically on two gynecology appointments with the highest no-show rate (PHR and RGL), which will be referred to as problem appointments.

3. **Compression** – One method of compression used consisted of compressing the kept appointments $T$ to $S$, so that $T = D(1-p)$, where $D$ is the appointment length in minutes and $p$ is the no-show probability. For example, if an appointment is scheduled for 20 minutes with a 25% no-show rate, then appointments would be scheduled every $T = 20(0.75) = 15$ minutes. This method allows for more appointments to be fit into a session.

4. **Sequencing** – Involves various sequencing strategies such as scheduling appointment types by highest to lowest show rate, blocking problem appointments together, and arranging them high, low, high, low, etc. based on show rate.
Recommended Design Concept

**Design Description**

The recommended design integrates strategic overbooking & sequencing with the following rules for scheduling:

1. Sequence Appointment types from lowest no-show rate to highest (see Figure 1.)
2. Double-book RGL and PHR appointment types
3. Double-book up to 2 of these appointment types per session
4. Begin by double-booking last block in the session, followed by double-booking a block in the middle of the session.

**Analytical Investigations**

Logistic Regression - From the logistic regression it was determined that the solution should focus on the two problem appointments which have the largest impact on the overall no-show rate of 18.2%. These were return gynecology patient (RGL) and gynecology annual physical (PHR), which have a 25.3% and 27.9% no-show rate respectively.

Simulation - A digital simulation of one four hour session of one doctor’s schedule was created to evaluate each overbooking method. A number of variations and combinations were simulated (21 in total), e.g. double-booking in different sequences, to determine the overall best solution. By experimenting with double-booking 1, 2, 3, or 4 problem appointments, it was concluded that double-booking 2 appointments per session was optimal. Through experimentation with sequencing of problem appointments, it was determined that it is best to sequence appointment types throughout the session from lowest no-show rate and combine this with double-booking. It was also determined that it’s best to schedule one double-booked appointment at the end of the session and a second one in the middle because it had the least impact on patient wait times if both patients show up.

Implementation includes a scheduling algorithm specified above in the Design Description. The t-test method was used to determine the numbers of days it will take HVMA to validate this solution (see figure 3). Equation 1 was used to solve for N.
Based on Equation 1, it was determined that the solution must be tested for approximately 26 days.

**Equation 1:**

$$n_2 = \left( \frac{s_2}{\Delta X} - \frac{s_1}{\sqrt{n_1}} \right)^2$$

Where $p = 0.05$ (95% Confidence Interval) and degrees of freedom is equal to 2. The customer will also be provided with instructions for measurement of key metrics and control charts to verify process performance.

**Experimental Investigations**

Utility Rate - The 12 overbooking strategies (that met customer-specified design requirements) were compared against each other by calculating a 'utility value'. By combining all three rationalized measurements in Table 1 using weighted factors, a single 'utility value' is assigned for a more linear comparison. Based on order of importance, utilization rate was attributed to be 70%, patient wait time was assigned 25% and provider overtime was weighted as 5%. Based on the utility value, it was determined that double-booking two problem appointments and sequencing them in the middle and at the end of the session yielded the maximum utility rate.

Testing – The group spent a day at the Quincy facility to verify the accuracy of the simulation. It was found that the actual utilization rate was .75 as compared to the predicted .83, the actual patient wait time was 13 minutes as compared to 10, and the actual overtime was 0 just as predicted. Therefore, we concluded that our simulation was, in fact, valid.

**Key Advantages of Recommended Concept**

The advantages of the strategic double booking and sequencing design include ease of implementation, maximizing provider utilization while minimizing patient wait time and provider overtime, and the solution comes at no cost to HVMA.

With the projected 10% increase in provider utilization from 83% to 93%, the approximated revenue loss is reduced from $446,000 to $184,000 annually.
Financial Issues

Implementation costs are directly related to the amount of provider overtime generated due to overbooking. Implementation costs are directly related to the amount of provider overtime generated due to overbooking. It was determined that the cost of overtime for all desk personnel is $50 per hour. Other costs such as provider overtime, implementation and training costs are less tangible. Providers do not get paid directly for overtime since they are on a fixed salary. Implementation and training costs are negligible, since implementation does not require any substantial effort. Costs of losing patients due to excessive wait time are also negligible, since the solutions considered minimize the wait time.

Recommended Improvements

The final deliverable includes implementation instructions for optimizing scheduling along with non-engineering suggestions to improve patient no-show rates.

Currently, appointments are scheduled three to thirteen months out so implementation will not be possible until August 2010. Incorporating patient attendance history is recommended so that patients with poor attendance record are overbooked initially. Patients with poor attendance should also be contacted twice before the appointment to increase their likelihood of showing up. The proposed solution will remain valid as long as the average no-show rate is 18 ± 2% and ratio of overbooking to under booking costs stay between 1.25 and 1.75. If future data shows a substantial change in the show rates or costs change, the overbooking practice can be adjusted based on the overbooking table shown in Table 3.

Table 3: Sample section of final solution table showing how many patients should be overbooked per session based on the average no-show rate and the cost ratio. Co is the cost of overbooking one patient and Cu is the cost of under booking a single patient.