Testing the assumptions of outpatient healthcare appointment scheduling

Linda R. LaGanga
Mental Health of Denver

Stephen R. Lawrence (Stephen.Lawrence@colorado.edu)
University of Colorado at Boulder (Corresponding Author)


Abstract
In this paper, we investigate and test the assumptions of outpatient healthcare appointments scheduling. There is significant interest in appointment scheduling in both healthcare practice and among scholarly researchers, but there has been little systematic comparison of practice versus theory. Using mathematical models and empirical surveys of clinic practitioners, we quantitatively and qualitatively evaluate the assumptions frequently made in the domain of outpatient healthcare appointment scheduling literature. Between theory and practice, we compare and contrast objective functions, relevant costs and benefits, appointment scheduling methods, prevalence and handling of no-show patients, schedule construction, and open access scheduling.

Keywords: Appointment scheduling, healthcare, operations

Introduction
Interest in healthcare appointment scheduling has exploded in recent years (Gupta and Denton, 2008). In early 2012, an online search of scholarly articles that include the words “healthcare appointment scheduling” generated 3,010 returns for the past 5 years (2007-2011), with 667 returns in 2011 alone. Clearly there is significant scholarly interest in the topic of healthcare appointment scheduling – ideally, this scholarly interest translates to practical benefit for healthcare professions and administrators. But while there is great interest in healthcare appointment scheduling among scholars, to our knowledge there has been no investigation that systematically tests the common assumptions of healthcare appointment scheduling research against practice. In this paper, we analytically and empirically evaluate the assumptions frequently made in the domain of outpatient healthcare appointment scheduling literature.

The broad research question that we seek to answer is if the assumptions of scholarly outpatient appointment scheduling research accurately reflects the actual practice and needs of actual healthcare clinics. Specifically, we are investigating:

- Choice of Objective Function. What implicit objective function(s) do healthcare clinics use when scheduling appointments? Is the common objective to minimize
costs (Ho and Lau 1992); optimize multiple objectives (White et al., 2011); maximize utility (LaGanga and Lawrence, 2012); or something else?

- **Clinic Benefits and Costs.** How do clinics evaluate the benefits and costs of scheduling decisions? Are provider expenses variable (Robinson and Chen, 2010) or sunk (LaGanga and Lawrence, 2012)? How do clinics evaluate the cost of patient waiting? Are trade-offs between different benefits and costs made strictly on a monetary basis or with consideration of other subjective measures (LaGanga and Lawrence, 2012)?

- **Use and Structure of Overbooking.** Do clinics typically overbook appointments to mitigate patient no-shows? If yes, do they double-book appointment slots (Liu et al., 2010) or do they compress the duration of slots in order to schedule more patients (LaGanga and Lawrence, 2007)? If they do not overbook, what other techniques are employed to mitigate no-shows?

- **Schedule Construction.** How do clinics fill open appointment slots when requests are received? Are they filled in order from the start to the end of the day (Robinson and Chen, 2010)? Are they randomly filled according to patient preference? Or are they filled according to some optimal or preferred pattern (LaGanga and Lawrence, 2012)?

- **Open Access Scheduling and Walk-Ins.** Are clinics explicitly using open access (same day) scheduling (Qu et al., 2007), or do they informally “squeeze in” patients who need to be quickly served? Are walk-in patients an issue, and if they are, how are they handled (Cayirli et al., 2012)?

As is evident from the citations above, there is a wide range of assumptions in the scholarly literature regarding the structure of outpatient healthcare appointment scheduling. While we do not expect to find consistency or unanimity in the operations of different clinics, we anticipate that patterns and preferences among clinics can be discerned. Answering these and similar questions will serve to guide future scholarly research and to help insure that the results of such research will be of utility to the healthcare profession.

We do not propose or try to prove empirical or theory-based hypotheses regarding healthcare clinic appointment scheduling. Instead, our objective is to compare the actual practice of appointment scheduling with the assumptions made in the scholarly literature in order to better guide future research and, ideally, to improve clinic practice. To achieve this objective, we employ both empirical and analytic methods.

Empirically, we have interviewed providers and administrators from a cross-section of outpatient healthcare clinics in order to better understand the outpatient scheduling problem from their perspective. Informed by the outcome of these interviews, we constructed a pilot survey instrument that was administered to a small test group of relevant healthcare professionals. Pilot survey results were collected and analyzed, and are reported in this paper.

Analytically, we have evaluated both the structure and practice of healthcare appointment scheduling. For example, our results compare cost-minimization to utility-maximization objective functions and show when each is most appropriate. Other analytic and numerical results compare the efficacy of appointment double-booking to slot-compression, and open-access vs. traditional scheduling. These analytic and numerical results serve as useful counterpoints to our empirical results and will help to further guide our understanding of the healthcare appointment scheduling problem.
Mathematical Analysis and Results

In the mathematical analysis of this section, we assume that a health care clinic sees patients on an appointment basis during a service session of duration $D$ time units (see Table 1 for a summary of notation). A clinic session is a period of time (e.g., a morning, a day) during which the clinic is in continuous operation. Each appointment during a session is of fixed duration $d$ and the length of a session is designed to be an integer multiple of $d$ so that $D=Nd$, where $N$ is the number of appointment “slots” scheduled in a session. Patients are scheduled to arrive at the start of an appointment slot and are assumed to be punctual (Blanco White and Pike, 1964; Soriano, 1966). However, some patients may not appear for their appointments (are no-shows) with frequency $0 \leq \eta \leq 1$. Since patients must either arrive for an appointment or are no-shows, the show rate is $0 \leq \sigma \leq 1$, where $\sigma = 1 - \eta$. A schedule $S$ for a clinic session is a vector of the number of patients scheduled to arrive at the start of each appointment slot.

Table 1 – Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>Duration of a session</td>
</tr>
<tr>
<td>$d$</td>
<td>Duration of an appointment (deterministic)</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of appointment slots in a session</td>
</tr>
<tr>
<td>$S$</td>
<td>An appointment schedule for a clinic session</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Average no-show rate of for all scheduled appointments (0 $\geq \eta \geq 1$)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Average show rate of for all scheduled appointments (0 $\geq \sigma \geq 1$)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Marginal net benefit parameter of one additional patient</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Marginal cost parameter for the clinic working one time unit of overtime</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Marginal cost parameter for a single patient waiting one time unit</td>
</tr>
<tr>
<td>$\Pi(S)$</td>
<td>Net benefit function</td>
</tr>
<tr>
<td>$\Omega(S)$</td>
<td>Patient waiting time cost function</td>
</tr>
<tr>
<td>$T(S)$</td>
<td>Clinic overtime cost function</td>
</tr>
<tr>
<td>$I(S)$</td>
<td>Provider idle time cost function</td>
</tr>
<tr>
<td>$\Gamma(S)$</td>
<td>Gross benefit function, where $\Gamma(S) = \Pi(S) + I(S)$</td>
</tr>
<tr>
<td>$U(S)$</td>
<td>Schedule utility function, where $U(S) = \Pi(S) - \Omega(S) - T(S)$</td>
</tr>
<tr>
<td>$C(S)$</td>
<td>Schedule cost function, where $C(S) = \Omega(S) + T(S) + I(S)$</td>
</tr>
</tbody>
</table>

The appointment scheduling problem of the clinic is to create a feasible schedule $S$ of patient appointments for a clinic session so that its chosen objective function is optimized, as defined below.

Utility Maximization versus Cost Minimization

We first examine two predominate objective functions used in appointment scheduling research: cost minimization models (e.g., Ho and Lau, 1992; Robinson and Chen, 2010) and utility maximization models (e.g., LaGanga and Lawrence, 2007). Using the notation shown in Table 1, a typical utility $U(S)$ maximization model is the sum of expected net benefits $\Pi(S)$ less expected patient waiting costs $\Omega(S)$ and expected clinic overtime costs $T(S)$:

$$\text{Max } U(S) = \Pi(S) - \Omega(S) - T(S)$$  \hspace{1cm} (1)
In turn, a typical expected total cost $C(S)$ model minimizes the sum of expected patient waiting costs $\Omega(S)$, expected clinic overtime costs $\Upsilon(S)$, and expected provider idle costs $\Iota(S)$:

$$\text{Min } C(S) = \Omega(S) + \Upsilon(S) + \Iota(S)$$

(2)

An obvious question is which objective function is most appropriate for use with health care appointment scheduling. To investigate, we study the following conjecture:

**Conjecture 1:** Cost minimization objective functions are a special case of utility maximization objectives.

To illustrate, let $\Gamma(S) = \Pi(S) + \Iota(S)$ be the gross benefit of a schedule $S$ (net benefit plus the explicit or implicit cost of any provider idle time) so that $\Pi(S) = \Gamma(S) - \Iota(S)$. Then it is easy to show that $U(S) = C(S)$ for all $S$ if and only if $\Pi(S) = 0$ (or equivalently, is a constant), so that $C(S)$ is a special case of $U(S)$, which occurs when there is no benefit (loss) for seeing more (fewer) patients. The implications of this result are several:

- Use of a utility maximizing function is preferred to a cost minimizing function since it includes the benefits of serving patients; not just the costs of service.
- A cost minimization function does not provide insights or methods to determine the optimal number of patients to be scheduled, so this number must be estimated or determined by other means. Without a constraint of number of patients to schedule, the optimal solution to an unconstrained cost-minimization function is to not schedule any patients – an absurd result.
- A utility maximization objective both determines the optimal number of patients to book in clinic session as well as the optimal appointment schedule.

**Overbooking vs. Slot Compression**

Another important aspect of clinic appointment scheduling is how a clinic handles patients who do not appear for scheduled appointments. Numerous studies have shown that no-shows are an endemic problem for outpatient clinics with no-show rates varying between 3 and 50% for many clinics (Rust et al., 1995). Two methods for mitigating the negative impacts of patient no-shows is by overbooking appointment slots or by slot compression of time between appointments (LaGanga and Lawrence, 2007). Both of these methods schedule patients into a clinic session beyond the capacity. With overbooking, one or more normal appointment slots are double or triple booked. With slot compression, the length of appointment slots is reduced, which allows more patients to be scheduled during a session without double-booking. Both methods work with the expectation that some scheduled patients will not show, but when too many patients do show, the results are increased patient waiting times and increased clinic overtime.

The question naturally arises as to which of overbooking or slot-compression is the better method for mitigating the impact of no-shows. To help answer this question, we investigate the following conjecture:
CONJECTURE 2: Overbooking will usually provide schedule utility that is equal to or greater than slot compression.

This conjecture is illustrated by a very simple example with one appointment slot of duration 1 time unit (e.g., an hour). Suppose the show-rate for patients is \( \sigma \) and the objective of scheduling is to maximize utility. First, consider an overbooking strategy that double-books the available slot. If one or both patients are no-shows, then there will be no patient waiting. If both patients show (probability \( \sigma \)), then the second arriving patient will wait for one slot duration. Therefore, total expected patient waiting time cost is \( \omega \sigma^2 \). Expected clinic overtime in this case is equal to the expected waiting time of the second patient, so that expected overtime cost is \( \tau \sigma^2 \). Finally, the expected benefit of 2 scheduled patients is \( 2\pi \sigma \), so the expected utility for this overbooking schedule \( S_{ob} \) is:

\[
U(S_{ob}) = 2\pi \sigma - \omega \sigma^2 - \tau \sigma^2
\] (3)

<table>
<thead>
<tr>
<th>Patient</th>
<th>Probability</th>
<th>Waiting</th>
<th>E[Wait]</th>
<th>Clinic OT</th>
<th>E[OT]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>( \eta^2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 0</td>
<td>( \sigma \eta )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0 1</td>
<td>( \sigma \eta )</td>
<td>0</td>
<td>0</td>
<td>( \sigma )</td>
<td>( \eta \sigma^2 )</td>
</tr>
<tr>
<td>1 1</td>
<td>( \sigma^2 )</td>
<td>( \eta )</td>
<td>( \eta \sigma^2 )</td>
<td>1</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

In the case of a slot-compression policy, the length of each slot is reduced by an amount proportional to the no-show rate \( \sigma \). For example, if \( \sigma = 70\% \), then the length of each slot would be reduced to 70\% of its original length. In our example, we schedule two patients into two compressed slots. As with case of double booking, the expected benefit of scheduling the two patients is \( 2\pi \sigma \). Calculations for expected waiting times and tardy times are shown in Table 2, with the result that total expected waiting costs are equal to \( \eta \sigma^2 \) and expected tardy costs are \( 2\eta \sigma^2 \) so total utility for this slot compression schedule \( S_{sc} \) is

\[
U(S_{sc}) = 2\pi \sigma - \omega \eta \sigma^2 - \tau \sigma^2 (1 + \eta) = U(S_{ob}) + \omega \sigma^3 - \tau \eta \sigma^2
\] (4)

From here it is straightforward to show that the utility of overbooking will be greater than utility using slot compression under the following condition:

\[
U(S_{ob}) \geq U(S_{sc}) \text{ when } \frac{\tau}{\tau + \omega} \geq \sigma
\] (5)

Note that when \( \tau = \omega \), then overbooking will be preferred if \( \sigma \geq 0.5 \). Show rates are usually higher than 0.5, and overtime costs are usually greater than patient waiting costs ( \( \tau > \omega \) ), so this result supports (but does not prove) the conjecture that overbooking is usually preferred to slot compression. In future work, we will generalize this result with the goal of providing a conclusive proof of Conjecture 3.
Traditional Scheduling versus Open Access

There is growing interest in Open Access (OA) scheduling in both healthcare practice and clinic scheduling literature (Qu et al., 2007). In contrast to traditional scheduling (TS) where patients book appointments days and weeks in advance, open access (OA) scheduling allows patients call in for a same-day appointment with a care provider. There has been great excitement in practice about OA scheduling since it seems to provide better and faster service to patients (Robinson and Chen, 2010), but some recent reports suggest that OA scheduling may not fulfill these expectations. We therefore posit that:

**CONJECTURE 3:** Traditional scheduling always provides greater schedule utility than open access scheduling.

We assume that traditionally scheduled (TS) patients show for their appointments with probability $\sigma$, and that open access (OA) patients always show ($\sigma = 1$) for their same-day appointments. We further assume the number of TS patients scheduled in a clinic session is $N = D / \sigma$, so that the expected number of arrivals is $D$. The expected number (mean) of OA patients calling in for an appointment on a given day is also assumed to be $D$. These assumptions provide a fair comparison between OA and TS clinics and insures that both serve the same number of patients over time. We assume that the number of TS patients is binomially distributed with parameters $N$ and $\sigma$, and that the number of OA patients calling in during a session is Poisson distributed with parameter $D$.

From these assumptions, we offer qualitative support for the conjecture. In the case of traditional scheduling, patients arrive according to a binomial distribution, so the maximum number arriving during a session is $N$, with an expected mean of $D$ and variance of $\sigma \eta N$. In contrast, the expected number of OA patients calling in during a session is also $D$, but the maximum number of call-ins is unbounded with variance $N$. So, the distribution of OA call-ins is unbounded above and has a higher variance than does traditional scheduling. This suggests that OA clinics will sometimes be starved for patients when an insufficient number call in, and will other times be swamped with too many patients. In contrast, TS clinics may have too many patients show, but this number is bounded at $N$. Sometimes TS patients will have excessive no-shows, but certainly with no greater frequency that with OA, given the higher distribution of OA arrivals. This analysis suggests that an OA clinic will experience higher levels of patient waiting, clinic overtime, and provider idleness than will a TS clinic. The observations described here have been empirically observed in prior research (LaGanga and Lawrence, 2009), but are contradicted by other published research (Robinson and Chen, 2010). Clearly, this is a topic in need of further analytic and empirical support.

Finally, an OA clinic may choose to truncate the number of arrivals by turning away excessive OA call-ins, but doing so will incur penalties with patients (e.g., lost goodwill, lost future business, etc.) and seems to run counter to the spirit of offering OA appointments in the first place.

**Survey Design and Results**

As the second part of this study, we are undertaking an extensive survey of healthcare clinic practitioners about their experiences and ideas regarding appointment
scheduling. We have completed a pilot survey of practitioners, the results of which
are reported here.

Survey Design
In preparation for a national survey, we interviewed several healthcare clinic
administrators about their appointment scheduling practices, and subsequently
designed an online survey based on their feedback. The objectives of the survey
include understanding about how active healthcare clinics currently:

- Schedule patients using traditional scheduling; open access; walk-ins; or a
  combination of these.
- Deal with no-shows using methods such as overbooking; slot compression; or
  other ad hoc methods.
- Handle patient service time variability.
- Deal with unpunctual patients.

The pilot survey was distributed to 18 clinics in a large metropolitan area, with 12
usable responses returned.

Survey Results
Responses to the pilot survey are summarized below. These results are intended to
guide the design of a larger national survey. While the statistics reported here are
interesting and informative, the sample size is small and the survey population is
relatively homogenous. Consequently, the results reported below are suggestive, but
not definitive.

Appointment Scheduling. Most respondents (83%) continue to use traditional
scheduling as the predominant method for booking appointments. Half of
respondents (50%) said they sometimes use Open Access scheduling; 25% often use
OA; and 25% seldom or never use OA. Most respondents (83%) report that they
sometimes allow walk-in patients, but none frequently allow walk-ins. These results
suggest that the clinics surveyed use open access scheduling and allow walk-ins on an
ad hoc basis as situations warrant, but they do not use them in a systematic way –
traditional scheduling remains the dominant scheduling method.

Service Variability. All respondents (100%) report that they vary the planned
length of appointments based on the purpose of the appointment. This has important
implications for appointment scheduling research in that assumptions of fixed
appointment durations used in many research papers may not accurately reflect the
variability in service durations experienced by practitioners.

Scheduling Objectives. Respondents indicated that the following factors where
most important on a 1-10 scale when constructing patient schedules:

- Minimize provider idle time (8.92)
- Minimize patient waiting time (8.58)
- Minimize clinic overtime (7.67)
- See as many patients as possible (7.75)
- Less important were meeting budget constraints (7.25); minimizing costs
  (7.17); and maximizing revenues or profits (6.83)

These results are interesting in that they support the use of both a cost minimization
objective (minimize provider idleness, patient waiting, and clinic overtime) and a
utility maximization objective (maximize patients seen, minimize patient waiting and
clinic overtime). These two objectives can be reconciled by noting that maximizing patients seen also works to minimize provider idle time (see Conjecture 1 above).

Patient No-Shows. Most respondents (75%) reported that no-shows are an ongoing problem, but none suggested that it is a critical problem. The mean reported frequency of no-shows was 20% with a standard deviation of 16% and a maximum no-show value of 50%. All respondents (100%) use telephone calls to remind patients of their scheduled appointments, with a smattering of other methods (postcards, email messages, online access) used by one or two clinics. These results are surprising in that while no-show rates were reported to be quite high (20%), respondents did not seem to see it as an important problem. Possible reasons for this dichotomy are explored below in the following section.

No Show Mitigation. Almost all (91%) of respondents reported that they never or rarely use overbooking to compensate for no shows – none reported that they use overbooking often or frequently. An almost identical response was obtained regarding over-scheduling a day or compressing appointment times. Again this is a surprising result given the relatively high no-show rates reported by respondents. We offer a possible explanation for this behavior with Conjecture 4 below.

Patient Timeliness. All (100%) reported that most patients arrive on time for their appointments. When asked how early/late patients arrive for appointments, respondents reported a range from 12 minutes early to 12 minutes late, with mean of 0. Patient timeliness was not a problem for respondents, which in part may be due to the small sample size of this pilot survey and the relative homogeneity of the sample population.

Discussion
In this paper, we have begun an investigation of a number of the explicit and implicit assumptions in extant appointment scheduling research. Our goal is to better understand the appointment scheduling problem space from both theoretical and practical perspectives. The results reported in this paper begin to answer a number of important questions about the assumptions and practice of healthcare appointment scheduling. Returning to the objectives outlined at the start of this paper, we offer the following:

Choice of Objective Function. Our pilot survey suggests that healthcare practitioners use multiple criteria when constructing appointment schedules that include elements of both cost minimization and utility maximization. In particular, minimizing provider idleness was reported to be the most important criteria by survey respondents – minimizing provider idleness is a feature in most cost minimization models.

Clinic Benefits and Costs. Our analytic results show that cost minimization is a special case or subset of utility maximization, but utility maximization does not explicitly include provider idleness in its construction. Instead, idleness is implicitly addressed in a utility function since increasing the number of patients seen will inevitably increase provider utilization (LaGanga and Lawrence, 2007). Further survey work is needed to better understand the motivation for focusing on provider idleness among clinic practitioners, since provider costs are usually fixed.

Use and Structure of Overbooking. Our pilot survey suggests that clinic administrators do not perceive no-shows to be a critical problem, although reported no-show rates averaged 20% (1 day in 5 lost to no-shows). None of the respondents reported that they used overbooking or slot compression to mitigate the impact of no-shows. This apparent contraction deserves further study. Respondents also reported
significant variability in the service time required for different patients and different procedures. As one responded reported, “You never know exactly what you will find when you prep [a patient].” Given the wide and uncertain variation in many clinic procedures, we hypothesize that clinics may use no-shows as a buffer against uncertainty. We therefore offer the following conjecture:

**Conjecture 4:** Clinics use patient no-shows as an informal way of buffering against uncertainty, and therefore rely on no-shows for effective operations of the clinic.

Schedule Construction. Half of respondents to the pilot survey reported that they allow patients to choose their own appointment slot, while the other half reported that they try to fill some appointments before others:

- To enhance efficiency, we will try to "cluster" appointments to preserve larger blocks...
- We have blocked off certain times for certain appointments. Harder things in the morning, easy in the afternoons.

A topic for future research is to empirically gather more detail about how practitioners assign patients to appoint slots, and to analyze how these policies positively or negatively impact the schedule utility of clinics.

Open Access Scheduling and Walk-ins. The respondents to our survey reported that they use open access scheduling and allow walk-in patients only on a limited ad hoc basis, and otherwise rely on traditional appointment scheduling. Our analytic research suggests that OA scheduling will usually provide lower schedule utility than will traditional scheduling, which may explain why these clinics do not use OA extensively. Topics for further research are to identify clinics that are actively using OA and to understand their experiences with OA vs. TS.

**Conclusions and Future Research**

There is significant interest in the problem of appointment scheduling in outpatient healthcare clinics, but the assumptions made in the appointment scheduling literature are widely divergent. Our research objectives are to better define the nature of the outpatient healthcare appointment scheduling problem, and to guide future research by suggesting new avenues for productive and useful inquiry. This paper begins to make headway on realizing these objectives as described above.

Moving forward, we are refining our survey instrument based on results and feedback from our pilot survey, and will soon distribute it to clinic practitioners nationally at a variety of outpatient healthcare clinics. Our intention is that the results of this national survey will provide an empirical foundation for ongoing work on clinic operations and will inform future research programs, both ours and those of other scholars.

We are also working on extending our appointment scheduling modeling analysis to support, modify, or disprove the conjectures we have proposed in this paper. The results of this work will also guide our future work and we hope that of others.
References
*Production and Operations Management* 12(4), pp. 519-549.
*IIE Transactions* 40 (9), pp. 800-819.
LaGanga, L. and Lawrence, S. (2007), “Clinic overbooking to improve patient access and increase 
provider productivity”, *Decision Sciences* 38(2), pp. 251-276.
LaGanga, L. and Lawrence, S. (2009), *Comparing Walk-In, Open Access, and Traditional Appointment 
Scheduling in Outpatient Health Care Clinics*. Production and Operations Management Annual 
Meeting, Orlando FL, May 3.
no-shows and cancellations,” *Manufacturing and Service Operations Management* 12(2), 
pp. 347-364.
to demand in advanced access scheduling systems,” *European Journal of Operations Research* 183, 
pp. 812-826.
pediatric resident continuity clinics,” *Archives of Pediatrics and Adolescent Medicine* 149(6), 
pp. 693-695.