Are Patients Patient? The Role of Time to Appointment in Patient Flow

Nikolay Osadchiy

Goizueta Business School
Emory University

Joint work with Diwas KC

May 7, 2016

Motivation

Practical problem

Large healthcare system: how to improve patient throughput and revenue?

- Long wait times lead to patient forgoing scheduling appointments
- Significant (up to 20%) fraction of no-shows
- Long waits increase noshows, and limit access to care for patients who really need it.
- Paradoxical situation emerges: waits are long, yet capacity is underutilized

Research questions

General

What is the effect of wait to appointment on capacity utilization, patient throughput, and revenue?

Sub-questions

- Unbooked appointments (lost demand)?
- Unutilized capacity (lost productive time)?
- Interaction between the two
  - Does a booking with a long wait tell us anything about likelihood of arrival?
- What is the overall effect of offered wait on capacity utilization?
  - Does wait monotonically reduce capacity utilization?

Literature

- Wait and customer queuing behavior
  - Maister (1985), Olivares et al. (2011), Musalem et al. (2012), Allon et al. (2011), Buell and Norton (2011), Debo and Kremer (2014)
- Wait and capacity utilization in healthcare
  - Gallucci et al. (2005), Cohen et al. (2007), Sherman et al. (2009)
- Methodology
  - Sample selection
    - Heckman (1978), Bloom and Killingsworth (1985), Wooldridge (2007)
  - Non parametric estimation
    - Mahajan and Van Ryzin (2001), Farias, Jagabathula, and Shah (2010); Van Ryzin and Vulcano (2015)

Setting: a two-stage service process with attrition

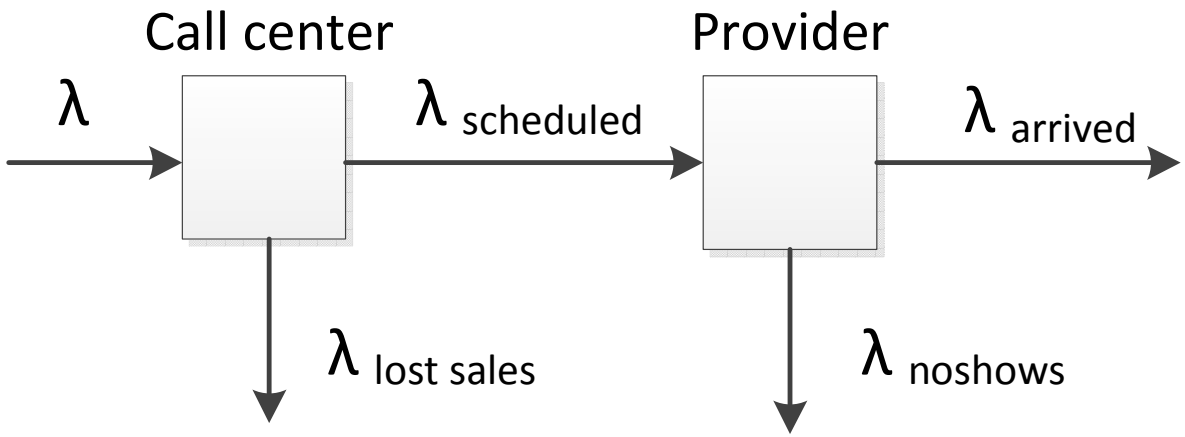


Figure : Patients flow: demand, scheduled appointments, lost sales, arrivals, and no-shows.

Simple example: Offered wait and the likelihood of arrival.

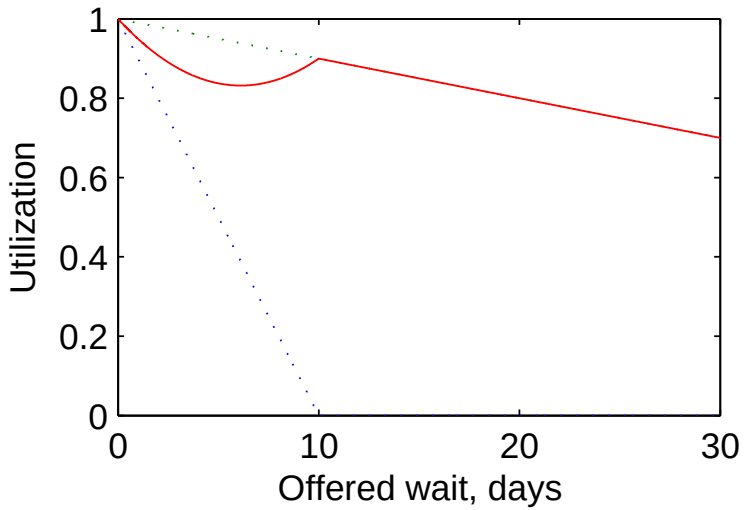


Figure : Ample demand, two patient types.

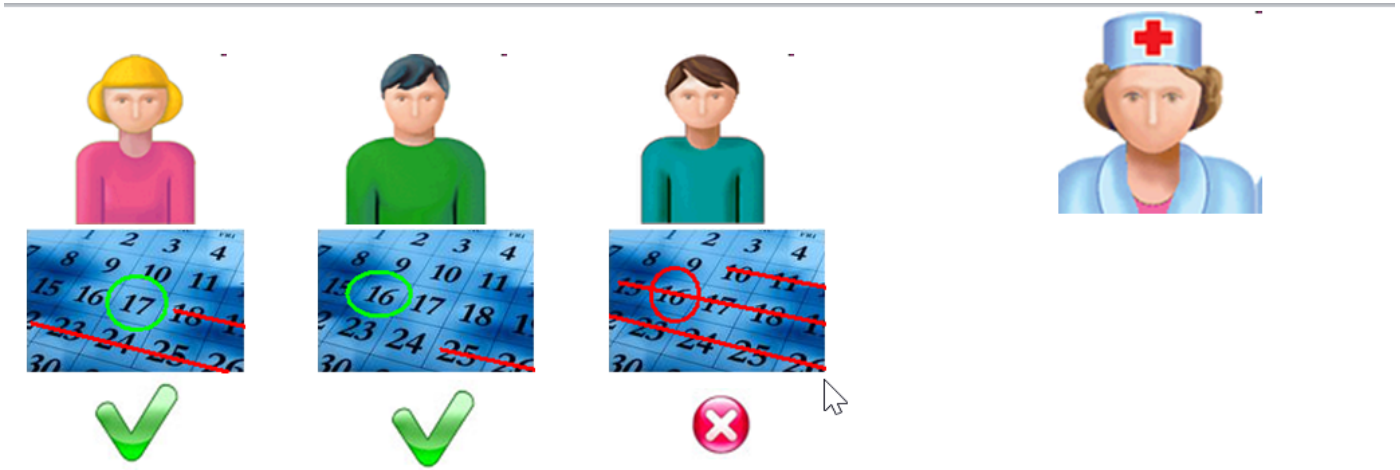
Type 0 is willing to book an appointment with any wait, and is likely to arrive for it.

Type 1 is willing to book an appointment if $w \leq 10$, and likely to be a no show.

Path forward

- Estimate willingness to book as a function of wait (WTW)
  - Estimates will be done by clinical specialty.
- Estimate likelihood of arrival as a function of booked wait, and WTW
  - Control for specialty specific WTW distribution.

Stage 1: Bookings



- For a given specialty, served by p providers, patients arrive according to a discrete time Bernoulli process, provider p rate λ_p.
  - $\lambda \ll 1$ - good approximation of the Poisson process.
  - The call center receives a steady volume of ≈ 15000 calls daily
- Each patient type is characterized by a willingness to wait (WTW) threshold τ_i
- S maximum willingness to wait
- Appointment availability (and, therefore, wait) changes over time
- Patient chooses appointment with the shortest wait, as long as it within his WTW

Estimation methodology

Non-parametric maximum likelihood estimation

- Bernoulli arrival process: rate λ
- Set of patient types: $\mathcal{S}$
- General distribution of patients over types
  - p.m.f. $f = \{x_i\}, i \in \mathcal{S}$
- Key construct: set of customer types compatible with period t observed booking (or no booking): M_t

Practical problem

Estimate f, λ given the observed booking decisions.

Illustrative example

Suppose the maximum booking horizon S=5, there are 6 time periods (I), and two customer types with WTW $\tau_i \in \{2, 5\}$

Observable data: Availability and appointment bookings

Waiting times	Period					
	1	2	3	4	5	6
0	Yes	Yes	No	Yes	No	No
1	Yes	Yes	No	Yes	No	No
2	Yes	Yes	No	No	No	No
3	Yes	Yes	No	No	Yes	No
4	Yes	Yes	Yes	No	No	Yes
5	Yes	No	Yes	Yes	No	Yes
Wait	0	-	4	0	-	-

Compatible customer types

$\{k\}$	{2,5}	{}	{5}	{2,5}	{2}{}	{2}{}

Unobservable data

τ_i	2	No arr.	5	2	No arr.	2

- Period 5 and 6

Likelihood function

$$\mathcal{L}(w|x,\lambda) = \sum_{p} \left\{ \sum_{i \in \mathcal{B}_p} \left(\log\lambda_p + \log \sum_{k \in \{\mathcal{S}: k \geq w_i\}} x_k \right) \right.$$

$$\left. + \sum_{i \in \mathcal{I}_p \setminus \mathcal{B}_p, \tilde{w}_i \geq 1} \log \left(\lambda_p \sum_{k \in \{\mathcal{S}: k < \tilde{w}_i\}} x_k + (1 - \lambda_p) \right) + \sum_{i \in \mathcal{I}_p \setminus \mathcal{B}_p, \tilde{w}_i = 0} \log(1 - \lambda_p) \right\}$$

- Accounts for bookings, no-bookings due to unacceptably long wait, and no-bookings due to non-arrivals

Optimization problem

$$\max_{x,\lambda} \mathcal{L}(w|x,\lambda) \tag{1}$$

$$\text{s.t.} \sum_{k=0}^{S} x_k = 1,$$

$$x_k \geq 0, \text{for all } k,$$

$$0 \leq \lambda_p \leq 1, \text{for all } p.$$

Notice that the lost sales corresponding to a waiting time of w is simply $LostSales_p(w) = \sum_{k=w+1}^{S} x_k$.

Stage 2: Likelihood of arrival

Goal: estimate the effect of the waiting time to appointment w on the likelihood that the patient arrives for the appointment.

- Selective sample: observe only patients with $\tau_i \leq w_i$
- Booked appointments can come from patients with high WTW and possibly low sensitivity of likelihood of arrival to wait
  - Need to control for patient's WTW
- Without correction estimates may not be representative of population, and would not allow a proper counterfactual
- Unlike Heckman (1978), the selection stage is latent
  - Bloom and Killingsworth (1985) study a similar problem in a fully parametric setting
- Our approach: explicitly control for expected WTW, and use inverse probability weighting (IPW) estimator (Wooldridge 2007)
  - Key idea: high WTW patients have higher probability of inclusion into the sample

Stage 2: Estimator

$$\min_{\theta \in \Theta} \sum_i -\frac{1}{p_i} \{y_i \log l_i + (1 - y_i) \log(1 - l_i)\}, \text{ where}$$

$$l_i = \frac{e^{\theta \mathbf{c_i}}}{1 + e^{\theta \mathbf{c_i}}},$$

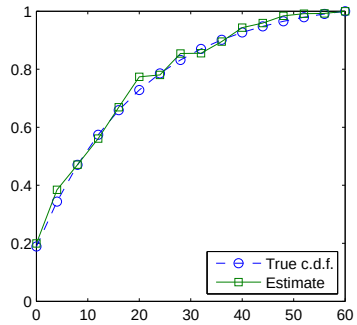
$$\theta \mathbf{c_i} = \theta_0 + \theta_1 w_i + \sum_{j=1}^{9} \theta_D(j)\ w\ \mathbb{1}\{Decile(E(\tau_i | \tau_i \geq w_i), Specialty_i) \leq j\} + \theta_c controls_i,$$

$$p_i = \Pr(\tilde{w} \leq \tau | \tau \geq w_i) = \sum_{t=w_i}^{\infty} \Pr(\tilde{w} \leq t) \Pr(\tau = t | \tau \geq w_i).$$

- l_i is the logistic probability, p_i is the probability of inclusion into the sample.
- Deciles are computed by specialty
- $\mathbf{c_i}$ includes patient's gender, ethnicity, marital status, age, payor; appointment scheduled time (AM/PM), urgent/routine indicator, weather, whether the appointment had the shortest time among available ones, distance from the patient's home zip to the providers' location, provider, booking DOW, and appointment DOW FE.

Performance on simulated data (mimics typical specialty)

Estimated WTW



Est. arrival rate λ	
True	0.25
Est.	0.2517(0.0195)

True model		Lin. mod. est.		Estimates by deciles	
				SS	SS IPW
Constant	2	1.9717(0.0160)		1.9747 0.0259	1.9805 0.0259
WTW		θ	E(WTW)	θ	θ
10%	0.05	-	10%	-0.0490(0.0118)	-0.0474(0.0118)
20%	0.05	-	20%	-0.0490(0.0118)	-0.0474(0.0118)
30%	0.05	-	30%	-0.0501(0.0109)	-0.0489(0.0109)
40%	0.05	-	40%	-0.0478(0.0071)	-0.0474(0.0071)
50%	0.05	-	50%	-0.0426(0.0055)	-0.0426(0.0055)
60%	0.03	-	60%	-0.0385(0.0049)	-0.0385(0.0049)
70%	0.03	-	70%	-0.0387(0.0037)	-0.0389(0.0037)
80%	0.03	-	80%	-0.0350(0.0026)	-0.0353(0.0026)
90%	0.03	-	90%	-0.0319(0.0018)	-0.0321(0.0018)
100%	0.03	-0.0295(0.0009)	100%	-0.0300(0.0011)	-0.0303(0.0011)

Figure : **Left:** Estimation results for the WTW distribution and the arrival rate using (1). **Right:** Estimation results for the sensitivity to wait using (2) on 100 simulated datasets, each containing approximately 550 booked appointments. Means and standard deviations over the 100 simulation runs are reported.

Data: large US East coast healthcare system

Appointments and arrivals

- Patient ID
- Specialty / condition
- Urgency / severity
- Category NPV/RPV
- Calling date, call order #
- Date of appointment
- Status (no-show, cancel, arrive)
- Demographics

- Waiting time
- No-show rate
- Utilization rate

Provider availability and ratings

- Provider ID
- Specialty
- Date
- Number of slots available
- Healthgrades.com rating
- Medicare reimbursement rate

For robustness tests

- Additional 4 months of data
- Jan-Feb of 2011 and 2010

- Focus on first time visits

Summary statistics

Variable	Mean	Standard Deviation	Median	Min	Max
Waiting time	20.83	31.34	11	0	363
Selected Specialties					
Family Practice	2.19	8.34	0	0	182
Orthopaedics	11.86	15.28	7	0	240
Neurology	67.91	70.17	49	0	238
Rheumatology	113.29	86.36	67.5	1	257
Waiting Time (if arrived)	18.1	26.05	9	0	363
Waiting Time (if no-show, or urc)	32.42	45.89	16	0	353
- among unrebooked cancellations (urc)	31	47.06	14	0	323
Number of patients scheduled (per provider)	49.56	65.91	27	1	482
Number of patients arrived (per provider)	40.11	54.45	22	0	440
% arrived (per provider)	79.82	18.84	83.52	0	100
% no show and urc (per provider)	20.18	18.84	16.48	0	100
– % unrebooked cancellations	10.63	17.02	1.98	0	100

29089 observations, 25114 patients for 587 providers, 107 clinical specialties in Jan-Feb 2012.

Notes:

Initial data: 237954 appointment records for 914 providers (all visits); 37688 NPV visits for 596 providers.

Less rescheduled, canceled and rebooked, pending: 157272 appointments for 827 providers

including 29089 NPV appointments for 587 providers remain.

Demographics

Gender	%	Ethnicity	%	Marital Status	%	Payer	%
Female	59.8	Hispanic or latino	1.4	Divorced	0.1	Blue Cross	11.6
Male	40.2	Non-hispanic or latino	54.1	Life partner	6.4	HMO	41.6
		Not reported	45.9	Married	41	Medicaid	9.5
				Not reported	19.9	Medicare	21.2
				Separated	0.8	Other	1.6
				Single	27.6	Outsourced	0.1
				Widow/er	4.2	PPO	10.7
						Self Pay	3.7

	Mean	St.Dev.	Median	Min	Max	N
Patients						
Age, years	48.81	20.16	50.37	0.01	100.01	29089
Distance, miles	35.63	113.8	13.1	0	4503	28993
Day of appointment						
Average temperature, F	49.06	8.07	49	28	67.5	29089
Average temperature departure, F	3.86	7.86	3.4	-15	18.4	29089
Precipitation, inches	0.07	0.13	0	0	1.19	29089
Providers						
Satisfaction rating, %	95.29	8.32	100	75	100	204
Average Medicare payment, $	71.46	41.05	56.97	15.38	279.86	361

Specialties with most NPVs		*Specialties with least NPVs*	
Orhtopaedics	15.1%	Pelvic reconst. surgery	0.03%
Spine care	9.3%	Lung transplantation	0.03%
Neurology	8.3%	Diabetes education	0.01%

Estimate of willingness to wait to appointment

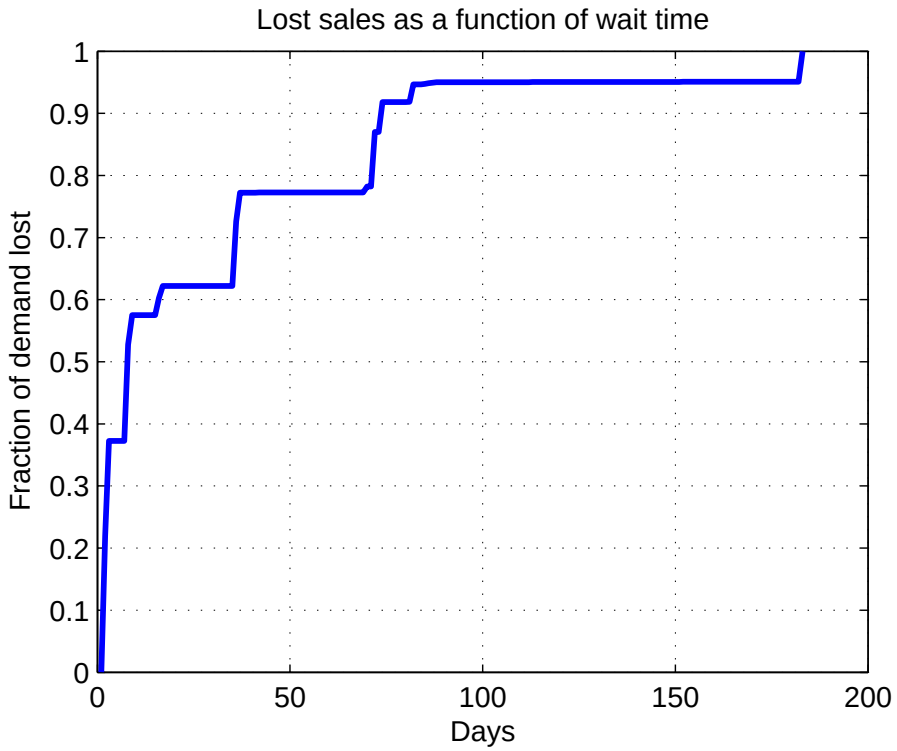


Figure : Clinical specialty: gastroenterology.

Effect of waiting time on no-shows

	SS	SS IPW
	(1)	(2)
Wait	-0.0139 (0.001)***	-0.0158 (0.0013)***
Wait*$\mathbb{1}\{Decile(EWTW) \leq 1\}$	-0.0081 (0.003)***	-0.0105 (0.0037)***
Wait*$\mathbb{1}\{Decile(EWTW) \leq 2\}$	-0.0165 (0.0089)*	-0.0197 (0.0092)**
Wait*$\mathbb{1}\{Decile(EWTW) \leq 3\}$	-0.0013 (0.003)	-0.0008 (0.0034)
Wait*$\mathbb{1}\{Decile(EWTW) \leq 4\}$	-0.0098 (0.0058)*	-0.0102 (0.0084)
Wait*$\mathbb{1}\{Decile(EWTW) \leq 5\}$	-0.0046 (0.0048)	-0.0062 (0.0056)
Wait*$\mathbb{1}\{Decile(EWTW) \leq 6\}$	-0.0055 (0.0021)***	-0.0081 (0.0027)***
Wait*$\mathbb{1}\{Decile(EWTW) \leq 7\}$	0.0012 (0.0014)	0.0016 (0.0017)
Wait*$\mathbb{1}\{Decile(EWTW) \leq 8\}$	-0.0066 (0.0031)**	-0.0071 (0.0038)*
Wait*$\mathbb{1}\{Decile(EWTW) \leq 9\}$	-0.0032 (0.0019)*	-0.0038 (0.0026)
FirstAvailable	0.2201 (0.0428)***	0.2367 (0.0489)***
Patient level controls: included		
Appointment level controls: included		
Medical condition controls: included		
Weather controls: included		
Provider & DOW reg/sched. F.E.	Yes	Yes
Wald Chi2(520).($Pr > ChiSq$)	$<$0.0001	$<$0.0001

27555 observations over 476 providers. 111 providers are excluded due to no variation in the dependent variable. Baseline: Gender (Female), Ethnicity (Hispanic or Latino), Marital status (Divorced), Payer (Blue Cross), Scheduled time (AM). Robust standard errors are in parentheses. *10% statistical significance; **5% statistical significance; ***1% statistical significance.

Notes:

1. In model (2) the baseline probability of arrival at $Wait = 0$ is 0.9451 (female, hispanic or latino, divorced, Blue Cross, age 49, distance 35 miles, Monday AM appointment with a radiation oncologyst, typical dry weather). 67 out of 476 providers have statistically significant fixed effects (FE), mean(FE)=-0.213, std(FE)=0.984.

2. The following modifications of model (2) with added interaction terms were also estimated for robustness: 1) with $Wait \times Urgent$, 2) $Wait \times Routine$, 3) $Wait \times Age$. All interactions were found insignificant.

Sensitivity of likelihood of arrival to offered wait

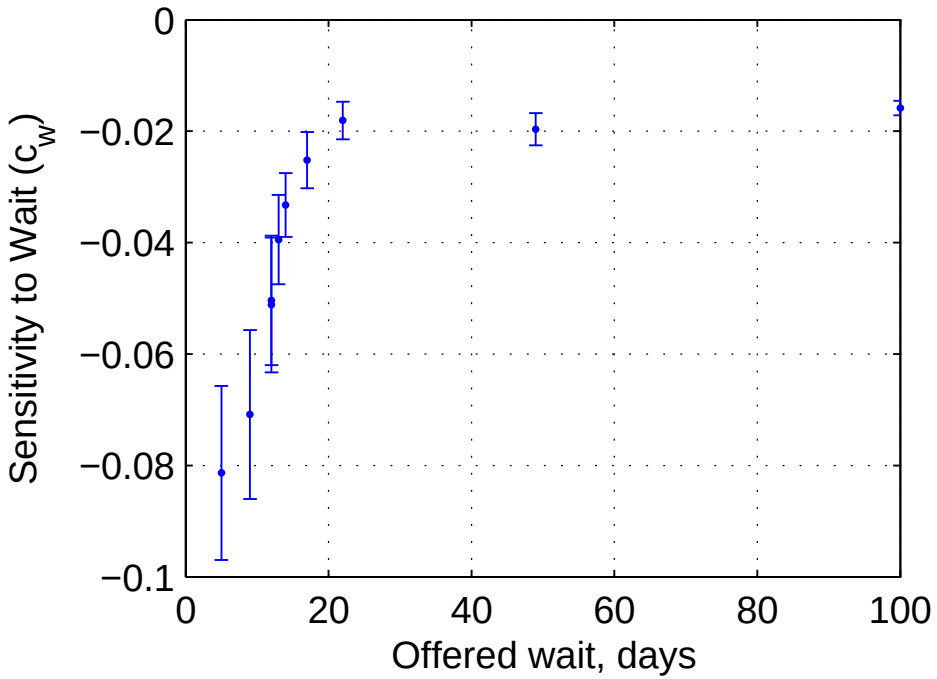


Figure : Sensitivity to wait as a function of offered wait. Offered waits are the median waits of the respective CWTW deciles.

The results are qualitatively similar on winsorized samples, specialty subsamples, and under alternative model specifications (log-waits, standardized waits, etc.).

Marginal effect of wait (keep mix of patients constant)

$$Throughput = \lambda \times Pr(Scheduled) \times Pr(Arrived)$$

Table : Average Marginal Effects with respect to wait time W (first time visits).

	Mean	St.Dev.	Median	Min	Max
W (days)	25.25	26.89	15.31	0.63	169.79
# of requests for appointments (calls/day)	4.01	5.15	2.64	0.00	73.00
Pr(Scheduled)	0.6907	0.2315	0.7521	0.0425	1.0000
Pr(Arrived)	0.7787	0.1466	0.8110	0.0532	0.9867
Throughput (%)	54.94%	20.55%	55.35%	1.18%	94.44%
Throughput (patients per day/day of wait)	2.1739	2.7533	1.3852	0.0000	27.9418
dPr(Scheduled)/dW	-0.0310	0.0396	-0.0228	-0.3724	0.0000
dPr(Arrived)/dW	-0.0025	0.0011	-0.0026	-0.0044	-0.0002
dThroughput/dW (patients per day/day of wait)	-0.1273	0.2848	-0.0460	-3.2999	0.0000
– incl from increased lost sales	78.38%	18.75%	81.15%	1.32%	100.00%
– incl from increased noshows	21.62%	18.75%	18.85%	0.00%	98.68%
dThroughput(%)/dW (% patients/day)	-5.47%	5.94%	-4.18%	-62.44%	-0.08%
Avg. Medicare payment per visit, $	74.59	42.30	65.12	15.76	279.86
Daily revenue, $	142.11	153.48	89.96	0.00	956.29
dRevenue/dW ($ per day/day of wait)	-9.07	20.19	-3.20	-168.50	0.00
dRevenue(%)/dW (% daily revenue/day of wait)	-5.47%	5.94%	-4.18%	-62.44%	-0.08%

1. Specialties of the 3 providers with the lowest $|dRevenue(\%)/dW|$: Cardiology (general), Cardiology (cardiac electrophysiology), and Neuro-ophtalmology.

2. Speciality of the 3 providers with the highest $|dRevenue(\%)/dW|$: General internal medicine (GIM).

What if a shorter wait offered to all (counterfactual)?

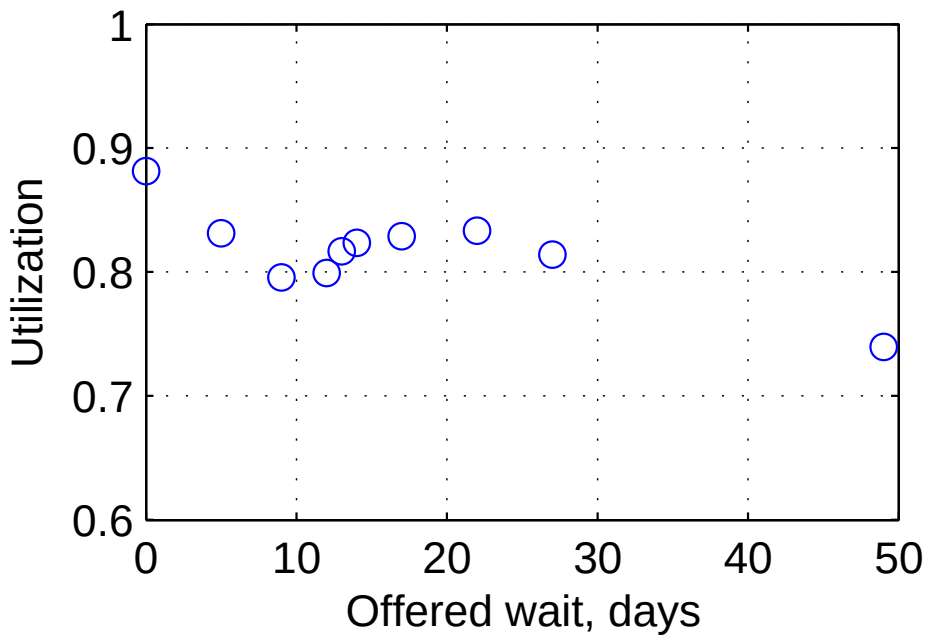


Figure : The counterfactual effect of *offered* wait on the likelihood of arrival for an appointment for a new patient visit.

Counterfactual by specialty

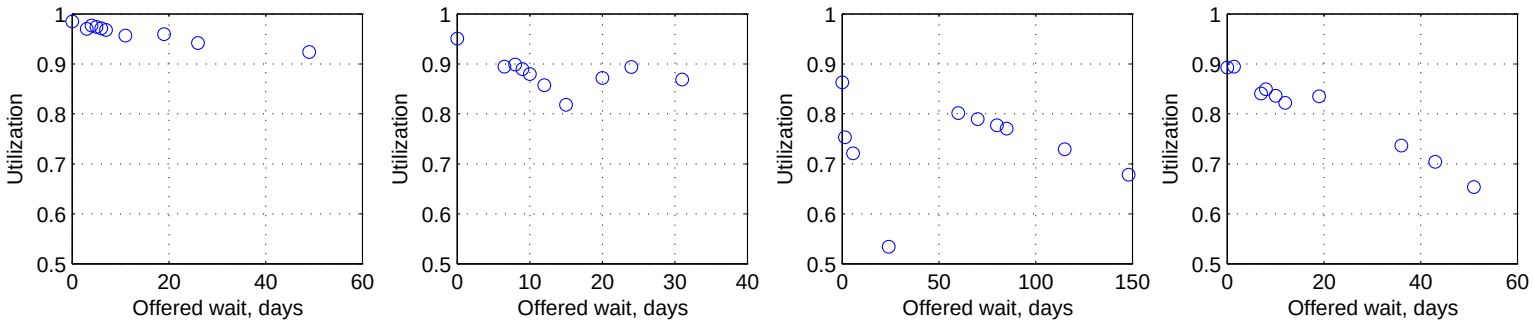


Figure : The counterfactual effect of *offered* wait on the probability of arriving for an appointment for a new patient visit by specialty (L to R): orthopaedics, spine center, neurology, otolaryngology.

Is the result robust? Two more years of data.

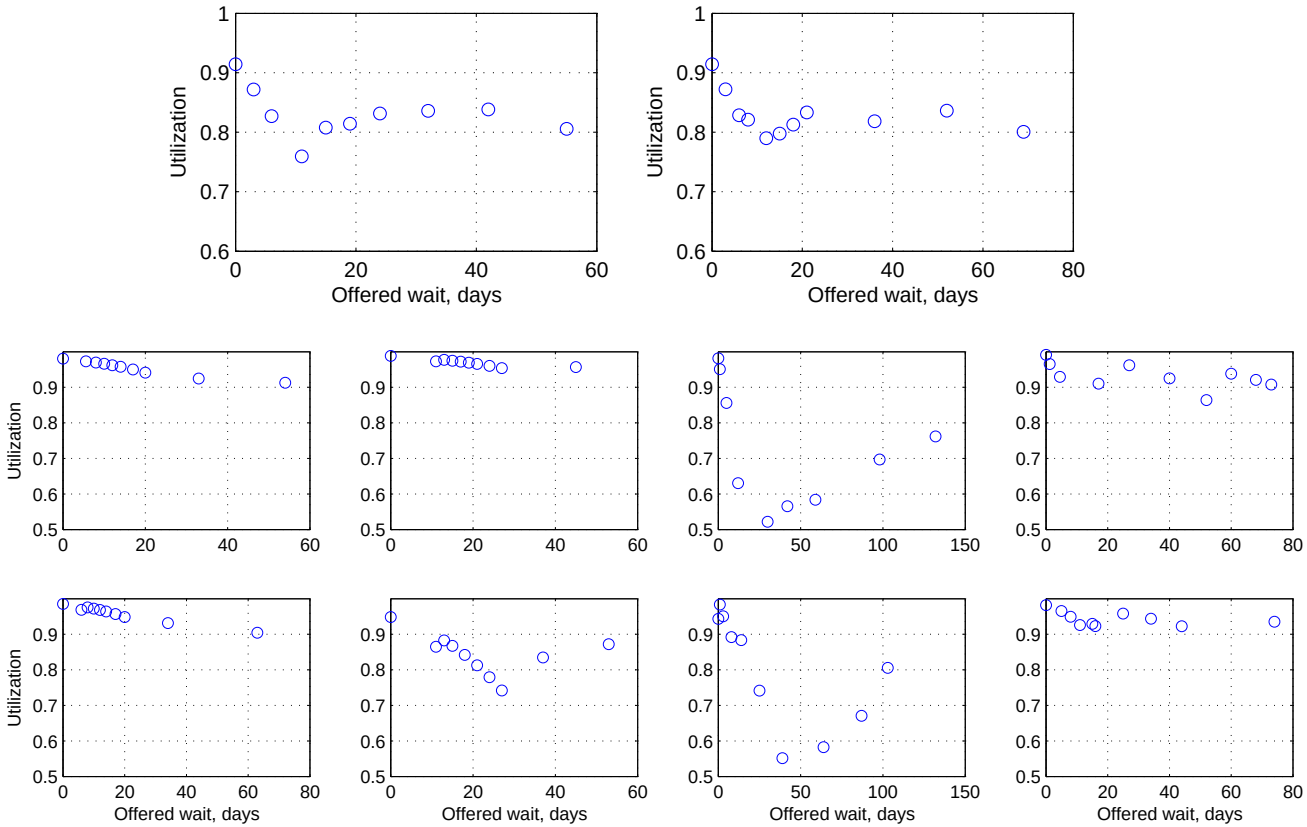


Figure : The counterfactual effect of *offered* wait on the probability of arriving for an appointment for a new patient visit for years 2011 and 2010, for all and specific specialties (L to R): orthopaedics, spine center, neurology, otolaryngology.

Conclusions

- Develop a two-stage model to examine a service system with multiple and interdependent attrition points
- Estimate the effects of wait reduction on capacity utilization
  - Increased wait can increase capacity utilization by screening out unreliable customers

Thank you!

Comments/questions: nikolay.osadchiy@emory.edu