Circulation: Cardiovascular Quality and Outcomes

CARE INNOVATIONS

Buffer or Suffer

Redesigning Heart Failure Postdischarge Clinic Using Queuing Theory

R. Kannan Mutharasan, MD
Faraz S. Ahmad, MD, MS
Itai Gurvich, PhD
Hannah Alphs Jackson, MD, MHSA
Jan A. Van Mieghem, PhD
Clyde W. Yancy, MD, MSc

Timely follow-up in clinic after heart failure hospitalization represents an evidence-based intervention associated with reduced rehospitalization.[1] Major cardiovascular societies endorse a 7-day follow-up visit as an appropriate target for quality.[2] Yet the rate of scheduled follow-up visits remains relatively low, at ≈65% in 2012 by registry data.[3] Even more striking is the rate of arrived follow-up visits: 30% in 7 days.[3] This represents a substantial missed opportunity to address and a likely explanation for ongoing avoidable readmissions. We took an unconventional approach to improving clinic scheduling policies by collaborating with our colleagues at the Northwestern University Kellogg School of Management to implement queuing theory as a novel approach to address a previously unyielding problem.

In 2015, as part of a multidisciplinary intervention to improve outcomes for hospitalized heart failure patients, we systematically identified all patients within our hospital at risk for heart failure-related readmissions through a daily enterprise data warehouse screen[4]; developed a multidisciplinary bridge and transition team to engage patients during the index hospitalization; and then deployed queuing theory to first assess and then improve clinic follow-up.

Queuing theory is the mathematical study of waiting times.[5,6] With roots in the telecommunications field, it has widespread applications in several processes such as understanding supermarket lines and managing factory inventory. A particularly powerful insight arising from queuing theory is the notion that extra capacity, or a capacity buffer, is necessary to ensure system performance when variable demand arises, such as for a hospital discharge clinic. We opted to use queuing theory to analyze hospital discharge load and understand the capacity needed in clinic to reduce wait times and improve access. Here, we provide our mathematical analysis based on real-world practice; the results of our intervention; and an online calculator (http://www.hfresearch.org) for other groups seeking to redesign access using our model.

GOALS AND VISIONS OF THE PROGRAM

The goal of this clinic redesign was to improve timely scheduling of hospital discharge patients into clinic, as part of comprehensive transitional care intervention for hospitalized heart failure patients.

LOCAL CHALLENGES IN IMPLEMENTATION

Several operational barriers impede timely access to heart failure postdischarge clinic at our institution. First, scheduling resources are available only during business hours Monday through Friday. Second, the culture of clinic utilization is to schedule

Key Words: goals ■ heart failure ■ hospitalization ■ transitional care

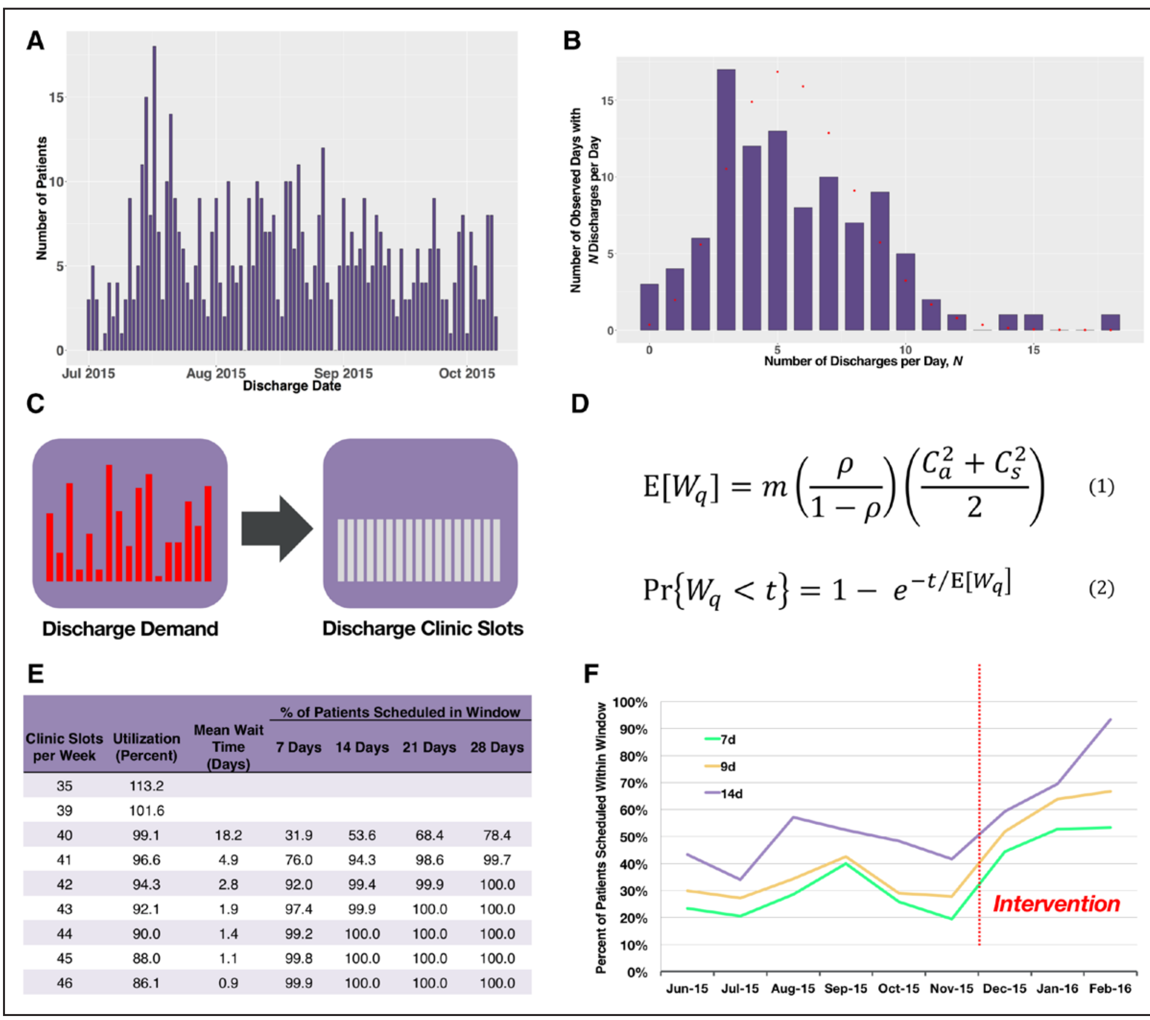
http://circoutcomes.ahajournals.org

to full capacity for each clinic with the intended goal of maximizing clinic resources. Third, the demand for hospital discharge visits varies daily. Fourth, the complexity of hospital follow-up for heart failure dictates significant consumption of resources, requiring up to twice as much clinician time as a standard return patient.

DESIGN OF THE INITIATIVE

To develop meaningful inputs into the queuing model, we first analyzed our hospital discharge volume, depicted in the Figure (A). During the study period of 100 days, 566 patients with hospitalized heart failure were discharged, for a mean of 5.66 discharges daily, or 39.6 discharges weekly. The distribution of daily discharges was skewed rightward. A postdischarge clinic sized to closely match this demand might provide 40 slots per week. The number of arrivals varies from day to day; these data are plotted in the frequency domain (Figure [B]). This distribution develops when discharges from the system occur at random times, as might be expected for hospitalized patients with variable lengths of stay. Hallmarks of this distribution include (1) a tail that skews rightward (mode=3, range=0–18, SD=3.3, coefficient of variation=0.58); (2) and a random interarrival time into the system. This random or Markovian distribution of arrivals is important, as systems exhibiting this characteristic are tractable to analytic solutions.[5] Based on our usual clinic performance, we determined at baseline that 23% of patients were scheduled within 7 days of hospital discharge.

Figure (C) represents a schematic of our queuing model wherein variable hospital discharges must be accommodated by fixed postdischarge clinic capacity.



$$\mathrm{E}[W_q] = m\left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2 + C_s^2}{2}\right) \quad (1)$$

$$\Pr\{W_q < t\} = 1 - e^{-t/\mathrm{E}[W_q]} \quad (2)$$

Clinic Slots per Week	Utilization (Percent)	Mean Wait Time (Days)	% of Patients Scheduled in Window			
			7 Days	14 Days	21 Days	28 Days
35	113.2					
39	101.6					
40	99.1	18.2	31.9	53.6	68.4	78.4
41	96.6	4.9	76.0	94.3	98.6	99.7
42	94.3	2.8	92.0	99.4	99.9	100.0
43	92.1	1.9	97.4	99.9	100.0	100.0
44	90.0	1.4	99.2	100.0	100.0	100.0
45	88.0	1.1	99.8	100.0	100.0	100.0
46	86.1	0.9	99.9	100.0	100.0	100.0

Figure. A, Heart failure discharge volume by day from our institution over a 100-day period. **B**, Overview of queuing theory analysis. Heart failure discharge volume when plotted in the frequency domain compared with a Poisson distribution (red). **C**, Schematic of queuing model. Variable discharge demand enters into a scheduling queue for a fixed number of postdischarge clinic slots. **D**, Key mathematical relationships that govern queues. Equation 1 demonstrates that estimated wait in a queue ($\mathrm{E}[W_q]$) is driven by 3 multiplicative effects: average service time (m), the utilization effect (ρ=utilization), and the variability effect (C_a=coefficient of variation of interarrival times; C_s=coefficient of variation of service times). Equation 2 describes the waiting time distribution function. **E**, Analytic results of the queuing model. At 99.1% utilization, only 31.9% of patients can successfully be scheduled into clinic at 7 d, and only 78.4% at 28 d. A modest capacity buffer of 12% (88% utilization) improves clinic performance such that virtually all patients can be successfully scheduled at 7 d. **F**, Performance of our postdischarge clinic after the queuing theory based intervention.

Daily discharge volume from the hospital was modeled as a Poisson distribution. We then approached the modeling of discharge clinic through a series of 2 refinements, each with complimentary insights into real-world behavior that allowed us to shape our intervention.

Figure (D) demonstrates the governing equations of the queuing model. We first modeled the clinic service time as also following an exponential distribution, an assumption that allows a full analytic solution to expected wait times. Patients waiting for the longest to be scheduled are scheduled first (first in, first out). Using the standard Kendall nomenclature in the field, this system may be termed an M/M/1 queue (Markovian arrival distribution, Markovian service time distribution, single server).

The table in Figure (E) demonstrates the key results of this model. At a service level of 40 discharge appointments a week (99.1% utilization), a capacity that closely approximates average demand, the queuing model predicts that the mean wait time to be scheduled into clinic is 18.2 days. In this condition, a mere 31.9% of patients can be scheduled within 7 days, and only 78.4% of patients can be scheduled within 28 days. This result seems surprising but can be understood by contemplating the following scenario: if a clinic has only 1% extra capacity (≈0.5 slots per week), a surge in demand for appointments, as often happens given the variability seen in the number of hospital discharges per week, will take weeks to accommodate. During this time, the number of patients who can be scheduled within 7 days will be close to 0.

The queuing model predicts that relatively modest capacity buffers can greatly improve clinic performance. By increasing clinic capacity in the model by 5 slots per week, thereby decreasing utilization to 88%, the average wait time is predicted to drop to 1.1 days, and virtually all patients can be accommodated within the prescribed 7-day postdischarge.

As a further check on the robustness of the modeling exercise, we generalized the model by withdrawing the assumption that the service times follow an exponential distribution. When modeling postdischarge clinic as deterministic service time—that is, each clinic appointment takes the same amount of time—we find that the expected mean wait for a clinic at 99.1% utilization is 9.1 days, and ≈53.6% of patients can be accommodated within the 7-day window. As a further generalization, we withdrew the assumption that the number of discharges from hospital follows a Poisson distribution. Using our real-world data, we found that the coefficient of variation in daily discharges was 0.58, in contrast to 1.0 for a true Poisson distribution. Using this more general G/D/1 model (general arrival distribution, deterministic service times, single server), we find that the expected mean wait for a clinic at 99.1% utilization is 4.9 days, and ≈75.8% of patients can be accommodated within the 7-day window. Thus, even these more stringent models point to unacceptably low performance of a postdischarge clinic used at 99% capacity. Furthermore, all 3 models predict that relatively modest increases in clinic capacity will yield marked improvements in scheduling performance. Further sensitivity analyses are shown in the Data Supplement.

IMPLEMENTATION OF THE INITIATIVE

By participating in the Centers for Medicare and Medicaid Services Bundled Payment for Care Improvement program for heart failure, we were afforded a unique opportunity to implement changes to our postdischarge clinic model in response to the queuing analysis. Under Bundled Payment for Care Improvement, our institution receives an increased payment from Centers for Medicare and Medicaid Services for patients assigned a heart failure diagnosis-related group, and in return is responsible for all healthcare costs during a 30-day episode of care posthospitalization. The institution is therefore strongly incentivized to avoid expensive posthospitalization costs, primarily by reducing readmissions and reducing skilled nursing facility length of stay.

As part of our strategy to realize these improvements in transitional care, we invested in a multidisciplinary care team comprising a physician champion, nurse practitioner, social worker, nurse educator, pharmacist, and transitional care liaison. We had latitude as a team to tailor work to enhance value most. As a result of the queuing analysis, we determined extra clinic capacity was needed for hospital discharge appointments and targeted an additional 15 postdischarge clinic slots per week with the nurse practitioner beyond the extant 32 slots. This was divided into 3 half-day sessions, each with 5, 40-minute appointments.

SUCCESS OF THE INITIATIVE

Figure (F) depicts our postdischarge clinic performance after the intervention was deployed. Within 1 month of the intervention, clinic access improved considerably. Before the intervention, with clinic utilization at 97%, 23% of patients were scheduled within 7 days and 43% of patients were scheduled within 14 days. After the intervention, with clinic utilization dropping to 84%, patients scheduled within 7 days rose to 53% and patients scheduled within 14 days rose to 93%. Moreover, although a Medicare payment paradigm change prompted postdischarge clinic redesign, we applied the intervention to all patients in our system regardless of payor, extending the reach of the intervention.

TRANSLATION TO OTHER SETTINGS

This analysis is readily translatable using a web calculator to other settings in which outpatient clinic capacity must be coupled to variable discharge demand from the hospital. With our web calculator (http://www.hfresearch.org), users input the average number of discharges per week, current clinic capacity, and proposed clinic capacity, and receive back an analysis of anticipated clinic performance.

A challenge in translation will be matching prescribed clinic capacity needs to available human resources. For example, it may be inefficient for a small group or hospital to hire a full-time clinician specifically for a heart failure postdischarge clinic. However, a clinician to staff a cardiology postdischarge clinic—or even more broadly a chronic disease postdischarge clinic—may be feasible.

A second challenge in translation will be for centers operating under a traditional fee-for-service model, as cost savings on readmissions cannot directly be recouped such as in a bundled payment model. However, under traditional fee-for-service, cost savings are still possible if heart failure admissions carry negative operating margins.[7]

SUMMARY OF THE EXPERIENCE, FUTURE DIRECTIONS, AND CHALLENGES

By applying queuing theory, we found that a modest capacity buffer is necessary to maintain performance of a hospital postdischarge clinic. An oft-cited mantra in the operations literature for addressing variable demand, as seen with hospital discharges, is buffer or suffer. In our experience, modest buffering represented a highly leveraged intervention for process improvement. The direct relationship between unused slot rate and availability when dealing with variable demand is indeed so robust, that unused slot rate can be used as an operational proxy for availability and therefore operational efficiency.

Through our work, we have identified several challenges. First, creating a buffer for a postdischarge clinic may be challenging in the absence of alternative payment models. In a fee-for-service payment structure, a clinic operating at full capacity is a more remunerative clinic. However, our model predicts that clinics running at 99% utilization will have insufficient capacity for timely follow-up. In contrast, a hospital postdischarge clinic with healthy availability will necessarily have an unused slot rate of ≈10%. In this manner, programs such as Bundled Payment for Care Improvement incentivize hospital systems to take a longer-term perspective by investing in a small increase in capacity and operating cost that may ultimately reduce rehospitalization and subsequent penalties.

Second, as expected when process improvement is undertaken within a complex system, these increases in clinic capacity have unveiled follow-on targets for intervention. The availability of scheduling resources 7 days a week would allow for inclusion of a confirmed postdischarge clinic appointment on discharge paperwork for all patients, not just patients leaving hospital during business days. Additionally, the no-show rate in our postdischarge clinic remains significant, at ≈30%. Further improvements here are necessary to allow for the follow-up visit to have a positive impact on care.

In summary, we found that a capacity buffer is necessary to ensure performance when there is variability in demand. By setting expected service levels, sizing calculations may be used to help ensure organizations meet quality metrics such as 7-day follow-up for hospitalized heart failure patients. In the future, we plan to externally validate this model in other hospitals, initially within our health system and then at other health systems. Further investigation will also evaluate the relationship between improved discharge clinic performance and the observed reduction in readmission rates. Redesigning care delivery systems can be challenging and resource-intensive; however, if administrators heed the principle of buffer or suffer, they may be able to increase operational efficiency and reap the financial rewards under alternative payment models and value-based care.

ARTICLE INFORMATION

The Data Supplement is available at http://circoutcomes.ahajournals.org/lookup/suppl/doi:10.1161/CIRCOUTCOMES.117.004351/-/DC1.

Correspondence

R. Kannan Mutharasan, Division of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, 676 N Saint Clair St, Arkes Pavilion, Suite 6-071, Chicago, IL 60611. E-mail kannanm@northwestern.edu

Affiliations

Division of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL (R.K.M., F.S.A., C.W.Y.). School of Operations Research and Information Engineering, Cornell University, New York, NY (I.G.). Department of Surgery, Northwestern Feinberg School of Medicine, Chicago, IL (H.A.J.). Department of Operations, Northwestern University Kellogg School of Management, Evanston, IL (J.A.V.M.).

Acknowledgments

We wish to acknowledge the dedicated care of the nurse practitioners and physician assistants who staff our postdischarge clinics.

Sources of Funding

Dr Ahmad is supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number T32HL069771 and by a grant from the Eleanor Wood-Prince Grants Initiative: A Project of The Woman's Board of Northwestern Memorial Hospital.

Disclosures

None.

REFERENCES

1. Hernandez AF, Greiner MA, Fonarow GC, Hammill BG, Heidenreich PA, Yancy CW, Peterson ED, Curtis LH. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *JAMA*. 2010;303:1716–1722. doi: 10.1001/jama.2010.533.
2. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL; WRITING COMMITTEE MEMBERS; American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation*. 2013;128:e240–e327. doi: 10.1161/CIR.0b013e31829e8776.
3. DeVore AD, Cox M, Eapen ZJ, Yancy CW, Bhatt DL, Heidenreich PA, Peterson ED, Fonarow GC, Hernandez AF. Temporal trends and variation in early scheduled follow-up after a hospitalization for heart failure: findings from Get With The Guidelines-Heart Failure. *Circ Heart Fail*. 2016;9:e002344. doi: 10.1161/CIRCHEARTFAILURE.115.002344.
4. Ahmad FS, Wehbe RM, Kansal P, Jackson HA, Anderson AS, Yancy CW, Mutharasan RK. Targeting the correct population when designing transitional care programs for Medicare patients hospitalized with heart failure. *JAMA Cardiol*. 2017;2:1274–1275. doi: 10.1001/jamacardio.2017.3089.
5. Anupindi R. *Managing Business Process Flows*. 3rd ed. Boston, MA: Prentice Hall; 2012.
6. Foster EM, Hosking MR, Ziya S. A spoonful of math helps the medicine go down: an illustration of how healthcare can benefit from mathematical modeling and analysis. *BMC Med Res Methodol*. 2010;10:60. doi: 10.1186/1471-2288-10-60.
7. Mihailoff M, Deb S, Lee JA, Lynn J. The effects of multiple chronic conditions on adult patient readmissions and hospital finances: a Management Case Study. *Inquiry*. 2017;54:46958017729597. doi: 10.1177/0046958017729597.

Circulation
Cardiovascular Quality and Outcomes

American Heart Association®

Buffer or Suffer: Redesigning Heart Failure Postdischarge Clinic Using Queuing Theory

R. Kannan Mutharasan, Faraz S. Ahmad, Itai Gurvich, Hannah Alphs Jackson, Jan A. Van Mieghem and Clyde W. Yancy

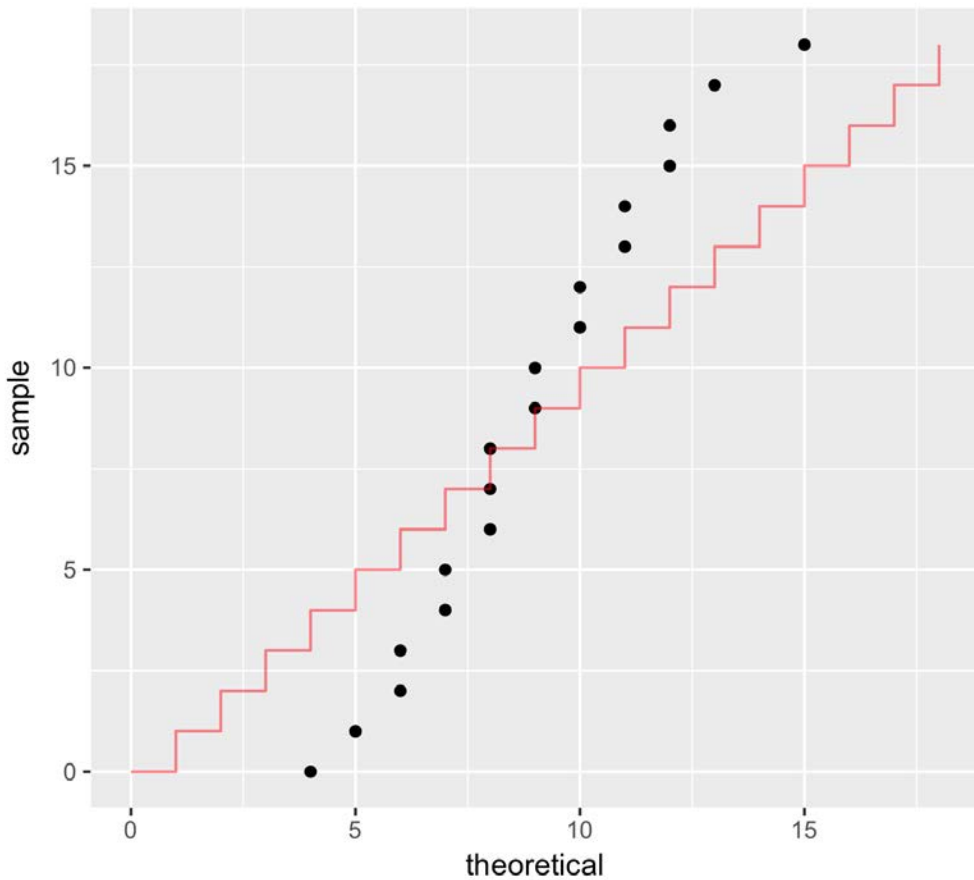
Circ Cardiovasc Qual Outcomes. 2018;11:
doi: 10.1161/CIRCOUTCOMES.117.004351
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1941-7705. Online ISSN: 1941-7713

Supplemental Material

Quintile-quintile plot analysis of the Poisson distribution assumption

The first queuing model presented in the manuscript assumes that the number of daily discharges follows follows a Poisson distribution. To evaluate this assumption, we constructed a quantile-quantile (q-q) plot of the observed sample distribution versus the expected distribution of number of discharges from hospital per day **(Supplemental Figure 1)**. The slope of the observed distribution is greater than that of the theoretical distribution, indicating that the observed distribution has somewhat less variance than a Poisson distribution would predict. This indicates that the M/M/1 model described in the first analysis in the manuscript represents an upper bound on the waiting times.



Supplemental Figure 1. Quintile-Quintile plot of the sample distribution versus the theoretical distribution of the number of discharges from the hospital per day.

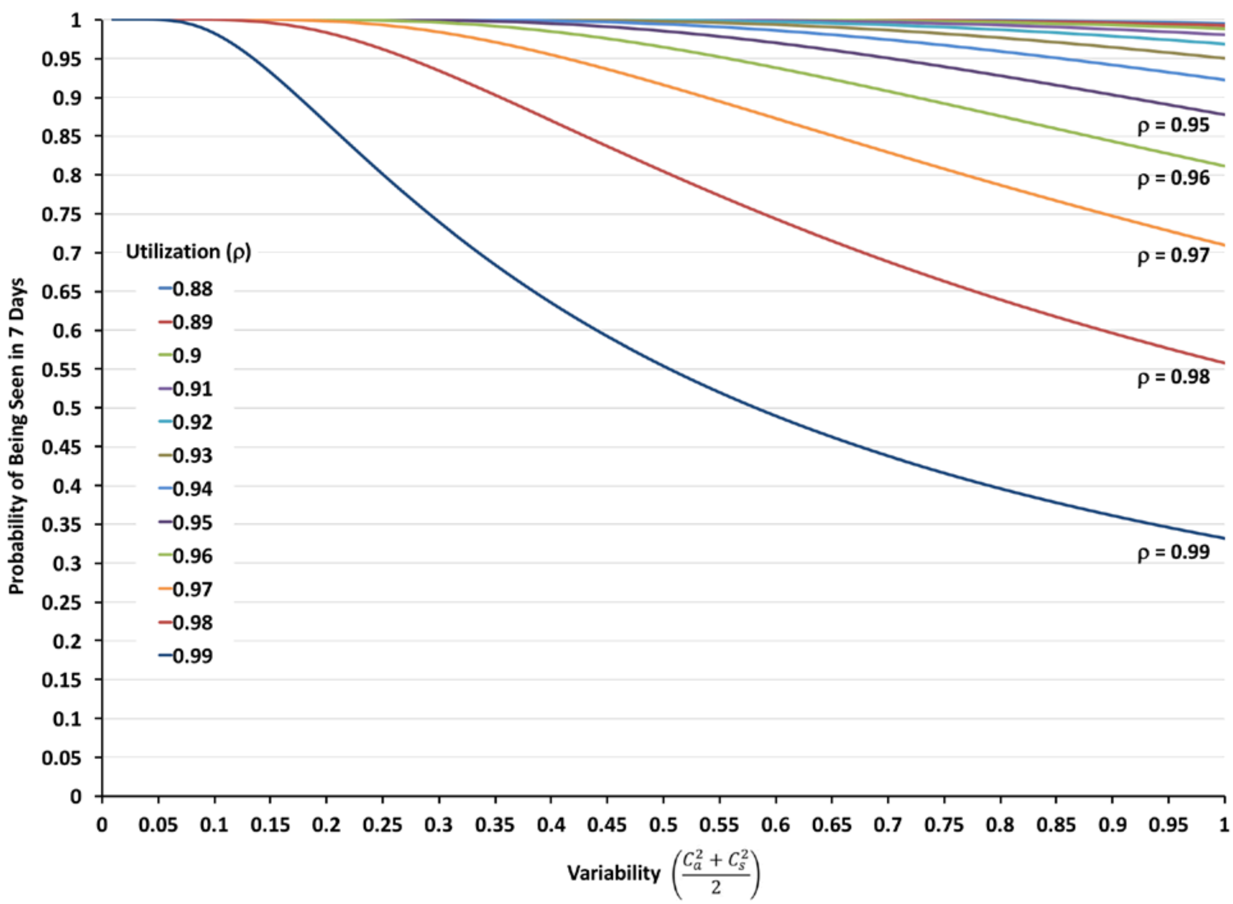
Sensitivity analysis of estimated wait time to utilization and variability terms

Equation 1 in the manuscript is reproduced below:

$$\mathrm{E}[W_q] = m\left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2 + C_s^2}{2}\right) \qquad (1)$$

This describes estimated wait time ($\mathrm{E}[W_q]$) as a function of 3 critical parameters: *m*, which represents average service time; the utilization effect term involving ρ which represents utilization, and the variability effect term involving C_a, the coefficient of variation of inter-arrival times, and C_s, the coefficient of variation of service times.

The key parameter we are optimizing for, the probability of being seen within 7 days of hospital discharge, is quite sensitive not only to utilization (ρ), but also to the properties of the discharge and arrival processes. The Poisson distribution has a coefficient of variation that equals 1. To the extent that the real world processes have less variance, waiting times for any level of utilization will be concomitantly reduced. Here we plot the probability of being seen within 7 days as a function of both utilization (ρ) and the variability term ($\frac{C_a^2+C_s^2}{2}$) . As variability in discharge and arrival distributions drops, relatively less spare capacity is needed to maintain adequate performance. However, as utilization rises to 0.99 and nears 1, even modest variability leads to major deterioration in performance.



Supplemental Figure 2. Probability of being seen in 7 days in discharge clinic as a function of utilization, ρ, and variability.