

RESEARCH

Uncertainty and Operational Considerations in Mass Prophylaxis Workforce Planning

Nathaniel Hupert, MD, MPH, Wei Xiong, PhD, Kathleen King, BS, Michelle Castorena, BS, Caitlin Hawkins, BS, Cindie Wu, BS, and John A. Muckstadt, PhD

ABSTRACT

Background: The public health response to an influenza pandemic or other large-scale health emergency may include mass prophylaxis using multiple points of dispensing (PODs) to deliver countermeasures rapidly to affected populations. Computer models created to date to determine “optimal” staffing levels at PODs typically assume stable patient demand for service. The authors investigated POD function under dynamic and uncertain operational environments.

Methods: The authors constructed a Monte Carlo simulation model of mass prophylaxis (the Dynamic POD Simulator, or D-PODS) to assess the consequences of nonstationary patient arrival patterns on POD function under a variety of POD layouts and staffing plans. Compared are the performance of a standard POD layout under steady-state and variable patient arrival rates that may mimic real-life variation in patient demand.

Results: To achieve similar performance, PODs functioning under nonstationary patient arrival rates require higher staffing levels than would be predicted using the assumption of stationary arrival rates. Furthermore, PODs may develop severe bottlenecks unless staffing levels vary over time to meet changing patient arrival patterns. Efficient POD networks therefore require command and control systems capable of dynamically adjusting intra- and inter-POD staff levels to meet demand. In addition, under real-world operating conditions of heightened uncertainty, fewer large PODs will require a smaller total staff than many small PODs to achieve comparable performance.

Conclusions: Modeling environments that capture the effects of fundamental uncertainties in public health disasters are essential for the realistic evaluation of response mechanisms and policies. D-PODS quantifies POD operational efficiency under more realistic conditions than have been modeled previously. The authors’ experiments demonstrate that effective POD staffing plans must be responsive to variation and uncertainty in POD arrival patterns. These experiments highlight the need for command and control systems to be created to manage emergency response successfully. (*Disaster Med Public Health Preparedness*. 2009;3(Suppl 2):S121–S131)

Key Words: emergency preparedness, antibiotic prophylaxis, decision support models, theoretical models, anthrax

The global public health response to the 2009 influenza A (H1N1) pandemic will include large-scale vaccination campaigns, which likely will involve the creation of vaccination centers at which a portion of or the entire population of a given geographical region will receive medications or vaccinations within a short period of time. For the past several years, US public health agencies have planned for rapid countermeasure dispensing campaigns in response to infectious disease and other emergency scenarios.^{1–3} These clinics, called points of dispensing, or PODs, are now an integral part of emergency response planning as part of the Centers for Disease Control and Prevention’s (CDC’s) Cities Readiness Initiative. Large metropolitan areas such as New York City, for example,

have made public plans to activate hundreds of PODs within hours of certain health emergencies.

The operational effectiveness of POD systems depends crucially on many factors.^{2,4} Public health planners must determine how many PODs to deploy, where to locate them, and how to design and staff each individual POD. In many POD layouts, the tasks associated with dispensing medications or other countermeasures (eg, collecting patient information, prescribing a dosage or medication, packaging drugs) are divided among several service stations. Patients move from station to station in 1 of a few possible paths, depending on their specific conditions and circumstances. The efficiency of POD-based mass

prophylaxis systems may be conceptualized as consisting of operations at 2 interrelated levels: at the level of individual PODs that make up a geographic network of dispensing facilities and at the level of the network as a whole. The efficiency of each POD depends on a number of factors, crucially whether staffing at each station is sufficient to meet patient demand. Staffing plans may be based on ad hoc estimates, past experience, or quantitative models, which in turn depend for their accuracy on assumptions made about the lengths of time required to process different patient types (eg, families of different sizes, elderly people, disabled individuals), the percentages of various patient types that follow each possible route through the POD, and the pattern of arrivals throughout the time the POD is in operation. A number of research groups have developed modeling environments for designing and evaluating the performance of alternative POD staffing plans. Three of the most detailed systems have been developed at Weill Cornell Medical College, Georgia Institute of Technology, and the University of Maryland. These research efforts are summarized below.

Hupert and colleagues at Weill Cornell Medical College developed the Bioterrorism and Epidemic Outbreak Response Model (BERM), a tool that has been widely used since 2004 to calculate POD staffing levels.⁵ Although early versions of the model used algebraic calculations in an Excel worksheet, the current version (<http://www.simfluena.org>) is a Web-based optimization and simulation tool that assumes patients arrive at each POD according to a stationary Poisson process, which can represent naturalistic variability around a constant mean. Using a modified Jackson queuing network model and an optimization algorithm, BERM calculates the minimum staffing levels needed at each station to achieve a desired level of performance. Although these results are beneficial as a starting point for planners, it is highly unlikely that in the absence of some external flow control, patients will arrive at PODs at a constant mean rate.

At the Institute for Systems Research of the University of Maryland, Hermann and others developed the Clinic Planning Model Generator (CPMG, downloadable at <http://www.isr.umd.edu/Labs/CIM/projects/clinic>), another queuing model-based capacity-planning tool that estimates POD performance based on user-specified inputs and allows patients to arrive either individually or in batches (as if people came by buses or children came with school classes). Aaby and colleagues⁶ from the Montgomery County (Maryland) Department of Health and Human Services used the results of a live exercise to validate their inputs for this simulation model, and then evaluated different clinic designs and operational policies using this model. CPMG was created to quickly provide quantitative estimates of appropriate POD staffing levels and to analyze the performance of the clinic based on those suggestions. One powerful aspect of CPMG is its ability to create a customized clinic layout; however, like the original BERM, the model assumes stationary arrival rates when it calculates the staff required for each station, patient

throughput, patient time spent in the system, and the effect of changing the number of stations. Thus, although CPMG is helpful in evaluating different clinic designs and suggesting layout and operational guidelines, its output is similarly potentially unreliable because stationary patient arrival rates are unlikely in a real emergency.

Partly in response to these limitations, Lee and colleagues at the Georgia Institute of Technology developed RealOpt, a tool that can be used to evaluate staffing levels and develop a customized layout for a POD.⁷ The model uses a heuristic-algorithmic design and may be used in real time as a decision support system during an actual dispensing campaign. RealOpt is intended by its creators to aid in strategic planning before and operational modeling during an emergency mass-dispensing event, because its primary aim is to rapidly recalculate locally optimal staffing levels during operations. This model's focus on local optimization poses an interesting question: does optimizing locally always translate into increased global efficiency of the POD network, and, if not, what general lessons for network design may be drawn from its results?

Although these 3 tools provide POD planners with the means to calculate human resources and to determine results of certain staffing level decisions, they do not sufficiently represent the fundamental uncertainties that would exist in any public health emergency situation, nor do they link the effects of that uncertainty to requirements for prophylaxis systems that quickly and reproducibly respond to changing operational environments. We sought to investigate these issues with a new model that more fully represents variation and uncertainty in patient arrivals, allows calculation of staff requirements to match patient demand over time, and considers resources to fulfill patient needs.

To accomplish these goals, we created a flexible Monte Carlo simulation model called the Dynamic POD Simulator (D-PODS) that represents the dynamics of patient flow within a POD in a probabilistic way. This experimental platform permits exploration of potential outcomes of different mass prophylaxis campaign designs under varying patient demand profiles. We used this model to attempt to answer 4 key operational questions about POD systems: What are the operational consequences of trying to match patient arrival patterns with time-varying staffing patterns? If staffing levels are changeable, how sensitive do they need to be to operating conditions in order to have an appreciable outcome benefit? Do staffing requirements for a multi-POD system depend on the number of PODs in the network, and if so, by how much? What type of command and control system would be required to achieve these modeled staff reassignment plans?

METHODS

Model Overview

D-PODS allows users flexibility in designing POD station layouts; patient types, routing, and care requirements (represented as probabilistic station- and type-specific processing times), and

non-stationary type-specific patient arrival processes. We used the output of this model to evaluate the consequences of operating the POD system in a variety of operating scenarios and under varying command and control regimes.

When determining the optimal staffing at each station in the POD, D-PODS uses a “waiting time tolerance” to determine staffing levels. This tolerance is the maximal average time spent waiting in each queue and can vary by station and time of day; however, in all of the runs reported here, it was set to 5 minutes. The D-PODS staffing calculator provides estimations that are based on standard G/G/s queuing approximations.⁸ These queuing approximation functions generate staffing levels based on 3 inputs: an expected patient arrival pattern, a probability distribution of processing times for each station, and a desired average patient waiting time for each station (W). If the patient arrival rate during the staffing interval is constant, then D-PODS will calculate staffing levels such that the average patient waiting time will approach W in the limit as the number of patients served increases to infinity. If patient arrival rates are not constant, then the staffing calculator conservatively estimates the arrival rate using the maximum patient arrival rate during the interval. This has the effect of making the average patient waiting time lower than the maximum allowed, W .

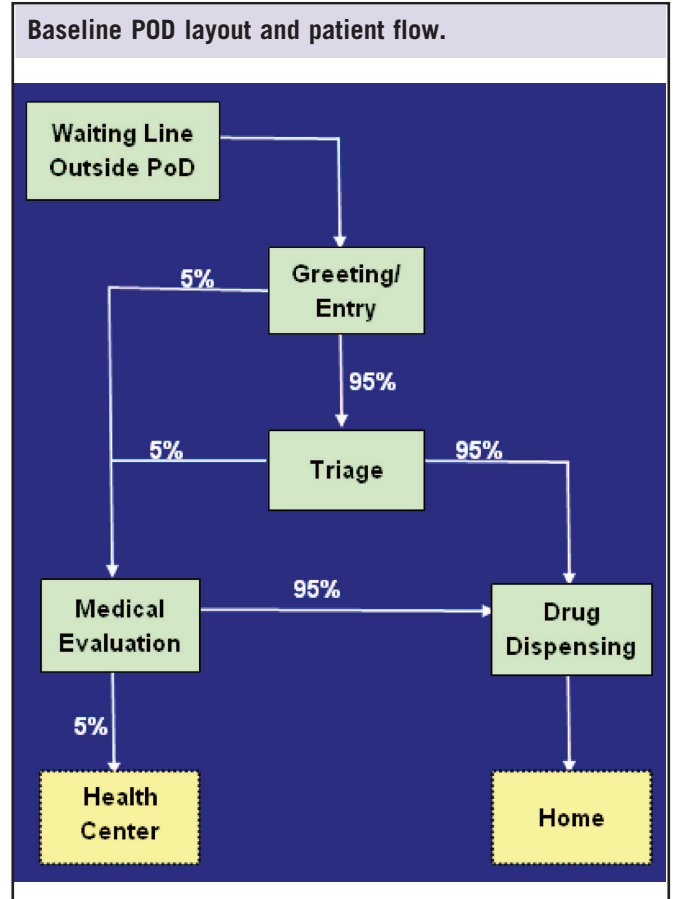
Individual POD

For the purpose of the analyses reported here, we chose to represent and evaluate the operations of a generic 4-station POD, including greeting, triage, medical evaluation, and drug dispensing.⁴ The station layout and routing probabilities are shown in Figure 1. Patients of various types (eg, single adult, family, limited mobility) arrive at the POD according to either a stationary (fixed average rate) or nonstationary (moving average rate) Poisson process and move from station to station according to transition probabilities. The number of staff at each station may vary over the simulated time horizon as described below. At each station, patients wait in a queue for an available staff; service times are assumed to be triangular (minimum-mode-maximum). Once they have received service, patients immediately join the queue for the next station or depart from the POD; we assume that patient travel time within a POD is negligible. For simplicity, in the analyses presented here we did not consider limitations on the POD’s total building capacity, although D-PODS does permit users to set such a limit. We assume that no patients abandon their queues, no matter how long they have waited.

Model Parameters

To illustrate how the simulation model can be used to assess the effectiveness of POD system management policies, we have chosen parameters to reflect a network that is set up to respond to an airborne anthrax attack. Studies have shown that antibiotic dispensing to prevent symptomatic inhalational anthrax is most effective if administered within 2 to 3 days of exposure.^{9–13} Given the expected delay for detection

FIGURE 1



and response to a covert anthrax release, current guidance from the CDC’s Cities Readiness Initiative calls for antibiotics to be dispensed to the entire affected population within 48 hours of the decision to do so.^{1,14} We, therefore, set the duration of the prophylaxis campaign to 48 hours and assumed that patients arrive for either 18 or 22 hours each day (with PODs closed at other times for restocking and cleaning). Under the assumption of a mean per-POD throughput rate of 500 patients per hour, each POD is expected to serve a total of 9000 to 11,000 people per day (individual patients or heads of household). To further simplify our analysis, we assumed that all arriving patients are of the same “type,” which means that the service time distributions are the same for all patients.

Staffing Experiments

We consider 2 types of staffing policies: a constant staffing policy and 1 of 4 time-varying policies (plans 1–4). Under the constant staffing policy, staff at each POD station remain unchanged throughout the simulated POD activation, whereas these levels change over time when a time-varying policy is used. Under plans 2 through 4, we assume a pool of available staff on-call who may be brought into the POD or sent away throughout the work day in 2-hour increments, corresponding to arrival rate. Plan 1 assumes that there is the perfect ability to forecast and adjust staff to accommodate the maximal arrival

rate seen in the next 2-hour shift; in this sense, it is the most fictional of the staffing plans. Plan 2 assumes that 30 minutes before each potential staffing change, POD managers can observe the average patient arrival rate to the POD (L) that is occurring at that moment. Using the Buzacott and Shanthikumar queuing formula mentioned above, a set of staffing levels can be calculated to keep the average patient waiting time to 5 minutes at each station, assuming that patients arrived at this projected rate of L patients per hour. The staffing level in the next 2-hour time segment is then adjusted to meet workforce requirements resulting from this projection.

To better understand staffing plan 2, consider the following example. Suppose that at 4 PM, we can change the number of staff working at each station in the POD; this may involve calling in new staff or sending people home, as well as shifting workers between stations. However, we need to declare our decisions to the workforce by 3:30, so that current workers can prepare to move around or go home and new staff have time to travel to the POD. So, just before 3:30, we determine the average current patient arrival rate for the past half hour (suppose that on average, we have been seeing 50 patients arrive every 5 minutes, which is equivalent to 600 patients per hour). We then use the Buzacott and Shanthikumar queuing formula to determine the optimal staffing levels for each station in the POD, assuming that patients continue to arrive at an average rate of 600 patients per hour. At 6 PM we will have a chance to reassign and rearrange staff once again, so we will repeat this estimation process at 5:30 PM.

Staffing plan 3 is similar to plan 2, except that staffing levels are based on patient arrival rates that occurred 1 hour, not 30 minutes, before each staffing change. In the example above, if we were set to have a staffing change at 4 PM, we would calculate staffing levels just before 3 PM. Therefore, if an average of 55 patients were arriving every 5 minutes from 2:30 to 3 PM, we would calculate new staffing levels assuming that patients continue to arrive at an average rate of 660 patients per hour (which is equivalent to 55 patients per 5 minutes).

Like plan 1, staffing plan 4 presumes that we can perfectly forecast the patient arrival rates at the beginning and the end of each staffing interval. In contrast to plan 1, however, plan 4's staffing level is based on the average (not the maximum) of these 2 numbers, again using the Buzacott and Shanthikumar formula to maintain the desired per-station waiting time. In the example above, this would mean that before the staffing change must be made at 4 PM, we could look at our forecasting mechanism and see exactly the average arrival rate from 4 PM to 4:30 PM and the average arrival rate from 5:30 PM to 6 PM. We would then average these 2 numbers and use the average to calculate staffing levels using the Buzacott and Shanthikumar formula.

Model Scenarios

Patient arrivals are modeled as a Poisson process, which is commonly used to represent random arrivals to a system clus-

tered around a mean value. In the base case, the mean arrival rate remained stationary at 500 patients per hour throughout the model run. We altered these arrival conditions to create time-varying (or nonstationary) Poisson processes as noted in Figures 2 and 3. Figure 2 shows progressively increasing and decreasing arrival rates during the operation of the POD. Figure 3 defines 3 scenarios (A, B, C) that exhibit complex time-varying properties that could mimic diurnal variation in patient arrival.

Implementation

We programmed the Monte Carlo simulation in Visual Basic for Applications within a Microsoft Excel workbook. Each reported scenario was run for 10 iterations for clarity of graphical presentation; results with higher numbers of replications did not produce quantitatively different outcomes (see Appendix for details).

RESULTS

A POD clinic designed to process a stationary mean arrival rate of 500 patients per hour within the target 5-minute per-station average waiting time required 42 active staff (5 assigned for greeting, 17 for triage, 9 for medical evaluation, and 10 for dispensing). In simulation runs, a total of 17,966 patients could be evaluated and treated during two 18-hour clinic days (SD 120.2), using a total of 738 total staff hours for a patient-to-staff-hour ratio of 24.3. Average transit time through this POD was 8.4 minutes (SD 0.5 min).

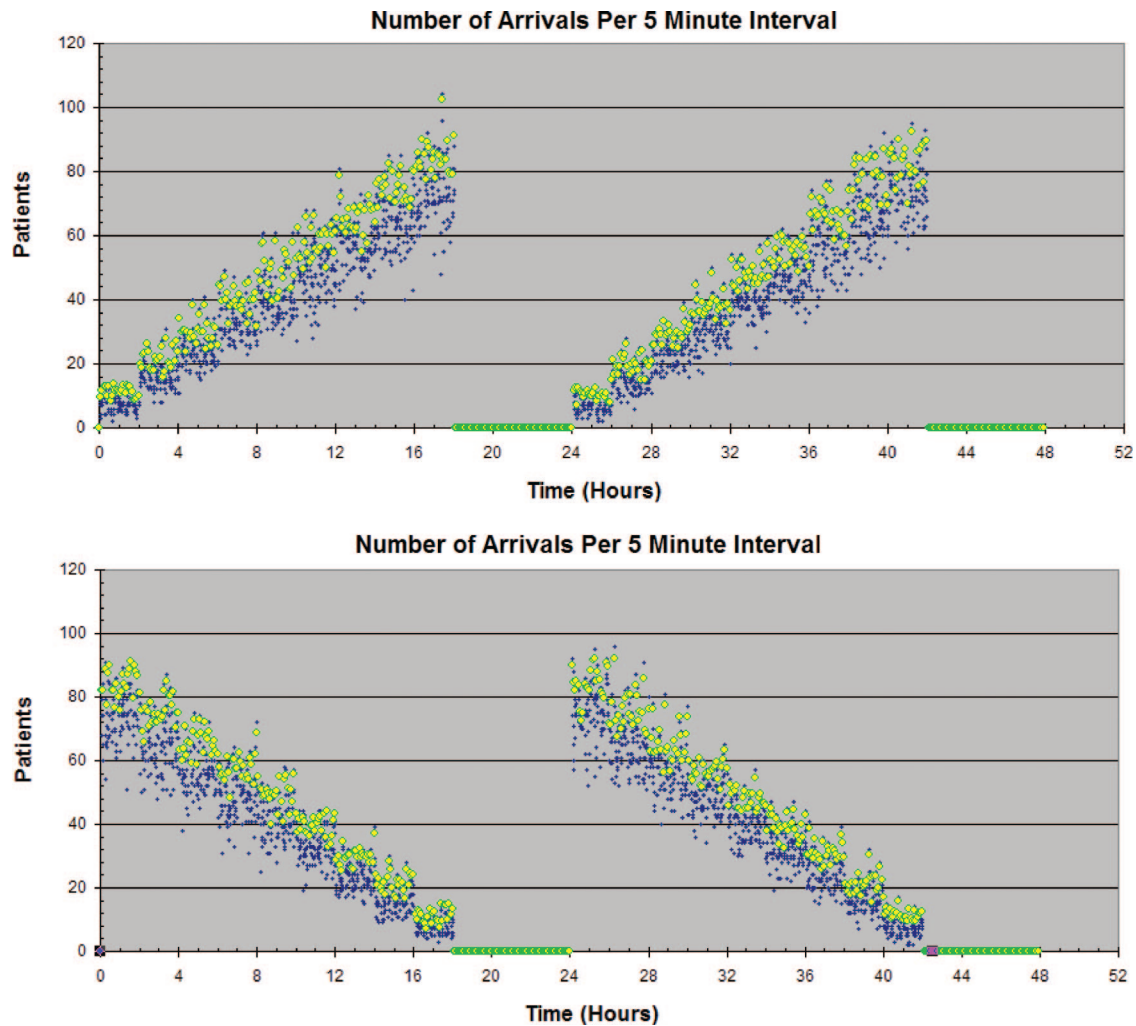
Nonstationary Arrival Rates: Simple Ascending and Descending

Our first experiment looked at the performance of this POD setup (ie, designed for a stationary mean of 500 patients per hour for 18 hours) under nonstationary arrival processes with ascending or descending arrival rates averaging between 100 and 900 patients per hour for 2-hour increments, but with an overall daily average of 500 patients per hour (Fig. 2). With steadily increasing arrivals, the daily throughput of the POD dropped by 21.7% from 9014 people (SD 124.4) to 7061 people (SD 33.2), with just over 1900 people left untreated on average at the end of each workday (Fig. 4, top). Greeting was the bottleneck station, with an average queue of 305.0 patients (SD 9.8) and staff utilization rising from 19% in the first interval to 100% for the final 6 hours of clinic operation. Because of significant overstaffing in the early hours of clinic operations, however, average staff utilization was only 76% at greeting, 75% at dispensing, 73% at triage, and 42% at medical evaluation.

In contrast, daily throughput in the descending arrival scenario was almost identical to that in the stationary arrival scenario (8996 people, SD 80.6), but with virtually no queue at the close of business (Fig. 4, bottom). This was because the front loading of arrivals led to considerably higher full-day staff utilization than in the ascending arrival scenario (greeting and dispensing 96%, triage 93%, and medical evaluation 52%), with a resulting midday average queue length of 1186.1 (SD 52.7) at the greeting station.

FIGURE 2

Increasing (top) and decreasing (bottom) nonstationary patient arrival patterns with 18-h POD (10 simulation runs, 5-min intervals, with upper 95% confidence bound highlighted by circles).



When staffing levels were adjusted according to model suggestions (ie, based on queueing equations), there were no bottlenecks or appreciable queues at any of the POD stations in either the increasing or decreasing arrival scenarios. Total 2-day throughput for the former was 17,929 patients (SD 58.2) and 17,991 patients (SD 186.9) for the latter. The variation in recommended staffing levels across arrival rates from 100 to 900 people per hour was 2 (40% of baseline staffing) to 9 staff (180% of baseline) at the greeting station; 4 (24%) to 29 (171%) for triage; 4 (44%) to 14 (156%) for medical evaluation; and 3 (30%) to 18 (180%) for dispensing. Because in these 2 scenarios patient arrival rates changed only in 2-hour increments, each of the staffing plans 1 through 4 yielded equivalent results.

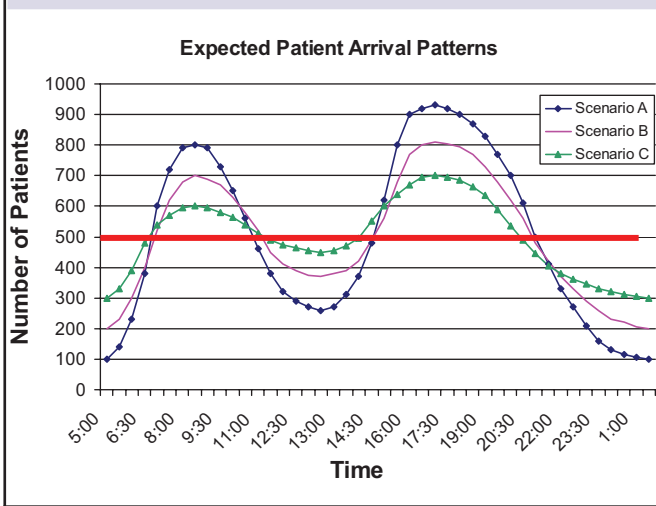
Nonstationary Arrival Rates: Diurnal Variation

We next simulated a 20-hour POD (open from 5 AM until 1 AM the following morning) with 1 of the 3 nonstationary

arrival scenarios that exhibit diurnal variation but average 500 patients per hour during a 24-hour period (Fig. 3, curves A, B, or C). When the POD operated with a constant staffing plan, average patient throughput time ranged from approximately 90 minutes for the most variable arrival regime (scenario A) to 50 minutes for the least variable (scenario C). In contrast, when the POD operated with flexible staffing plan 1, this average throughput time dropped to approximately 5 minutes under each arrival scenario (Fig. 5, top). Plan 1 also performed better than the other flexible plans, although even the worst performing of these plans (plan 3, with adjustments based on arrivals observed 1 hour before staffing interval) was able to achieve a 30% reduction in throughput time compared to invariant staffing (Fig. 5, bottom). The “cost” of these performance improvements is increased staff requirements, ranging from approximately 10% to 14% additional staff for plan 1 across the different arrival scenarios (data

FIGURE 3

Nonstationary patient arrival patterns: diurnal variation (with baseline).



shown) to approximately 5% to 7% additional staff for plan 3 (data not shown).

Of the 9 potential combinations of nonstationary arrival scenarios and flexible staffing plans, several comparisons illustrate well the tradeoffs involved in designing effective POD staffing plans under conditions of uncertainty. First, we evaluated POD functioning under arrival scenario A (with the greatest diurnal variation) comparing the constant staffing plan against plans 1 and 4, which are similar in their requirement of knowledge of both current and future patient demand at the POD. Figure 6 (top panel) shows that the queue at the greeting station under the constant staffing plan has a daily bimodal distribution reaching maxima of approximately 750 and 1500 to 2000 people. When staff is adjusted to meet maximum anticipated arrival in each 2-hour period (plan 1), this queue is reduced by almost 100-fold (to intraday peaks of approximately 20 [Fig. 6, middle]). In contrast, adjusting the staff to meet the average 2-hour arrival rate (plan 4) resulted in a 10-fold decrease in queue length and smoothing of the diurnal variation in queue size (although as noted in Figure 5, bottom, this led to only a minimal difference in combined POD service and waiting times between plans 1 and 4). Queues at downstream stations (triage and dispensing) were minimal (maximum <5 people) with the constant staffing plan (because greeting formed a bottleneck); under the flexible plans these queues were minimally larger (maximum <20).

Because it would be difficult if not impossible to accurately forecast future arrival rates to a POD except when this component of POD system design is under direct control (ie, if patients are based on a schedule to the POD), staffing plans 1 and 4 may be thought of as hypothetical experiments. In contrast, plans 2 and 3 are more realistic and could be implemented in any POD with the proper communications

system and on-call staffing arrangement. In fact, the on-the-fly staffing adjustments considered in these plans are similar to the tactical information the RealOpt program is explicitly designed to provide to POD managers.

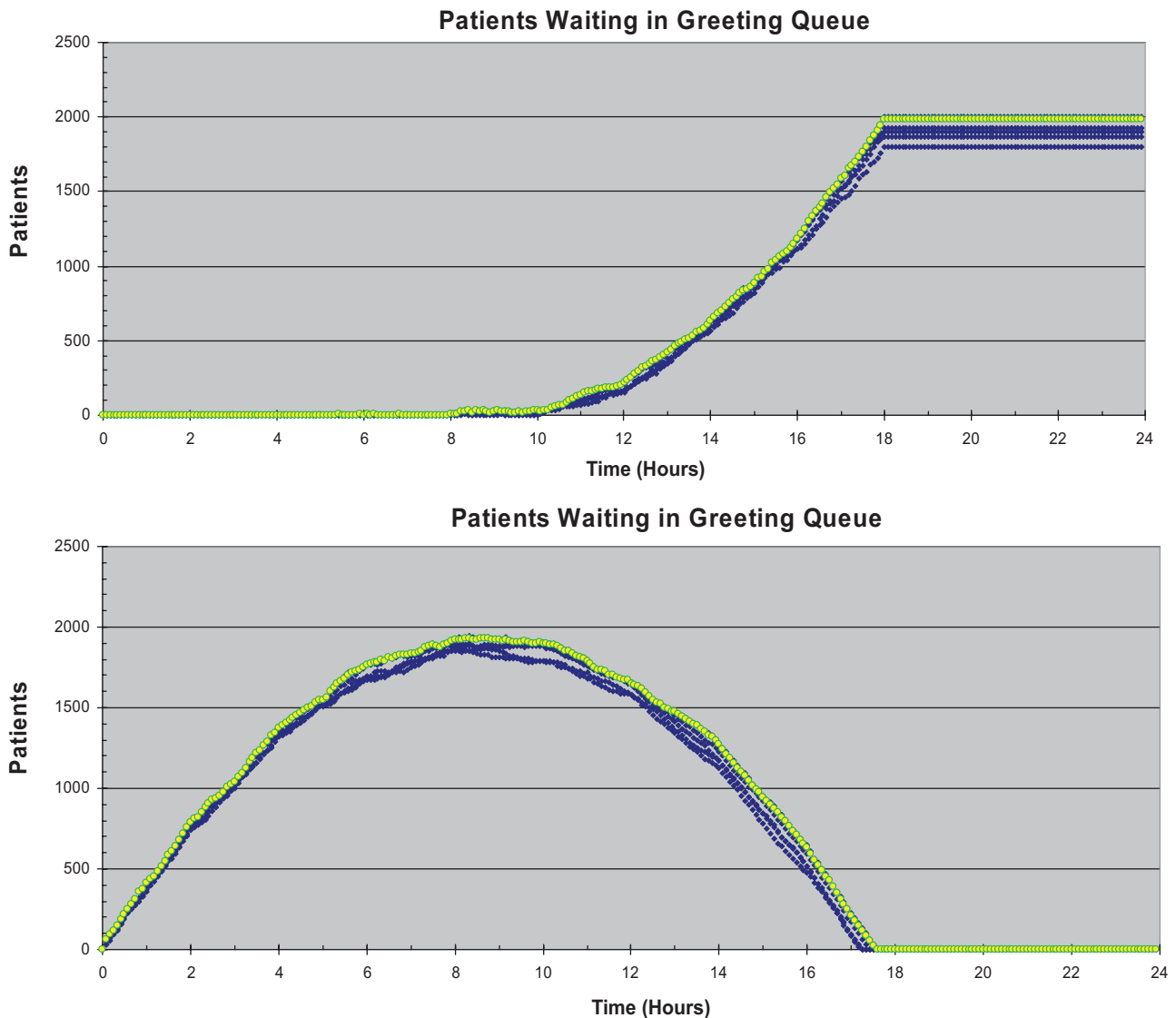
When we applied staffing plan 3 (adjustment based on arrival rate 1 hour before work shift) to arrival scenario C, which has the least diurnal variation of the nonstationary arrival scenarios, the results were notable for 3 findings (Fig. 7). First, adjustment according to this more realistic staffing plan yields only a 5-fold reduction in maximal queue length (to approximately 400 people—far larger than for plans 1 or 4), but succeeds like the hypothetical plans in smoothing out the large diurnal variation in entry queues seen under a constant staffing regime. Second, the queues at downstream stations (Fig. 7, middle and bottom) are not simple recapitulations of the greeting queue, but instead have both different maxima (averaging approximately 250 across the simulation runs shown for triage and 50 for the dispensing station) and different peak times (generally appearing in the second half of each day after the initial greeting bottleneck resolves). These queues are notable as well for their rapid appearance and disappearance as staff utilization varies due to the combined but asynchronous effect of changing arrivals and staff numbers. Third, these downstream queues demonstrate dramatic run-to-run variation, giving the graphs a scattershot appearance that only worsens with increasing numbers of simulation iterations. This means that the “mean performance” of a POD system likely will not reflect the experience of individual PODs, making coordination between incident commanders and POD managers all the more critical for managing patient movement between and through PODs.

CONCLUSIONS

POD-based mass prophylaxis operations initiated in response to public health emergencies are almost certain to be conducted against a backdrop of considerable uncertainty regarding population behavior, availability of countermeasures, and operational capability of dispensing centers. In this context, there are 2 important implications of our experimental results that show that patient waiting times and queue lengths at PODs are minimized in an environment in which staffing levels are appropriately calculated to match average patient arrival rates. In a straightforward sense, these results strongly suggest that a responsive and robust POD system can exist only if systems are in place to dynamically assess and then continuously recalculate staff requirements throughout a mass prophylaxis campaign. This requires 2 separate technical capabilities: situational awareness of queue dynamics both outside and inside PODs¹⁵ and computer modeling-based or protocol-based capability to assess how to appropriately adjust staffing and resource levels in the face of these dynamic changes. If this type of response capability is not created, then 1 of 2 outcomes will occur: Either significantly larger numbers of staff will have to be put into operation to handle unanticipated peaks in patient arrival, or patients will have long

FIGURE 4

Queue at bottleneck station with increasing (top) and decreasing (bottom) arrival rates.



waits on average to be served. Neither of these 2 outcomes is desirable. Models like RealOpt aim to accomplish just this sort of on-the-fly optimization of staffing in relation to patient influx, but such optimization tools are only as good as the command and control system that can provide an integrated picture of staffing requirements across a POD network.

These observations may be extended beyond the workforce to management of critical resources such as consumable medical supplies and durable equipment. Thus, we can extend our conclusion by stating that creating a command and control system that balances loads throughout the POD network by moving patients or resources among PODs is essential for managing system operations effectively. An extension of this

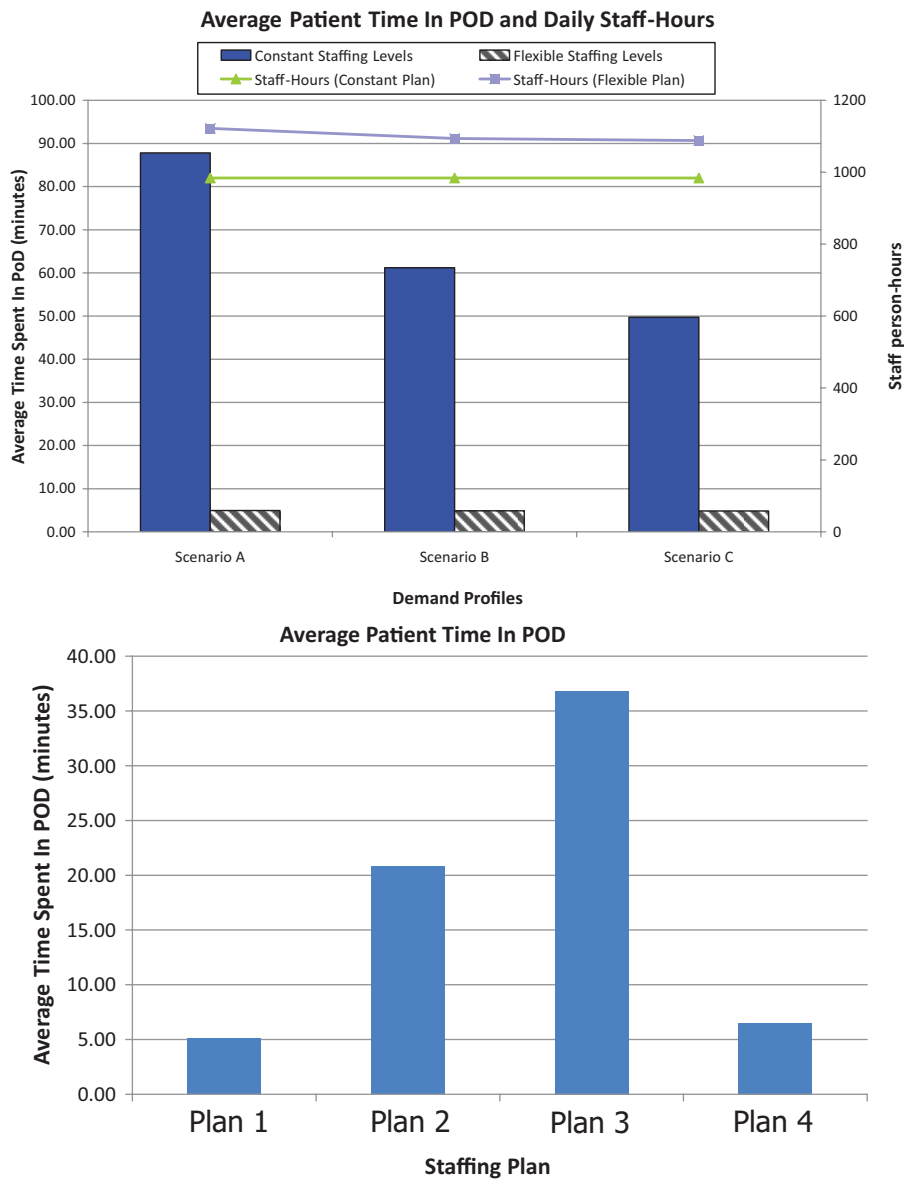
modeling approach can be used to demonstrate that to serve the same number of patients, fewer larger PODs require a smaller total staff than many smaller PODs (see Appendix). Building larger PODs would also help mitigate the impact of uncertainty by decreasing variance across POD sites. Using larger PODs may help planners create command and control environments in which demand for services at each POD can be assessed and responded to in a reasonable time frame without information overload. Determining the optimal size of a POD for a particular scenario is the focus of future work.

APPENDIX: MONTE CARLO SIMULATION MODEL

Standard Monte Carlo simulations use a single future event list (FEL), which keeps track of the events that are scheduled

FIGURE 5

POD performance metrics for nonstationary arrival scenarios A–C with constant versus flexible staffing plan 1. Comparison of POD throughput time under arrival scenario C for staffing plans 1–4.



to occur, such as arrivals and departures from different queues and service locations. To decrease run time of the model, we chose to create 5 FELs, 1 for each service station (arrivals to outside the POD, greeting, triage, medical evaluation, and drug dispensing). With 5 FELs, each list of scheduled activities in the simulation is much smaller and quicker to search through and insert a new item in sorted order. For the 5 FELs to operate in time sequence, an extra step must be added to the simulation method to determine which FEL has the earliest event at its head. The simulation compares the heads of each FEL to determine the list that contains the earliest event. The simulation then processes that event. The simu-

lation continues to operate as a typical Monte Carlo simulation by performing the tasks relating to the event found. Thus, changing a single FEL to 5 multiple FELs reduces the time for inserting an event, with a slight compromise in additional time for removing an event.

POD Size

Consider an extremely simple example: Suppose that there are 10 small PODs, to which an average of 500 patients arrive each hour. Compare this network to 1 large POD that sees an average arrival rate of 5000 patients per hour. Both systems will experience the same total expected daily patient de-

FIGURE 6

Queues at greeting for arrival scenario A (extreme diurnal variation) under constant staffing, staffing plan 1 (forecasting maximal arrival rate in next period), and staffing plan 4 (forecasting average arrival rate in next period).

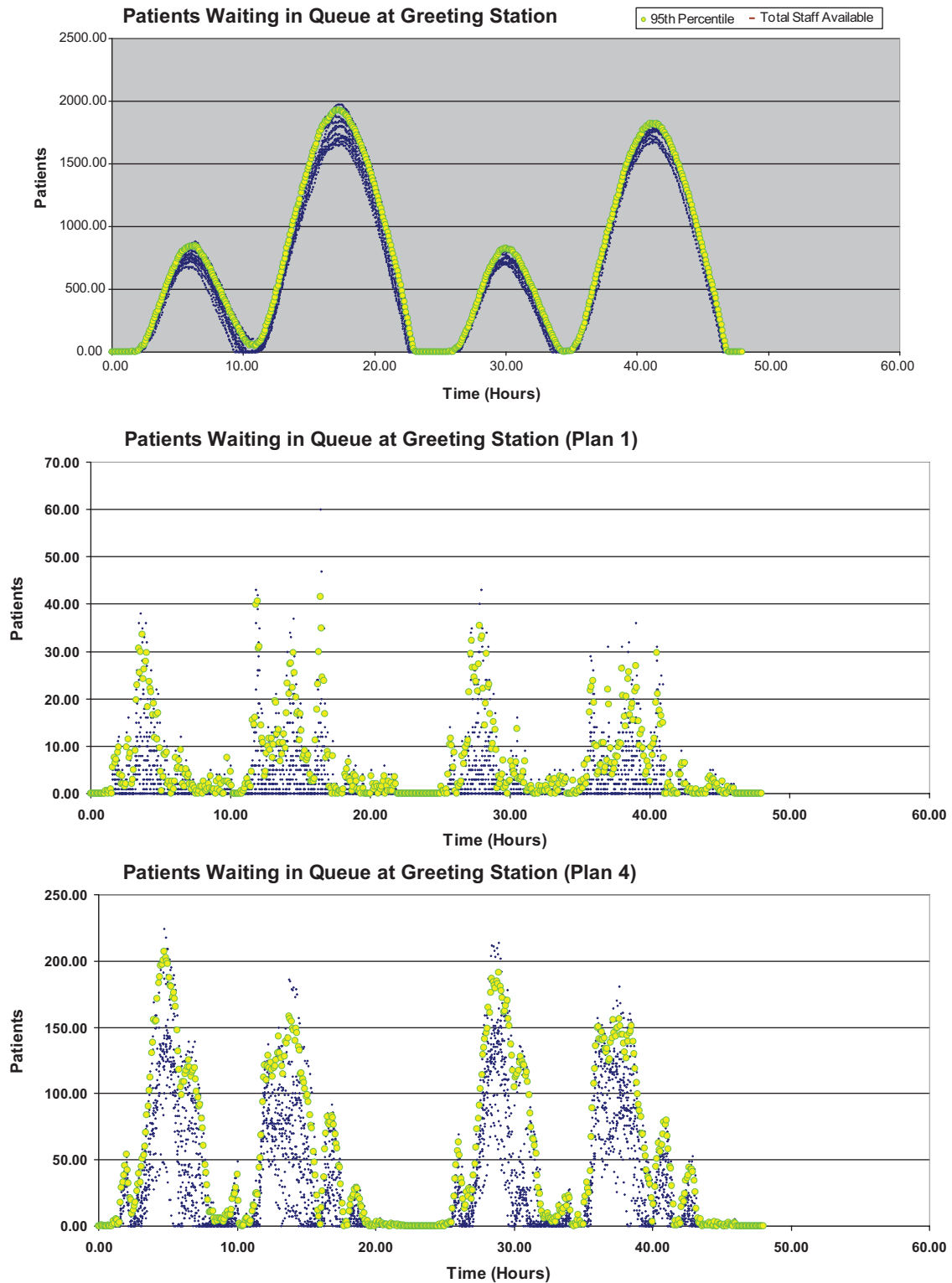


FIGURE 7

Queues at greeting, triage, and dispensing for arrival scenario C (mild diurnal variation) and staffing plan 3 (forecasting demand 1 hour before staffing change).

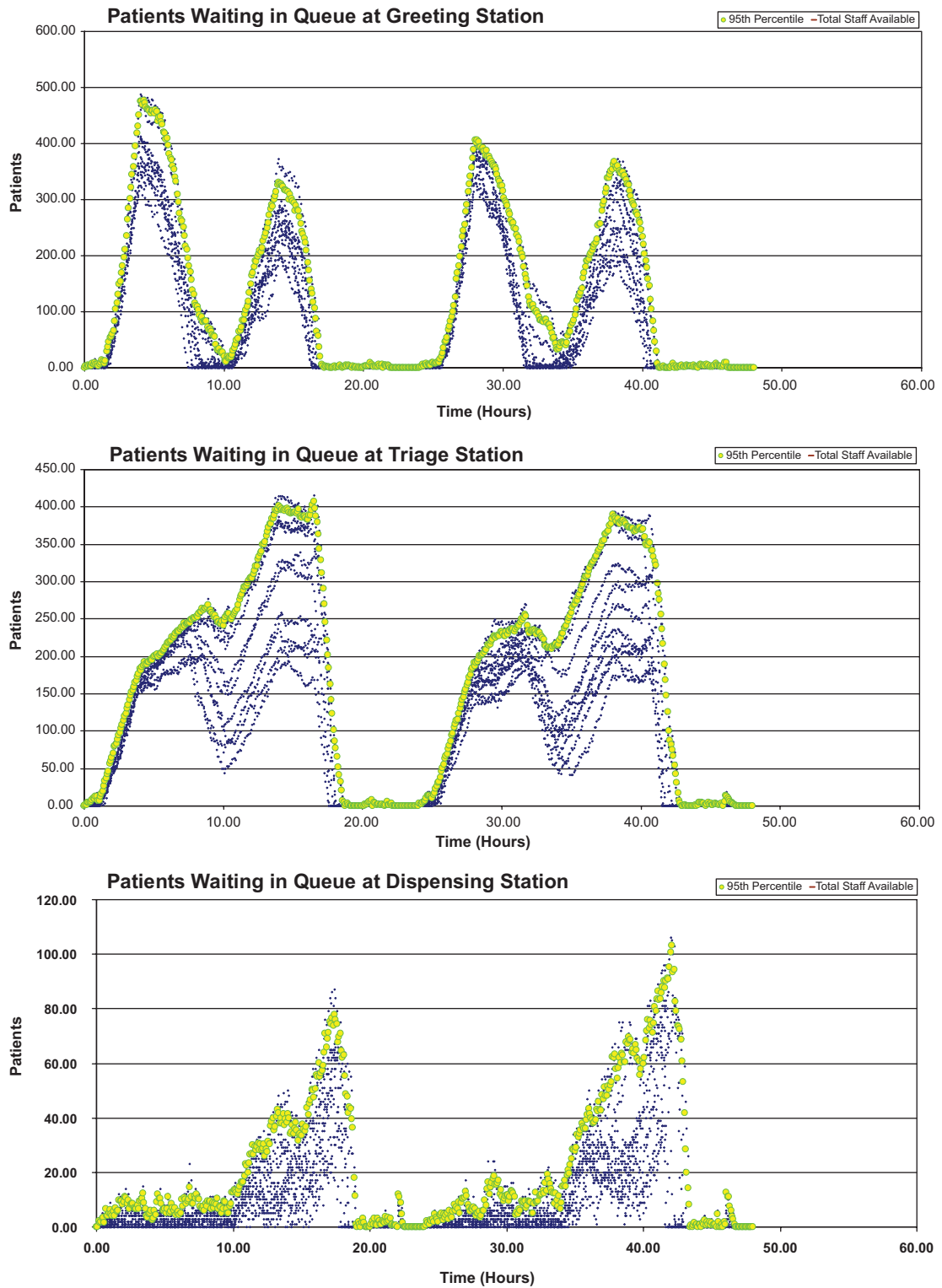


TABLE 1

Model Parameters		
Parameter	Value	Unit
General point of dispensing operations		
Duration of campaign	2	Days
Hours of operation per day	22	Hours
Baseline arrival rate	500	Per hour
Maximum average allowable queue per station	5	Minutes
Patient routing		
Probability that a patient at the greeting station will require medical evaluation	0.05	NA
Probability that a patient at the triage station will require medical evaluation	0.05	NA
Probability that a patient will be sent to a health center after medical evaluation	0.05	NA
Processing time		
Service time distribution for greeting/entry station	Triangular (0.25–0.50–1)	Minutes
Service time distribution for triage station	Triangular (1–2–3)	Minutes
Service time distribution for medical evaluation station	Triangular (2.5–5–10)	Minutes
Service time distribution for drug dispensing station	Triangular (0.5–1–2)	Minutes
Arrival distributions	Variable (Fig. 3)	
Simulation iterations per scenario	10	

mand. For each system, we calculated the staffing levels necessary to limit average patient waiting times to 5 minutes at each station within the PODs.

We find that each of the small PODs uses 984 staff-hours to provide an average patient waiting time of 6.72 minutes; this gives a total of 9840 staff-hours to operate the POD network. The large POD requires only 8664 staff-hours to provide an average patient waiting time of 6.28 minutes. Thus, almost 15% more staff-hours are required to operate a system of small PODs, which provides slightly worse service to patients. These staff-hour calculations do not even account for the additional overhead that would likely be required to operate a POD network that is more widely distributed over many locations, as seen in Table 1.

About the Authors

Drs Hupert and Xiong are with Weill Medical College; Ms King, Ms Castorena, Ms Hawkins, Ms Wu, and Dr Muckstadt are with the School of Operations Research and Information Engineering, College of Engineering, Cornell University.

Address correspondence and reprint requests to Nathaniel Hupert, MD, MPH, Department of Public Health, 402 E 67th St, New York, NY 10065 (e-mail: nah2005@med.cornell.edu).

Received for publication August 8, 2009; accepted August 28, 2009.

Dr Hupert currently serves as Director of the Preparedness Modeling Unit of the CDC.

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position(s) of the CDC.

Authors' Disclosures

The authors report no conflicts of interest.

Acknowledgments

This work was generously supported by an unrestricted grant from the Stafford Family Foundation to the Cornell Institute for Disease and Disaster Preparedness. Additional support came from the Offices of the Deans of the Weill Cornell School of Medicine and College of Engineering, Cornell University, as well as from NewYork-Presbyterian Hospital.

ISSN: 1935-7893 © 2009 by the American Medical Association and Lippincott Williams & Wilkins.

DOI: 10.1097/DMP.0b013e3181be9c39

REFERENCES

- Koh HK, Elqura LJ, Judge CM, et al. Implementing the cities readiness initiative: lessons learned from Boston. *Disaster Med Public Health Preparedness*. 2008;2:40–49.
- Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. *Proc Natl Acad Sci U S A*. 2003;100:4346–4351.
- Hupert N, Mushlin AI, Callahan MA. Modeling the public health response to bioterrorism: using discrete event simulation to design antibiotic distribution centers. *Med Decis Making*. 2002;22(5 Suppl):S17–S25.
- Hupert N, Cuomo J, Callahan M, Mushlin A, Morse S. *National Guidelines for Mass Prophylaxis Using Dispensing/Vaccination Centers (DVCs)*. Washington, DC: Dept of Health and Human Services, Agency for Healthcare Research and Quality; 2003.
- Hupert N, Cuomo J. New Model Helps Hospitals and Health Systems Better Respond to Potential Bioterrorism. Agency for Healthcare Research and Quality Web site. <http://www.aHRQ.gov/news/press/pr2003/btmodpr.htm>. Accessed September 11, 2009.
- Aaby K, Abbey RL, Herrmann JW, Treadwell M, Jordan CS, Wood K. Embracing computer modeling to address pandemic influenza in the 21st century. *J Public Health Manag Pract*. 2006;12:365–372.
- Lee E, Maheshwary S, Mason J, Glisson W. Large-scale dispensing for emergency response to bioterrorism and infectious-disease outbreak. *Interfaces*. 2006;36:591–607.
- Buzacott J, Shanthikumar J. *Stochastic Models of Manufacturing Systems*. Englewood, NJ: Prentice-Hall; 1993.
- Baccam P, Boechler M. Public health response to an anthrax attack: an evaluation of vaccination policy options. *Biosecur Bioterror*. 2007;5:26–34.
- Brookmeyer R, Johnson E, Bollinger R. Public health vaccination policies for containing an anthrax outbreak. *Nature*. 2004;432:901–904.
- Wein LM, Craft DL. Evaluation of public health interventions for anthrax: a report to the secretary's Council on Public Health Preparedness. *Biosecur Bioterror*. 2005;3:348–356.
- Wilkening DA. Sverdlovsk revisited: modeling human inhalation anthrax. *Proc Natl Acad Sci U S A*. 2006;103:7589–7594.
- Wilkening DA. Modeling the incubation period of inhalational anthrax. *Med Decis Making*. 2008;28:593–605.
- Hupert N, Wattson D, Cuomo J, Hollingsworth E, Neukermans K, Xiong W. Predicting hospital surge after a large-scale anthrax attack: a model-based analysis of CDC's Cities Readiness Initiative prophylaxis recommendations. *Med Decis Making*. 2009;29:424–37.
- Gross D, Harris CM. *Fundamentals of Queueing Theory*. New York: Wiley; 1998.