Decision Sciences
Volume 36 Number 3
August 2005
Printed in the U.S.A.

Schedule Recovery: Unplanned Absences in Service Operations*

Fred F. Easton†
Robert H. Brethen Operations Management Institute, Syracuse University, Syracuse, NY 13244-2130, e-mail: ffeaston@syr.edu

John C. Goodale
Lundquist College of Business, University of Oregon, Eugene, OR 97403-1208, e-mail: jgoodale@LCBmail.uoregon.edu

ABSTRACT

The U.S. service sector loses 2.3% of all scheduled labor hours to unplanned absences, but in some industries, the total cost of unplanned absences approaches 20% of payroll expense. The principal reasons for unscheduled absences (personal illness and family issues) are unlikely to abate anytime soon. Despite this, most labor scheduling systems continue to assume perfect attendance. This oversight masks an important but rarely addressed issue in services management: how to recover from short-notice, short-term reductions in planned capacity.

In this article, we model optimal responses to unplanned employee absences in multi-server queueing systems that provide discrete, pay-per-use services for impatient customers. Our goal is to assess the performance of alternate absence recovery strategies under various staffing and scheduling regimes. We accomplish this by first developing optimal labor schedules for hypothetical service environments with unreliable workers. We then simulate unplanned employee absences, apply an absence recovery model, and compute system profits.

Our absence recovery model utilizes recovery strategies such as holdover overtime, call-ins, and temporary workers. We find that holdover overtime is an effective absence recovery strategy provided sufficient reserve capacity (maximum allowable work hours minus scheduled hours) exists. Otherwise, less precise and more costly absence recovery methods such as call-ins and temporary help service workers may be needed. We also find that choices for initial staffing and scheduling policies, such as planned overtime and absence anticipation, significantly influence the likelihood of successful absence recovery. To predict the effectiveness of absence recovery policies under alternate staff ing/scheduling strategies and operating environments, we propose an index based on initial capacity reserves.

Subject Areas: Labor and Staff Planning, Mathematical Programming/ Optimization, Service Operations, Staff Planning, and Workforce Scheduling.

---

*This research was partially sponsored by the Center for the Creation and Management of Digital Ventures at Syracuse University. We gratefully acknowledge the helpful suggestions from an associate editor and two anonymous referees to improve an earlier version of this article.

†Corresponding author.

INTRODUCTION

Unplanned employee absences in the U.S. service sector consume 2.3% of all scheduled work hours (Bureau of Labor Statistics, 2001) and according to some surveys, 15–20% of total payroll expense (Mitchell, 2001; Anonymous, 2002; Robinson, 2002). Absence costs include direct expenses such as sick leave and salary continuation benefits (Ferguson, Ferguson, Muedder, & Fitzgerald, 2001). They also include indirect expenses such as the wages and overtime paid to co-workers or temporary workers to replace absent employees, the value of lost production, the administrative effort to secure qualified replacements, and the productivity/quality losses that occur when employees are reassigned to unfamiliar positions (Allen, 1981; Hinkin & Tracey, 2000). These costs motivate many of the more than 500 scholarly papers and books written about employee absenteeism in recent years (Harrison & Martocchio, 1998).

The absenteeism literature focuses largely on the explanation (Steers & Rhodes, 1978; Steel & Rentsch, 1995), measurement (Watson, Driver, & Watson, 1985), and mediation (Buschak, Craven, & Ledman, 1996) of unplanned absenteeism. However, policies for managing the immediate operational consequences of unplanned absences, which occur whenever employees are not present during their scheduled work hours, are rarely discussed. Unplanned absences temporarily reduce effective service capacity, often with minimal advance notice. Capacity losses from even very short-term unplanned absences, such as tardy arrivals or early departures, have the potential to disrupt operations. Operationally, it matters little whether an unplanned absence is voluntary (culpable) or involuntary; the impact on capacity is the same.

We consider absence recovery to be the managerial response to short-notice capacity losses due to unplanned absences. The recovery process first acknowledges unplanned capacity losses, then generates and evaluates feasible alternatives to replace the lost capacity, and finally selects and implements an appropriate option. In this article, our goal is to characterize the absence recovery problem and analyze both active and passive absence recovery options. With passive absence recovery, organizations simply disperse the work of missing employees among the workers who did report for duty as scheduled. Active absence recovery can be accomplished with holdover overtime (keeping people on duty a few extra hours beyond the scheduled end of their shift) and/or call-ins (activating an on-call employee enjoying a scheduled day off), or by contracting with temporary external workers to cover unplanned absences (Moore, 1965; CTI, 2002). Each absence recovery option provides different capabilities, availability constraints, and relevant costs.

Absence recovery effectiveness may often be affected by the firm's short-term staffing and scheduling decisions, which govern the size and capabilities of the workforce and how they are deployed over time. In the call-center industry, for example, some managers inflate the forecasts of hourly labor requirements by an allowance factor for anticipated absenteeism (Gans, Koole, & Mandelbaum, 2003). To effectively counter capacity losses due to attendance problems and ensure that enough scheduled workers are present to satisfy target service levels, some consultants suggest that the minimum hourly labor requirements should be grossed up by as much as 10–40% (Chadwick-Jones, 1981; Durr & Matan, 2002). The main advantage of this anticipatory staffing and scheduling strategy is that it simplifies the administrative effort needed to recover from unplanned absences and reduces the need for active absence recovery efforts such as overtime, call-ins, or temporary workers. However, it also significantly increases per-capita costs such as fringe benefits and training expenditures.

Other staffing strategies may exacerbate the impact of unplanned absences and constrain recovery options. For example, in the U.S., 29% of all employees work more than 40 hours each week (U.S. Census Bureau, 2002b). Their employers often rely on scheduled overtime to help avoid hiring additional staff. Although planned overtime helps reduce total wage and benefit expenses, it tends to result in larger voids whenever employees assigned to overtime schedules are absent. Furthermore, planned overtime staffing often leaves fewer workers who are available and willing to accept additional, short-notice overtime assignments to cover unexpected capacity shortages.

Ideally, absence recovery mitigates the effects of unplanned absences without serious harm to customers, employees, or profits. In most service organizations, the basic absence recovery process is very similar. After learning of an unplanned absence, managers first evaluate feasible alternatives for covering the loss of productive capacity (including "do nothing") by weighing the costs and benefits of each, then selecting the most promising option and implementing it. Although estimating the costs of each absence recovery option tends to be straightforward in most organizations, estimating the benefits of service recovery is often more complicated and perhaps even industry-specific.

In this article, we model absence recovery decisions for multiserver queueing systems that provide discrete, pay-per-use services for impatient customers. We assume that system revenue varies with the number of completed service transactions and that customers tend to abandon the queue when waiting times become excessive. Fast-food establishments, retail stores, walk-in medical clinics, inbound retail call centers, transportation networks, and automobile repair businesses are a few examples of the kinds of enterprises that often have these characteristics. An excellent, detailed example of such a system is described in the Sof-Optics, Inc. (A) case (Sasser et al., 1991). However, our approach may not be appropriate for systems in which it is difficult to estimate the expected marginal revenue produced by successive units of labor. Examples of such systems include those that maintain a continuous, long-term relationship with their customers and impose high switching costs (an insurance company or police protection), systems in which waiting costs are borne primarily by the customer (some government services), or in which monetary opportunity costs for service delays are difficult to measure (e.g., a nursing unit in a hospital).

Our purpose is to characterize and evaluate the performance of alternate absence recovery strategies under various staffing and labor scheduling policies. In addition to the important insights about the process of absence recovery that are suggested by our experimental data, we believe this project provides at least three other important contributions to the literature. First, to the emerging literature dealing with real-time schedule adjustment (Thompson, 1999; Hur, Mabert, & Bretthauer, 2004) we formally introduce and model the absence recovery problem for service operations. Second, we identify worker absenteeism as a heretofore overlooked source of uncertainty for labor scheduling decisions and propose a new labor scheduling model that represents that uncertainty. Third, we propose a new, marginal revenue-based method for determining the ideal staffing levels for each planning interval. This approach, based on estimates of losses due to reneging, is a distinct alternative to simpler, fixed-response time approaches that are often used to determine minimum staffing requirements for multiserver queueing systems providing pay-per-use services for impatient customers.

In the following sections, we:

(i) Review the economic consequences of absenteeism, describe popular absence recovery policies, and discuss staffing strategies that may affect absence recovery performance.
(ii) Review staffing/scheduling models, describe our labor scheduling model for systems facing absenteeism, and introduce our absence recovery model. This model reduces to an easily solved integer generalized network for certain recovery policies.
(iii) Describe our experiments to test staffing and recovery policies and present our results.
(iv) Present our conclusions, discuss some limitations of our study, and suggest avenues for future research.

ECONOMIC CONSEQUENCES AND RECOVERY FROM ABSENTEEISM

Widely reported surveys by Commerce Clearinghouse (CCH, 2003) and the Integrated Benefits Institute (IBI, 2000) estimate employers spend \$750–\$800 per employee per year for unplanned absences. However, these figures do not include the indirect costs of unscheduled absences: overtime pay for other employees, costs of temporary workers who replace absent employees, productivity losses, and supervisory time spent rearranging work schedules. In their survey of 11 major telecommunications firms, for example, IBI (2000) found that absence-related benefit costs and productivity losses cost these employers \$11.5 billion in 1999, 8% of their total revenue.

Out-of-pocket benefit costs are easy to identify; the measurement of productivity losses due to absenteeism is more complex. Unplanned absences reduce service capacity and often increase waiting time. The utility of the service usually decreases with waiting time, especially when the expected duration of the wait cannot be estimated (Osuna, 1985). In many systems, increased disutility reduces the likelihood of subsequent trials (Anderson & Sullivan, 1993) and the average purchase amount (Andrews & Parsons, 1993). Increased waiting time also increases the likelihood that arrivals will abandon the queue before service is initiated (Ittig, 2002).

In this article, we focus on systems that incur an expected economic loss equal to the average revenue per transaction whenever a customer abandons the queue. To estimate these losses, we must first characterize reneging behavior in queues. Suppose patience (or willingness to wait) is unique for each customer and independent of other customers. In such cases, the probability that customers will continue to wait tends to resemble an exponential process (Carmon, Shanthikumar, & Carmon, 1995; Whitt, 1999). This assertion may not hold for all queueing situations, however (Zohar, Mandelbaum, & Shimkin, 2002).

Abandonment or reneging reduces the effective arrival rate (λeff_t), and sales/period, but it also speeds up the line for those still waiting. To estimate the potential for revenue loss in a multiserver queue with impatient customers, Easton and Goodale (2002) computed the number of arrivals that remained in the system until service completion. They assumed that during period t, customers arrived at a system at rate λ_t (Poisson-distributed) and abandoned the queue after waiting an average of α^{-1} (exponentially distributed). With s identical servers working during period t, each operating at mean service rate μ (Poisson-distributed), they showed that server utilization U_{eff} is:

$$
\begin{aligned}
U_{eff} = \frac{\lambda eff_t\left(\lambda_t, \mu, s, \alpha^{-1}\right)}{s\mu} &= \sum_{i=0}^{s}\left(\frac{i}{s}\right)F_i + \sum_{j=s+1}^{\infty}\frac{s}{s}F_j \\
&= \sum_{i=0}^{s}\left(\frac{i}{s}\right)F_i + \left(1 - \sum_{j=0}^{s}F_j\right).
\end{aligned} \tag{1}
$$

Thus, the effective arrival rate, $\lambda eff_t(\lambda_t, \mu, s, \alpha^{-1}) = s\mu U_{eff}$, can be determined from the state probabilities F_k for an M/M/S +M queueing system, where the "+M" suffix to the standard queueing description denotes an exponential reneging distribution (Bacelli & Hebuterne, 1981). That is, for any period t,

$$
\lambda eff_t\left(\lambda_t, \mu, s, \sigma^{-1}\right) = \mu\left[\sum_{i=1}^{s} iF_i + s\left(1 - \sum_{k=0}^{s}F_k\right)\right]. \tag{2}
$$

Gross and Harris (1998) outlined the derivation of the stationary state probabilities F_k for M/M/1 +M (a single-server queueing model in which the probability of waiting patiently decreases exponentially with queue time). Whitt (1999) described state-dependent transition rates for the multiserver birth-death process with exponential reneging and balking. Garnett, Mandelbaum, & Reiman (2002) derived the state probabilities F_k for finite multiserver queues with blocking and exponential reneging (M/M/S/N +M). Simultaneously, Easton and Goodale (2002) derived the state probability equations for M/M/S/∞ +M and suggested an approximate procedure to evaluate state probabilities F_k for an arbitrary level of precision.

Because queue time generally falls with increased staffing levels, we also expect reneging to diminish (and the effective arrival rate to increase to λ_t). Let M_t be the staffing level where $\lambda_{eff}(\lambda_t, \mu, M_t, \alpha^{-1}) \cong \lambda_t$ (that is, no reneging occurs). If R_t is the average revenue per completed transaction during period t, the incremental revenue generated by the ith employee servicing customers during period t, or r_{ti}, is:

$$
r_{ti} = R_t\left[\lambda_{eff}\left(\lambda_t, \mu, i, \alpha^{-1}\right) - \lambda_{eff}\left(\lambda_t, \mu, i-1, \alpha^{-1}\right)\right], \quad \text{for } i = 1, \ldots, M_t. \tag{3}
$$

It follows from the asymptotic behavior of $\lambda_{eff}(\lambda_t,\ \mu,\ i,\ \alpha^{-1})$ that r_{ti} is non-increasing in i, the number of servers.

Table 1: Effective arrival rate and marginal revenue for various staffing levels s when $\lambda = 50$/hour, $\mu = 12$/hour, $\alpha^{-1} = 5/60$ hours, and $R = \$10$/completed transaction.

s	$\lambda eff(\lambda, \mu, \mathrm{s}, \alpha^{-1})$	r_{ti}
1	11.814	\$118.14
2	22.853	110.39
3	32.276	94.23
4	39.457	71.81
5	44.301	48.44
6	47.199	28.98
7	48.744	15.45
8	49.484	7.40
9	49.805	3.21
10	49.932	1.27
11	49.978	.46
12	49.993	.15
13	49.998	.05
14	49.999	.01
15	50.000	.01

We apply this concept in a brief example to illustrate the value of capacity lost due to unplanned absence. In Table 1, we assume the average arrival rate $\lambda = 50$ customers/hour, the mean service rate $\mu = 12$/hour, the average patience $\alpha^{-1} = 5$ minutes, and the average revenue/transaction $R = \$10$. Suppose nine workers were scheduled for duty during this period, but one was absent. The expected value of the lost production, the marginal revenue generated by the ninth worker (r_{t9}), is \$3.21 (i.e., $\lambda_{eff}(\lambda_t, \mu, 9, \alpha^{-1}) - \lambda_{eff}(\lambda_t, \mu, 8, \alpha^{-1}) = .321$ more customers renege), a useful measure for evaluating the value of alternate recovery options.

We can also use the marginal returns in Table 1 to assess the economic impact of unplanned absences. For example, suppose four employees are scheduled for duty during period t and absenteeism averages $A = 20\%$ of scheduled hours. That is, of the four workers scheduled we expect that only $(1 - .2) \times 4 = 3.2$ employees will actually report for duty during period t. The first three employees will produce \$118.14, \$110.39, and \$94.23 per hour, respectively. The fourth (fractional) employee will produce $.2 \times \$71.81$ per hour, so the expected hourly income with four scheduled employees is \$337.12. The expected hourly cost of lost productivity due to absences is the difference between the revenue produced by four reliable employees (\$394.57) and the expected income, or \$57.45.

Active recovery from an unplanned absence requires flexible, reserve capacity. In Table 2, we summarize the major options that service organizations use to recover from unplanned absences. The simplest absence recovery option is passive recovery, in which the workload of absent coworkers is dispersed among the employees who report for duty. This may be appropriate if, due to scheduling rules or the vagaries of demand, the unit's original labor schedule provided excess coverage

Table 2: Recovery options for unplanned absences.

Recovery Option	Indirect Costs	Recovery Constraints
Do nothing	- Lost sales, revenue, and goodwill - Employee resentment	Not applicable
Transfer cross-trained employees from lower-priority assignments	- Training expense - Reduced productivity - Wages paid to cover transferred absence	- Match employee skills with assignment - Cover absence in sending department
Holdover employee (extend regular shift)	- Additional wages (often overtime) - Reduced productivity	- Match shift-end with time of absence - Observe meal and rest breaks - Maximum shift and tour lengths
Call-in employee	- Standby expense - Additional wages (often overtime) - Show-up pay	- Observe minimum and maximum shift and tour lengths for call-ins - Cover absences on days off
Temporary (THS) worker	- Temporary wages and markup - Training expenses - Reduced productivity - Administrative costs	- Observe minimum assignment duration for THS workers - Recognize reduced productivity

during the periods of the unplanned absence. For smaller service units with high skill requirements, this may be the only practical option. However, the "do nothing" option may adversely affect employee morale, customer service, and the prospect for future sales (Buschak et al., 1996).

Some multitask operations invest in cross-training and recover from unplanned absences by reassigning personnel to more urgent tasks when shortages arise. However, this approach merely shifts the location of the unplanned absence to the department from which the reassigned worker was sent. Ultimately, the department with the personnel shortage must recover (often with overtime work). In addition, reassigning personnel to cover shortages in other areas may not be feasible if the sending and receiving units experience similar demand patterns. A more popular option is holdover overtime, in which employees work beyond the scheduled end of their shift to cover an emergency staffing need. Holdover overtime is the preferred absence recovery technique for 41% of the CTI (2002) survey respondents. However, some employers find that short-notice overtime increases workplace stress and exacerbates absenteeism problems. This may explain why 28% of the firms in the CTI (2002) survey cover unplanned absences exclusively from an on-call pool of trained, off-duty employees who are willing to report for duty on short notice. Work rules governing call-in assignments typically require minimum work stretches (four to eight hours), plus additional "standby" wage payments (Hirschman, 1999).

Finally, some organizations cover unplanned absences by outsourcing. Temporary help service (THS) firms (i.e., nursing registries or firms like Manpower, Inc., or Kelly Temporary Services) are labor market intermediaries, employing per diem workers whose services are sold to other employers. The THS industry exceeds $40 billion per year in sales, and provides a wide range of staffing options (Greble, 1997). However, our interest is with the coverage of unplanned absences. Surveying 1,200 firms in industries that rely heavily on THS workers (health care, business services, finance, and insurance), Mangum, Mayall, and Nelson (1985) found 23.4% of all THS appointments were for durations of one to two days, near the modal duration reported for unplanned absences (Globerson & Nagarvala, 1974).

Segal and Sullivan (1997) found negligible differences in the wages paid to temporary and permanent workers after adjusting for skills and experience. Furthermore, they report that clients pay an average markup equivalent to 40% of the THS worker's wages to cover the agency's fringe benefits, labor and administrative costs, and contribution to profit. Like call-in employees, THS workers are usually assured a minimum work stretch. Finally, with limited training, THS workers may be confined to entry-level assignments (Davis-Blake & Uzzi, 1993).

The set of absence recovery options available to front-line supervisors may be limited by their organization's staffing and scheduling strategies. In addition to staffing strategies related to the use of THS workers, at least two other strategic staffing decisions may affect a firm's ability to recover from unplanned absences. One is the adoption of anticipatory staffing strategies, in which some absenteeism is assumed and the firm staffs accordingly. For example, some firms overstaff each work period by 10–40% above estimated needs (Chadwick-Jones, 1981; Durr & Matan, 2002). That way, when some employees fail to report for duty as scheduled, extra personnel are already present to cover the shortage.

Anticipatory staffing strategies may reduce administrative effort but may also require a larger workforce and greater recruiting, training, and per-capita expenses. Per-capita expenses vary with the number of employees rather than the number of hours each employee works. Accounting for 25.8% of total labor costs (U.S. Census Bureau, 2002a), these expenses include pension and retirement benefits, life insurance and health care benefits, unemployment insurance premiums, paid vacations, sick leave, and paid holidays (Ehrenberg, 1989). Still, overstaffing may be an effective strategy if per-capita costs are low and workers are not indemnified for unplanned absences (Easton & Goodale, 2002). If workers do receive sick pay, however, optimal staffing levels may be reduced because effective hourly wages per worker increase by $1/(1 - \text{absence rate})$.

Another staffing strategy that may exacerbate the impact of unplanned absenteeism and inhibit the firm's ability to recover from unplanned absences is the use of planned overtime. Many firms substitute planned overtime for hiring (Hetrick, 2000). Although overtime work usually requires premium pay, it helps reduce the total number of permanent employees needed to staff the system, and thus helps reduce total per-capita labor expenses. Unfortunately, if an employee scheduled for planned overtime is absent, a larger block of time is affected. With fewer employees working longer hours, there are fewer off-duty employees available to respond to unplanned absenteeism.

STAFFING/SCHEDULING AND ABSENCE RECOVERY

Many service organizations with absenteeism problems face consumer demands that vary from hour-to-hour and day-to-day. Because service capacity usually cannot be stored, they risk lost sales if their services cannot be produced upon demand. Furthermore, their service capacity decisions, in the form of employee work schedules, are usually made a week or more before either customer demand or employee absenteeism is realized. These decisions are often constrained by work rules that limit the types of acceptable multiday work schedules for their employees with respect to minimum and maximum shift lengths, tour lengths, rest periods, days off, allowable shift start times and start-time variation, the proportion of full-time to part-time workers, etc. Labor scheduling decisions have intrigued researchers for half a century.

Most models for labor scheduling decisions assume 100% attendance. As defined earlier, we consider absence recovery to be the process of acknowledging an unplanned absence, planning a recovery strategy, and implementing it through changes in the labor schedule. Because labor staffing and scheduling decisions control the size of the workforce and how employees are deployed over time, they may also affect the operational and economic impact of unplanned absences and the effectiveness of absence recovery policies.

Models for labor scheduling decisions were originally cast as deterministic generalized set covering problems (Dantzig, 1954; Baker, 1976; Bailey, 1985; Gans & Zhou, 2002; Gans, Koole, & Mandelbaum, 2003). Deterministic goal programs (Keith, 1979; Andrews & Parsons, 1989, 1993) were later developed to address concerns of economic asymmetries between under- and overstaffing. Bechtold and Jacobs (1990), Mabert and Watts (1982), and others proposed heuristics to reduce the state space and improve the tractability of these problems.

Most early labor scheduling models were constrained by a vector of minimum labor requirements. Based on demand forecasts (in units characteristic of the service), these parameters were often estimated by varying target staffing levels for each period to find the best balance between expected labor costs and shortage expenses (Baker, 1976). Each period was regarded as an independent epoch, using the average hourly wage as an approximation of marginal labor costs. Implicitly, an employee could be scheduled for as little as a single period of work, an assumption that usually violates local scheduling rules (e.g., "all part-time employees scheduled to work during a particular day will be guaranteed a shift of at least four hours"). As noted by Baker (1976), when true incremental labor expenses differ from the assumed value, the labor requirements parameters determined by marginal analysis may be suboptimal. To counter this limitation, Thompson (1995) and Easton and Rossin (1996) integrated labor requirements planning with staff ing and scheduling decisions to optimize service levels. Their approaches, specifically designed for systems with stochastic service demand, avoided the need to approximate marginal labor costs. Finally, Pinker and Larson (2003) proposed a staffing model for uncertain demand that optimized the mix of regular employees and contingent (THS) workers for single-shift, back-office service operations.

Although most existing labor scheduling models still assume 100% attendance, Gans et al. (2003) reported that some call-center managers create labor schedules that have capacity cushions to help protect the system from unplanned absences. Easton and Goodale (2002) proposed an anticipatory labor staffing/scheduling model for service operations with stochastic service demand, customer patience, and employee turnover. However, neither of these approaches considers active absence recovery strategies such as holdover overtime, call-ins, or THS workers.

Because the performance of alternate absence recovery approaches may depend on the staffing and scheduling policies adopted by an organization, and we wish to consider both active and passive recovery approaches, we begin this section by adapting Easton and Goodale's (2002) staffing/scheduling model to include relevant features such as sick pay, per-capita costs, and the ability to accommodate alternate staffing strategies such as planned overtime. Like Easton and Goodale, we assume that the number of customer arrivals/hour and hourly service rates per employee are Poisson-distributed random variables with means λ_t and μ, respectively. We also assume that completed transactions earn an incremental contribution (in dollars) before labor expense, and that customers have limited patience and will abandon the queue after waiting an average of α^{-1} (exponentially distributed). Finally, we assume that absenteeism is uniformly distributed over the planning horizon, with each employee shift equally susceptible to loss. Unlike Easton and Goodale (2002), we ignore employee turnover and the effects of learning. In addition, we assume that fringe benefits and sick pay accrue to employees whether they report for duty or not. However, we exclude wage payments for the expected portion of planned overtime that will be lost to unplanned absences.

We implement alternate staffing strategies such as planned overtime by specifying the work rules that constrain the set of feasible employee schedules or tours. A tour can be uniquely defined by specifying the workdays in the planning horizon, the shift starting time and ending time for each workday, and the timing of meal and rest breaks during each shift. In general, the cardinality of the set of feasible tours increases with the amount of scheduling flexibility permitted by the organization's work rules. Mechanisms for efficiently generating feasible employee schedules from these parameters are discussed in Mabert and Watts (1982); Burns and Koop (1987); Easton and Rossin (1991); Jarrah, Bard, and deSilva (1994); Jacobs and Brusco (1996); and Brusco and Jacobs (1998); among others.

We implement staffing strategies that anticipate absences by adjusting the expected absenteeism rate parameter used in our model. This parameter helps the staffing/scheduling model manage the expected effective marginal labor cost versus marginal revenue trade-off to determine the optimal number of employees and how they should be deployed over time to maximize expected profit. Our adaptation of Easton and Goodale's (2002) staffing and scheduling model uses the following notation.

Model Parameters

Workforce Characteristics

$A =$ average absenteeism rate (total hours lost due to unexpected absences $\div$ total scheduled hours).

$F =$ average employee fringe benefit or per-capita costs.

$\mu =$ average service rate per employee.

Customer and Market Characteristics

$T =$ the number of time periods in planning horizon, indexed $t = 1, \ldots, T$.

$\lambda_t =$ forecasted mean arrival rate for period t (Poisson-distributed), $\forall\, t$.

$\alpha^{-1} =$ average customer patience, exponentially distributed.

$\lambda_{eff}(\lambda_t, \mu, i, \alpha^{-1}) =$ effective arrival rate (# arrivals − # reneging) for period t, given gross arrival rate λ_t, average service rate μ, staff ing level i, and average patience α^{-1}.

$R_t =$ average contribution per completed customer transaction, before labor costs, for $t = 1, \ldots, T$.

$M_t =$ minimum staffing level s such that $\lambda_{eff}(\lambda_t, \mu, s, \alpha^{-1}) \cong \lambda_t$.

$r_{ti} =$ the incremental revenue generated by the ith employee working during period t, or $R_t[\lambda_{eff}(\lambda_t, \mu, i, \alpha^{-1}) - \lambda_{eff}(\lambda_t, \mu, i-1, \alpha^{-1})]$.

Schedule Characteristics

$K =$ the cardinality of the set of allowable work schedules for employees, a function of the work rules governing the construction of schedules, indexed $k = 1, \ldots, K$.

$a_{tk} =$ 1 if t is a working period in schedule k, 0 otherwise, for $t = 1, \ldots, T$; $k = 1, \ldots, K$.

$O_k =$ number of scheduled overtime work periods in work schedule k (total number of work hours in tour – 40 hours).

$W =$ regular wage rate (\$/period), with overtime work compensated at rate $1.5W$.

Decision Variables and Consequence Variables

$X_k =$ number of employees assigned to schedule k, where $k = 1, \ldots, K$.

$Y_{ti} =$ the fraction of the ith employee who reports for duty during period t, where $0 \leq Y_{ti} \leq 1$, $i = 1, \ldots, M_t$, and $t = 1, \ldots, T$.

The model determines the ideal number of employees, and their schedules, to maximize expected profits. It accounts for expected revenue, scheduled labor costs, and the direct and indirect costs of unplanned absenteeism. Mathematically, the goal is to:

$$\text{Maximize} Z = \sum_{t=1}^{T} \sum_{i=1}^{M_t} r_{ti} Y_{ti} - \left(\sum_{j=1}^{K} X_j \left[F + W \left(\sum_{t=1}^{T} a_{tj} - O_j \right) + (1 - A)\, 1.5 W O_j \right] \right), \tag{4}$$

subject to:

$$(1 - A)\sum_{k=1}^{K} a_{tk}X_k - \sum_{i=1}^{M_t} Y_{ti} \geq 0, \quad \forall t \in T, \tag{5}$$

$$X_k \geq 0 \quad \text{and integer}, \quad k = 1, \ldots, K, \tag{6}$$

$$0 \leq Y_{ti} \leq 1, \quad i = 1, \ldots, M_t, \quad \forall t \in T. \tag{7}$$

The objective function (4) computes the expected contribution for the solution by determining expected revenues and expenses. The first term sums the incremental revenue r_{ti} generated by each employee expected to report for duty during each period of the planning horizon. The second term computes the expected labor expenses for the solution. Summing over all feasible tours, it adds the per-capita expenses and regular wages for all employees scheduled for duty during the planning horizon, whether or not they report for duty. Although we assume the wages for scheduled regular time work are guaranteed, the objective excludes wage payments for the portion of scheduled overtime expected to be lost due to unplanned absences.

Because the labor staffing/scheduling problem is usually solved one to six weeks in advance of the service date, managers will not know with certainty whether everyone scheduled for duty will actually show up. However, the number of scheduled employees who are actually present during period t can be modeled as a series of $n = (\sum_{k=1}^{K} a_{tk}X_k)$ independent Bernoulli trials (Feller, 1968). To illustrate, let A be the expected absenteeism rate and assume that n employees have been scheduled for duty during period t. If period t is independent of all other periods, and each absence is an independent event, the probability that exactly $0 \leq k \leq n$ workers will report for duty during period $t = p_k = \binom{n}{k}(1 - A)^k A^{n-k}$. The expected coverage during period t will be $\sum_{k=0}^{n} kp_k = (1 - A)n$, given by the first term of constraint (5).

The purpose of constraint (5) is to force the summed Y_{ti} terms, which are used to compute expected revenue in the objective function, to equal the expected number of employees who actually report for duty during period t. Because the values of r_{ti} are non-increasing in expected coverage i (i.e., total revenue increases with i, but at an decreasing rate), this constraint, along with the bounds on Y_{ti} (constraint (7)) and the direction of the objective function, will force the continuous variables Y_{ti}, $i = 0, \ldots, \lfloor n(1 - A)\rfloor$, to their upper bounds (i.e., 1). When the expected coverage for period t is fractional (i.e., $Y_{t,\lceil n(1-\mathrm{A})\rceil} = n(1 - A) - \lfloor n(1 - A)\rfloor > 0$), we assume that the expected marginal revenue produced by the final (fractional) worker is $Y_{t,\lceil n(1-A)\rceil} \times r_{t,\lceil n(1-A)\rceil}$. Like Gans et al. (2003), we assume the number of workers scheduled for duty is grossed up from the number actually needed by parameter A, a “shrinkage factor” that reflects average losses to production capacity due to absenteeism. Finally, we assume that the proportion of absenteeism is the same for all periods in the planning horizon.

Constraint (6) requires the number of employees assigned to each schedule to be non-negative integers. Constraint (7) limits the Y_{ti} variables to the interval 0–1, allowing fractional expected coverage for period t. The non-increasing nature of the marginal revenue function (r_{ti}) and the direction of the optimization ensure that $Y_{ti} \geq Y_{ti+1}$, a characteristic first exploited in the shift-scheduling model by Goodale, Verma, and Pullman (2003).

Typically, labor schedules are usually developed one to six weeks in advance of the service date. An optimal solution specifies the profit-maximizing number of employees assigned to each feasible schedule or tour (represented by X_j, for $j = 1, \ldots, K$ feasible tours). Each tour is a T-element binary vector, where $a_{tj} = 1$ if period t is a duty period for tour j, or 0 otherwise. Assume that employee i is initially assigned to active tour $S_i = \{a_{1i}, a_{2i}, \ldots, a_{Ti}\}$. Suppose that before the schedule is implemented, we also construct a set of alternate recovery schedules S'_i, to which, as in Hur et al. (2004), the employee agrees to work in the event of a staffing emergency. Let C'_{ik} be the cost for employee i to work schedule $S_k \in S'_i$. Note that the cardinality of set S'_i, n_i, increases with the short-notice scheduling flexibility for employee i.

For this study, assume S'_i includes initial tour S_i and all tours that can be constructed from S_i by appending additional work periods (as holdover and/or call-in overtime) allowed under the applicable absence recovery policy. However, if responding to demand forecast errors it may also be useful to create recovery tours with fewer work periods (and lower wage costs) than employee i's initial schedule. Finally, for absence recovery policies that allow the use of THS workers, assume that the set of allowable work schedules S'' and labor charges C''_j for THS worker schedule $S''_j \in S''$ have been negotiated with the THS agency. Additional notation needed to characterize the absence recovery problem is defined in the following:

Parameters

$abs(t,i)$ = 0 if employee i will be absent during duty period t, or 1 otherwise, for $t = 1, \ldots, T$ and $i = 1, \ldots, W$. Updated absence vectors materialize a short time before each shift begins.

a'_{tij} = 1 if period t is a work period for employee i when assigned to recovery tour j, for $t = 1, \ldots, T$; $i = 1, \ldots, W$; and $j = 1, \ldots, n'_j$.

a''_{tj} = 1 if period t is a work period for THS schedule j, or 0 otherwise.

τ = the average proficiency of THS workers; $\tau = 1$ implies the same proficiency as permanent employees.

Decision Variables

V_{jk} = 1 if existing employee j is assigned to recovery tour $S_k \in S'_j$, or 0 otherwise, for $j = 1, \ldots, W$ and $k = 1, \ldots, n_j$.

X''_j = number of THS workers assigned to temporary worker schedule $S_j, S''_j \in S''$.

Y_{ti} = the effective fraction of the ith worker who reports for duty during period t, where $0 \leq Y_{ti} \leq 1$.

The object of the absence recovery problem is to redeploy employees and appoint THS workers to maximize expected contribution after realizing unplanned absences, or:

$$\text{Maximize} \sum_{t=1}^{T} \sum_{w=0}^{M_t} r_{tw} Y_{tw} - \sum_{i=1}^{W} \sum_{k=1_i'}^{n_i} V_{ik} C_{ik}' - \sum_{j \in S''} C_j'' X_j''. \tag{8}$$

Subject to:

$$\sum_{i=1}^{M_t} \sum_{j \in S_i} \mathrm{V}_{ij} a_{tij}' \mathrm{abs}(t, i) + \tau \sum_{k \in S''} a_{tk}'' X_k'' - \sum_{w=0}^{M_t} Y_{tw} \geq 0, \forall t, \tag{9}$$

$$0 \leq Y_{tk} \leq 1, \forall t \quad \text{and } k = 0, \ldots, M_t, \tag{10}$$

$$\sum_{k=1}^{n_i} V_{ik} = 1, \quad \text{for } i = 1, \ldots, W, \tag{11}$$

$$0 \leq V_{ik} \leq 1, \quad i = 1, \ldots, W, \quad \forall S_k \in S_i', \tag{12}$$

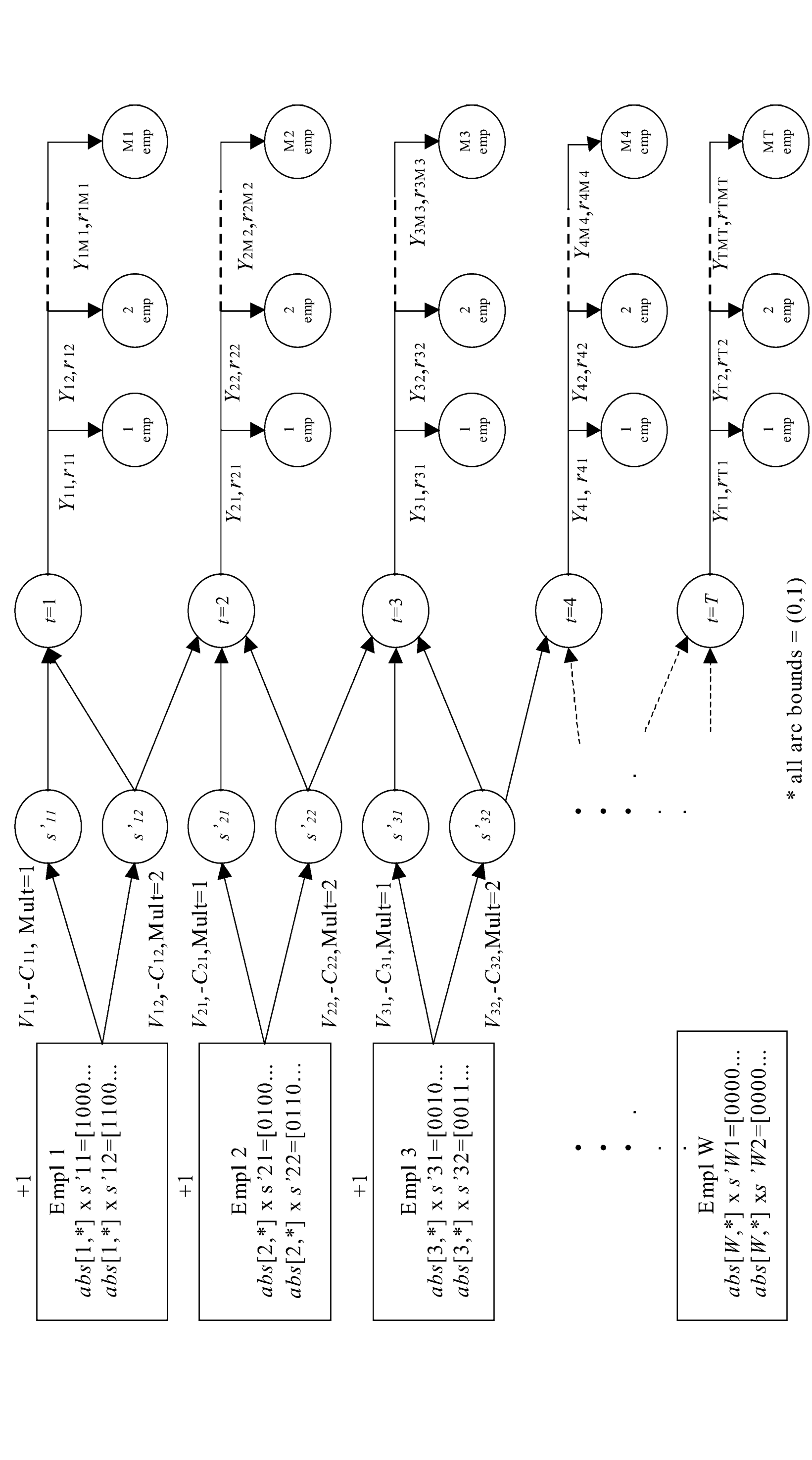
$$X_j'' \geq 0, \quad \text{integer}, \forall j \in S''. \tag{13}$$

Equation (8) computes the contribution to profit following absence recovery. Summing over all periods and relevant staffing levels, the first sum determines the total revenue generated by the recovery schedules during each period. The second and third sums subtract the wage costs after employees are assigned recovery schedules and the fees for THS workers, respectively. Per-capita costs are not affected by the recovery decision and, therefore, are excluded from the model. Constraint (9) sets the upper bound for the sum of the Y_{ti} variables to the number of regular employees who report for duty during period t (i.e., $a_{ti}' \times abs(t,i) = 1$) plus the proficiency-adjusted number of THS workers scheduled for duty during that period. In constraint (10), the Y_{ti} variables are limited to values on the interval 0–1, similar to constraint (7).

Constraints (11) and (12) serve as multiple-choice constraints that ensure exactly one schedule extension from set S_j' is selected for employee j. For recovery policies that preclude the use of THS workers (i.e., $S'' = \{\}$), the recovery model can be implemented as an integer generalized network (Glover, Klingman, & Hultz, 1978). An example of such a network is depicted in Figure 1, which shows the Y_{ti} nodes represent the ith employee equivalent in period t. This property ensures integer values for the V_{jk} variables when they are constrained to the interval 0–1, and makes it possible to solve the absence recovery problem as a linear program. When $S'' \neq \{\}$, constraint (13) ensures non-negative integer values for the number of THS workers assigned to each feasible THS schedule. Unfortunately, this constraint changes the model's mathematical structure to a mixed integer program.

Although we expect the staffing/scheduling model (equations (4)–(7)) will be used to produce employee schedules once every one to six weeks, the absence recovery model (equations (8)–(13)) can be applied whenever new absence information is received. Shift supervisors need only update the absence vector [$abs(t,i)$] for the affected employees and rerun the absence recovery model. Its computational efficiency makes it an ideal tool for dealing with absence information in real time. After implementing an absence recovery solution, however, employee schedules

Figure 1: Generalized network* for absence recovery with internal resources.



will also need to be updated to reflect any changes before the recovery model is next applied.

EXPERIMENTAL DESIGN AND RESULTS

The economic impact of unplanned absences may be mitigated by choosing appropriate staffing/scheduling policies and absence recovery measures. Staffing policies, including those pertaining to absence anticipation and planned overtime, influence both labor costs and the revenue impact of unplanned absences. Recovery options (i.e., holdover or call-in overtime, THS workers, or absorbing the absence with the existing staff) may be constrained by the choice of staffing policies and may differ in cost and effectiveness. In this section, we describe our experiments to assess the effectiveness of each combination. Our principal performance measure is absence cost, or the expected profit before absences are realized minus expected profit after the absence recovery plan is implemented.

Our experiments simulate the effects of absenteeism on expected profits for a hypothetical M/M/S/∞ +M system experiencing unplanned absences at various rates (0%, 5%, and 10%). We assume a planning horizon of $T = 168$ consecutive periods (7 days $\times$ 24 hours), with hourly forecasted customer arrival rates λ_t (for $t = 1, \ldots, 168$) that vary from hour to hour in a sinusoidal pattern. The mean, amplitude, and frequency of the pattern can be adjusted to represent different service industries. We assume that employees process transactions at the rate of μ/hour per employee, and that parameter μ may be adjusted to reflect different processes. We also assume that when THS workers are used, they complete transactions at 80% of the regular service rate μ (Campbell, 1999). Lastly, like Andrews and Parsons (1993) we assume that 80% of all completed service transactions during period t produced revenue $= R_t$; the remaining transactions are informational and produce no revenue.

We compared staffing and scheduling policies with and without planned overtime and anticipated absences. The specific factor levels were anticipated absenteeism rates of 0% and 5% and planned overtime of zero hours/week (i.e., no planned overtime) and planned overtime of up to 20 hours per week. The assumed absenteeism rates straddle the U.S. service sector's average, and should differ enough to reveal possible performance advantages for anticipatory staffing policies. For scheduled overtime strategies, we designed weekly tours of 40–60 hours with five consecutive shifts of up to 12 hours per shift. However, shifts longer than 10 hours require a second unpaid one-hour meal break at the start of the overtime portion of the shift. Standard (40 hours/week) tours provide five consecutive nine-hour shifts each week, with an unpaid one-hour meal break in the middle of each shift. The shifts in each tour could begin at any hour of the day, but had to be the same on each regular workday. Under standard work rules, there are 24 possible shift-starting times and seven possible days-off pairs, or 168 unique, feasible tours. With overtime, each shift in a tour can be increased by zero to four hours. With five shifts per tour, the number of feasible tours increases to $168 \times 5^5 = 525{,}000$ unique feasible tours.

For each staffing/scheduling problem we used a custom C++ program to create the set of all feasible tours and the marginal revenue matrix r_{ti}. Due to the large number of feasible tours, we formulated each staffing/scheduling problem with a working subset of feasible tours that were selected by Easton and Rossin's (1991) column generation procedure. IBM's (2000) OSL version 3.0 provided the shadow prices for the column generation algorithm and also solved the final integer program. We configured OSL to converge to within .25% of theoretical maxima, which it did within five minutes on a 2.5 GHz workstation.

We next arbitrarily assigned each of the W employees required by the staffing/scheduling solution to a specific schedule, then simulated employee absences. For each employee shift, we randomly (with probability = 0%, 5%, or 10%) determined if the employee would be absent, assuming the duration of a simulated absence equals the length of the scheduled work shift. We repeated this for all scheduled employee shifts, then updated the absence indicators $abs(t,i)$ for each employee and each period. The results were passed to the absence recovery model. We repeated the simulation 30 times for each staffing/scheduling solution.

We began the recovery phase by computing expected profits with staffing levels depleted by the simulated absences, reflecting a passive absence recovery policy. We then evaluated the remaining five active absence recovery policies: holdover (H), call-in (C), THS workers (T), $H + C$, and $H + C + T$. For each policy, we created a set of allowable absence recovery tours for each employee (S'_j) using a custom C++ matrix generator. For holdover overtime, the recovery tours were created by appending overtime to the end of each existing employee's scheduled shift, provided the resulting schedule conformed to the overtime work rules described above (i.e., ≤12 hours/shift and ≤60 hours/week). For the call-in policy, recovery tours for employee j included all possible four-hour assignments, separated by at least one rest period, that allow at least 12 hours of rest between the call-in assignment and the employee's regular assignments. In addition to confining call-in assignments to scheduled days off, we required them to conform to the above overtime work rules. Potentially, there could be as many as 121,875 different recovery options to consider for each employee ((up to 4^5 possible holdover overtime alternatives) × (up to 39 possible call-in overtime start times)). Typically, however, only a few of these options were ever needed to cover absence-related capacity shortages. Finally, we assumed that THS employees worked standard nine hour/day shifts (eight paid hours plus a one-hour, unpaid meal break) and were paid 1.4 times the base hourly wage of permanent employees (including the THS markup). These assumptions allow 168 unique feasible tours for temporary workers.

For absence recovery problems using either holdover or call-in overtime, the SAS/OR (SAS Institute Inc, 1999) linear programming procedure provided optimal schedule assignments in an average of 18 seconds. For absence recovery with THS workers, the SAS/OR branch-and-bound procedure found integer solutions within .15% of optimal in an average of 60 CPU seconds (on a 2.3 GHz workstation).

We evaluated the staffing/scheduling and absence recovery combinations in a variety of hypothetical service environments with factors likely to influence the impact of absenteeism. The factors and the levels we selected are classified as either customer or employee factors, and are described in Table 3. In an earlier study, Easton and Goodale (2002) found that the economic impact of unplanned

Table 3: Experimental factors and factor levels.

Experimental Factor	Number of Levels	Level Values
Employee Characteristics		
Average service rate for fully-trained employees	2	$\mu = 4$ and 16 transactions per hour (.8 μ for THS workers)
Quasi-fixed cost/employee	2	$F =$ \$100/week and \$200/week; for THS workers, daily markup = \$32.
Market Characteristics		
Daily demand pattern	2	Unimodal and trimodal sinusoidal patterns, averaging 100 arrivals/hour
Amplitude of demand	2	Coefficient of variation of .25 and .50; sine function amplitudes of .353 and .706
Unit contribution before labor	2	$R_t =$ \$14.40 and \$28.80 per order (80% of all arrivals are orders; 20% are non-revenue producing)
Average customer patience	2	$\sigma^{-1} = 60$ seconds and 240 seconds, exponentially distributed
Staffing Options		
Planned overtime	2	0 hours/week and up to 4 hours/day, 20 hours/week, per employee
Anticipated absence rate	2	$A = 0\%$ and 5%
Absence Recovery		
Recovery policies	6	Passive, Holdover, Call-in, THS workers, Holdover + Call-in, Holdover + Call-in + THS workers.
Simulated absence levels	3 × 30	0%, 5%, and 10%, 30 replications each
Additional Parameters		Hourly wage rate = \$10/hour regular, \$15/hour overtime, \$14/hour for THS employees.

absenteeism and employee turnover tends to be greater when scheduled workforce utilization is high. Market factors favoring high server utilization include slim operating margins, greater customer patience, high transaction volumes, and demand patterns with fewer peaks but greater amplitude. They also found that systems with many servers and low service rates tend to be quite vulnerable to economic losses due to absenteeism. In addition to these factors, we believe the effectiveness of absence recovery may be influenced by per-capita employee expenses. With higher per-capita costs, firms may be reluctant to adopt anticipatory staffing and will be more likely to staff with planned overtime. This action is likely to increase workforce utilization and absence costs. Accordingly, we simulated service environments with per-capita expenses of \$100 and \$200 per employee (25–50% of the regular weekly wages earned by full-time employees). The former value is slightly less than the U.S. average. The higher value is likely to favor more planned overtime and encourage absence recovery options involving overtime or THS workers.

A full factorial study, with 64 possible environmental factor combinations and 24 different staffing/scheduling/recovery policies (1,536 combinations in all), would allow us to explore all main effects and interactions. However, our main interest lies with the relative performance of staffing/scheduling and recovery policy combinations. Using an orthogonal fractional factorial experimental design (oa8.7.2.2) by Hedayat, Sloane, and Stufken (1999), we were able to reduce the number of different environmental configurations we simulated for each staffing/recovery policy from 64 to 8 and still identify the main effects for the environmental variables. The entire experimental design involved: (four staffing strategies) $\times$ (six recovery policies) $\times$ (three simulated absence levels) $\times$ (eight environmental configurations) $\times$ (30 replications) $=$ 17,280 test problems.

Results

In Table 4, we examine the average performance of alternate recovery policies with different staffing strategies. We report the average expected profit with zero absenteeism for each staffing strategy, the reduction in expected profit after absenteeism is realized (absence cost), and the percent of absence cost recovered by each recovery policy. In general, passive absence recovery is dominated by all active recovery policies, regardless of the staffing strategy adopted. Furthermore, aggressive absence recovery policies (like $H + C$ and $H + C + T$) recapture on average about half the profits lost to unplanned absences. All of the absence recovery policies reduced the cost of absenteeism, and they seemed to perform better at higher simulated absence rates. However, they were not equally effective.

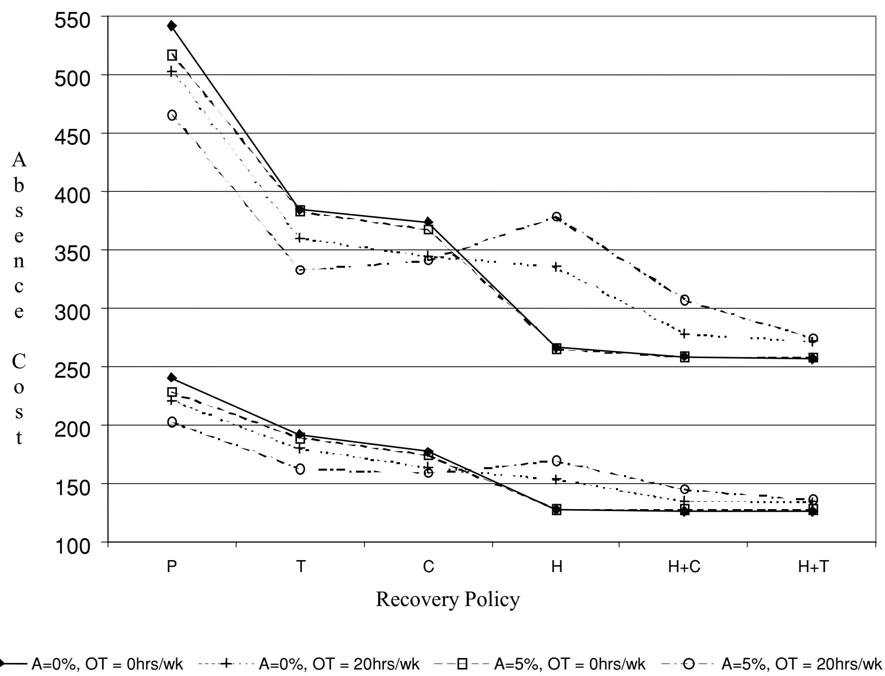
Holdover overtime generally outperforms the other two single-method policies (except under high absenteeism with staffing strategies that use planned overtime), because it allows managers to apply precise amounts (increments as small as one worker-hour) of overtime labor at the most advantageous times. With both call-in and THS worker assignments, resources must be activated for four to eight hours to recover from absences, even if only a portion of that time is needed to cover staff ing shortages. Holdover plus call-in recovery policies generally provide a modest improvement over holdover absence recovery alone, and even smaller incremental gains are achieved when all three absence recovery policies are combined.

The rows of Table 4 show the average performance of each staffing strategy at their high and low levels of intensity. THS and call-in recovery performance appears to be relatively independent of the staffing strategy in use. For holdover absence recovery, however, significant performance differences are apparent between the low and high settings for absence anticipation and planned overtime. Analysis of the data reveals several causes for this behavior. With up to 20 hours of planned overtime per week, absent workers assigned to overtime schedules left larger holes in the schedule. In addition, planned overtime staffing strategies consumed a significant portion of the firm's available overtime that would otherwise be available for absence recovery. Here, available overtime is the difference between the upper bound on the length of recovery tours (shifts $\leq$12 hours and tours $\leq$60 hours/week) and the hours consumed by the original employee schedules. Finally, planned overtime staffing usually resulted in fewer workers on duty each day, so there were fewer shifts available for holdover assignments. As a result, it was

Table 4: Recovery of absence costs by recovery policy and staffing strategy.

		Expected Weekly Profit, 0% Realized Absenteeism	Absence Cost (Passive Recovery)	Average Percent Absence Cost Recaptured by Recovery Policy				
				THS (T)	Call-ins (C)	Holdover (H)	$H + C$	$H + C + T$
Simulated Absence = 5%								
Staffing Strategy								
Anticipated	0%	$113,939	$2,312	20.63%	28.91%	45.65%	49.96%	50.23%
Absence	5%	$113,923	$2,161	20.14%	25.85%	35.51%	42.20%	44.99%
Planned	0 hours/week	$113,689	$2,351	20.03%	29.00%	54.09%	54.53%	54.55%
Overtime	20 hours/week	$114,173	$2,121	20.75%	25.76%	27.07%	37.64%	40.67%
Simulated Absence = 10%								
Staffing Strategy								
Anticipated	0%	$113,939	$5,225	28.86%	32.21%	45.44%	51.48%	52.37%
Absence	5%	$113,923	$4,914	28.27%	29.38%	36.60%	44.92%	48.94%
Planned	0 hours/week	$113,689	$5,298	27.86%	31.68%	54.23%	55.47%	55.69%
Overtime	20 hours/week	$114,173	$4,841	29.27%	29.90%	27.82%	40.92%	45.62%

Figure 2: Absence cost by staffing strategy/recovery policy.



often impossible to cover absences. The latter reason also helps explain the poor performance of holdover absence recovery with staffing strategies that anticipate absences. Absence anticipation tends to reduce the workforce size when employees are indemnified for time lost to illness, because it increases effective hourly wages.

Additional insights emerge when we consider the two staffing strategies jointly. In Figure 2, we plot average absence costs for four anticipated absence—planned overtime staffing strategy combinations. In this case, the baseline profit used to compute absence cost is the highest expected profit achieved for each operating environment before realizing absenteeism. For convenience, we plot experimental results for 5% and 10% realized absenteeism on the same chart. The top-performing staffing/recovery combination in Figure 1, for both 5% and 10% realized absenteeism (the lower and upper parts of the chart, respectively), are standard, 40-hour-per-week employee schedules with or without anticipated absenteeism, combined with the most aggressive absence recovery policy (holdover + call-ins + THS). However, nearly the same performance can be achieved by combining these staffing strategies with holdover overtime alone. As a bonus, the staff ing and absence recovery models for these combinations are usually easier to solve because they are integer generalized networks.

Environmental Variables

We next examine the interaction between staffing strategies and various environmental variables on absence recovery efforts. Before reporting recovery performance with the environmental variables at their high and low levels, however, it is useful to introduce the concept of unscheduled initial service capacity per hour (slack per hour or slack/hour, the number of employees $\times$ service rate $\times$ difference between maximum recovery-mode hours/tour and average scheduled hours/tour, divided by the number of periods in the planning horizon). Essentially, slack/hour measures the reserve internal service capacity available to respond to unplanned absences. Accompanied by appropriate coefficients for the environmental variables and the operant staffing strategy in use, slack/hour predicted the percentage of absence costs recovered in our experiments with an adjusted R^2 of .88 and standard error of .06. In general, our data show the more slack/hour, the more successful the recovery with internal resources.

In Table 5a, we show the average percentage of absence costs recovered for six environmental variables and four staffing strategies, at realized absenteeism of 5% and 10%, using holdover overtime. Once again, holdover recovery efforts appear to be less successful with staffing plans that anticipate absenteeism or utilize planned overtime. This relationship persists for all environmental variables at all levels considered in this study. However, Table 5a also shows significant differences in absence recovery between the high and low levels for service rate, per-capita costs, demand patterns, unit revenue, and customer patience.

Consider first the employee-related experimental factors, which appear to have strong main effects. Absence recovery efforts were consistently more successful in environments with high service rates ($\mu = 16/\text{hour}$). At low service rates ($\mu = 4/\text{hour}$) gross margins tend to be lower due to higher labor costs per transaction. This encourages staffing and scheduling solutions that attain high resource utilization. In Table 5b, which reports the average initial slack for the staffing solutions, slack/hour for the low-service-rate solutions were consistently lower; perhaps dangerously so for staffing strategies that anticipate absenteeism and permit planned overtime. Higher per-capita costs also tend to favor staffing/scheduling solutions with a smaller workforce, resulting in less slack/hour. However, the opportunity to significantly reduce workforce size arises only when planned overtime is a staffing option. Our staffing model exploited planned overtime for high-per-capita-cost environments whenever the option was available. When combined with staffing strategies that allow planned overtime and absence anticipation, high per-capita costs drove average slack to the lowest point of all the cells in Table 5b, and also produced the poorest absence recovery record (Table 5a).

The limitations of our orthogonal experimental design become apparent when we consider absence recovery success with respect to market-related factors. In general, when customers are more *patient*, queues can be longer without risking revenue losses due to reneging. This tends to favor lower staffing levels and reduced slack/hour. Similarly, lower revenue/transaction reduces margins, making it more difficult to support a large workforce when demand is finite. Again, this should yield staffing/scheduling solutions with less slack. We also expected that demand patterns with high amplitude and more peaks per day would require more employees, and generally result in greater slack. Table 5b confirms our expectations with regard to slack/hour (except for tri-modal demand with 0% anticipated absenteeism and planned overtime). Although in most cases, the general relationship between

Table 5a: Percent of absence cost recovered using holdover overtime.

Factor	Levels	Realized $A = 5\%$				Realized $A = 10\%$			
		Projected $A = 0\%$		Projected $A = 5\%$		Projected $A = 0\%$		Projected $A = 5\%$	
		$OT = 0$	$OT = 20$	$OT = 0$	$OT = 20$	$OT = 0$	$OT = 20$	$OT = 0$	$OT = 20$
Service rate (customer/hour)	4	64.1%	29.3%	53.2%	9.6%	55.7%	29.2%	49.8%	10.7%
	16	76.1	47.5	74.3	30.5	67.2	44.4	66.1	30.8
Quasi-fixed ($/employee)	100	67.5	39.9	55.9	37.5	59.0	40.5	53.0	39.0
	200	72.7	36.9	71.6	2.6	63.8	33.1	63.0	2.5
Daily demand pattern	Unimodal	53.8	45.9	48.7	21.6	52.3	45.3	49.9	22.9
	Trimodal	86.4	30.9	78.9	18.5	70.5	28.3	66.1	18.6
Amplitude of demand	.25	66.3	46.5	56.1	21.8	59.9	42.9	54.0	22.7
(Coefficient of variation)	.50	73.9	30.3	71.4	18.3	63.0	30.7	61.9	18.9
Unit contribution before labor ($/customer)	14.40	74.6	29.1	74.1	11.7	61.4*	26.9	61.1	12.7
	28.80	65.6	47.8	53.4	28.4	61.5*	46.7	54.9	28.8
Average customer patience (seconds)	60	62.3	31.8	61.3	24.2	57.1	32.1	56.6	25.4
	240	77.9	45.0	66.2	15.9	65.8	41.5	59.4	16.1

*Difference not significant at $p < .05$.

Table 5b: Average slack/hour (unscheduled initial capacity in transactions/hour) by environmental factor and staffing strategy.

		Average Slack (transactions/hour)			
		$A = 0\%$		$A = 5\%$	
Environment Variable	Level	$OT = 0$	$OT \leq 20$	$OT = 0$	$OT \leq 20$
Service Rate	4	33.8	22.5	29.4	11.4
(transactions/hour)	16	52.9	41.1	44.9	26.3
Per-capita Cost	$100	44.0	40.0	38.2	32.8
	$200	42.6	23.6	36.1	5.0
Demand Pattern	Unimodal	42.6	34.6	36.4	17.5
	Trimodal	44.1	29.0	37.9	20.3
Demand Amplitude	.25	41.2	31.5	35.7	17.6
(coefficient of variation)	.50	45.5	32.2	38.6	20.2
Revenue/Transaction	$14.40	41.7	27.0	35.3	15.0
	$28.80	45.0	36.7	39.0	22.8
Customer Patience	60	45.5	34.5	38.6	25.5
(seconds)	240	41.2	29.2	35.7	12.3

slack/hour and absence recovery success holds for the different market-related factors,

Slack/hour is a useful predictor for recovery success with the market-related factors. Comparing Tables 5a and 5b, firms with tri-modal demand, high demand amplitude, low revenue/transaction, or high customer patience should avoid staff ing plans with low slack/hour. However, we were unable to reliably predict relative absence recovery success for high- and low-market attribute levels if the difference between the high and low slack/hour values was less than 10%. This lack of precision may be due to untested interactions between market-related factors.

CONCLUSIONS

In this study we explored the relationship between staffing strategies and real-time schedule recovery policies in service organizations that experience unplanned absenteeism and indemnify their employees for lost time. We defined the total cost of absenteeism in terms of profit reduction, which considers the impact of unplanned absences on both revenue and labor expense. Rather than measure or prevent absenteeism (which we consider inevitable), our goal was to identify robust staffing, scheduling, and absence recovery practices that allow firms to effectively recover from absenteeism when it occurs.

We believe that this work provides four significant contributions to the relevant literature. First, we introduce a labor staffing and scheduling model that formally recognizes absenteeism. Second, we present an absence recovery model that provides an analytical coping mechanism for short-term absences. Third, we establish a link between staffing and scheduling policies and the effectiveness of alternate absence recovery strategies. Finally, we establish the operating conditions in which absence recovery approaches such as holdover overtime tend to be more effective than other methods.

Our study was based on more than 17,000 simulation experiments with different staffing strategies, absence recovery policies, and environmental factors. To facilitate the comparison of these combinations, we devised a flexible and efficient absence recovery model. This model, new to the literature, optimizes short-term schedule adjustment decisions. For recovery policies using only internal resources, the model is an integer-generalized network—a linear program. Although not attempted in this study, the model could also be used to recover from demand forecast errors.

We found the reduction in total profits due to absenteeism is strongly influenced by the staffing strategies and absence recovery policies that firms adopt to cope with absenteeism. Forty-hour work weeks and zero anticipated absenteeism with holdover absence recovery appears to be a fairly robust combination, on average reclaiming nearly 60% of the profit consumed by unchecked absenteeism. For firms unwilling or unable to implement active absence recovery policies, however, planned overtime staffing strategies with absence anticipation appear less vulnerable to absenteeism.

We devised a measure, unscheduled initial service capacity/hour (slack/hour), which accurately predicts absence recovery success. Higher values of slack/hour indicate a greater capability to respond to unplanned absences with internal resources. This measure could serve as a useful risk index for firms attempting to improve productivity through the use of planned overtime or other staffing strategies that may exacerbate unplanned absenteeism.

Our conclusions may be affected by certain limitations in this study, such as the long-term consequences of reneging. We assumed that reneging customers have a short memory and remain in the customer population. If this is not the case, future arrival rates and revenues may decline. A related limitation concerns the limited memory of our employees. Some experts believe the use of short-notice overtime tends to increase absenteeism in future periods. We assumed a constant rate of absenteeism. Next, our staffing and scheduling model assumes that expected absences in any given period are the result of a series of independent Bernoulli trials (Feller, 1968). In practice, an absence at time t is likely to persist for several periods into the future, especially when the duration of each planning interval is less than a full shift. Put another way, an employee who is absent at the beginning of his or her shift is likely to remain absent for several consecutive planning intervals. Therefore, the assumption of independence that is implicit for the Bernoulli trials or capacity cushion approach will usually be violated in most realistic situations. Another limitation is that we applied our recovery model after simulating absences for an entire week. In practice, absence information materializes only a short time—a few hours at best—before the absence occurs. Thus, our approach may induce a favorable bias to all of our results. Finally, we assumed that employees maintained the same service rate over time, ignoring possible effects due to congestions (employees work faster when the queue is large) or fatigue near the end of a long shift.

There are also at least three possible avenues for future research in this area. One is to address other domains. In this study, unplanned absences threatened response time and revenue. In the public sector, which tends to experience higher absenteeism than the private sector, absence costs may be more difficult to measure and may have public safety as well as economic implications. Second, slack/hour appears to be an effective way to predict the expected cost of absenteeism when choosing staffing/scheduling strategies. This work could be extended by estimating absence costs and exploiting slack/hour to evaluate new staffing proposals with a single, integrated model. Third, short-term schedule adjustments may be required for forecast errors as well as unplanned absences. A useful extension would be to consider both absences and forecast errors simultaneously when developing robust staffing and scheduling solutions. [Received: January 2004. Accepted: March 2005.]

REFERENCES

Allen, S. (1981). An empirical model of work attendance. *Review of Economics and Statistics, 62*(1), 77–87.

Anderson, E., & Sullivan, M. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science, 12*(2), 125–143.

Andrews, B., & Parsons, H. (1993). Establishing telephone-agent staffing levels through economic optimization. *Interfaces, 23*(2), 14–20.

Andrews, B., & Parsons, H. (1989). L.L. Bean chooses a telephone agent scheduling system. *Interfaces, 19*(6), 1–9.

Anonymous. (2002). Real cost of absence up to L23 billion. *Management Services, 46*(10), 6.

Bacelli, F., & Hebuterne, G. (1981). On queues with impatient customers. In F. J. Kylstra (Ed.), *Performance'81*. Amsterdam: North-Holland, 159–179.

Bailey, J. (1985). Integrated days off and shift personnel scheduling. *Computers and Industrial Engineering, 9*(4), 395–404.

Baker, K. (1976). Workforce allocation in cyclical scheduling problems: A survey. *Operational Research Quarterly, 27*(1), 155–167.

Bechtold, S., & Jacobs, L. (1990). Implicit modeling of flexible break assignments in optimal shift scheduling. *Management Science, 36*(11), 1339–1351.

Brusco, M., & Jacobs, L. (1998). Personnel tour scheduling when starting time restrictions are present. *Management Science, 44*(4), 534–547.

Bureau of Labor Statistics. (2001). Absence from work of employed full-time wage and salary workers by occupation and industry (Table 47). *Current Population Survey—Employment and Earnings*, Washington DC: Superintendent of Documents, U.S. Government Printing Office.

Burns, R., & Koop, G. (1987). A modular approach to optimal multiple-shift manpower scheduling. *Operations Research, 35*(1), 100–110.

Buschak, M., Craven, C., & Ledman, R. (1996). Managing absenteeism for greater productivity. *S.A.M. Advanced Management Journal, 61*(1), 26–31.

Campbell, G. (1999). Cross-utilization of workers whose capabilities differ. *Management Science, 45*(5), 722–732.

Carmon, Z., Shanthikumar, J., & Carmon, T. (1995). A psychological perspective on service segmentation models: The significance of accounting for consumers' perceptions of waiting and service. *Management Science, 41*(11), 1806–1815.

CCH. (2003). Unscheduled absence survey. *Human Resources Management Ideas & Trends*, *569*, 153.

Chadwick-Jones, J. (1981). Renegotiating absence levels. *Journal of Occupational Behavior*, *2*(4), 255–266.

CTI. (2002). Executive summary: Shiftwork practices 2002, accessed March 18, 2003, available at www.circadian.com.

Dantzig, G. (1954). A comment on Edie's 'Traffic delays at toll booths.' *Operations Research*, *2*(3), 339–341.

Davis-Blake, A., & Uzzi, B. (1993). Determinants of employment externalization: A study of temporary workers and independent contractors. *Administrative Sciences Quarterly*, *32*(2), 195–223.

Durr, W., & Matan, O. (2002). Workforce optimization: The new workforce management (WP_WFO.pdf), accessed March 18, 2003, available at www.bluepumpkin.com.

Easton, F. F., & Goodale, J. C. (2002). Labor scheduling with employee turnover and absenteeism (working paper), accessed March 18, 2003, available at www.som.syr.edu/facstaff/ffeaston/research/bandaid030401.pdf.

Easton, F. F., & Rossin, D. F. (1991). Sufficient working subsets for the tour scheduling problem. *Management Science*, *37*(11), 1441–1451.

Easton, F. F., & Rossin, D. F. (1996). A stochastic goal program for employee scheduling. *Decision Sciences*, *27*(3), 541–568.

Ehrenberg, R. (1989). *Fringe benefits and overtime behavior*. Lexington, MA: Lexington Books.

Feller, W. (1968). *An lntroduction to probability theory and its applications* (Vol. 1, 3rd ed.). New York: Wiley.

Ferguson, T., Ferguson, W., Muedder, K., & Fitzgerald, M. (2001). The case for total absence management and integrated benefits. *HR Human Resource Planning, 24*(3), 36–46.

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management, 5*(2), 79–141.

Gans, N., & Zhou, Y. P. (2003). Managing learning and turnover in employee staff ing. *Operations Research, 50*(6), 991–1007.

Garnett, O., Mandelbaum, A., & Reiman, M. (2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management, 4*(3), 208–227.

Globerson, S., & Nagarvala, P. (1974). Unexplained absenteeism: A simulation approach. *Winter Simulation Conference Annual Proceedings*. Washington DC, 54–55.

Glover, F., Klingman, D., & Hultz, J. (1978). Generalized networks: A fundamental computer-based planning tool. *Management Science, 24*(12), 1209–1220.

Goodale, J. C., Verma, R., & Pullman, M. E. (2003). A market utility-based model for capacity scheduling in mass services. *Production and Operations Management, 12*(2), 165–185.

Greble, T. (1997). A leading role for HR in alternative staffing. *HRMagazine, 42*(2), 99–105.

Gross, D., & Harris, C. (1998). *Fundamentals of queueing theory*, 3rd ed. New York: Wiley.

Harrison, D., & Martocchio, J. (1998). Time for absenteeism: A 20-year review of origins, offshoots, and outcomes. *Journal of Management, 24*(3), 305–350.

Hedayat, A. S., Sloane, N. J. A., & Stufken, J. (1999). *Orthogonal arrays: Theory and applications*. New York: Springer-Verlag; also accessed March 29, 2003, available at www.research.att.com/∼njas/oadir/index.html.

Hetrick, R. (2000). Analyzing the recent upward surge in overtime hours. *Monthly Labor Review, 123*(2), 30–33.

Hinkin, T., & Tracey, J. B. (2000). The cost of turnover. *Cornell Hotel and Restaurant Administration Quarterly, 41*(3), 14–21.

Hirschman, C. (1999). Paying for waiting. *HRMagazine, 44*(8), 98–105.

Hur, D., Mabert, V., & Bretthauer, K. (2004). Real time schedule adjustments: A case study. *Omega, 32*(5), 333–344.

IBI. (2000). Integrated Benefits Institute reports on health and productivity benchmarks: Telecommunications industry, accessed March 18, 2003, available at www.ibiweb.org/publications/research/17/.

IBM. (2000). Optimization solutions and library (OSL v3), accessed March 29, 2003, available at www-3.ibm.com/software/data/bi/osl/index.html.

Ittig, P. (2002). The real cost of making customers wait. *International Journal of Service Industry Management, 13*(3), 231–241.

Jarrah, A., Bard, J., & deSilva, A. (1994). Solving large-scale tour scheduling problems. *Management Science, 40*(9), 1124–1144.

Jacobs, L., & Brusco, M. (1996). Overlapping start-time bands in implicit tour scheduling. *Management Science, 42*(9), 1247–1259.

Keith, G. (1979). Operator scheduling. *AIIE Transactions, 11*(1), 37–41.

Mabert, V., & Watts, C. (1982). A simulation analysis of tour-shift construction procedures. *Management Science, 28*(5), 520–532.

Mangum, G., Mayall, D., & Nelson, K. (1985). The temporary help industry: A response to the dual internal labor market. *Industrial and Labor Relations Review, 38*(4), 599–611.

Mitchell, R. (2001). Employers taking steps to reduce costs of absenteeism. *National Underwriter, 105*(15), 57.

Moore, M. (1965). The temporary help service industry: Historical development, operation, and scope. *Industrial and Labor Relations Review, 18*(4), 554–569.

Osuna, E. (1985). The psychological cost of waiting. *Journal of Mathematical Psychology, 29*(1), 82–105.

Pinker, E., & Larson, R. (2003). Optimizing the use of contingent labor when demand is uncertain. *European Journal of Operational Research, 144*(1), 39–55.

Robinson, B. (2002). An integrated approach to managing absence supports greater organizational productivity. *Employee Benefits Journal, 27*(2), 7–11.

SAS Institute Inc. (1999). *SAS/OR user's guide: Mathematical programming, version 8*. Cary, NC: SAS Institute.

Sasser, W. E., Jaikumar, R., & Rikert, D. C. (1991). Sof-Optics (A), Harvard Business School Case No. 9-681-052. Boston: Harvard Business School Publishing.

Segal, L., & Sullivan, D. (1997). The growth of temporary services work. *Journal of Economic Perspectives, 11*(2), 117–136.

Steel, R. P., & Rentsch, J. R. (1995). Influence of cumulation strategies on the long-range prediction of absenteeism. *Academy of Management Journal, 38*(6), 1616–1634.

Steers, R. M., & Rhodes, S. R. (1978). Major influences on employee attendance: A process model. *Journal of Applied Psychology, 63*(4), 391–407.

Thompson, G. M. (1995). Labor scheduling using NPV estimates of the marginal benefit of additional labor capacity. *Journal of Operations Management, 13*(1), 67–86.

Thompson, G. M. (1999). Labor Scheduling Part 4. *Cornell Hotel and Restaurant Administration Quarterly*, 40(3), 85–96.

U.S. Census Bureau. (2002a). Employer costs for employee compensation per hour worked (Table 618). *Statistical Abstract of the United States*. Washington, DC: Government Printing Office.

U.S. Census Bureau. (2002b). Persons at work by hours worked: 2001 (Table 575). *Statistical Abstract of the United States*. Washington, DC: Government Printing Office.

Watson, C. J., Driver, R. W., & Watson, K. D. (1985). Methodological issues in absenteeism research: Multiple absence measures and multivariate normality. *Academy of Management Review, 10*(3), 577–586.

Whitt, W. (1999). Predicting queueing delays. *Management Science, 45*(6), 870–888.

Zohar, E., Mandelbaum, A., & Shimkin, N. (2002). Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science, 48*(4), 566–583.

Fred F. Easton is an associate professor in the School of Management at Syracuse University, where he teaches operations management, supply chain management, service systems management, management science, and total quality management. His research focuses on capacity management issues in both service and manufacturing organizations, and has been published in *Management Science* and other scholarly journals. He directs the Robert H. Brethen Operations Management Institute and holds editorial review board positions with two journals. Prof. Easton advises executives in the health services, auto manufacturing, and emergency services industries about capacity management issues.

John C. Goodale is an assistant professor of Decision Sciences in the Lundquist College of Business where he teaches courses in operations management. He has been an active member of the Decision Sciences Institute, the Production and Operations Management Society, and INFORMS. In addition, he is a member of the editorial board for the *Cornell Quarterly*. His research is focused on scheduling and quality in service operations, and papers on recent projects have appeared or are forthcoming in *Production and Operations Management*, *Journal of Operations Management*, *European Journal of Operational Research*, and other journals.