



## Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Nurse Absenteeism and Staffing Strategies for Hospital Inpatient Units

Wen-Ya Wang, Diwakar Gupta

To cite this article:

Wen-Ya Wang, Diwakar Gupta (2014) Nurse Absenteeism and Staffing Strategies for Hospital Inpatient Units. *Manufacturing & Service Operations Management* 16(3):439-454. <http://dx.doi.org/10.1287/msom.2014.0486>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Nurse Absenteeism and Staffing Strategies for Hospital Inpatient Units

Wen-Ya Wang

Department of Marketing and Decision Sciences, San José State University, San Jose, California 95192,  
[wenya.wang@sjsu.edu](mailto:wenya.wang@sjsu.edu)

Diwakar Gupta

Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, Minnesota 55455,  
[gupta016@umn.edu](mailto:gupta016@umn.edu)

Inpatient staffing costs are significantly affected by nurse absenteeism, which is typically high in U.S. hospitals. We use data from multiple inpatient units of two hospitals to study which factors, including unit culture, short-term workload, and shift type, explain nurse absenteeism. The analysis highlights the importance of paying attention to heterogeneous absentee rates among individual nurses. We then develop models to investigate the impact of demand and absentee rate variability on the performance of staffing plans and obtain some structural results. Utilizing these results, we propose and test three easy-to-use heuristics to identify near-optimal staffing strategies. Such strategies could be useful to hospitals that periodically reassign nurses with similar qualifications to inpatient units in order to balance workload and accommodate changes in patient flow. Although motivated by staffing of hospital inpatient units, the approach developed in this paper is also applicable to other team-based and labor-intensive service environments.

**Keywords:** absenteeism; staffing; hospital operations

**History:** Received: June 5, 2012; accepted: January 28, 2014. Published online in *Articles in Advance* May 15, 2014.

## 1. Introduction

This paper focuses on the problem of assigning nurses with similar qualifications to their home units to balance workload and reduce understaffing costs due to unplanned absences. In particular, we study how hospitals may periodically reassign nurses who have heterogeneous absentee rates to multiple inpatient units that face stochastic demand. The nurse assignment problem is formulated as a stochastic integer program from which we obtain analytical results that serve as a rigorous proof of a high-level concept that many nurse managers will find intuitive—evenly distribute high and low absentee rate nurses among different units as much as possible.

Inpatient units are often organized by nursing skills required to provide care. A typical classification of inpatient units includes the following tiers: intensive care (ICU), step-down, and medical/surgical. Multiple units may exist within a tier, each with a somewhat different specialization. For example, different ICUs may take care of a common pool of patients and in addition have a few beds that are reserved for subpopulations such as cardiac and vascular, surgical, trauma, and pediatric patients. Nurses whose skills are adequate for assignment to a particular type of unit may be assigned to any one of the multiple units of that type, which is then called their home unit. When

nurses take care of a special subpopulation of patients within their home unit, they undertake training that is specific to the needs of that patient group. Hospitals also provide periodic learning opportunities to nurses to keep their skills up to date.

The need to assign (or reassign) nurses to their home units arises periodically. Among the hospitals we have interacted with, nurses' home unit assignment is typically revisited annually.

Some hospitals create a pool of nurses, referred to as the float pool, who may be assigned to different units depending on short-term variation in nursing needs. However, hospitals generally do not dynamically assign non-float-pool nurses to different units to balance nursing needs in the short term because of the potential negative impact on patient safety (especially when patient requirements are different) and increased stress on nurses from working in unfamiliar physical environments, using unfamiliar equipment, and having unfamiliar coworkers (Ferlise and Baggot 2009). Some contract rules in fact prevent nurse managers from temporarily reassigning nurses to work in different units (California Nurse Association and National Nurses Organizing Committee 2012). Hence, an important long-term staffing decision is to find an allocation of non-float-pool nurses to inpatient units to minimize overall costs including those from unplanned absences.

Nurses' absentee rates are high in U.S. hospitals. According to the U.S. Department of Veterans Affairs' Outcomes Database, the average unplanned absentee rate across all hospitals in the Veterans Health Administration's hospitals, which includes sick leaves and leaves without pay, was 6.4% for a 24-week period between September 2011 and February 2012. This statistic is significantly higher than absentee rates among healthcare practitioner/technical occupations and all occupations in the United States, which happen to be 3.7% and 3%, respectively (Bureau of Labor Statistics 2011), underscoring the importance of considering absenteeism in nurse staffing models. We argue in this paper that in addition to balancing supply and demand, careful assignment of nurses to their home units can help lower staffing costs due to unplanned nurse absenteeism.

Hospitals face significant variation in demand for nursing services; see §2.1 for empirical evidence from two hospitals. Nurse absenteeism exacerbates the difficulty of matching supply and demand for nursing services in an economical manner. Nurse shortage jeopardizes quality of care and patients' safety, increases length of stay, and lowers nurses' job satisfaction (e.g., Unruh 2008, Aiken et al. 2002, Needleman et al. 2002, Cho et al. 2003, Lang et al. 2004, Kane et al. 2007). For these reasons, hospitals generally find replacements to fill in absent shifts as much as possible even if this practice is costly. For example, in one of the hospitals discussed in this paper, the total overtime cost was about 2.3 million per two-week pay period. Therefore, implementable strategies that could lower even a fraction of understaffing costs would be attractive to hospitals.

We formulate the problem of assigning a cohort of nurses to their home units as a discrete stochastic optimization problem, which is a combinatorially hard problem. The practical usefulness of our model depends on the characterization of a function that maps assignment of nurses to the random number of nurses who show up in each unit. To characterize such mapping, we utilize computerized records from multiple nursing units of two health systems and develop a statistical model to explain nurse absenteeism.

Because variability in absenteeism and the extent to which some nurses may be similar to others in terms of their pattern of absenteeism will vary from one hospital to another, we consider three different functional forms of absenteeism within the stochastic integer program: (1) deterministic, (2) random aggregate, and (3) nurse-specific Bernoulli. The first model assumes that the number of absences is a deterministic function of the number of nurses assigned to a unit. This model will be appropriate when nurse absences are predictable and it is the variability in demand that drives staffing decisions. In this paper, it serves largely

as a benchmark and as a mean-value approximation of the underlying stochastic optimization problem. It also serves to highlight the importance of making sure that assignments match average supply to mean demand in each unit. The second model groups nurses into subsets such that nurses within a subset are homogeneous, but each subset has a different absentee rate distribution. The third model can be viewed as a special case of the second model in which each subset is of size 1, i.e., each nurse is different and the overall absentee rate distribution of a unit is the convolution of absentee rate distributions of particular nurses assigned to that unit.

Using stochastic orders, we show that greater variability in demand and attendance patterns increases a hospital's costs, and that for a fixed overall absentee rate, each unit's cost is minimized by choosing a more *heterogeneous* cohort of nurses. This suggests that a desirable staffing plan will maximize heterogeneity within a unit but create demand-adjusted uniform plans across units. Whereas the effect of increased variability is consistent with intuition, the desirability of a heterogeneous cohort of nurses is not. To explain this finding, we also show that greater heterogeneity leads to smaller variance in attendance patterns of the cohort of nurses assigned to a unit. This qualitative insight can be extrapolated to other settings in which work is performed by teams of employees.

In addition to establishing the above qualitative staffing guidelines, we also establish that the hospital's objective function is supermodular. Because greedy heuristics generally work well when objective function is supermodular (Topkis 1998, Il'ev 2001, Asadpour et al. 2008, Calinescu et al. 2011), we explore greedy and one other heuristic for solving the staffing-plan optimization problem. The two heuristics are compared to each other and a straw assignment heuristic. Both heuristics are shown to work well in numerical experiments. The heuristics ensure a heterogeneous combination of nurses within each nursing unit. We report the results of simulation experiments using historical data, which realize savings of about 3%–4% of understaffing costs upon following our recommended heuristics relative to the straw heuristic. These experiments suggest that hospitals can reduce staffing costs by utilizing historical attendance data and relatively easy-to-use heuristic approaches for nurse assignment.

This paper makes a contribution to both operations management (OM) and health services research (HSR) literatures. Its contribution to the OM literature comes from establishing certain stochastic order relations and from obtaining qualitative staffing guidelines as well as easy-to-use heuristics for making nurse assignments under different functional forms of absenteeism. Its contribution to the HSR literature comes from testing a variety of potential predictors of nurse absenteeism

in a statistical model and corroborating survey-based results reported in the HSR literature. We elaborate on these points below.

The staffing problem studied in this paper is similar to the random yield problem studied in the OM literature in the sense that the realized staffing level (equivalently, the yield of good items produced) may be lower than the planned staffing level (production lot size) because of nurses' show uncertainty (random yield). Random yield models characterize yield uncertainty in one of the following ways: (1) For any given lot size  $n$ , the yield  $Q(n)$  is a binomial process with a yield rate  $p$ ; e.g., Grosfeld-Nir et al. (2000). (2) The yield is a product of the lot size and a random yield rate (i.e.,  $Q(n) = n \cdot \xi$ , where  $n$  is the lot size, and  $\xi$  is the random yield rate); e.g., Gerchak et al. (1988), Bollapragada and Morton (1999), Araman and Popescu (2010). (3) The production process is in control for a period of time followed by a period when it is out of control (i.e., yield  $Q$  has a geometric distribution); e.g., Grosfeld-Nir et al. (2000). (4) Yield is the result of having random capacity to produce good units (i.e.,  $Q = \min\{n, C\}$ , where random  $C$  captures the unreliability of the equipment); e.g., Ciarallo et al. (1994). (5) The distribution of yield is known (i.e.,  $p(q | n)$  is the probability of  $q$  good units given a lot size  $n$ ); e.g., Gurnani et al. (2000). In addition, Yano and Lee (1995) and Grosfeld-Nir and Gerchak (2004) contain comprehensive reviews of random yield models. These models assume independent and homogeneous yield rates with the result that the problem of determining how to put together production lots (equivalent to the composition of nurses in our setting) does not arise in such settings. Our effort may be viewed as a generalization of random yield models to a situation in which yield rate is different for each item.

Green et al. (2013) formulate a model with endogenous yield rates for medium-term nurse staffing problems that occur at the unit level. The authors use data from one emergency department (ED) of a single hospital and observe that nurses' anticipated workload (measured by the ratio of staffing level in a shift and the long-term average census) is positively correlated with their absentee rate. The results from our data analysis are significantly different from those reported in Green et al. (2013). In particular, short-term workload does not have a consistent effect on absenteeism in our data. The differences arise because of the fundamental differences in the types of data and problem scenarios modeled, which we explain next. First, inpatient units and EDs face different demand patterns and patients' length of stay with patients staying significantly longer in inpatient units. According to surveys done in 2006 and 2010, average length of stay in emergency departments (delay between entering emergency and being admitted or discharged) was 3.7 hours and 4.1 hours,

respectively, (Fuson 2006, Press Ganey Associates 2010). In contrast, the average length of inpatient stay in short-stay hospitals was 4.8 days according to 2007 data (National Center for Health Statistics 2011, Table 99, part 3). Second, it may be argued that EDs present a particularly stressful work environment for nurses, and therefore ED nurses may react differently to workload variation than nurses who work in inpatient units. Third, we use data from multiple units, which allows us to examine whether the impact of workload and other covariates on absentee rate is unit specific or consistent across units, whereas Green et al. (2013) examine data from a single ED.

Much of the OM literature dealing with nurse staffing has focused on developing nurse schedules to minimize costs while satisfying nurses' work preferences; see Lim et al. (2011) for a recent review. These works are not closely related to our paper. There are numerous papers that are motivated by applications outside the healthcare domain that take into account staff absenteeism; see, e.g., Hur et al. (2004), Whitt (2006), and Blumenfeld and Inman (2009). However, to the best of our knowledge, they do not focus on identifying predictors of absenteeism or assigning personnel with heterogeneous absentee rates to different work units.

Some papers in the HSR literature explain why nurses take unplanned time off; see Davey et al. (2009) for a systematic review. This literature concludes that nurse absences may be associated with nurses' personal characteristics (e.g., Laschinger and Grau 2012), prior attendance records (e.g., Roelen et al. 2011), organizational norms or leadership styles of nurse managers (e.g., Schreuder et al. 2011), and chronic work overload and burn out (e.g., Rajbhandary and Basu 2010). Our results are consistent with this literature and contribute to it by presenting quantitative evidence that each nurse's attendance history is a predictor of his or her future absences. Note that the majority of previous studies cited above are based on nurses' self-reported surveys.

The remainder of this paper is organized as follows. Section 2 formulates the nurse assignment problem and presents statistical analyses to support the proposed absenteeism models. Section 3 discusses the properties of an optimal staffing strategy that leads to the proposed staffing heuristics and provides numerical examples comparing performance of different staffing strategies. Section 4 concludes the paper.

## 2. Model Formulation and Analysis

We begin with a list of notation in Table 1, some of which are not needed until §3.

There are  $u$  inpatient units that require nurses with a particular skill set,  $n$  nurses with this skill set are available for a particular shift type (e.g., day shift),



**Table 1** Notation

Indices	
$j$	nurse index, $j = 1, \dots, n$
$t$	nurse type index, $t = 1, \dots, m$ ; $m \leq n$
$i$	unit index, $i = 1, \dots, u$
Parameters	
$\mathbf{p}$	$(p_1, \dots, p_m)$ , absent probabilities by nurse type
$\mathbf{n}$	$(n_1, \dots, n_m)$ , number of nurses by type
$\mathbf{X}$	$(X_1, \dots, X_u)$ , (random) nursing needs vector
$c(\cdot)$	an increasing convex shortage cost function
Decision variables	
$\mathbf{a}^{(i)}$	$(a_1^{(i)}, \dots, a_m^{(i)})$ , number of nurses assigned to unit $i$ by type
$\mathbf{a}$	$(\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(u)})$ , staffing plan
$\mathcal{A}$	set of all possible staffing plans $\mathbf{a}$
Calculated quantities	
$\phi_i(\mathbf{a})$	the number of absent nurses in unit $i$ given assignment $\mathbf{a}$
$\phi_i^{[k]}(\mathbf{a})$	$\phi_i(\mathbf{a})$ under absentee model $k$ , $k \in \{d, r, b\}$
$Q_i(\mathbf{a})$	$= \sum_{t=1}^m a_t^{(i)} - \phi_i(\mathbf{a})$ = number of nurses who show up in unit $i$
$\pi_i(\mathbf{a})$	total expected shortage cost for unit $i$ when staffing plan $\mathbf{a}$ is used
$\pi(\mathbf{a})$	total expected shortage cost when staffing plan $\mathbf{a}$ is used $= \sum_{i=1}^u \pi_i(\mathbf{a})$

and these nurses can be divided into  $m$  types based on their absentee rates. In particular, nurses that belong to the same type have the same probability or pattern of being absent in an arbitrary shift. At one end of the spectrum when  $m = n$ , every nurse is different. At the other end, when  $m = 1$  all nurses are identical. In general, a nurse manager may choose  $1 \leq m \leq n$  to classify nurses. The objective is to minimize expected total cost and the staffing-plan optimization problem can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \pi(\mathbf{a}) = \sum_{i=1}^u E(c(X_i - Q_i(\mathbf{a}))^+) \\ \text{subject to} \quad & \sum_{i=1}^u a_t^{(i)} \leq n_t, \quad t \in \{1, \dots, m\} \text{ and } \mathbf{a} \in \mathcal{A}, \end{aligned} \quad (1)$$

where  $c(\cdot)$  is a function that captures the cost of understaffing resulting from utilizing overtime/float/agency nurses, or reduced service and cancellations,  $X_i$  is unit  $i$ 's demand for nursing services for the chosen shift type,  $Q_i(\mathbf{a})$  denotes the random number of nurses who show up for work in unit  $i$  under assignment  $\mathbf{a}$ , and  $\mathcal{A}$  is the set of all possible nurse assignments. The decision variables are the components of matrix  $\mathbf{a}$ , where the  $(t, i)$ th element is the number of type  $t$  nurses assigned to work in unit  $i$ . The expectation is taken over all possible realizations of  $X_i$  and  $Q_i(\mathbf{a})$  and  $(x - q)^+ = \max\{x - q, 0\}$ . This staffing problem is non-trivial because there are  $n^u$  possible nurse assignments, each of which could lead to a different cost.

The model specification is not complete until we specify the functional form of  $Q_i(\mathbf{a})$ , which drives nurse allocation decisions. For example, if nurses are homogeneous, it does not matter which nurses are assigned to which units. In contrast, if nurses are heterogeneous,

the particular cohort of nurses assigned to a unit does matter. Therefore, we first study staffing and absence data from two health systems to ascertain the functional form of  $Q_i(\mathbf{a})$  in §2.1. This allows us to choose factors that potentially constitute  $Q_i(\mathbf{a})$  including unit index, nurses who are assigned, shift time, month, holiday, weather, and workload. The choice of such factors are also based on interactions with nurse managers and findings in previous studies. Then, we test which factors should be included in the stochastic optimization model by analyzing a statistical model.

## 2.1. Evidence from Data

We studied de-identified census and absentee records from two hospitals located in a large metropolitan area. Census and absentee data from hospital 1 were for the period January 3, 2009, through December 4, 2009, whereas hospital 2's data were for the period September 1, 2008, through August 31, 2009. Basic information about these hospitals from fiscal year 2009 is summarized in Table 6 in Online Supplement A. (The online supplement is available at <http://dx.doi.org/10.1287/msom.2014.0486>.) The differences between the maximum and the minimum patient census in 2009 were 52.7% and 49.5% of the average census for hospitals 1 and 2, respectively, indicating that the overall variability in nursing demand was high. Patients' average lengths of stay were 4.0 and 4.9 days and registered nurse (RN) salary accounted for 17.9% and 15.5% of total operating expenses of the two hospitals.

Hospital 1 had five shift types. There were three eight-hour shifts designated as day, evening, and night, which operated from 7 A.M. to 3 P.M., from 3 P.M. to 11 P.M., and from 11 P.M. to 7 A.M., respectively. There were also two 12-hour shifts, which were designated as day 12 and night 12. These operated from 7 A.M. to 7 P.M. and from 7 P.M. to 7 A.M., respectively. Hospital 2 had only three shift types, namely, the eight-hour day, evening, and night shifts. Hospital 1's data pertained to three step-down (telemetry) units labeled T1, T2, and T3 with 22, 22, and 24 beds; and hospital 2's data pertained to two medical/surgical units labeled M1 and M2 with 32 and 31 beds. The common data elements were hourly census, admissions, discharges, and transfers, planned and realized staffing levels, and the count of absentees for each shift. Hospital 1's data also contained individual nurses' attendance history.

Both hospitals used target nurse-to-patient ratios to describe the desirable workload for their inpatient units. The target ratios were determined by panels of experts (mostly experienced nurse managers) in each hospital. Hospital 1's target nurse-to-patient ratios for telemetry units were 1:3 for day and evening shifts during week days and 1:4 for night and weekend shifts. Hospital 2's target nurse-to-patient ratios for medical/surgical units were 1:4 for day and evening

shifts and 1:5 for night shifts. Hospital 1's planned staffing levels were based on the mode of the midnight census in the previous planning period. Nurse managers would further tweak the staffing levels up or down to account for holidays and to meet nurses' planned-time-off requests and shift preferences. Hospital 2's medical/surgical units had fixed staffing levels based on the long-run average patient census by day-of-week and shift. In both cases, staff planning was done in four-week increments and planned staffing levels were posted two-weeks in advance of the first day of each four-week plan. Consistent with the fact that average lengths of stay in these hospitals were between four and five days, staffing levels were not based on a projection of short-term demand forecast. When the number of patients assigned to some nurses exceeded the target nurse-to-patient ratios, nurse managers attempted to increase staffing by utilizing extra-time or overtime shifts, or calling in agency nurses. Similarly, when census was less than anticipated, nurses were assigned to indirect patient care tasks or education activities, or else asked to take voluntary time off. These efforts were not always successful and realized nurse-to-patient ratios often differed from the target ratios. For example, Hospital 1's unit T3 on average staffed lower than the target ratios, whereas hospital 2's unit M2 on average staffed higher than the target ratios during weekends; see Figure 1 where the solid dots were plotted at (mean, standard deviation) pairs of number of patients per nurse corresponding to the horizontal (respectively, vertical) axes by shift type and the 95% confidence intervals of the mean were drawn along the horizontal lines.

In addition to the two hospitals whose data are analyzed in this paper, the authors have interacted with nurse managers at numerous other urban hospitals. Staffing practices do vary from one hospital to another. However, the range of practices prevalent at these two hospitals are representative of many other hospitals. That is, models based on such practices should be useful for other hospitals as well.

**Table 2** Summary Statistics for Individual Nurses' Absentee Rates in Hospital 1's Telemetry Units

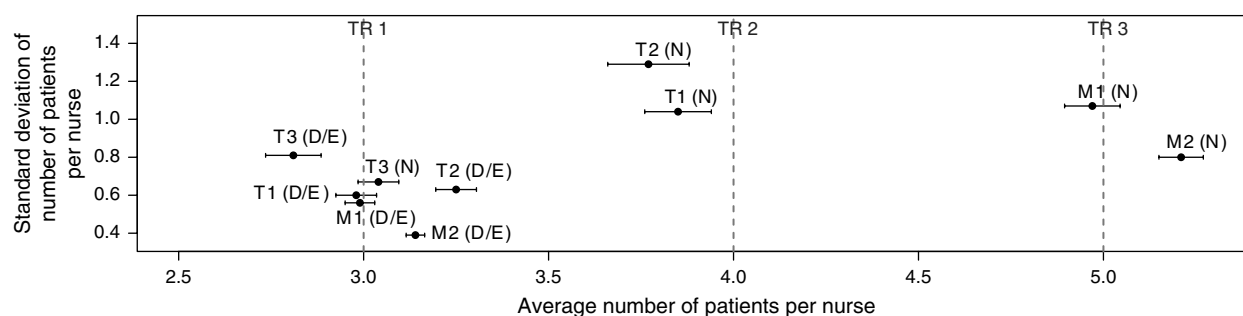
	T1	T2	T3
Number of nurses	50	45	58
Mean (%)	12.6	12.5	8.4
Median (%)	7.9	9.5	6.8
Standard deviation (%)	12.9	14.9	6.1
Minimum (%)	0	0	0
Maximum (%)	63	32	27
25th percentile (%)	4.6	5.3	4.7
75th percentile (%)	16.5	14.9	11.1
Coefficient of variation	1.02	1.2	0.7

The two hospitals' data were analyzed independently because (1) the data pertained to different time periods, (2) the target nurse-to-patient ratios were different for the two types of inpatient units, and (3) the two hospitals used different staffing strategies. Based on nurse managers' suggestion and our initial analysis of data, which we describe in the next three paragraphs, we considered the following potential predictor of nurse absenteeism: (1) individual nurses' attendance history, (2) unit index (which potentially captures factors that were unit specific, e.g., unit culture), (3) day of week, (4) shift time, (5) month of year (season), (6) holiday, and (7) extreme weather. In addition, we also considered workload because workload has been found to be correlated with absenteeism in previous studies (e.g., Green et al. 2013).

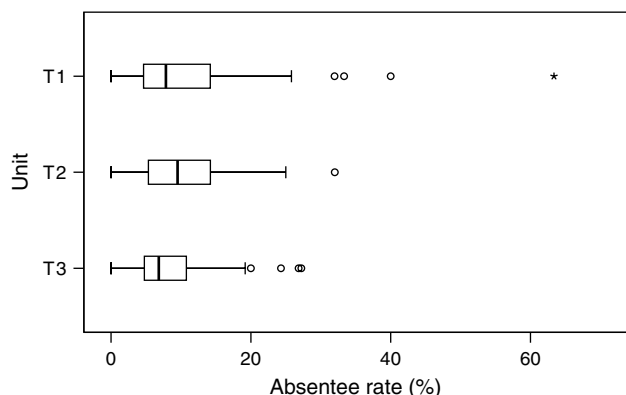
Table 2 summarizes individual nurses' absentee rate statistics for hospital 1's three telemetry units. With a coefficient of variation of 1.02, 1.2, and 0.70 in units T1, T2, and T3, it is reasonable to conclude that the absentee rates among nurses are highly variable and nurses are heterogeneous. The high variability among individual nurses' absentee rates can also be seen in Figure 2.

We calculated each hospital-1 nurse's quarterly absentee rate and found that there was not a statistically

**Figure 1** Average Number of Patients per Nurse by Shift Type



*Notes.* D, day; E, evening; N, night. TR 1, target ratio for T1, T2, and T3 in D and E shifts during weekdays. TR 2, target ratio for T1, T2, and T3 in N shifts or weekends; TR 2 is also the target ratio for M1 and M2 in D and E Shifts. TR 3, target ratio for M1 and M2 in N shifts. A numerical summary of these statistics is included in Table 8 of Online Supplement A.

**Figure 2** Absentee Rate Distribution (Boxplots) for Nurses in Hospital 1's Telemetry Units

significant difference in nurses' absentee rates across different quarters. This suggests that although absentee rates were highly variable among hospital 1's nurses, each individual nurses' absentee rate was relatively stable across time. This observation supports our choice of nurses' attendance history as a potential predictor of future absences.

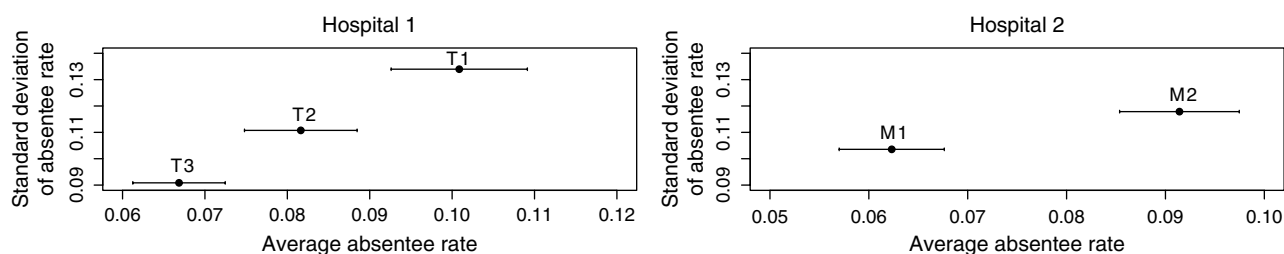
The absentee rate among hospital 1's three telemetry units varied from 3.4% (T1, Sunday, day shift) to 18.3% (T1, Saturday, night shift) depending on unit, shift time, and day of week. Similarly, the absentee rate for hospital 2's medical/surgical units varied from 2.99% (M1, Wednesday, evening shift) to 12.98% (M2, Tuesday, evening shift) depending on unit, shift time, and day of week. The 95% confidence intervals for absentee rates did not overlap across all units and shift types

(Figures 3 and 4) but overlapped across all days of week (see Table 8 in Online Supplement A). These statistics suggest that absentee rates were significantly different by unit and shift, but not by the day of week. We also compared absentee rates for regular days and holidays, and fair-weather days and storm days. At a 5% significance level, holidays had a lower average absentee rate than nonholidays for hospital 2 and storm days had a higher average absentee rate for hospital 1. We kept month as a potential predictors as there may be seasonality in absentee rates. A numerical summary of absentee rates is included in Table 8 in Online Supplement A.

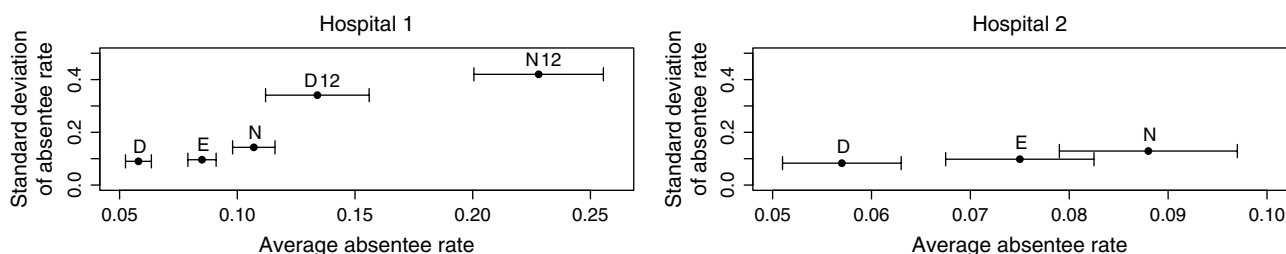
## 2.2. The Statistical Model

The observations in §2.1 suggest that absentee rates may vary by nurse, shift type, holiday, weather, month of year, and workload. Although some of these factors will not be included in the stochastic optimization model because either they are not actionable for long-term nurse assignment decisions or their use is prevented by contract rules, we will include them in the statistical model to obtain a more complete analysis. Note that only hospital 1's data contained individual nurses' history of schedules. Hospital 2's individual nurse records were not available.

To separate individual nurse effect from group-level effects, we fitted hospital 1 data to the following generalized linear mixed model (Equation (2)). The model assumes that each nurse  $j$  has his or her own personal traits that contribute to absenteeism (a random-intercept model). In addition, because the vast majority of the nurses (with only a few exceptions) work only

**Figure 3** Absentee Rate by Unit

Note. A numerical summary of these statistics is included in Table 8 of Online Supplement A.

**Figure 4** Absentee Rate by Shift Type

Note. A numerical summary of these statistics is included in Table 8 of Online Supplement A.

in one nursing unit during the time period covered by our data set, the nurses are considered nested within the inpatient units and we also treated unit effect as random. Note that the potential interaction between unit and nurse assignment could not be evaluated with this data because very few nurses were scheduled to work in more than one unit during the period. Our generalized linear mixed model is the following:

$$\log\left(\frac{p_{ijt}}{1-p_{ijt}}\right) = \mathbf{X}_{ijt}\boldsymbol{\beta} + \mathbf{L}_i + \mathbf{B}_{ij}, \quad (2)$$

where  $p_{ijt}$  is the absence probability for unit  $i$ 's nurse  $j$  when his or her  $t$ th scheduled shift has characteristics  $\mathbf{X}_{ijt}$ ,  $\boldsymbol{\beta}$  represents the coefficients for factors described in  $\mathbf{X}_{ijt}$ ,  $\mathbf{L}_i$  is the random effect explained by unit  $i$ , and  $\mathbf{B}_{ij}$  represents the nurse- $j$ -specific intercept—i.e., nurse  $j$ 's impact on his or her absenteeism that is not captured by  $\mathbf{X}_{ijt}$  and the unit effect. Note that  $\mathbf{X}_{ijt}$  is a vector that captures information concerning the shift type (i.e., D, E, N, D12, or N12); standardized workload (relative to the unit's target workload, which we explain in next paragraphs); month (i.e., January, ..., December); holiday (i.e., nonholiday or holiday); weather (i.e., normal or extreme); and the interaction between shift type and workload. We include the interaction between shift type and workload because the target nurse to patient ratios for different shifts are not the same and thus the impact of workload for different shifts may not be the same.

We measured unit  $i$ 's and shift  $t$ 's anticipated workload in three different ways: (1)  $w_{it}^{(1)} = n_{it}/E[C_{it}]$ , (2)  $w_{it}^{(2)} = \sum_{m=1}^m (c_{t-m}/m)(1/n_{it})$ , and (3)  $w_{it}^{(3)} = \sum_{m=1}^m (c_{i(t-m)}/m)$ , where  $n_{it}$  is the planned staffing level,  $c_{it}$  is unit  $i$ 's start-of-shift census for shift  $t$ , and  $E[C_{it}]$  is the long-run expected census for unit  $i$ . Put differently,  $w_{it}^{(1)}$  equals the anticipated nurse-to-patient ratio,  $w_{it}^{(2)}$  equals the  $m$ -period moving average of estimated number of patients per nurse, and  $w_{it}^{(3)}$  equals the  $m$ -period moving average of census. The choice of  $w_{it}^{(1)}$  is appropriate for units with stable nursing demand,  $w_{it}^{(2)}$  for units in which both census and staffing levels vary from shift to shift, and  $w_{it}^{(3)}$  for units that have constant staffing levels (such as in hospital 2).

In Equation (2), we used standardized workload  $\tilde{w}_{it}^{(k)}$ , which was calculated as follows:  $\tilde{w}_{it}^{(k)} = ((w_{it}^{(k)} - \bar{w}_{it}^{(k)})/\bar{w}_{it}^{(k)}) \times 100\%$ , where  $\bar{w}_{it}^{(k)}$  is the workload at the target nurse-to-patient ratio for the shift type to which shift  $t$  belongs for  $k=1, 2$ , and the average census for the shift type for  $k=3$ . If shift  $t$ 's anticipated workload equals its target nurse-to-patient ratio, then  $\tilde{w}_{it}^{(i)} = 0$ . This allows us to obtain a more intuitive interpretation of the coefficients that are related to shift and workload: the shift coefficients can be explained as the shift effect at the target workload (when  $\tilde{w}_{it}^{(k)} = 0$ ), and the coefficients for  $\tilde{w}_{it}^{(k)}$  are the impact of a 1% increase or decrease in workload on absentee rates.

### 2.3. Statistical Analysis Results

The model, regardless of workload variants, leads to a conclusion that individual nurses' random effect is much stronger than unit random effect. For the sake of brevity, complete details are presented in Online Supplement B.1 and here we highlight our key findings. The standard deviation of  $B_{ij}$  is three to seven times higher than the standard deviation of  $L_i$ . In particular, the standard deviation of  $B_{ij}$  and  $L_i$  are (0.75, 0.24), (0.75, 0.20), and (0.75, 0.11) for the three workload model variants, respectively. This result supports the inference that unit effect is less important than individual nurse effect, and hospitals may achieve greater efficiencies by periodically reallocating some nurses to different units.

Our goal is to identify actionable factors that can be utilized in the nurse assignment decisions. For this reason, we check whether the fixed effects considered in model (2) are consistent across units by fitting the model to each unit separately. We drop the random unit effect  $L_i$  in (2) and fit the model to each unit one by one. The models of three workload variants lead to the same conclusions. Therefore, we only report the case with  $\tilde{w}_{it}^{(1)}$  in Table 3 for the sake of brevity. Results for models with  $\tilde{w}_{it}^{(2)}$  and  $\tilde{w}_{it}^{(3)}$  when  $m=6$  are available in Tables 10 and 11 in Online Supplement B.2. Note that for workload models 2 and 3, we varied  $m$  (the moving average parameter) between 4 and 10 and the conclusion from the statistical analysis remains the same.

From Table 3, we observe that nurses' random effect was strong (see the standard deviations of the random intercepts). For example, when in unit T1 a D12 shift's workload was 10% higher than its average workload (i.e.,  $\tilde{w}_{it}^{(1)} = -10\%$  because a lower  $\tilde{w}_{it}^{(1)}$  indicates a higher workload based on the definition of  $w_{it}^{(1)}$ ), the log odds of absentee rate of a given nurse would decrease by  $10 \times (0.007 + 0.07) = 0.77$ , which was smaller than the standard deviation among nurses' random intercepts (0.873). The coefficients for month effects were also much smaller than the standard deviation of the random intercepts for nurses. This suggests that nurses' random effect on absenteeism was much stronger than workload and month effects. The strong nurse random effects can also be observed in units T2 and T3.

Among the fixed effects, we found that extreme weather was significant for unit T1, but not T2 and T3. Holiday effect was not significant. Month effect was inconsistent across units—a particular month that has a higher or lower absentee rate for one unit does not have the same effect for a different unit. In addition, even for the same unit, month effect is not consistent when we use different workload measures in the model. It is possible that some units in some months of the year have significantly higher or lower absentee rates.



**Table 3** Result of the Statistical Model When Workload Is Evaluated by  $\tilde{w}_{it}^{(1)}$ 

Unit:	T1			T2			T3		
Random effect:	Standard deviation			Standard deviation			Standard deviation		
Nurse (intercept)	0.873			0.621			0.638		
Fixed effects	Estimate	S.E.	$P(>  z )$	Estimate	S.E.	$P(>  z )$	Estimate	S.E.	$P(>  z )$
(Intercept)	−2.578	0.261	<0.001*	−2.448	0.216	<0.001*	−3.144	0.267	<0.001*
Shift E	0.068	0.241	0.779	0.398	0.173	0.022*	0.521	0.255	0.041*
Shift N	0.893	0.287	0.002*	0.401	0.206	0.052	0.468	0.287	0.103
Shift D12	0.376	0.601	0.531	0.547	0.280	0.051	−1.666	0.558	0.003*
Shift N12	1.489	0.578	0.010*	0.321	0.271	0.236	−0.011	0.400	0.977
$\tilde{w}_{it}^{(1)}$	0.007	0.005	0.158	0.007	0.005	0.198	0.011	0.005	0.028
Month Feb	−0.002	0.234	0.995	−0.112	0.227	0.620	−0.001	0.228	0.995
Month Mar	−0.388	0.243	0.110	−0.142	0.228	0.534	−0.215	0.236	0.362
Month Apr	−0.056	0.228	0.807	−0.501	0.246	0.042*	−0.115	0.230	0.618
Month May	−0.228	0.238	0.339	−0.239	0.241	0.322	−0.345	0.255	0.176
Month Jun	−0.192	0.238	0.419	−0.235	0.243	0.334	−0.257	0.238	0.279
Month Jul	−0.567	0.257	0.028*	−0.022	0.240	0.926	0.232	0.225	0.302
Month Aug	−0.451	0.249	0.071	0.032	0.232	0.889	−0.066	0.227	0.770
Month Sep	−0.426	0.247	0.085	−0.568	0.267	0.033*	−0.189	0.239	0.429
Month Oct	0.156	0.224	0.485	0.271	0.215	0.206	0.405	0.208	0.052
Month Nov	−0.328	0.241	0.173	−0.076	0.232	0.744	−0.130	0.229	0.569
Month Dec	−1.191	0.570	0.037*	−0.081	0.440	0.854	−0.991	0.620	0.110
Holiday	−0.126	0.361	0.726	−0.107	0.326	0.744	0.581	0.300	0.053
Weather	0.704	0.325	0.030*	0.346	0.376	0.357	−0.372	0.449	0.407
Shift E: $\tilde{w}_{it}^{(1)}$	<0.001	0.007	0.965	0.001	0.007	0.883	−0.002	0.006	0.674
Shift N: $\tilde{w}_{it}^{(1)}$	−0.016	0.006	0.015*	0.002	0.007	0.775	−0.005	0.006	0.369
Shift D12: $\tilde{w}_{it}^{(1)}$	0.070	0.033	0.033*	0.063	0.014	<0.001*	0.046	0.011	<0.001*
Shift N12: $\tilde{w}_{it}^{(1)}$	0.006	0.015	0.667	0.011	0.009	0.181	0.009	0.008	0.245

Notes. Standard deviation is of the random intercepts. S.E., standard error of the estimate,  $P(> |z|)$ :  $p$ -value of Wald statistics.

\*: Significant at 0.05 level.

For example, it can happen that several nurses in some months take unplanned sick days. However, without a consistent pattern, such effects cannot be the basis for long-term staffing decisions. Data show that monthly variations in absentee rates require a short-term mitigation strategy because of the lack of a consistent pattern.

The impact of workload on absentee rates was not consistent across shifts and units for the three variants of model (2) mentioned above, and the impact (if statistically significant) was small. Furthermore, different  $\tilde{w}_{it}^{(k)}$  led to different conclusions for the same nursing unit. Take unit T1 in Table 3 as an example. The anticipated workload was negatively correlated with D12 shifts, but positively correlated with night shifts. Workload, when evaluated by  $w_{it}^{(2)}$  was significant for N12 shift for T1, but not significant in T2 and T3 (Table 10 of Online Supplement B.2). Workload, when evaluated by  $\tilde{w}_{it}^{(3)}$  was not significant for T2 (see Table 11 of Online Supplement B.2). Finally, if we take into account the magnitude of random nurse effect, we see that the standard deviation of the random intercept is much higher than the fixed effects due to workload. For these reasons, we model heterogeneous nurses and

do not incorporate workload in the long-term nurse assignment problem.

When we look at shift effect at the target workload (i.e.,  $\tilde{w}_{it}^{(1)} = 0$ ), shift effects were not consistent across units. T1's night and N12 shifts had significantly higher absentee rates than day, evening, and D12 shifts; T2's evening shifts had higher absentee rates than day shifts; T3's evening shifts had a higher absentee rate and D12 shifts had a lower absentee rate than the other shifts. This inconsistency may come from unit-specific reasons that were not included in the model. When we refer back to the unit-pooled model (Table 9 in Online Supplement B.1) in which we evaluated shift effect for all three units together while allowing random unit effects, we find that although some shift types might have higher or lower absentee rates, shift effects were weaker than the variability of nurse random effects.

The above analysis shows that nurses have heterogeneous absentee rates and shift effect may exist. We analyze the nurse assignment problem in (1) for each shift type. This approach is taken because nurses' shift work patterns are governed by their individual contracts and most nurses have a primary work shift. We assume that when we reassign nurses to different units, those nurses will still be scheduled to work

for the same shift time dictated by the contract rule. Although this leaves open the possibility that shift effect may be confounded with individual nurses' absentee pattern, our data does not allow us to tease out shift and nurse effects separately and the actionable factor considered in §3 is the nurse-specific absentee rate.

### 3. Staffing Strategies

As mentioned in the introduction section we consider three functional forms of  $Q_i(\mathbf{a})$ : (1) deterministic, (2) random aggregate, and (3) nurse-specific Bernoulli. These models are denoted by letters  $d$ ,  $r$ , and  $b$ , respectively, and notation needed for further analysis is summarized in Table 1. Note that some of this notation was introduced earlier as part of model formulation. Although we have identified that  $Q_i(\mathbf{a})$  is driven by heterogeneous nurses who are assigned to the unit in §2, we evaluate these three functional forms of  $Q_i(\mathbf{a})$  in the staffing problem for the following reasons. The deterministic functional form of  $Q_i(\mathbf{a})$  is used to obtain a benchmark performance when a nurse manager has perfect information about which nurses are going to be absent. The random aggregate functional form of  $Q_i(\mathbf{a})$  is used to evaluate the impact of variability in nurses' absentee rates. Finally, the nurse-specific Bernoulli model is used to capture the feature of heterogeneous absentee rates to evaluate its impact on staffing. These three functional forms help us identify characteristics of optimal staffing assignments.

The first model assumes that there is a deterministic mapping from  $\mathbf{a}^{(i)}$  to the number of absentees, i.e.,  $\phi_i(\mathbf{a}) = \phi_i^{[d]}(\mathbf{a}) \in [0, \sum_{t=1}^m a_t^{(i)}]$ . The deterministic model does not represent reality but serves as a benchmark mean value approximation of the underlying stochastic optimization problem (Birge and Louveaux 2011). In the random aggregate model, nurse absence is viewed as a group characteristic such that there is a type-specific random absence uncertainty  $\xi_t \in [0, 1]$  for each type  $t$  nurse that is independent of  $a_t^{(i)}$ , and  $\phi_i^{[r]}(\mathbf{a}, \xi) = \sum_{t=1}^m a_t^{(i)} \xi_t$ . Note that this is similar to the multiplicative random yield model studied extensively in the OM literature (see, e.g., Yano and Lee 1995). The third model is the closest to the functional form identified from the analysis of hospital 1 data. It considers each nurse as an independent decision maker with no-show probability  $p_t$ . In this case  $m = n$  and  $n_t = 1$  because each nurse is a type. The number of nurses who are absent equals  $\phi_i(\mathbf{a}) = \phi_i^{[b]}(\mathbf{a}) = \sum_{t=1}^n a_t^{(i)} B(p_t)$ , where  $B(p_t)$  is a Bernoulli random variable with parameter  $p_t$ , and each  $a_t^{(i)}$  is either 0 or 1.

In what follows, we index nurse types such that  $p_1 \geq p_2 \geq \dots \geq p_m$  and evaluate the effect of demand variability and absentee rate variability on the performance of an assignment. For this purpose, we use certain

definitions and established results from the theory of stochastic orders and majorization. These definitions and results are presented next. Further details can be found in Shaked and Shanthikumar (2007), Marshall et al. (2011), and Liyanage and Shanthikumar (1992).

**DEFINITION 1.** Given nursing requirement vectors  $\mathbf{X}$  and  $\mathbf{X}'$ ,  $\mathbf{X}$  is said to be smaller than  $\mathbf{X}'$  in the increasing convex order ( $\mathbf{X} \leq_{icx} \mathbf{X}'$ ) in a component-wise manner if  $E[g(X_i)] \leq E[g(X'_i)]$  for every increasing convex function  $g$  for which the expectations exist and every  $i = 1, \dots, u$ . Intuitively, if  $\mathbf{X} \leq_{icx} \mathbf{X}'$ , then  $\mathbf{X}$  is both smaller and less variable than  $\mathbf{X}'$  in some stochastic sense.

**DEFINITION 2.** Given vectors  $\mathbf{p}$  and  $\mathbf{p}'$ , where the components of these vectors are indexed such that  $p_1 \geq p_2 \geq \dots \geq p_m$ ,  $p'_1 \geq p'_2 \geq \dots \geq p'_m$ , and  $\sum_{t=1}^m p_t = \sum_{t=1}^m p'_t$ , we say that vector  $\mathbf{p}$  is majorized by  $\mathbf{p}'$ , written  $\mathbf{p} \leq_M \mathbf{p}'$  if  $\sum_{t=1}^l p_t \leq \sum_{t=1}^l p'_t$  for every  $l \leq m$ . When the requirement of the equality  $\sum_{t=1}^m p_t = \sum_{t=1}^m p'_t$  is dropped, we say that  $\mathbf{p}'$  weakly sub-majorizes  $\mathbf{p}$  and is denoted  $\mathbf{p} \leq_{WM} \mathbf{p}'$ .

Majorization compares the dissimilarity of the components of a vector. Intuitively, if  $\mathbf{p} \leq_M \mathbf{p}'$ , then components of  $\mathbf{p}'$  are more heterogeneous than  $\mathbf{p}$ .

**DEFINITION 3.** Let  $\{Z(\theta), \theta \in \Theta\}$  be a family of random variables with survival functions  $\bar{F}_\theta(z) = P(Z(\theta) > z)$ ,  $\theta \in \Theta$ . The family  $\{Z(\theta), \theta \in \Theta\}$  is said to be stochastically increasing and linear in the sense of the usual stochastic order, denoted  $SIL(st)$ , if  $E[g(Z(\theta))]$  is increasing linearly in  $\theta$  for all increasing functions  $g$ .

**THEOREM 1** (SHAKED AND SHANTHIKUMAR 2007, THEOREM 8.C.1). *The family  $\{Z(\theta), \theta \in \Theta\}$  satisfies  $\{Z(\theta), \theta \in \Theta\} \in SIL(st)$  if and only if  $\bar{F}_\theta(z)$  is linear in  $\theta$  for each fixed  $z$ .*

**COROLLARY 1** (SHAKED AND SHANTHIKUMAR 2007, THEOREM 8.A.3).  *$B(p)$  is  $SIL(st)$ .*

**DEFINITION 4.** A real-value function  $g$  defined on a set  $\Theta \subset R^n$  is said to be Schur convex on  $\Theta$  if  $\theta \leq_M \theta' \Rightarrow g(\theta) \leq g(\theta')$ .

In other words, functions that preserve the ordering implied majorization are said to be Schur convex.

**DEFINITION 5.** A real-valued random variable  $Z(\theta)$  parameterized by  $\theta \in \Theta$  is stochastically Schur convex in the sense of the usual stochastic [increasing convex] ordering if for any  $\theta, \theta' \in \Theta$ ,  $\theta \leq_M \theta'$  implies  $Z(\theta) \leq_{st} [\leq_{icx}] Z(\theta')$ . We denote this  $\{Z(\theta), \theta \in \Theta\} \in S - SchurCX(st) [S - SchurCX(icv)]$ . If in addition  $Z(\theta)$  is stochastically increasing in  $\theta$ , then  $Z(\theta)$  is stochastically increasing and Schur convex. We denote this  $\{Z(\theta), \theta \in \Theta\} \in SI - SchurCX(st) [SI - SchurCX(icx)]$ .

**DEFINITION 6.** A function  $f: R^k \rightarrow R$  is supermodular if  $f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) + f(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in R^k$ , where  $\mathbf{x} \vee \mathbf{y}$  denotes the component-wise maximum and  $\mathbf{x} \wedge \mathbf{y}$  denotes the component-wise minimum of  $\mathbf{x}$  and  $\mathbf{y}$  (Topkis 1998). If a function  $f$  is supermodular, then function  $-f$  is submodular.

**THEOREM 2** (LIYANAGE AND SHANTHIKUMAR 1992, THEOREM 2.12). Suppose  $\{Z_i(\theta), \theta \in \Theta\} \in \text{SIL}(st)$ . Then for any increasing symmetric submodular [supermodular] function  $h$  and  $Z(\theta) = h(Z_1(\theta_1), \dots, Z_m(\theta_m))$  one has  $\{Z(\theta), \theta \in \Theta^m\} \in \text{SI} - \text{SchurCX}(icv)[\text{SI} - \text{SchurCV}(icx)]$ . That is for any  $\theta, \theta' \in \Theta^m$ , increasing and symmetric submodular [supermodular] function  $h$  and increasing and concave [convex] function  $g$ ,

$$\begin{aligned} \theta \leq_{\text{WM}} \theta' &\Rightarrow E g \circ h(Z_1(\theta_1), \dots, Z_m(\theta_m)) \\ &\leq E g \circ h(Z_1(\theta'_1), \dots, Z_m(\theta'_m)). \end{aligned}$$

Now we are ready to analyze the properties of optimal nurse assignments. In what follows, we use the asterisk notation to denote optimal quantities. In particular,  $\mathbf{a}_\alpha^*$  denotes an optimal assignment of nurses when the problem is characterized by parameter  $\alpha$ . For example, suppose  $\mathbf{X}$  and  $\mathbf{X}'$  denote two different nurse requirement vectors. Then,  $\mathbf{a}_\mathbf{X}^*$  and  $\mathbf{a}_{\mathbf{X}'}^*$  are used to denote optimal staffing plans with nurse requirements  $\mathbf{X}$  and  $\mathbf{X}'$ , respectively. With these notations in hand, we carry out stochastic comparisons to obtain the following results.

**PROPOSITION 1.** If  $\mathbf{X} \leq_{\text{icx}} \mathbf{X}'$  in a component-wise sense, then  $\pi(\mathbf{a}_\mathbf{X}^*) \leq \pi(\mathbf{a}_{\mathbf{X}'}^*)$ .

**PROOF FOR PROPOSITION 1.** Note that  $X_i \leq_{\text{icx}} X'_i$  for each  $i = 1, \dots, u$ . Therefore,  $E[c(X_i - Q_i(\mathbf{a}))^+] \leq E[c(X'_i - Q_i(\mathbf{a}))^+]$  for each  $i$  and  $\pi(\mathbf{a}_\mathbf{X}^*) = \sum_{i=1}^u E[c(X_i - Q_i(\mathbf{a}_\mathbf{X}^*))^+] \leq \sum_{i=1}^u E[c(X_i - Q_i(\mathbf{a}_{\mathbf{X}'}^*))^+] \leq \sum_{i=1}^u E[c(X'_i - Q_i(\mathbf{a}_{\mathbf{X}'}^*))^+] = \pi(\mathbf{a}_{\mathbf{X}'}^*)$ .  $\square$

Proposition 1 states that larger and more variable demand leads to greater expected shortage costs. This statement applies to all three models. The result in Proposition 1 is consistent with intuition because for a fixed level of available staff, shortage costs increase when demand is larger and more variable.

**PROPOSITION 2.** Let  $\phi_i(\mathbf{a}) = \phi_i^{[d]}(\mathbf{a})$  be a deterministic mapping from  $\mathbf{a}^{(i)}$  to  $[0, \sum_{t=1}^m a_t^{(i)}]$ . If  $\{X_i\}$  are independent and identically distributed (i.i.d.), then an optimal staffing plan is realized upon making  $Q_i(\mathbf{a}^*)$  equal for all  $i = 1, \dots, u$ .

**PROOF OF PROPOSITION 2.** The statement of Proposition 2 comes from a property of Schur-convex functions. This property says that if a vector  $\mathbf{q} = (q_1, \dots, q_u)$  is majored by a vector  $\mathbf{q}' = (q'_1, \dots, q'_u)$  where  $\sum_{i=1}^u q_i = \sum_{i=1}^u q'_i = \bar{q}$  and  $g(\cdot)$  is a convex

function, then  $\sum_{i=1}^k g(q_i) \leq \sum_{i=1}^k g(q'_i) \forall k = 1, \dots, u$  (Marshall et al. 2011).

Let  $g(q_i) = E[c(X_i - q_i)^+]$ , where  $q_i$  is the (deterministic) number of nurses who show up in unit  $i$ . It is straightforward to verify that  $g(\cdot)$  is convex in  $q_i$ . Also, let  $\bar{q} = (1/u) \sum_{i=1}^u q_i$ . Because  $(\bar{q}, \dots, \bar{q}) \leq_M (q_1, \dots, q_u)$ , the inequality  $\sum_{i=1}^u g(\bar{q}) \leq \sum_{i=1}^u g(q_i)$  holds for all convex functions.

Note that inpatient units' demands are independent and identically distributed, therefore the cost function  $\pi_i(\mathbf{a}) = E[c(X_i - q_i)^+]$  is the same for all  $i$  and  $q_i$ . Based on the above Schur-convex function property, the best expected shortage cost ( $\pi(\mathbf{a}^*) = \sum_{i=1}^u \pi_i(\mathbf{a}^*)$ ) is achieved when each unit's staffing level is equal (i.e.,  $q_i = Q_i(\mathbf{a}) = \sum_{t=1}^m a_t^{(i)} - \phi_i(\mathbf{a}) = q \forall i = 1, \dots, m$ ).  $\square$

Proposition 2 states that if absenteeism can be predicted reasonably well and inpatient units have i.i.d. demand, then the nurse manager should assign nurses such that each unit has the same realized staffing level. This result makes sense on an intuitive level because balanced staffing across identical units can avoid scenarios in which some units frequently face excess staffing and others frequently face understaffing.

For the random aggregate absence model, we assume that the number of absentees upon scheduling  $a_t^{(i)}$  type  $t$  nurse in unit  $i$  is determined by a function  $g_t(a_t^{(i)}, \xi_t)$ , where  $\xi_t$  represents the uncertainty regarding type  $t$  absentees. We also assume that  $a_t^{(i)}$  and  $g_t(a_t^{(i)}, \xi_t)$  could take any value between 0 and the total number of nurses assigned to unit  $i$ . In particular,  $\phi_i^{[r]}(\mathbf{a}, \xi) = \sum_{t=1}^m g_t(a_t^{(i)}, \xi_t) = \sum_{t=1}^m [g_{t,1}(a_t^{(i)}) + g_{t,2}(a_t^{(i)})\xi_t]$ , and for each  $\xi_t$ ,  $g_t(a, \xi_t) \leq a$ ,  $0 \leq g'_t(a, \xi_t) \leq 1$ , and  $g''_t(a, \xi_t) \leq 0$ . Relaxing the discrete nurse staffing level assumption helps to identify high-level nurse assignment strategies. In the following analysis, we consider the case in which the shortage cost is linear in the number of shifts short, i.e.,  $c(x) = c_0 \cdot x$ , for every  $x$  for ease of exposition.

Let  $\psi_i(\{a_t^{(i)}\}) = x_i - \sum_{t=1}^m a_t^{(i)} + \sum_{t=1}^m g_t(a_t^{(i)}, \xi_t)$  denote the difference between demand and availability of nurses given realizations  $x_i$  and  $\xi_t$ . Then, the staffing problem can be written as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^u E(c_0 \cdot (X_i - Q_i(\mathbf{a}))^+) \\ & = \sum_{i=1}^u E(c_0 \cdot (\psi_i(\{a_t^{(i)}\}))^+) \end{aligned} \quad (3)$$

$$\text{subject to } \sum_{i=1}^u a_t^{(i)} = n_t, \quad \text{for each } t = 1, \dots, m. \quad (4)$$

Each function  $\psi_i(\{a_t^{(i)}\})$  is decreasing concave in each  $a_t^{(i)}$ . Furthermore, because  $\psi_i(\{a_t^{(i)}\})$  consists of functions that are separable in  $a_t^{(i)}$ , it is also jointly concave in  $a_t^{(i)}$ . The function  $(\cdot)^+$  is increasing convex in its



argument. The composition of an increasing function and a decreasing function is a decreasing function. From above, we infer that  $(\psi_i(\{a_t^{(i)}\}))^+$  is decreasing in  $a_t^{(i)}$ . Furthermore, realizations  $x_i$  and  $\xi_t$  are independent of the choice of  $a_t^{(i)}$ . Therefore,  $E(c_o \cdot (\psi_i(\{a_t^{(i)}\}))^+)$  is also decreasing in  $a_t^{(i)}$ . Finally,  $\sum_{i=1}^u E(c_o \cdot (\psi_i(\{a_t^{(i)}\}))^+)$  is decreasing in  $a_t^{(i)}$  because the sum of decreasing functions is a decreasing function.

Because the problem in (3)–(4) is to minimize a decreasing function under linear constraints, the first-order necessary Karush-Kuhn-Tucker (KKT) conditions imply that there exist  $\gamma_k$ 's that are unrestricted in sign and

$$c_o \cdot E\left(\frac{\partial(\psi_i(\{a_t^{(i)}\}))^+}{\partial a_k^{(i)}}\right) + \gamma_k = 0, \quad \text{for each } k \text{ and each } i. \quad (5)$$

The derivative of  $\psi_i(\{a_t^{(i)}\})^+$  with respect to  $a_k^{(i)}$  is zero if  $x_i \leq \sum_{t=1}^m a_t^{(i)} - \sum_{t=1}^m g_t(a_t^{(i)}, \xi_t)$  and  $-1 + g'_k(a_k^{(i)}, \xi_t)$  otherwise. Therefore, Equation (5) can be written in the following equivalent way:

$$\gamma_k = c_o \cdot E_{X_i} \left[ E_{\xi_t} \left( 1 - g'_k(a_k^{(i)}, \xi_t) \mid X_i \leq \sum_{t=1}^m a_t^{(i)} - \sum_{t=1}^m g_t(a_t^{(i)}, \xi_t) \right) \right]. \quad (6)$$

The advantage of writing Equation (5) in this equivalent form is that (6) can be interpreted. The right-hand side of Equation (6) is the expected rate of decrease in unit  $i$ 's shortage cost as a function of the staffing level of type  $k$  nurses when unit  $i$  experiences nurse shortage. Note that the left-hand side does not depend on unit index  $i$ . This means that under an optimal allocation, the expected rate of decrease in each unit's shortage cost as a function of the staffing level of each nurse type when that unit experiences nurse shortage should be the same. When  $X_i$ 's are i.i.d., one way to achieve this equality is to set  $a_t^{(i)} = a_t^{(j)}$  for each pair  $(i, j)$ , and  $\sum_{t=1}^m a_t^{(i)} = n_t$  for every  $t$ . That is, to staff such that each unit has the same number of type  $t$  nurses, for each  $t$ .

A special case of the random aggregate model arises when  $g_{t,1}(a_t^{(i)}) = 0$  and  $g_{t,2}(a_t^{(i)}) = a_t^{(i)}$ . In this case, for a given  $\mathbf{a}$ , we show in Proposition 3 that a nurse manager would prefer a cohort of nurses with a stochastically smaller absentee rate (i.e., nurses with smaller mean and/or variance of aggregate absentee rate).

**PROPOSITION 3.** Let  $\phi_i(\mathbf{a}) = \sum_{t=1}^m a_t^{(i)} \xi_t$ , where  $\xi_t \in [0, 1]$  is the random absentee rate for type  $t$  nurses and  $\xi = (\xi_1, \dots, \xi_m)$ . For each fixed  $\mathbf{a}$ , if  $\xi \leq_{icx} \xi'$  in a component-wise manner, then  $\pi(\mathbf{a}_\xi^*) \leq \pi(\mathbf{a}_{\xi'}^*)$ .

**PROOF OF PROPOSITION 3.** Let  $\pi(\mathbf{a}) = \sum_{i=1}^u \pi_i(\mathbf{a}^{(i)})$ . Because  $\xi_t \leq_{icx} \xi'_t$ ,  $\pi_i(\mathbf{a}^{(i)} \mid \xi) = E[c(X_i - \sum_{t=1}^m a_t^{(i)} + \sum_{t=1}^m a_t^{(i)} \xi_t)^+] \leq E[c(X_i - \sum_{t=1}^m a_t^{(i)} + \sum_{t=1}^m a_t^{(i)} \xi'_t)^+] = \pi_i(\mathbf{a}^{(i)} \mid \xi')$  for each  $i$ . Therefore,  $\pi(\mathbf{a}_\xi^*) = \pi(\mathbf{a}_\xi^* \mid \xi) \leq \pi(\mathbf{a}_{\xi'}^* \mid \xi) \leq \pi(\mathbf{a}_{\xi'}^* \mid \xi') = \pi(\mathbf{a}_{\xi'}^*)$ .  $\square$

So far we have shown that a nurse manager prefers smaller and less variable absentee rates. Next, we address a different question. Given a cohort of heterogeneous nurses, which nurses should a nurse manager choose to realize less variable absentee rates for his or her unit? We use the concept of majorization to answer this question. To avoid situations where the manager of each unit would want only nurses that rarely take unplanned time off, we fix the total aggregate absentee rate that any choice of nurses must satisfy for that unit. We define  $\mathbf{p}^{(i)}$  to be the absentee probabilities of nurses assigned to unit  $i$ . That is, components of  $\mathbf{p}^{(i)}$  contain information about only those nurses that are assigned to unit  $i$ . Furthermore, let  $\pi_i(\mathbf{p}^{(i)})$  denote the expected shortage cost incurred in unit  $i$  with no-show probability vector  $\mathbf{p}^{(i)}$ . Then, with an individual Bernoulli no-show model, we can prove that a nurse manager would prefer a more variable mix of absentee rates.

**PROPOSITION 4.** If  $\mathbf{p}^{(i)} \leq_M \mathbf{p}'^{(i)}$ , then  $\pi_i(\mathbf{p}'^{(i)}) \leq \pi_i(\mathbf{p}^{(i)})$ .

**PROOF OF PROPOSITION 4.** Suppose  $\{B(p_t), p_t \in (0, 1)\}$  is a family of independent random variables parameterized by  $p_t$ . By Corollary 1  $B(p_t)$  is SIL(st). Next, we observe that the function  $h(\mathbf{b}) = x + \sum_{k=1}^{m_i} b_k^{(i)} - m_i$ , where  $x$  is a realization of random demand  $X_i$ ,  $b_k^{(i)}$  is a realization of  $B(p_k^{(i)})$ ,  $p_k^{(i)}$  is the  $k$ th element of vector  $\mathbf{p}^{(i)}$ , and  $m_i$  is the cardinality of the vector  $\mathbf{p}^{(i)}$ , is an increasing valuation in  $\mathbf{b}$ . A function is said to be a valuation if it is both submodular and supermodular (Topkis 1998). Define  $Z(\mathbf{p}^{(i)}) = h(B(p_1^{(i)}), \dots, B(p_{m_i}^{(i)}))$ . Then, by Theorem 2, it follows immediately that  $\{Z(\mathbf{p}^{(i)}), \mathbf{p}^{(i)} \in (0, 1)^{m_i}\}$  is  $SI - SchurCX(icx)$ . In addition, for an increasing convex function  $g$ ,  $\mathbf{p}^{(i)} \leq_M \mathbf{p}'^{(i)}$  implies that  $E[g(h(B(p_1^{(i)}), \dots, B(p_{m_i}^{(i)})))] \leq E[g(h(B(p_1^{(i)}), \dots, B(p_{m_i}^{(i)})))]$ . The statement of the proposition then follows from the fact that  $c(\cdot)$  is an increasing convex function.  $\square$

Proposition 4 shows that between two cohorts of nurses that have same head counts and average group-level absentee rates, the unit's nurse manager would prefer to utilize the more heterogeneous cohort of nurses (i.e., the cohort of nurses with more variable absentee probabilities  $p^{(i)}$ ). However, this may not be the best overall strategy when costs across different units need to be balanced. In fact, it is the difficulty of balancing staffing across units while maximizing heterogeneity within a unit that makes it difficult to identify an optimal assignment strategy.

Next, we show that the results in Propositions 3 and 4 are related. If we view each nurse as a type,



then Corollary 2 establishes that greater heterogeneity within a cohort of nurses leads to smaller variability in attendance patterns of that cohort taken together, which makes it more desirable according to Proposition 3.

**COROLLARY 2.** Let  $\bar{B}_i(\mathbf{p}) = (\sum_{t=1}^n a_t^{(i)} B(p_t)) / (\sum_{t=1}^n a_t^{(i)})$ . If  $\mathbf{p}^{(i)} \leq_M \mathbf{p}'^{(i)}$ , then  $\bar{B}_i(\mathbf{p}') \leq_{icx} \bar{B}_i(\mathbf{p})$ .

**PROOF OF COROLLARY 2.** Recall that  $\mathbf{p}^{(i)}$  contains only the absentee rates of the  $m_i = \sum_{t=1}^n a_t^{(i)}$  nurses who are assigned to unit  $i$ . Also, the nurse index in vector  $\mathbf{p}^{(i)}$ , which we denote by  $k$ , reindexes nurses who are assigned to unit  $i$ ; therefore, it is different from index  $t$ .

Following the logic in the proof of Proposition 4, and let  $h(\mathbf{b}) = \sum_{k=1}^{m_i} b_k^{(i)}$ . It is straightforward to argue that  $h(\mathbf{b})$  is an increasing valuation in  $\mathbf{b}$ . Let  $Z(\mathbf{p}^{(i)}) = \sum_{t=1}^n a_t^{(i)} B(p_t^{(i)}) = h(B(p_1^{(i)}), \dots, B(p_{m_i}^{(i)}))$ . Then, it can be argued that  $\{Z(\mathbf{p}^{(i)}), \mathbf{p}^{(i)} \in (0, 1)^{m_i}\}$  is  $SI - SchurCX(icx)$ . By Theorem 2, for any increasing convex function  $g$ ,  $\mathbf{p}^{(i)} \leq_M \mathbf{p}'^{(i)}$  implies that  $E[g(h(B(p_1^{(i)}), \dots, B(p_{m_i}^{(i)})))] \leq E[g(h(B(p_1^{(i')}), \dots, B(p_{m_i}^{(i')})))]$ . In other words, if  $\mathbf{p}^{(i)} \leq_M \mathbf{p}'^{(i)}$ , then  $\sum_{k=1}^{m_i} B(p_k^{(i)}) \leq_{icx} \sum_{k=1}^{m_i} B(p_k^{(i')})$ , which is equivalent to  $\sum_{t=1}^n a_t^{(i)} B(p_t^{(i)}) \leq_{icx} \sum_{t=1}^n a_t^{(i)} B(p_t^{(i')})$ . Let  $\bar{B}_i(\mathbf{p}) = (\sum_{t=1}^n a_t^{(i)} B(p_t)) / (\sum_{t=1}^n a_t^{(i)}) = (\sum_{k=1}^{m_i} B(p_k^{(i)})) / m_i$  be the random proportion of nurses who are absent from work. Then, from arguments presented above,  $\mathbf{p}^{(i)} \leq_M \mathbf{p}'^{(i)}$  also implies  $\bar{B}_i(\mathbf{p}') \leq_{icx} \bar{B}_i(\mathbf{p})$ .  $\square$

The structural results in Propositions 2–4 and Corollary 2 show that for inpatient units with identical demand patterns, hospitals' costs are lower when nurse assignments are as heterogeneous as possible within a unit, but uniform across units. That is, it is desirable to split high absentee rate nurses across units, and it is desirable to ensure a balanced realized staffing level when demand patterns are identical. This result is consistent with intuition: it will be difficult to manage a unit if all nurses have high absentee rates, and a solution is to mix high and low absentee rate nurses while balancing demand and supply. However, it is not obvious whether simple rules that ensure expected supply equal expected demand would perform well. Also, solving the model formulated in (1) is computationally demanding when the number of nurses is large. Therefore, in §4, we present three heuristics that attempt to achieve assignments consistent with the principles identified in this section.

## 4. Heuristics and Performance Comparisons

Because the formulation in (1) with nurse-specific Bernoulli representation of absenteeism is a combinatorially hard problem, we focus in this section on developing implementable heuristics. To explain the difficulty, consider the fact that for a 12-nurse and two-unit problem, there are 4,096 possible assignments. For

each assignment, there are 4,096 possible realizations of nurses' absentee patterns. To calculate the expected shortage cost, one would need to consider  $4,096 \times 4,096$  assignments and nurse attendance combinations for each demand realization. This makes the search for the optimal nurse assignment difficult for larger-sized problems.

We propose three heuristics in this section, which can be used to obtain assignment of nurses to inpatient units. We also test these heuristics in numerical experiments. In this section, we allow multiple nurses to have the same no-show rate. In particular, this means that  $n_t$  could be greater than 1, and therefore  $a_t^{(i)} \leq n_t$  could be greater than 1 as well. Note that this includes the model in which  $a_t^{(i)} \leq 1$  as a special case. The number of absent nurses in unit  $i$  given assignment  $\mathbf{a}$  can be expressed by  $\phi_i^{[b]}(\mathbf{a}) = \sum_{t=1}^m \sum_{j=1}^{a_{ti}^{(i)}} B_j(p_t)$ , where  $B_j(p_t)$  are i.i.d. Bernoulli random variables with parameter  $p_t$ . It should be clear that allowing  $n_t \geq 1$  does not result in a loss of generality because we will obtain the Bernoulli absence model by setting  $n_t = 1$ .

First, we show that the objective function in (1) is supermodular, which helps to motivate the heuristics in the sequel. We define  $\delta_t^{(i)}(\mathbf{a}) = \pi(\mathbf{a}) - \pi(\mathbf{a} + \mathbf{e}_{ti})$ , where  $\mathbf{e}_{ti}$  is a  $m \times u$  matrix with the  $(t, i)$ th component equal to 1 and the remaining components equal to 0, as the incremental benefit of adding a type  $t$  nurse to unit  $i$ . Then, it can be shown that

- $\delta_t^{(i)}(\mathbf{a}) \geq 0 \forall t = 1, \dots, m$ , and  $\forall i = 1, \dots, u$ . In addition, when no-show probabilities are ordered such that  $p_1 \geq p_2 \geq \dots \geq p_m$ ,  $\delta_t^{(i)}(\mathbf{a}) \leq \delta_{t'}^{(i)}(\mathbf{a})$  for  $t \leq t'$ .

- $\delta_t^{(i)}(\mathbf{a}) - \delta_t^{(i)}(\mathbf{a} + \mathbf{e}_{t'i}) \geq 0 \forall t' = 1, \dots, m$  and  $\forall i = 1, \dots, u$ . Note that  $\delta_t^{(i)}(\mathbf{a}) - \delta_t^{(i)}(\mathbf{a} + \mathbf{e}_{t'i})$  is the difference in incremental benefits of adding a type  $t$  nurse to unit  $i$  under two situations: one in which this unit had  $a_{t'i}^{(i)}$  type  $t'$  nurses and another in which the number of type  $t'$  nurses was  $a_{t'i}^{(i)} + 1$ .

These results have a straightforward intuitive explanation. The first bullet confirms that it is better to add a nurse with a lower absentee rate. The second bullet says that the benefit (reduction in cost) of adding one more type  $t$  nurse to a particular unit diminishes in the number of type  $t'$  nurses in the unit when the staffing levels of the remaining groups are held constant (note that  $t'$  is an arbitrary type, which includes  $t$ ). Together, these observations imply that the objective function in (1) is supermodular. A more formal argument can be constructed as follows— $\delta_t^{(i)}(\mathbf{a}) \geq \delta_t^{(i)}(\mathbf{a} + \mathbf{e}_{t'i})$  indicates that  $\pi(\mathbf{a}) + \pi(\mathbf{a} + \mathbf{e}_{t'i} + \mathbf{e}_{ti}) \geq \pi(\mathbf{a} + \mathbf{e}_{t'i}) + \pi(\mathbf{a} + \mathbf{e}_{ti})$ . Let  $\mathbf{x} = \mathbf{a} + \mathbf{e}_{ti}$  and  $\mathbf{y} = \mathbf{a} + \mathbf{e}_{t'i}$ , then we have  $\pi(\mathbf{x} \vee \mathbf{y}) + \pi(\mathbf{x} \wedge \mathbf{y}) \geq \pi(\mathbf{x}) + \pi(\mathbf{y})$  and  $\pi$  is supermodular.

The reason why supermodularity is relevant is that greedy heuristics have been shown to work well when the objective function is supermodular; see Il'ev (2001), Calinescu et al. (2011), Asadpour et al. (2008) for

examples of using greedy algorithms for minimizing (respectively, maximizing) supermodular (respectively, submodular) functions. Our objective is to minimize a supermodular cost function; therefore, the nurse manager may wish to sort nurses by increasing absentee rates and assign them to different units sequentially to maximize the marginal benefit from each assignment (i.e., in a greedy fashion) until all nurses are assigned. We call this strategy the greedy assignment. It can be argued that when nurses have identical no-show probabilities, the greedy strategy results in an optimal assignment. We omit the details of that argument in the interest of brevity. We present the greedy assignment approach more formally and propose two other assignment strategies in the sequel.

*H<sub>1</sub>: Greedy Assignment.* Assign one nurse at a time to a unit that generates the highest expected marginal benefit. If there are multiple units with the same expected marginal benefit, then randomly pick one unit. It can be shown that  $H_1$  will assign a nurse to the unit that has the highest expected shortage prior to the new addition. Therefore, when nurses are assigned to identical units in a sequence sorted by absentee rates,  $H_1$  is likely to result in heterogeneous absentee rates for nurses assigned to the same unit while maintaining a similar expected shortage level across units in the spirit of Propositions 2 and 4.

*H<sub>2</sub>: Balanced Assignment.* This strategy sorts nurses by their absentee rates, then sequentially assigns the nurses to the unit that has the smallest expected supply to expected demand ratio. This ratio is calculated by (expected number of assigned nurses who show up to work)/(expected nurse requirement for the unit). If more than one unit has the smallest ratio, then we randomly assign to one of those units. Once a nurse is assigned to a unit, we update that unit's ratio and repeat the assignment process until all nurses are assigned. This heuristics is also motivated by Propositions 2 and 4—it maintains a heterogeneous cohort of nurses within a unit while ensuring that each unit's expected supply expected demand ratio is similar.

*H<sub>3</sub>: Homogeneous Assignment.* This strategy sorts nurses from the lowest absentee rate to the highest absentee rate, then fills the units one by one using the same nurse-to-expected demand ratio. The homogeneous assignment ( $H_3$ ) is used to mimic the scenarios where either (1) nurses choose a home unit and those with similar absentee rates choose the same unit, or (2) the nurse manager deliberately assigns nurses with similar absentee rates to the same unit. This is a straw policy that will be used to demonstrate the benefit of using either  $H_1$  or  $H_2$ .

All ensuing comparisons are performed using the assumption that each of the multiple nursing units has identically distributed nursing requirements and

that nurses' absentee rates are independent and heterogeneous. For large-sized problems, the computational effort required to search for an optimal assignment becomes prohibitive. For this reason, we evaluate the performances of heuristics by comparing their expected shortage cost relative to a lower bound.

*Lower Bound.* Consider a particular nursing unit  $i$ . For a given nurse allocation  $\mathbf{a}$ , this unit's shortage cost is  $E[c_o \cdot (X_i - Q_i(\mathbf{a}))^+]$ . We know from Jensen's inequality that  $E[c_o \cdot (X_i - E(Q_i(\mathbf{a})))^+] \leq E[c_o \cdot (X_i - Q_i(\mathbf{a}))^+]$  because  $(\cdot)^+$  is a convex function. This lower bound for a particular unit's shortage cost is dependent on the particular allocation of nurses to this unit through  $E(Q_i(\mathbf{a}))$ . Next, we derive an allocation-free lower bound.

Let  $q_i = E(Q_i(\mathbf{a}))$  denote the mean number of nurses who show up in unit  $i$  given its allocation. Furthermore  $\sum_i q_i = q$ , where  $q$  is the expected number of nurses who will show up for all units. Then an allocation-free lower bound for the problem in expressions (3)–(4) can be written as follows:

$$\min_{q_i} \sum_{i=1}^u E[c_o \cdot (X_i - q_i)^+] \quad (7)$$

subject to  $\sum_{i=1}^u q_i = q$  and  $q_i \geq 0$ . This is a convex minimization problem with linear constraint. It can be solved easily using KKT first-order necessary conditions (which are also sufficient). The first order necessary condition states that there exists a  $\gamma$  (unrestricted in sign) such that  $c_o \cdot \bar{F}_i(q_i) = \gamma$  for each  $i$ . Therefore, when demand in each unit is i.i.d., it follows that  $q_i^* = q/u$ . When demand in each unit does not have the same distribution, then the optimal allocation is achieved at  $\mathbf{q}^* = (q_1^*, \dots, q_u^*)$  such that  $q_i^* = \bar{F}_i^{-1}(\gamma/c_o) \forall i = 1, \dots, u$  and  $\sum_{i=1}^u q_i^* = q$ . In summary, for i.i.d. demand across units, an allocation-free lower bound is obtained by assigning to each unit an equal proportion of the mean number of nurses who show up in the cohort of all  $n$  nurses and assuming no uncertainty in nurses' attendance. Similarly, when inpatient units do not have i.i.d. demand, a lower bound is obtained by staffing at the same percentile of the demand distributions for each unit such that the aggregate staffing level sums to the expected number of nurses who will show up.

Next, we present a numerical study to evaluate the three heuristics' performance relative to this lower bound using hospital 1's staffing data. The examples will show that  $H_1$  and  $H_2$  consistently outperform  $H_3$ , regardless of the size of the problem. Consider a pool of  $8u$  nurses who will be assigned to  $u$  identical units. Hospital 1's target staffing level for each telemetry unit was eight nurses per shift. This motivated us to pick eight nurses per unit in a typical shift. To ensure a balanced design such that the expected demand equals the expected staffing level, we assume each

**Table 4** Average Relative Performance

Heuristics	Number of units ( $u$ )									
	3	4	5	6	7	8	9	10	11	12
$H_1$	1.0161	1.0166	1.0178	1.0182	1.0187	1.0191	1.0188	1.0196	1.0201	1.0193
$H_2$	1.0161	1.0166	1.0178	1.0182	1.0187	1.0191	1.0188	1.0196	1.0201	1.0193
$H_3$	1.0429	1.046	1.0519	1.0531	1.0541	1.056	1.0575	1.058	1.0602	1.0591

unit's demand follows a Poisson distribution with a mean of  $\lambda = 8(1 - \bar{p})$ , where  $\bar{p}$  is the average absentee rate for all nurses. We draw 500 random samples of  $8u$  nurses from the pool of nurses who worked in hospital 1's three telemetry units and calculate each samples' expected shortage under  $H_1$ ,  $H_2$ ,  $H_3$ , and the lower bound for  $u = 3, \dots, 12$ . The relative performance for each heuristic reported in Table 4 is the ratio of its expected shortage cost and the lower bound cost.

The first two heuristics,  $H_1$  and  $H_2$ , are consistently about 3%–4% better than  $H_3$  on average. In this numerical example, over all three shifts, a 4% savings for the four-unit instance will be about \$73,052 per year when hourly shortage cost is calculated as 1.5 times 2012's nurse mean hourly wage of \$32.67 (Bureau of Labor Statistics 2012). When extrapolated to the entire hospital of a size similar to hospital 1, savings could be as high as \$1.8–2.4 million per year. To further illustrate the benefit of ensuring a similar unit-level absentee rate and heterogeneous cohorts of nurses (properties captured by  $H_1$  and  $H_2$ ), we show the percentage of cases in which  $H_1$  or  $H_2$  are better than  $H_3$  by  $x\%$  of the lower bound in Table 5.

The results in Table 5 show that the benefit of  $H_1$  and  $H_2$  as compared to  $H_3$  becomes more significant as the total number of nurses increases because with more nurses, unit-level absentee rate can be more variable.

**Table 5** Percentage of Cases in Which  $H_1$  or  $H_2$  Is Better Than  $H_3$  by  $x\%$  of the Lower Bound

Example 1: i.i.d demand across units (%)										
$H_1$ vs. $H_3$	Number of units ( $u$ )									
<b>x</b>	3	4	5	6	7	8	9	10	11	12
1	74	99.2	100	99	100	100	100	100	100	100
2	36.8	78	93.2	83	88.2	94.4	100	100	99.8	100
3	15	50.8	62.8	53.6	61.8	73.6	97.2	97.8	98.4	96
4	7.2	29.6	38.2	36.2	39.8	50.8	81.2	79.8	81.2	76.2
5	4	17.6	23.8	18.6	24.8	24.8	57.6	54.8	57.6	50.6
$H_2$ vs. $H_3$	Number of units ( $u$ )									
<b>x</b>	3	4	5	6	7	8	9	10	11	12
1	90	95.6	99.4	99.8	99.6	100	99.8	100	100	100
2	52.8	65.2	78.2	81.2	88.8	89.2	95.4	95.6	97	98.8
3	30.2	39	48.8	54	58.6	62.4	72.4	71.2	79.4	79.6
4	16.6	21.8	33.8	31.2	31.6	37.2	43.2	41.8	48.6	49.8
5	11.4	12.6	18.8	17	16.8	20.2	20.8	18.2	23	20.2

These results also confirm Propositions 2 and 4 that it is desirable to utilize heterogeneous cohorts of nurses for individual units while maintaining a similar aggregate absentee rate among different units.

We also tested the performance of the three heuristics when unit demands were heterogeneous, but other parameters remained unchanged. In particular, we assumed that roughly one third of the units had mean demand  $0.8\lambda$ , one third of the units had mean demand  $\lambda$ , and the remaining one third had mean demand equal to  $1.2\lambda$ . When the number of units could not be evenly divided into three parts, we chose equal first and third cohorts such that the total expected supply and demand remained balanced. We reached a similar conclusion— $H_1$  and  $H_2$  outperformed  $H_3$  by 2%–5%. We report the detailed performance comparison for this numerical example in Online Supplement C.

## 5. Concluding Remarks

This paper shows that nurse managers may use a nurse's attendance history to predict his or her likelihood of being absent in a future shift. This information can be utilized within easy-to-implement staffing heuristics, e.g., heuristics labeled  $H_1$  and  $H_2$ , to reduce staffing costs. The use of such heuristics does not require much effort on the part of nurse managers.

The contribution of this paper lies in (1) developing detailed analyses of data from multiple inpatient units to identify observable predictors of nurse absenteeism, (2) establishing structural properties of optimal assignment strategies, and (3) developing and testing easy-to-implement heuristics for use by nurse managers. The structural properties we establish provide insights for developing staffing strategies in environments where work is performed by teams of employees.

The savings are of the order of 3%–4% of overtime costs, which may be considered small by some. However, it is important to note that our approach reduces variance in attendance rate and that the approach is easy to implement. Therefore, nurse managers do not need to exert much effort to realize the benefit of reduced cost and reduced day-to-day variability in the number of additional nurses they will need to find to meet requirements in each shift. In addition, the proposed strategies utilize information that is easy to track from historical data—nurses' absentee rates and inpatient units' demand distributions. The former may



be estimated from nurses' attendance records whereas the latter may be estimated from census data and target nurse-to-patient ratios.

A limitation of our mathematical model comes from an implicit assumption that nurses' absentee rates will remain unchanged with reassignments. Many hospital environments are highly dynamic, and nurses' absenteeism may be affected by unit-specific norms, management style, and esprit de corps. Whereas behavioral factors and people management skills are important in hospital settings, such factors are beyond the scope of this paper. Many papers in the HSR literature discuss possible strategies to mitigate such operational challenges. See Seago and Faucett (1997), Fletcher (2001), Andrews and Dziegielewska (2005), and San Park and Kim (2009) for further information.

### Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/msom.2014.0486>.

### References

- Aiken LH, Clarke SP, Sloane DM, Sochalski J, Siber JH (2002) Hospital nurse staffing and patient mortality, nurse burnout and job dissatisfaction. *J. Amer. Medical Assoc.* 288(16):1987–1993.
- Andrews DR, Dziegielewska SF (2005) The nurse manager: Job satisfaction, the nursing shortage and retention. *J. Nursing Management* 13(4):286–295.
- Araman VF, Popescu I (2010) Media revenue management with audience uncertainty: Balancing upfront and spot market sales. *Manufacturing Service Oper. Management* 12(2):190–212.
- Asadpour A, Nazerzadeh H, Saberi A (2008) Stochastic submodular maximization. Papadimitriou C, Zhang S, eds. *Internet and Network Economics* (Springer, Berlin), 477–489.
- Birge JR, Louveau F (2011) *Introduction to Stochastic Programming*, 2nd ed. (Springer, New York).
- Blumenfeld DE, Inman RR (2009) Impact of absenteeism on assembly line quality and throughput. *Production Oper. Management* 18(3):333–343.
- Bollapragada S, Morton TE (1999) Myopic heuristics for the random yield problem. *Oper. Res.* 47(5):713–722.
- Bureau of Labor Statistics (2011) Household data annual averages. Current population survey, Bureau of Labor Statistics, Washington, DC. Accessed April 28, 2014, <http://www.bls.gov/cps/cpsaat47.pdf>.
- Bureau of Labor Statistics (2012) Occupational employment and wages for registered nurses. Occupational employment statistics, Bureau of Labor Statistics, Washington, DC. Accessed April 29, 2014, <http://www.mynextmove.org/profile/ext/oesmaps/29-1141.00>.
- California Nurse Association, National Nurses Organizing Committee (2012) Floating according to the rules. Nursing Practice and Patient Advocacy Alert. Accessed April 19, 2012, <http://www.nationalnursesunited.org/page/-/files/pdf/nursing-practice/advo-alerts/float-to-rules.pdf>.
- Calinescu G, Chekuri C, Pál M, Vondrák J (2011) Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.* 40(6):1740–1766.
- Cho S-H, Ketefian S, Barkauskas VH, Smith DG (2003) The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs. *Nursing Res.* 52(2):71–79.
- Ciarallo FW, Akella R, Morton TE (1994) A periodic review, production planning model with uncertain capacity and uncertain demand—Optimality of extended myopic policies. *Management Sci.* 40(3):320–332.
- Davey MM, Cummings G, Newburn-Cook CV, Lo EA (2009) Predictors of nurse absenteeism in hospitals: A systematic review. *J. Nursing Management* 17(3):312–330.
- Ferlise P, Baggot D (2009) Improving staff nurse satisfaction and nurse turnover: Use of a closed-unit staffing model. *J. Nursing Admin.* 39(7/8):318–320.
- Fletcher CE (2001) Hospital RNs' job satisfactions and dissatisfactions. *J. Nursing Admin.* 31(6):324–331.
- Fuson K (2006) "Patient" says it all. *USA Today* (May 31), [http://usatoday30.usatoday.com/news/health/2006-05-31-er-waits\\_x.htm](http://usatoday30.usatoday.com/news/health/2006-05-31-er-waits_x.htm).
- Gerchak Y, Vickson RG, Parlar M (1988) Periodic review production models with variable yield and uncertain demand? *IIE Trans.* 20(2):144–150.
- Green LV, Savin S, Savva N (2013) "Nurse vendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Sci.* 59(10):2237–2256.
- Grosfeld-Nir A, Gerchak Y (2004) Multiple lotsizing in production to order with random yields: Review of recent advances. *Ann. Oper. Res.* 126(1–4):43–69.
- Grosfeld-Nir A, Gerchak Y, He Q-M (2000) Manufacturing to order with random yield and costly inspection. *Oper. Res.* 48(5):761–767.
- Gurnani H, Akella R, Lehoczy J (2000) Supply management in assembly systems with random yield and random demand. *IIE Trans.* 32(8):701–714.
- Hur D, Mabert VA, Bretthauer KM (2004) Real-time work schedule adjustment decisions: An investigation and evaluation. *Production Oper. Management* 13(4):322–339.
- Il'ev VP (2001) An approximation guarantee of the greedy descent algorithm for minimizing a supermodular set function. *Discrete Appl. Math.* 114(1):131–146.
- Kane RL, Shamliyan TA, Mueller C, Duval S, Wilt TJ (2007) The association of registered nurse staffing levels and patient outcomes. Systematic review and meta-analysis. *Medical Care* 45(12):1195–1204.
- Lang TA, Hodge M, Olson V, Romano PS, Kravitz RL (2004) Nurse-patient ratios: A systematic review on the effects of nurse staffing on patient, nurse employee, and hospital outcomes. *J. Nursing Admin.* 34(7–8):326–337.
- Laschinger HKS, Grau AL (2012) The influence of personal dispositional factors and organizational resources on workplace violence, burnout, and health outcomes in new graduate nurses: A cross-sectional study. *Internat. J. Nursing Stud.* 49(3):282–291.
- Lim GJ, Mobasher A, Kardar L, Côté MJ (2011) Nurse scheduling. Hall RW, ed. *Handbook of Healthcare System Scheduling: Delivering Care When and Where It Is Needed*, Chap. 3 (Springer, New York), 31–64.
- Liyanage L, Shanthikumar JG (1992) Allocation through stochastic schur convexity and stochastic transposition increasingness. Lecture Notes-Monograph Series, Vol. 22, 253–273.
- Marshall AW, Olkin I, Arnold BC (2011) *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. (Springer, New York).
- National Center for Health Statistics (2011) Health, United States, 2010: With special feature on death and dying. U.S. Department of Health and Human Services Publication No. 2011-1232. Accessed April 28, 2014, <http://www.cdc.gov/nchs/data/health/2010.pdf>.
- Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K (2002) Nurse-staffing levels and the quality of care in hospitals. *New England J. Medicine* 346(22):1715–1722.
- Press Ganey Associates (2010) 2010 emergency department pulse report—Patient perspectives on American health care. Report, Press Ganey Associates, South Bend, IN. Accessed April 28, 2014, [http://www.pressganey.com/Documents\\_secure/Pulse%20Reports/2010\\_ED\\_Pulse\\_Report.pdf](http://www.pressganey.com/Documents_secure/Pulse%20Reports/2010_ED_Pulse_Report.pdf).
- Rajbhandary S, Basu K (2010) Working conditions of nurses and absenteeism: Is there a relationship? An empirical analysis using national survey of the work and health of nurses. *Health Policy* 97(2–3):152–159.
- Roelen CAM, Koopmans PC, Schreuder JAH, Anema JR, van der Beek AJ (2011) The history of registered sickness absence predicts future sickness absence. *Occupational Medicine* 61(2):96–101.



- San Park J, Kim TH (2009) Do types of organizational culture matter in nurse job satisfaction and turnover intention? *Leadership in Health Services* 22(1):20–38.
- Schreuder J, Roelen C, van Zweeden N, Jongsma D, van der Klink J, Groothoff J (2011) Leadership styles of nurse managers and registered sickness absence among their nursing staff. *Health Care Management Rev.* 36(1):58–66.
- Seago JA, Faucett J (1997) Job strain among registered nurses and other hospital workers. *J. Nursing Admin.* 27(5):19–25.
- Shaked M, Shanthikumar JG (2007) *Stochastic Orders* (Springer, New York).
- Topkis D (1998) *Supermodularity and Complementarity*, 2nd ed. (Princeton University Press, Princeton, NJ).
- Unruh L (2008) Nurse staffing and patient, nurse, and financial outcomes. *Amer. J. Nursing* 108(1):62–71.
- Whitt W (2006) Staffing a call center with uncertain arrival rate and absenteeism. *Production Oper. Management* 15(1):88–102.
- Yano CA, Lee HL (1995) Lot sizing with random yields: A review. *Oper. Res.* 43(2):311–334.