



Management Science

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

“Nursevendor Problem”: Personnel Staffing in the Presence of Endogenous Absenteeism

Linda V. Green, Sergei Savin, Nicos Savva

To cite this article:

Linda V. Green, Sergei Savin, Nicos Savva (2013) “Nursevendor Problem”: Personnel Staffing in the Presence of Endogenous Absenteeism. *Management Science* 59(10):2237-2256. <http://dx.doi.org/10.1287/mnsc.2013.1713>

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.

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 © 2013, 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>

“Nursevendor Problem”: Personnel Staffing in the Presence of Endogenous Absenteeism

Linda V. Green

Graduate School of Business, Columbia University, New York, New York 10027, lvg1@columbia.edu

Sergei Savin

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, savin@wharton.upenn.edu

Nicos Savva

London Business School, London NW1 4SA, United Kingdom, nsavva@london.edu

The problem of determining nurse staffing levels in a hospital environment is a complex task because of variable patient census levels and uncertain service capacity caused by nurse absenteeism. In this paper, we combine an empirical investigation of the factors affecting nurse absenteeism rates with an analytical treatment of nurse staffing decisions using a novel variant of the newsvendor model. Using data from the emergency department of a large urban hospital, we find that absenteeism rates are consistent with nurses exhibiting an aversion to higher levels of anticipated workload. Using our empirical findings, we analyze a single-period nurse staffing problem considering both the case of constant absenteeism rate (exogenous absenteeism) as well as an absenteeism rate that is a function of the number of nurses scheduled (endogenous absenteeism). We provide characterizations of the optimal staffing levels in both situations and show that the failure to incorporate absenteeism as an endogenous effect results in understaffing.

Key words: healthcare; hospitals; nursing; newsvendor; econometrics; organizational studies; manpower planning

History: Received October 24, 2011; accepted December 12, 2012, by Yossi Aviv, operations management.

Published online in *Articles in Advance* June 14, 2013.

1. Introduction and Literature Review

In recent years hospitals have been faced with increasing pressure from their major payers—federal and state governments, insurers, and large employers—to cut costs. Because nursing personnel account for a very large portion of expenses, the response in many instances has been reductions of the nursing staff. Nurse workloads have been further increased by shorter hospital lengths of stay and increasing use of outpatient procedures, resulting in sicker hospitalized patients who require more nursing care. The adverse impact of these changes has been documented by a number of studies and include increases in medical errors, delays for patients waiting for beds in emergency rooms, and ambulance diversions (Needleman et al. 2002, Aiken et al. 2002, Cho et al. 2003). In response, a number of state legislatures, for example, Victoria in Australia and California in the United States, have mandated minimum nurse staffing levels (Gordon et al. 2008).

Establishing the appropriate nursing level for a particular hospital unit during a specific shift is complicated by the need to make staffing decisions well in advance (e.g., six to eight weeks) of that shift, as

well as labor constraints dealing with the number of consecutive and weekend shifts worked per nurse, vacation schedules, personal days, and preferences (see, e.g., Miller et al. 1976, Wright et al. 2006). Adding to these complexities is the prevalence of nurse absenteeism. According to the U.S. Bureau of Labor Statistics (2008), in 2008 U.S. nurses exhibited the highest number of incidents of illness or injury involving days away from work, 7.8 per 100 FTEs, substantially higher than the national average of 2.1 incidents per 100 FTEs. The goal of this paper is to construct a model for nurse staffing that includes the impact of absenteeism and investigate whether and how clinical units should take it into consideration when they make staffing level decisions. To do this, we must first understand whether and how absenteeism is affected by staffing levels themselves. Because, to the best of our knowledge, there is no literature that adequately addresses this linkage, we first conduct an empirical hospital-based study to understand how to incorporate absenteeism into an analytical model.

Most studies on nurse absenteeism find that it is positively related to levels of work-related stress (Shamian et al. 2003) and nurse workload (Bryant

et al. 2000, Tummers et al. 2001, McVicar 2003, Unruh et al. 2007). However, at least one study examining absenteeism among trainee nurses finds a negative correlation between nurse absenteeism and workload (Parkes 1982). Nearly all of the existing studies of nurse absenteeism use qualitative or self-reported measures of workload and cross-sectional analyses of this phenomenon (see Rauhala et al. 2007 for an exception). In particular, they compare the long-run absenteeism behavior of nurses across different clinical units, rather than tracking the same set of nurses over time. As such, these studies are of limited value to managers responsible for day-to-day staffing decisions.

The first part of our paper presents a longitudinal investigation of the link between nurse absenteeism and workload using data from the emergency department (ED) of a large New York City hospital. Rather than relying on subjective self-reported workload measures, we use patient census values to calculate nurse-to-patient ratios that are treated as proxies for the workload experienced by nurses working a particular shift. Nurse shortages and other organizational limitations (such as union rules) provide the necessary exogenous variation in staffing decisions that we exploit to identify the impact of workload on absenteeism. We hypothesize that fluctuations in nurse workload due to irregularities in scheduling and/or unpredictable demand have a direct impact on absenteeism. An anticipated increase in workload could result in an increase in nurses' motivation to help their fellow nurses and, therefore, in decreased absenteeism, or, given generally high workloads and adverse working conditions, in increased absenteeism. In either case, the resulting behavior may result in optimal nurse staffing levels that are different from those predicated by models that do not consider this workload-induced behavior. Therefore, our empirical investigation aims to provide support for one of these specific hypotheses so that we can incorporate the appropriate assumption into our analytical modeling. Our main finding is that absenteeism increases when there is a higher anticipated workload. Specifically, we find that for our data set with an average absenteeism rate of 7.3%, an extra scheduled nurse is associated with an average reduction in the absenteeism rate of 0.6%. As such, our paper is related to the growing body of literature on the effects of workload on system productivity (KC and Terwiesch 2009, Powell and Schultz 2004, Powell et al. 2012, Schultz et al. 1999).

The second goal of our paper is to develop a model of optimal staffing in service environments with workload-dependent absenteeism. Extant literature either ignores absenteeism or treats it as an exogenous phenomenon (Bassamboo et al. 2010, Easton and

Goodale 2005, Harrison and Zeevi 2005, and Whitt 2006 provide examples of call-center staffing, whereas Fry et al. 2006 and He et al. 2012 provide examples of firefighter and hospital operating room staffing, respectively). The uncertain supply of service capacity created by nurse absenteeism connects our work with a stream of literature focused on inventory planning in the presence of unreliable supply/stochastic production yield (Yano and Lee 1995) with two important distinctions. First, the overwhelming majority of papers on stochastic supply yields model them as being either additive or multiplicative (Henig and Gerchak 1990, Ciarallo et al. 1994, Bollapragada and Morton 1999, Gupta and Cooper 2005, Yang et al. 2007), a justifiable approach in manufacturing settings. The supply uncertainty in our model has a binomial structure, a more appropriate choice in personnel staffing settings. Binomial yield models are a relative rarity in the stochastic yield literature, perhaps due to their limited analytical tractability (Grosfeld-Nir and Gerchak 2004, Fadiloglu et al. 2008). Most importantly, to the best of our knowledge, our analysis is the first to introduce and analyze endogenous stochastic yields.

Using our model we characterize the optimal staffing levels under exogenous and endogenous absenteeism. We show that the failure to incorporate absenteeism as an endogenous effect results in understaffing, which leads to a higher-than-optimal absenteeism rate and staffing costs. Specifically, for model parameters that closely match the hospital we study, we find that ignoring the endogenous nature of absenteeism can lead to a staffing cost increase of 2% to 3%. In addition to the cost impact, understaffing associated with ignoring absenteeism may result in an increase in medical errors, particularly in the pressured and sensitive environment of an ED. Considering that nursing costs is one of the biggest components of overall hospital operating costs, more accurate nurse staffing based on endogenous absenteeism constitutes a substantial opportunity for hospitals to simultaneously reduce costs and improve quality of care.

Finally, we show that despite understaffing, the exogenous-absenteeism model will appear to be self-consistent in the sense that the assumed exogenous absenteeism rate will be equal to the observed (endogenous) absenteeism rate. This is particularly worrisome for staffing managers because it implies that it is impossible to tell whether the model is well specified just by examining the observed absenteeism rate. In this regard, our paper contributes to the literature on model specification errors in operations (e.g., Cachon and Kök 2007, Cooper et al. 2006, Lee et al. 2012, Mersereau 2012). In this literature, as in our paper, the model specification error cannot be

detected by studying data as the misspecified model will produce consistent outcomes. To the best of our knowledge, our paper is the first to study model specification error in the context of staffing, and the first in the model specification literature to start from an empirical observation.

2. Endogeneity in Nurse Absenteeism Rates: An Empirical Study

Our study is based on nurse absenteeism and patient census data from the ED of a large New York City hospital. Nurses employed in this unit are full-time employees, each working on average 3.25 shifts per week. The unit uses two primary nursing shifts: the day shift starts at 8:00 A.M. and ends at 8:00 P.M.; the night shift starts at 8:00 P.M. and ends at 8:00 A.M. Another (evening) shift is also operated from 12:00 P.M. to 12:00 A.M. For each shift, for a period of 10 months starting on July 1, 2008 (304 day shifts, 304 evening shifts, and 303 night shifts), we collected the following data: the number of nurses scheduled, the number of nurses absent, the number of patient visits, and the patient census data recorded every two hours. The resulting descriptive statistics for three shifts are presented in Table 1.

In our analysis of absenteeism, we limit our attention to the nurses on the day and night shifts because the evening shift is fundamentally different from the other two. First, the nurses working on this shift do not work on the other two shifts. Thus, it is less likely that they are informed about the nurse staffing schedules for the day and night shifts. Second, the evening shift consists of fewer nurses who are more experienced and exhibit less absenteeism than the other two shifts, as shown in Table 1. However, we do take into

account the evening shift when measuring workload because the evening shift overlaps with both the day shift and the night shift.

The nurse scheduling process starts several weeks before the actual work shift when the initial schedule is established. This initial schedule often undergoes a number of changes and corrections due to, for example, family illnesses, medical appointments, and jury duty obligations, which may continue until the day before the actual shift. In our study, we have used the *final* schedules, that is, the last schedules in effect before any "last minute" absenteeism is reported for the shift. We record as absenteeism any event where a nurse does not show up for work without giving sufficiently advance notice for the schedule to be revised. In the clinical unit we study, nurses are allowed to use up to 10 personal days per year, which do not require any significant advance notice. To compensate for absenteeism, the management of the ED uses either agency nurses or nurses from the previous shift to work overtime. Therefore, the number of nurses present to work during any given shift is generally equal to the number of nurses in the final schedule.¹

The average patient census during a shift varies substantially from day to day. Some of this variation (52.2% for day shifts and 32.6% for night shifts) can be explained by day-of-week and weekly fixed effects. Furthermore, the patient census exhibits significant serial autocorrelation ($\rho = 30.6\%$) with the values recorded during the previous shift. The number of nurses scheduled for a particular type of shift, for example, day shift on a Wednesday, is highly variable. Approximately 25% of the variation in the number of nurses scheduled can be explained by day-of-week and weekly fixed effects (adjusted $R^2 = 27.5\%$ for the day shift and adjusted $R^2 = 24.1\%$ for the night shift). Also, after controlling for fixed effects, the number of nurses scheduled for a shift shows little dependence on either the average patient census during that shift or on the average census values for the 14 previous shifts, which correspond to one calendar week. This indicates that the unit's nurse staffing policy on each shift does not seem to be affected in any significant way by the realized patient census, over and above the day-of-week fixed effects.

Our discussions with the ED nurse manager indicated that there are two main factors driving the significant variations in the number of scheduled nurses. First, personnel scheduling is subject to numerous constraints (e.g., union rules) that often prevent the

Table 1 Descriptive Statistics for Nurse and Patient Data

Measure	Mean	Std. dev.	Minimum	Maximum
Day shift				
Nurses scheduled	11.4	1.07	8	16
Absenteeism rate	0.0762	0.0799	0	0.4
Patient visits	141	20.1	77	188
Average census	116	17.1	56.5	158
Maximum census	136	20.8	64	182
Night shift				
Nurses scheduled	10.5	0.849	9	14
Absenteeism rate	0.0707	0.0829	0	0.4
Patient visits	66.0	9.45	40	95
Average census	102	14.2	54.3	142
Maximum census	127	20.4	57	174
Evening shift				
Nurses scheduled	3.63	0.756	2	5
Absenteeism rate	0.0589	0.119	0	0.5
Patient visits	137	16.2	75	196
Average census	125	18.6	58.2	164
Maximum census	137	20.7	64	182

¹ We note that because of budget constraints, the ED is only allowed to use agency/overtime nurses to cover for absenteeism and is not allowed to make use of such expensive resources for routine staffing. Inquiries with staffing managers in other hospital settings have suggested that this practice is by no means unique to this specific ED.

manager from assigning the number of nurses desired for a particular shift. Second, as mentioned earlier, initial schedules often undergo a series of changes before they are finalized. Such scheduling variations are not desirable from the point of view of managing the match between the demand for nursing services and the supply of nursing capacity, but they provide an opportunity to examine how absenteeism rates are related to the numbers of scheduled nurses.

2.1. Nurse Workload and Absenteeism: Empirical Results

We model the phenomenon of nurse absenteeism as follows. We treat all nurses as being identical and independent decision makers and focus on a group of y_t nurses scheduled to work during a particular shift t ($t = 1$ for the first shift in the data set, $t = 2$ for the second shift, etc., up to $t = 607$). For nurse n , $n = 1, \dots, y_t$, the binary variable $Y_{n,t}$ denotes her decision to be absent from work ($Y_{n,t} = 1$), or to be present ($Y_{n,t} = 0$). We assume that this absenteeism decision is influenced by a number of factors expressed by the vector x_t which include parameters related to workload as well as fixed effects such as the day of the week or the shift. Each nurse compares the utility she receives from being absent from work to the utility she receives from going to work. The difference in these utility values is given by $U_{n,t}^* = x_t'\beta + \epsilon_{n,t}$, where $\epsilon_{n,t}$ are, for each n and t , independent and identically distributed random variables with mean zero. Although the utility difference $U_{n,t}^*$ is an unobservable quantity, we can potentially observe each nurse's decision to show up for work. The decision is such that $Y_{n,t} = 1$ if $U_{n,t}^* > 0$, and $Y_{n,t} = 0$ otherwise. Assuming that $\epsilon_{n,t}$ follow the standardized logistic distribution (the standard normal distribution), we obtain the logit (probit) model (Greene 2005). Note that our empirical data do not record the attendance decisions of individual nurses. Rather, we measured the aggregate absenteeism behavior of a group of nurses scheduled for a particular shift. Consequently, we treat all nurses scheduled for a given shift as a homogeneous group and build the model for the corresponding group behavior. We examine the impact of relaxing this assumption in §2.2. We focus on the maximum-likelihood-based logit estimation of the probability of absenteeism γ_t during shift t .

Because our goal is to study how the nurse absenteeism rate is affected by workload, we need to measure and quantify nurse workload for each shift. We use the nurse-to-patient ratio as a proxy for the workload nurses experience during a particular shift. For shift t , we define the nurse-to-patient ratio variable, denoted as NPR_t , as the ratio of the number of nurses working during a particular shift and the patient census averaged over the duration of that

shift. Therefore, the number of nurses present during each 24-hour period varies as follows: Between 8:00 A.M. and 12:00 P.M., it is equal to the number of nurses scheduled for the day shift (y_t); between 12:00 P.M. and 8:00 P.M., it is equal to the number of nurses scheduled for the day shift (y_t) plus the number of nurses scheduled for the evening shift (e_t); between 8:00 P.M. and 12:00 A.M., it is equal to the number of nurses scheduled for the evening shift (e_t) plus the number of nurses scheduled for the night shift; and between 12:00 A.M. and 8:00 A.M., it is equal to the number of nurses scheduled for the night shift (y_t). Thus, we estimate NPR_t as follows:

$$NPR_t = \frac{y_t + \frac{2}{3}e_t}{C_t} \text{ for the day shift,} \quad (1)$$

$$NPR_t = \frac{y_t + \frac{1}{3}e_t}{C_t} \text{ for the night shift,}$$

where C_t is the patient census averaged over the duration of shift t .

In making their attendance decisions for shift t , nurses may be influenced by the anticipated workload for shift t . The impact of *anticipated* workload arises because nurses are informed in advance of their schedule and are aware of how many (and which) other nurses are scheduled to work on the same shift. Because nurses anticipate a certain patient census $E[C_t]$, consistent with their past experience of working in the ED, nurses form an expectation about the anticipated workload for that shift. Naturally, if fewer (more) nurses are scheduled on that particular shift than the nurses deem appropriate, they will anticipate a higher (lower) workload than normal. The group attendance data do not present a measurement challenge because nurses scheduled for the same shift are subjected to the same anticipated workload value.

The anticipated workload can have a dampening effect on absenteeism through the "pressure-to-attend" mechanism (Steers and Rhodes 1978) or can enhance absenteeism by encouraging "withdrawal behavior" (Hill and Trist 1955, Hobfoll 1989). To the best of our knowledge, the impact of anticipated workload on absenteeism has not been previously studied. We test this potential impact in our setting by including in the vector of covariates x_t the anticipated value of nurse-to-patient ratio,

$$ENPR_t^1 = \frac{y_t + \frac{2}{3}e_t}{E[C_t]} \text{ for the day shift,} \quad (2)$$

$$ENPR_t^1 = \frac{y_t + \frac{1}{3}e_t}{E[C_t]} \text{ for the night shift,}$$

where e_t is the number of nurses scheduled in the evening shift which overlaps with $\frac{2}{3}$ ($\frac{1}{3}$) of the duration of the day (night) shift in question. Day- and

night-shift nurses are fully informed about the schedule for their shifts, but it is not clear that they would be as familiar with the schedule of the evening shift staffed by a different pool of nurses. Motivated by this observation, we estimate two models based on alternative definitions of the expected nurse-to-patient ratio. In the first definition ($ENPR_t^1$ of Equation (2)), we use the exact number of evening nurses scheduled (e_t), whereas in the second definition ($ENPR_t^2$), we use the average value of e_t (averaged over all evening shifts in our sample). The latter formulation reflects the situation where day- and night-shift nurses do not know precisely how many evening nurses will be present but form a rational expectation about this value. In other words, in the second model, day- and night-shift nurses behave as if they ignore any variation in the number of nurses scheduled for the evening shift that overlaps with their own shift. The expected patient census values $E[C_t]$ are computed by averaging the patient census over all shifts in our sample. This formulation reflects an assumption that nurses, when making their attendance decisions, use a mental model that captures any potential difference occurring on different days/shifts with a fixed effect and, therefore, focus on expected patient census. We also assume that the nurses form rational expectations about the patient census consistent with empirically observed patient census data. In addition to the models based on (2), we have also estimated several alternative variants, which we discuss in 2.2.

In addition to the anticipated nurse-to-patient ratio ($ENPR_t^i$, $i = 1, 2$), the vector of covariates x_t includes a number of controls. In particular, we include a day-of-the-week dummy variable to capture any systematic variation in absenteeism across days, a day/night-shift fixed effect to capture variations between day and night shifts, and a weekly fixed effect to capture any systematic variations that remain constant over one week and affect absenteeism, but are otherwise unobservable. Also, we include a holiday fixed effect that takes the value of 1 on national public holidays and 0 on any other day. This last variable is designed to deal with a potential endogeneity problem because nurses may be inherently reluctant to work on some select days, such as public holidays. These days are known to the management of the clinical unit which tries to accommodate the nurses' aversion by staffing fewer nurses on such days. Nevertheless, the nurses that are scheduled to work on these "undesirable" days are still more likely to be absent, irrespective of the chosen staffing levels. By including the holiday variable we are trying to explicitly account for this effect. There may exist other correlated variables that we omit, but to the extent that they do not vary drastically over a period of one week, the weekly fixed effect should capture the

influence of those variables. Finally, to account for the possibility that absenteeism might be a delayed response to past workloads, we include the values of 14 lagged nurse-to-patient ratios NPR_{t-j} , $j = 1, \dots, 14$, which correspond to one calendar week. Because the number of past shifts we use is rather arbitrary, we conducted our statistical analysis for several different values to make sure the results are not sensitive to the number we choose.

Specifically, the models we estimate are

$$\begin{aligned} \logit(\gamma_t) = & \beta_0^i + \beta_{ENPR}^i \times ENPR_t^i + \sum_{j=1}^{14} \beta_{NPR,j}^i \times NPR_{t-j} \\ & + \sum_{d=2}^7 \beta_{DAY,d}^i \times DAY_{d,t} + \sum_{f=2}^{44} \beta_{W,f}^i \times W_{f,t} \\ & + \beta_{DAYSHIFT}^i \times DAYSHIFT_t \\ & + \beta_{HOLIDAY}^i \times HOLIDAY_t, \end{aligned} \quad (3)$$

where γ_t is the probability that a nurse is absent in shift t ; $i = 1, 2$ refers to the definition of $ENPR^i$ used, $DAY_{d,t}$ and $W_{f,t}$ are the day and week fixed effects, and $DAYSHIFT_t$ and $HOLIDAY_t$ are the shift and holiday fixed effects. The estimation results for Equation (3) are presented in Table 2. Model I uses the first definition of the anticipated nurse-to-patient ratio ($ENPR_t^1$), whereas Model III uses the second definition ($ENPR_t^2$). To test whether the observed effect of anticipated nurse-to-patient ratio on absenteeism is robust, we also estimated the restricted versions of Models I and III (which we denote as Models II and IV), where we omit the 14 lagged nurse-to-patient ratio variables. If the lagged nurse-to-patient ratios are not related to absenteeism (i.e., if $\beta_{NPR,j}^i = 0$ for all $j = 1, \dots, 14$), omitting these variables will not introduce any bias even if the lagged nurse-to-patient variables are correlated with the variables included in the model. Model II uses the first definition of the anticipated nurse-to-patient ratio ($ENPR_t^1$), whereas Model IV uses the second definition ($ENPR_t^2$).

As can be seen from Table 2, the anticipated nurse-to-patient ratio has a significant effect (at the 5% or 10% level) on absenteeism rates in Models I, III, and IV. In Model II the p -value of the anticipated nurse-to-patient ratio is 10.7%. The more nurses scheduled for a particular shift, the less likely each nurse is to be absent. In particular, according to the first model, we estimate that the marginal effect of staffing an extra nurse (calculated at the mean values of all remaining independent variables and using the expected patient census value of 109) on the individual absenteeism rate is around $0.572\% = 0.623/109$. In other words, the absenteeism rate would decrease from its average value of 7.34% to about 6.78% when

Table 2 Estimation Results for Logit Models

Variable	Model I		Model II		Model III		Model IV	
	Coefficient (robust error)	Marginal effect (robust error)						
<i>ENPR1</i>	-10.01* (5.669)	-0.623* (0.352)	-8.534 (5.291)	-0.533 (0.330)				
<i>ENPR2</i>					-15.24** (6.355)	-0.946** (0.393)	-13.25** (5.892)	-0.827** (0.367)
<i>NPR1</i>	-0.248 (3.061)	-0.0154 (0.190)			-0.474 (3.071)	-0.0294 (0.191)		
<i>NPR2</i>	2.463 (3.438)	0.153 (0.214)			3.024 (3.459)	0.188 (0.215)		
<i>NPR3</i>	-0.492 (2.989)	-0.0306 (0.186)			-0.712 (2.982)	-0.0442 (0.185)		
<i>NPR4</i>	3.314 (3.265)	0.206 (0.203)			3.673 (3.275)	0.228 (0.203)		
<i>NPR5</i>	-2.620 (3.020)	-0.163 (0.188)			-2.917 (3.023)	-0.181 (0.188)		
<i>NPR6</i>	5.912* (3.249)	0.368* (0.201)			5.837* (3.251)	0.362* (0.201)		
<i>NPR7</i>	0.518 (3.155)	0.0322 (0.196)			0.615 (3.157)	0.0382 (0.196)		
<i>NPR8</i>	1.556 (3.245)	0.0968 (0.202)			1.570 (3.254)	0.0974 (0.202)		
<i>NPR9</i>	-4.112 (2.998)	-0.256 (0.186)			-4.252 (3.002)	-0.264 (0.186)		
<i>NPR10</i>	5.656* (3.205)	0.352* (0.199)			5.746* (3.207)	0.357* (0.198)		
<i>NPR11</i>	-3.320 (2.980)	-0.207 (0.186)			-3.263 (2.991)	-0.203 (0.186)		
<i>NPR12</i>	2.462 (3.271)	0.153 (0.204)			2.707 (3.276)	0.168 (0.203)		
<i>NPR13</i>	-2.934 (3.016)	-0.183 (0.188)			-2.724 (3.018)	-0.169 (0.187)		
<i>NPR14</i>	4.934* (2.987)	0.307* (0.185)			4.823 (2.978)	0.299 (0.184)		
<i>DAYSHIFT</i>	0.112 (0.151)	0.00694 (0.00937)	0.233 (0.143)	0.0145 (0.00893)	0.211 (0.163)	0.0130 (0.0101)	0.327** (0.155)	0.0203** (0.00959)
<i>HOLIDAY</i>	-0.853* (0.440)	-0.0382*** (0.0134)	-0.787** (0.401)	-0.0364*** (0.0131)	-0.848* (0.440)	-0.0380*** (0.0135)	-0.783* (0.401)	-0.0362*** (0.0131)
<i>Week FE</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Day-of-Week FE</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Constant</i>	-2.381* (1.426)		-2.171*** (0.721)		-1.834 (1.449)		-1.726** (0.752)	
Observations		6,481		6,631		6,481		6,631
Log-likelihood		-1,654		-1,690		-1,653		-1,688
LR-test		102.7		92.35		105.3		95.13

Notes. Standard errors are shown in parentheses. Marginal effects are calculated at the mean values of the independent variables. For the dummy variables, the marginal effect shows the difference caused by changing the variable from 0 to 1.

*, **, and *** denote statistical significance at the 10%, 5%, and 1% confidence levels, respectively.

an extra nurse is added to the schedule. The log-likelihood is (marginally) higher and the coefficient of the anticipated nurse-to-patient ratio is larger in Models III and IV, where the variation in the number of scheduled evening nurses is ignored. This might suggest that when nurses decide whether to show up for work, they place greater emphasis on the number of nurses working in their shift (i.e., $ENPR_t^2$) rather than both the number of nurses working in their shift

and the evening shift that overlaps with their own (i.e., $ENPR_t^1$).

As with any empirical finding, one might argue that the relationship we find between absenteeism and anticipated workload is a result of reverse causality (i.e., it is not absenteeism that reacts to anticipated workload but instead staffing, and thus workload, that reacts to absenteeism). However, in the ED study site, we know that absenteeism was not considered

by the nurse manager making staffing decisions. More generally, in discussions with managers responsible for nurse staffing in other hospitals, absenteeism patterns were not tracked or used in staffing decisions.

Note that the lag 6, lag 10, and lag 14 (lag 6 and lag 10) of the nurse-to-patient ratio variables have positive coefficients in Model I (Model III) which are individually significant at the 10% level. This seems to imply that the probability of a nurse being absent on any shift would increase if the shift occurring three, five, or seven days ago had a higher nurse-to-patient ratio. Although these lags are individually significant, the Wald test statistic and the likelihood ratio test statistic for joint significance of all 14 lagged workload variables in Model I as well as Model III reject the hypothesis (even at the 10% level) that lagged nurse-to-patient ratios have any joint explanatory power. In the absence of any plausible explanation as to why a lighter workload on a similar shift occurring three, five, or seven days ago might increase absenteeism, and in light of the weakness of this statistical relationship, we are inclined to treat this result as spurious.

Interestingly, the holiday variable's effect is significant (at 10% confidence level) and negative, thus suggesting that nurses are about 3.8% less likely to be absent on public holidays. Week fixed effects are jointly significant (at the 1% level) in all four models. One of the effects that week dummies seem to detect reasonably well is the impact of weather (in particular, heavy snow conditions) on absenteeism. For example, week 35 of our data set includes March 2–4, 2009. During six day and night shifts corresponding to these dates, snow on the ground in New York City was recorded to be more than five inches,² the level identified by the New York Metropolitan Transportation Authority as the one at which the public transportation disruptions are likely to set in.³ The week-35 fixed effect is positive and significant (at the 5% level) in Models II and IV with a marginal effect equal to 8.75% in Model I and 9.80% in Model IV. Only three other days in our data set had as much snow on the ground (December 3, December 21, and January 20). Turning to the impact of day-of-week fixed effect, there is some (weak) evidence that nurses are more likely to be absent on weekends. They are also more likely to be absent during a day shift, when conflicting family obligations is often cited as an important reason behind nurse absenteeism (Erickson et al. 2000, Nevidjon and Erickson 2001) are likely to be more prevalent.

² <http://www.accuweather.com/>.

³ <http://www.mta.info/news/stories/?story=173> (last accessed January 23, 2011).

2.2. Verification Tests

To check the validity of our model estimation procedure, we have also conducted several verification tests as described below. (Estimation details are available from the authors.)

2.2.1. Alternative Specifications. To ensure the robustness of our results, we estimated a number of alternative modeling specifications. Namely, we estimated the models of Equation (3) under the probit specification. We also estimated variants of our models that use month fixed effect variables instead of week fixed effect variables. The results were almost identical in terms of variable significance, model significance, and magnitude of marginal effects. The model of Equation (3) assumes that any difference between the night shift and the day shift is completely captured by the dummy variable *DAYSHIFT*. However, it is possible that the two shifts are inherently different, and so are their best-fit coefficients. To test the hypothesis that the coefficients for two shifts differ, we fitted an unrestricted binary choice model to the data for each of the two shifts separately and compared the fit with the restricted model by constructing a likelihood ratio test. The test does not reject the hypothesis that the coefficients are the same at the 5% confidence level in all models. Finally, we estimated models with alternative definitions of the expected nurse-to-patient ratio. More specifically, when estimating the expected patient census during the upcoming shift in Equation (2), nurses (a) distinguish between day and night shifts, but not between days of the week; (b) distinguish between days of the week, but not between day shifts or night shifts; and (c) distinguish between both the shift type and the day of the week. Our main finding that higher (lower) anticipated nurse-to-patient ratios decrease (increase) nurse absenteeism is robust to these alternative modeling specifications.

2.2.2. Nurse Heterogeneity and Aggregation Bias.

The aggregate nature of our data does not permit an exact characterization of how workload impacts the absenteeism behavior of each individual nurse. Indeed, it would be interesting to measure whether aversion to anticipated increased workloads is a commonly occurring nurse characteristic or limited to a relatively small subset of nurses. However, our aggregate results are not invalidated by the lack of such a characterization. Through a simulation study, we find that estimating a homogeneous logit model when nurses are in fact heterogeneous does produce biased estimation coefficients. However, we find that the bias on the *ENPR* coefficient is *positive*, that is, it would bias our (negative) coefficient toward zero and not away from it. Therefore, our estimate should be treated as a lower bound on how workload affects

nurse absenteeism. Furthermore, we find that for most reasonable assumptions about nurse heterogeneity, the magnitude of the bias is small compared to the estimated value of the coefficient. This result is similar to that of Allenby and Rossi (1991), whose analysis of marketing data lead the authors to conclude that in most realistic settings there exists no significant aggregation bias in logit models.

2.2.3. Within-Shift Correlation. One of the assumptions we have made when estimating our models is that, conditional on the vector of covariates \mathbf{x}_i , the nurses are independent decision makers. In this section, we relax this assumption and assume that nurses can exhibit within-shift correlation. We estimate a model within the generalized estimating equations framework, which allows the individual nurse decisions $Y_{n,t}$ to exhibit within shift correlation. The correlation structure estimated is of the "exchangeable form," that is, $\text{Corr}(Y_{n,t}, Y_{m,i}) = \rho$ for $t = i$, and is 0 otherwise. The least-square estimation of these models yields very similar results to those presented in Table 2, and the estimated correlation ρ is around negative 1%.

2.3. Discussion and Implications of the Empirical Study

Our empirical investigation has uncovered a link between absenteeism and workload at the shift level. We find that nurses are rational and forward looking in their decision whether to attend or to be absent from work. When an extra nurse is added to the schedule, the absenteeism rate decreases on average from 7.34% to 6.78%, a relative decrease of 7.8%. This particular behavior, where nurses are more likely not to show up when the workload per nurse is already high, if not appropriately countered, will exacerbate an already difficult situation for the hospital by creating more workload for the nurses that do show up. Such a positive feedback in workload (i.e., high workload generates even higher workload) can have dire consequences for patient safety (see Needleman et al. 2002) unless mitigating actions take place. However, there is a silver lining to our finding in the sense that the endogeneity of absenteeism might represent an opportunity for hospitals. In the presence of endogenous absenteeism, scheduling an extra nurse might be more beneficial and even more cost effective than hospital managers might realize because this extra nurse increases the probability that all of her colleagues will show up for work, therefore reducing the costs associated with absenteeism. Failing to account for this fundamental additional marginal benefit, which is purely due to the endogeneity of absenteeism, might lead to chronic understaffing.

To further investigate the implications of absenteeism and, in particular, its endogenous nature with

regard to staffing level decisions, we construct a stylized model of nurse staffing. The aim of our model is to generate managerial insights on the impact of nurse absenteeism, in general, and the endogenous nature of absenteeism, in particular, as it affects the decision of how many nurses to staff. Although our model is parsimonious, we believe that an appropriately calibrated version of it can be used by nursing management in making tactical staffing decisions. In particular, by periodically running the model for all types of shifts (which differ in the distribution of patient census and in nurse absenteeism propensity), the nurse manager can decide how many nurses the unit will need for each shift. Thus, coupled with rostering considerations, the model can help the manager decide the appropriate aggregate staffing level for the unit. In addition to providing insight on the aggregate level of permanent nurses needed by observing how many agency or overtime nurses are required on average to cover for last minute mismatches between demand and supply, our model can help to determine how many such "flexible" nurses the hospital will need to maintain adequate nurse staffing levels.

3. Endogenous Nurse Absenteeism: Implications for Nurse Staffing

We begin our nurse staffing model by outlining the key assumptions. We assume that a clinical unit uses the primary nursing care (PNC) mode of nursing care delivery (Seago 2001), which was used in the ED we studied. Under the PNC mode, the nursing staff includes only registered nurses (as opposed to licensed practical nurses or unlicensed nursing personnel) who provide all direct patient care throughout the patients' stay in the clinical unit. The nurse staffing process starts several weeks in advance of the actual shift for which planning is performed. It is then that a hospital staff planner needs to decide how many nurses (y) to schedule for that particular shift. Because of the phenomenon of absenteeism, the actual number of nurses who show up for work on that shift, N , is uncertain. We model N as a binomial random variable $B(y, 1 - \gamma(y))$, where $\gamma(y)$ is the probability that any scheduled nurse will be absent from work:

$$\begin{aligned} \text{Prob}(N = k | y, \gamma(y)) \\ = p(k; y, \gamma(y)) = \begin{cases} \frac{y!}{k!(y-k)!} (\gamma(y))^{y-k} (1 - \gamma(y))^k & \text{for } 0 \leq k \leq y, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (4)$$

We assume that the clinical unit follows a policy of specifying, for each value of the average patient census during a shift, C , a target integer number of nurses $T = R(C)$ required to provide adequate patient

care during a particular shift. We assume that C takes on discrete values and that $R(C)$ is a monotone increasing function with $R(0) = 0$. A simple example of $R(C)$ is provided by a "ratio" approach under which $R(C) = \lceil \alpha C \rceil$, with $\alpha \in [0, 1]$ representing a mandated nurse-to-patient ratio. Alternately, if a clinical unit is modeled as a queueing system in which patients generate service requests and nurses play the role of servers, as was done in Yankovic and Green (2011), $R(C)$ can take a more complex form to ensure that certain patient service performance measures, such as the expected time patients wait to be served, conform to prespecified constraints.

At the time of the nurse staffing decision, we assume that the decision maker uses a known probability density function of the average patient census C during the shift for which personnel planning is conducted, $\text{Prob}(C = n) = p_C(n)$, $n \in \mathcal{N}^+$, $\sum_{n=0}^{\infty} p_C(n) = 1$. We treat the demand uncertainty expressed by the patient census C and the supply uncertainty expressed by N as being independent and assume that the realized values of C and N become known shortly before the beginning of the shift. Any nursing shortage $(R(C) - N)^+$ is covered by either hiring agency nurses or asking nurses who have just completed their shift to stay overtime. We further assume that nurses that do show up are paid $\$w_r$ per shift, whereas nurses that do not show up are paid $\$w_n$, where $w_r \geq w_n$. Setting $w_r = w_n$ represents a clinical unit where nurses are paid the full wage whether or not they actually show up for work. This setting is consistent with the PNC mode of nursing care delivery where nurses are salaried employees who can only be scheduled to work on a fixed number of shifts per week. When a scheduled nurse does not show up for work, she cannot be rescheduled in lieu of the shift she missed. Thus, in effect, nurses are paid for each shift for which they are scheduled and are not penalized for being absent, as long as their absenteeism does not exceed the annual limit of 10 personal days. In contrast, setting $w_n = 0$ represents a setting where nurses receive no pay when absent. In addition, we assume that if more nurses show up for work than required given the number of patients present ($N > R(C)$) they all have to be paid and cannot be sent home. The per-shift cost of extra/overtime nurses is $\$w_e$, where $w_e \geq w_r$.

The goal of the decision maker is to choose a nurse staffing level y that minimizes the expected cost $W(y)$ of meeting the target $R(C)$:

$$W(y) = w_n y + (w_r - w_n) \mathbf{E}_N[N | y] + w_e \mathbf{E}_{C,N}[(R(C) - N)^+ | y], \quad (5)$$

where \mathbf{E}_N denotes expectation taken with respect to the number of nurses who show up for work and $\mathbf{E}_{C,N}$ denotes expectation taken with respect to both

the number of patients and the number of nurses who show up for work. Note that because there is a one-to-one correspondence in our model between the patient demand C and the number of required nurses T , we can recast the calculation of the expectation with respect to the demand value in terms of an equivalent calculation over the distribution of T using the corresponding probability distribution function. In particular, let \mathcal{S}_n be the set of average patient census values all corresponding to the same number of required nurses n : $\mathcal{S}_n = \{C \in \mathcal{N}^+ | R(C) = n\}$. Then the probability distribution for T is given by $\text{Prob}(T = n) = p_T(n)$, $n \in \mathcal{N}^+$, $p_T(n) = \sum_{l \in \mathcal{S}_n} p_C(l)$, $\sum_{n=0}^{\infty} p_T(n) = 1$. In turn, the cost minimization based on (5) becomes

$$\min_{y \in \mathcal{N}^+} (w_n y + (w_r - w_n)(1 - \gamma(y))y + w_e \mathbf{E}_{T,N}[(T - N)^+ | y]). \quad (6)$$

Note that with no absenteeism ($\gamma(y) = 0$), the number of nurses showing up for work N is equal to the number of scheduled nurses y , and the nurse staffing problem reduces to a standard newsvendor model with the optimal staffing level given by

$$y_0^* = \min \left(y \in \mathcal{N}^+ \mid F_T(y) \geq 1 - \frac{w_r}{w_e} \right), \quad (7)$$

with $F_T(y) = \sum_{n=0}^y p_T(n)$ being the cumulative density function of the demand function evaluated at y , and the value $1 - w_r/w_e$ playing the role of the critical newsvendor fractile. Below we present an analysis of the staffing decision (6) starting with the case of exogenous absenteeism, which we use as a benchmark.

3.1. Optimal Nurse Staffing Under Exogenous Absenteeism Rate

Consider a clinical unit that experiences an endogenous nurses' absenteeism rate $\gamma(y)$, but treats it as exogenous. For example, the schedule planner uses the average value of all previously observed daily absenteeism rates, γ_{ave} . The cost function to be minimized under this approach is given by

$$W_{\text{ave}}(y) = y(w_r(1 - \gamma_{\text{ave}}) + w_n \gamma_{\text{ave}}) + w_e \sum_{k=0}^y \sum_{n=0}^{\infty} (n - k)^+ p_T(n) p(k; y, \gamma_{\text{ave}}) = w y + w_e \sum_{k=0}^y q(k) p(k; y, \gamma_{\text{ave}}), \quad (8)$$

where $w = w_r(1 - \gamma_{\text{ave}}) + w_n \gamma_{\text{ave}}$ is the effective cost per scheduled nurse, and

$$q(k) = \sum_{n=0}^{\infty} (n - k)^+ p_T(n) \quad (9)$$

represents an expected nursing shortage given that k regular nurses show up for work. The optimal staffing level in this case is expressed by the following result.

PROPOSITION 1. (a) The minimizer of (8) is given by

$$y_{ave}^* = \min \left(y \in \mathcal{N}^+ \left| \sum_{k=0}^y F_T(k) p(k; y, \gamma_{ave}) \geq 1 - \frac{w_r(1 - \gamma_{ave}) + w_n \gamma_{ave}}{w_e(1 - \gamma_{ave})} \right. \right), \quad (10)$$

and is a nonincreasing function of w_r/w_e and w_n/w_e .

(b) Consider two cumulative distribution functions for the required number of nurses T , $F_T^1(k)$ and $F_T^2(k)$ such that $F_T^1(k) \geq F_T^2(k)$ for all $k \in \mathcal{N}^+$, and let $y_{ave}^{*,i}$ be the optimal staffing levels corresponding to $F_T^i(k)$, $i = 1, 2$. Then, $y_{ave}^{*,1} \leq y_{ave}^{*,2}$.

We relegate all proofs to the appendix. Note that (10) represents a generalization of the expression for the optimal staffing levels without absenteeism (7). As in the no-absenteeism setting, it is never optimal to decrease staffing levels when the target nursing level increases or when the cost advantage associated with earlier staffing becomes more pronounced. While this behavior of the optimal policy is intuitive, the dependence of the optimal staffing levels on the value of the absenteeism rate is not as straightforward. In particular, depending on the interplay between the ratios of the cost parameters w_r/w_e and w_n/w_e , the characteristics of the target nursing level distribution, and the absenteeism rate, the increase in the absenteeism rate can increase or decrease the optimal staffing level. The following result describes the properties of the optimal staffing levels in general settings.

PROPOSITION 2. (a) There exists γ_{ave}^u such that the optimal staffing level y_{ave}^* is a nonincreasing function of the absenteeism rate γ_{ave} for all $\gamma_{ave} \in [\gamma_{ave}^u, 1]$.

(b) For

$$w_n \leq w_e \left(\left[F_T^{-1} \left(1 - \frac{w_r}{w_e} \right) \right] \right) p_T \left(\left[F_T^{-1} \left(1 - \frac{w_r}{w_e} \right) \right] \right), \quad (11)$$

there exists γ_{ave}^l such that the optimal staffing level y_{ave}^* is a nondecreasing function of the absenteeism rate γ_{ave} for all $\gamma_{ave} \in [0, \gamma_{ave}^l]$.

A more detailed characterization of the optimal staffing levels can be obtained for some target nursing level distributions, for example, for a discrete uniform distribution.

COROLLARY 1. Let

$$F_T(k) = \begin{cases} \frac{k+1}{T_{max}+1} & \text{for } 0 \leq k \leq T_{max}, \\ 1 & k \geq T_{max}. \end{cases} \quad (12)$$

Then, for $w/w_e \geq \frac{1}{4}$, the optimal nurse staffing level is given by

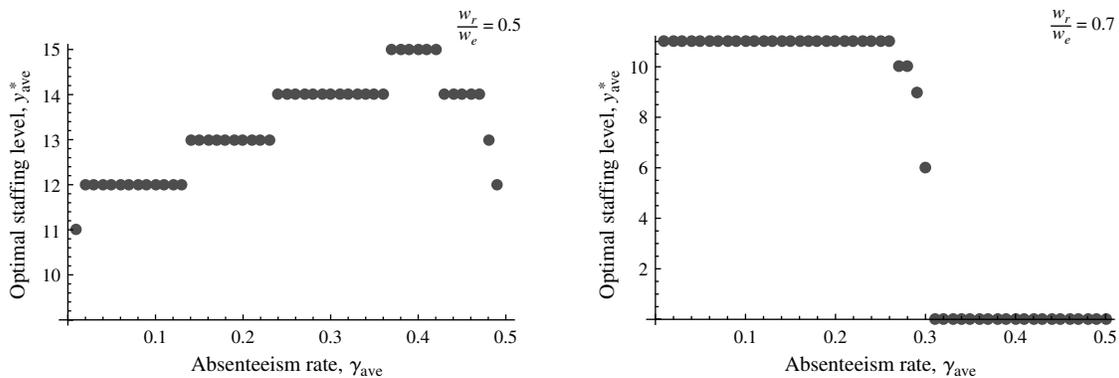
$$y_{ave}^* = \left\lceil \left(\frac{T_{max}}{1 - \gamma_{ave}} - \frac{T_{max} + 1}{(1 - \gamma_{ave})^2} \frac{w}{w_e} \right) \right\rceil, \quad (13)$$

and is a nondecreasing (nonincreasing) function of γ_{ave} for $\gamma_{ave} \leq \gamma_{ave}^u$ ($\gamma_{ave} > \gamma_{ave}^u$), where

$$\gamma_{ave}^u = \max \left(0, 1 - \max \left(0, \frac{2(T_{max} + 1)w_r}{T_{max}w_e - (T_{max} + 1)(w_r - w_n)} \right) \right). \quad (14)$$

To illustrate the monotonicity properties of the optimal staffing levels formalized in Proposition 2, we use the distribution for the number of required nurses obtained from our empirical data for the average patient census using 1-to-10 nurse-to-patient ratio. For this distribution, Figure 1 shows the dependence of the optimal staffing level on the absenteeism rate for $w = w_r = w_n$. For a given value of the cost ratio w/w_e , there exists a critical value of the absenteeism rate γ_{ave}^u for which the optimal response to an increase in absenteeism switches from staffing more nurses to staffing fewer. Note that irrespective of the distribution for targeted nursing level, for high values of the absenteeism rate or high values of the cost ratio w/w_e (to be precise, for $\gamma_{ave} \geq 1 - w/w_e$), it is more cost-effective not to staff any nurses in advance and to

Figure 1 Optimal Staffing Level as a Function of the Absenteeism Rate for Different Values of the Cost Ratio w/w_e for $w = w_r = w_n$ and the Empirical Targeted Nursing Level Distribution



rely exclusively on the extra/overtime mechanisms of supplying the nursing capacity. For low values of the absenteeism rate and low values of the cost ratio w/w_e , higher absenteeism can induce an increase in staffing levels, as it is cheaper to counter the increased absenteeism by staffing more nurses. However, as the cost ratio w/w_e increases, it becomes more cost effective to staff fewer nurses, relying increasingly on the extra/overtime supply mechanism.

3.2. Endogenous Absenteeism: Optimal Staffing

In the endogenous absenteeism setting, the expected staffing cost (6) becomes

$$W(y) = yw_r - (w_r - w_n)a(y) + w_e L(y, \gamma(y)), \quad (15)$$

where

$$a(y) = y\gamma(y) \quad (16)$$

is the expected number of absent nurses, and

$$L(y, \gamma(y)) = \sum_{k=0}^y q(k)p(k; y, \gamma(y)), \quad (17)$$

where $q(k)$ is defined by (9), $p(k; y, \gamma(y))$ is the probability mass function of the binomial distribution, where k nurses show up for work when y are scheduled and $\gamma(y)$ is the (endogenous) probability of a nurse being absent. Note that for general absenteeism rate function $\gamma(y)$, the increasing marginal property of the "exogenous" staffing cost function (8) with respect to the number of scheduled nurses may not hold. Below we formulate a sufficient condition for this property to be preserved under endogenous absenteeism. First, for a given distribution of the targeted nursing level $p_T(k)$, $k \geq 0$, we introduce the following quantity:

$$\gamma_T(y) = 1 - \min\left(1, \left(\frac{\sum_{k=y-2}^{\infty} p_T(k)}{yp_T(y-1) + p_T(y-2)}\right)^{1/(y-1)}\right), \quad y \in \mathcal{N}^+, y \geq 2. \quad (18)$$

As shown below, (18) represents one of the bounds on the absenteeism rate function that ensures the optimality of the greedy-search approach to finding the optimal nurse staffing level.

PROPOSITION 3. *Let $\gamma(x) \in C^2(0, \infty)$, $0 \leq \gamma(x) \leq 1$ be a nonincreasing, convex function defined on $x \geq 0$. Consider an endogenous absenteeism setting characterized by the absenteeism rate $\gamma(y)$ for $y \in \mathcal{N}^+$ scheduled nurses. Then, the optimal staffing level is given by*

$$\begin{aligned} y^* &= \min\left(y \in \mathcal{N}^+ \mid L(y+1, \gamma(y+1)) - L(y, \gamma(y)) \right. \\ &\quad \geq -\frac{w_r}{w_e}(1 - \gamma(y)) - \frac{w_n}{w_e}\gamma(y) \\ &\quad \left. + \frac{w_r - w_n}{w_e}y(\gamma(y+1) - \gamma(y))\right) \quad (19) \end{aligned}$$

and is a nonincreasing function of w_r/w_e , provided that

$$\gamma(y) \leq \min\left(\frac{2}{y}, \gamma_T(y)\right) \quad (20)$$

and

$$\frac{d^2 a(y)}{dy^2} \leq 0 \quad (21)$$

for any $y \geq y^*$. In addition, consider two cumulative distribution functions for the required number of nurses T , $F_T^1(k)$ and $F_T^2(k)$ such that $F_T^1(k) \geq F_T^2(k)$ for all $k \in \mathcal{N}^+$, and let $y^{*,i}$ be the optimal staffing levels corresponding to $F_T^i(k)$, $i = 1, 2$. Then, $y^{*,1} \leq y^{*,2}$, provided that (20) holds for any $y \geq y^{*,2}$.

Proposition 3 essentially describes a greedy-search algorithm for finding the optimal number of nurses to schedule. Starting from a chosen number of nurses y , the nursing manager can start adding (or subtracting) nurses as long as the total costs continue decreasing. At y^* , where the costs stop decreasing, the nursing manager can stop the search. While this is a one-dimensional optimization problem, we would argue that the greedy-search approach is useful because the sums over the probability mass of the patient census (see (9)) and nurses present (see (17)) required to calculate the costs of this problem can be time consuming to estimate.

The fact that such a greedy-search approach is optimal requires the sufficient condition (20) that states, intuitively, that the increasing marginal shape of the staffing cost function with respect to the number of scheduled nurses is preserved under endogenous absenteeism if the absenteeism rate is not too high. Thus, (15) is not too different from the cost function in (6). More specifically, this sufficient condition requires that the absenteeism rate function is limited from above by two separate bounds. The first bound implies that the expected number of absent nurses does not exceed two irrespective of the number of nurses actually scheduled for work. The sufficient condition (21) requires that the expected number of absent nurses exhibits nonincreasing returns to scale.

To study the endogenous absenteeism case further, we use a parametric specification consistent with our empirical findings, in particular with the logit model specification. We specify that

$$\gamma(y) = \frac{1}{1 + e^{\alpha + \beta y}}, \quad (22)$$

where both α and β are positive constants. The assumption about positive values for these absenteeism rate parameters is plausible in a wide range of settings; $\beta > 0$ implies that the absenteeism rate declines with the number of scheduled nurses, whereas $\alpha > 0$ ensures that the absenteeism rate is

not too high even when the number of scheduled nurses is low and the anticipated workload is high. In particular, evaluating the best-fit logit model in (3), using the estimates reported in Table 2, we obtain $\beta = -\beta_{ENPR}^1/E[C_i] = 0.092$, with $\beta_{ENPR}^1 = -10.01$, $E[C_i] = 109.0$. For the average absenteeism rate to match our sample average of 7.34%, we set $\alpha = 1.533$. Note that the endogenous absenteeism rate $\gamma(y)$ characterized by the logistic function given by (22) with $\alpha \geq 0$ and $\beta \geq 0$ is a monotone decreasing convex function. Thus, the result of Proposition 3 is ensured by the following restrictions on the values of α and β :

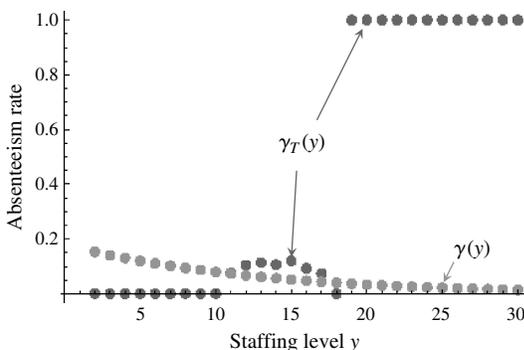
LEMMA 1. For $\gamma(y) = 1/(1 + e^{\alpha + \beta y})$, with $\alpha, \beta \geq 0$, $\beta e^{1 + \alpha + 2\beta} \geq \frac{1}{2}$ implies (20), and $\beta \leq 2$ implies (21).

In the ED we studied, the estimated values of $\alpha = 1.533$ and $\beta = 0.092$ satisfy Lemma 1. In particular, the maximum value of the product of number of scheduled nurses y and the estimated absenteeism rate calculated using these values is equal to 0.804, well below 2. The second bound on the right-hand side of (20) takes the form of an effective absenteeism rate function that depends exclusively on the distribution of the targeted nursing level. Note that $\gamma_T(y) \geq 0$ if and only if

$$\frac{p_T(y-1)}{\sum_{k=y-1}^{\infty} p_T(k)} \geq \frac{1}{y}. \tag{23}$$

The expression on the left-hand side of (23) is the hazard rate function for the distribution of the targeted nursing level. Thus, (23) stipulates that the bound described by (20) is meaningful only in settings where such a hazard rate evaluated at y exceeds $1/(y+1)$. For the absenteeism rate function given by (22), the constraint $\gamma(y) \leq \gamma_T(y)$ implies, in the same spirit as Lemma 1, the lower-bound restriction on the values of α and β : $\gamma(y) \leq \gamma_T(y) \Leftrightarrow \alpha + \beta y \geq \log((1 - \gamma_T(y))/\gamma_T(y))$. Figure 2 compares the absenteeism rate (22) computed for $\alpha = 1.533$ and $\beta = 0.092$, with the effective absenteeism rate $\gamma_T(y)$ from (18)

Figure 2 Absenteeism Rate $\gamma(y)$ Computed for $\alpha = 1.533$ and $\beta = 0.092$ and the Effective Absenteeism Rate $\gamma_T(y)$ Computed Using the Empirical Targeted Nursing Level Distribution



computed using the empirical targeted nursing level distribution. Note that the sufficient condition of Proposition 3 ($\gamma(y) \leq \gamma_T(y)$) is satisfied for virtually any staffing level above 10 nurses, which is approximately equal to the expected number of required nurses in our setting.

3.3. Endogenous Absenteeism: Implications of Model Misspecification on Staffing Decisions

In this section, we compare the optimal nurse staffing levels with those made by a clinical unit that incorrectly treats the absenteeism rate as exogenous and uses a trial-and-error procedure under which the assumed exogenous absenteeism rate is updated every time a new staffing decision is made. In this latter case, which we label “misspecified-with-learning,” the clinical unit selects staffing level y^{ML} such that

$$y^{ML} = \min \left(y \in \mathcal{N}^+ \mid \sum_{k=0}^y F_T(k)p(k; y, \gamma(y^{ML})) \geq 1 - \frac{w_r(1 - \gamma(y^{ML})) + w_n\gamma(y^{ML})}{w_e(1 - \gamma(y^{ML}))} \right), \tag{24}$$

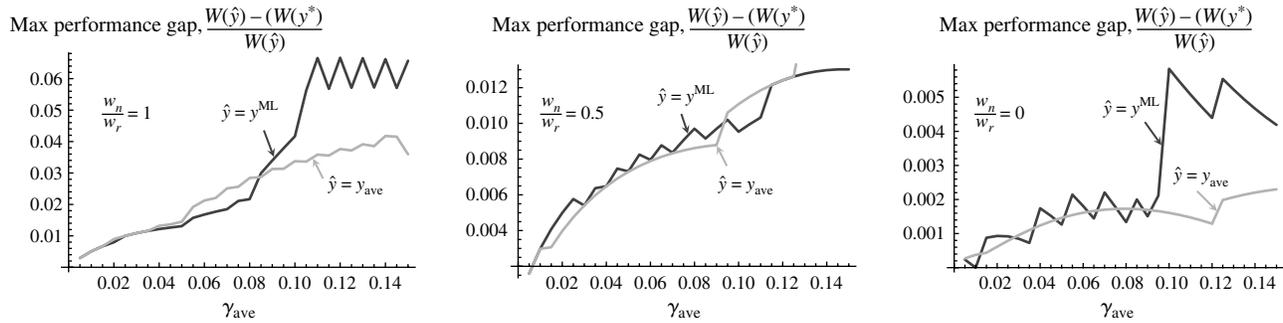
with $\gamma(y)$ denoting the endogenous absenteeism rate. Note that (24) reflects a self-consistent way of selecting the staffing level; y^{ML} is the best staffing decision in the setting where the absenteeism rate is exogenous and determined by $\gamma(y^{ML})$. In other words, a clinical unit assuming that the absenteeism rate is given by constant value $\gamma(y^{ML})$ will respond by scheduling y^{ML} nurses and, as a result, will observe exactly the same value of the absenteeism rate, even if the true absenteeism process is endogenous and described by $\gamma(y)$. An intuitive way of rationalizing the choice of y^{ML} is to consider a sequence of exogenous staffing levels y_n , $n \in \mathcal{N}^+$, such that

$$y_{n+1} = \min \left(y \in \mathcal{N}^+ \mid \sum_{k=0}^y F_T(k)p(k; y, \gamma(y_n)) \geq 1 - \frac{w_r(1 - \gamma(y_n)) + w_n\gamma(y_n)}{w_e(1 - \gamma(y_n))} \right), \quad n \in \mathcal{N}^+. \tag{25}$$

Equation (25) reflects a sequence of repeated adjustments of staffing levels, starting with some y_0 , each based on the value of the absenteeism rate observed after the previously chosen staffing level is implemented. In this updating scheme, y^{ML} can be thought of as the limit, $\lim_{n \rightarrow \infty} y_n$, if such a limit exists. Note that for a general demand distribution $F_T(k)$ and a general absenteeism rate function $\gamma(y)$, the set of staffing levels E satisfying (24) may be empty or may contain multiple elements. The analysis of existence and uniqueness of y^{ML} is further complicated by the discrete nature of staffing levels. In the following discussion, we bypass this analysis and assume that there exists at least one staffing level satisfying (24).

Downloaded from informs.org by [163.119.96.41] on 24 March 2016, at 11:01. For personal use only, all rights reserved.

Figure 3 Worst-Case Performance Gaps Between the Optimal Staffing Policy and the Misspecified-with-Learning (ML) and Exogenous Policies as Functions of the Average Absenteeism Rate γ_{ave} Under the Empirical Targeted Nursing Level Distribution for $w_n/w_r = 1, 0.5, 0$



As the following result shows, even if E contains multiple elements, each of them is bounded from above by the optimal endogenous staffing level in settings where the expected number of absentees decreases with the number of scheduled nurses.

PROPOSITION 4. *Suppose that the conditions of Proposition 3 hold and that the set of staffing levels satisfying (24), E , is nonempty. Then, $y^{ML} \leq y^*$, for any $y^{ML} \in E$, provided that, at the optimal staffing level y^* , the expected number of absent nurses decreases with the number of nurses scheduled, $a(y^* + 1) < a(y^*)$.*

Proposition 4 implies that ignoring the endogenous nature of absenteeism can lead to understaffing in settings where both the endogenous absenteeism rate and the expected number of absent nurses decline with the number of scheduled nurses. Figure 3 illustrates the results of the numerical experiment designed to quantify a potential cost impact of using heuristic staffing policies for realistic values of problem parameters. In our study, we have varied the cost ratio w_r/w_e from $w_r/w_e = 0.5$ to $w_r/w_e = 0.9$. The lower limit of this interval, $w_r/w_e = 0.5$, corresponds to the setting in which use of agency nurses carries a 100% cost premium, a realistic upper bound on the values encountered in practice. The upper limit, $w_r/w_e = 0.9$, reflects the use of overtime to compensate for absenteeism, with w_e at about 10% premium with respect to w_r . In the ED we studied, absent nurses were paid at the same rate as the nurses who showed up for work (so that $w_n = w_r$). To investigate the effect of lower compensation levels for absent nurses, we have included the cost ratios $w_n = 0.5w_r$ and $w_n = 0$. As the absenteeism rate function, we have used $\gamma(y) = 1/(1 + e^{\beta y})$ with the value of β varied to explore different average absenteeism rates calculated as

$$\gamma_{ave} = \sum_{y=y_{min}}^{y_{max}} \frac{p_T(y)}{1 + e^{\beta y}}, \quad (26)$$

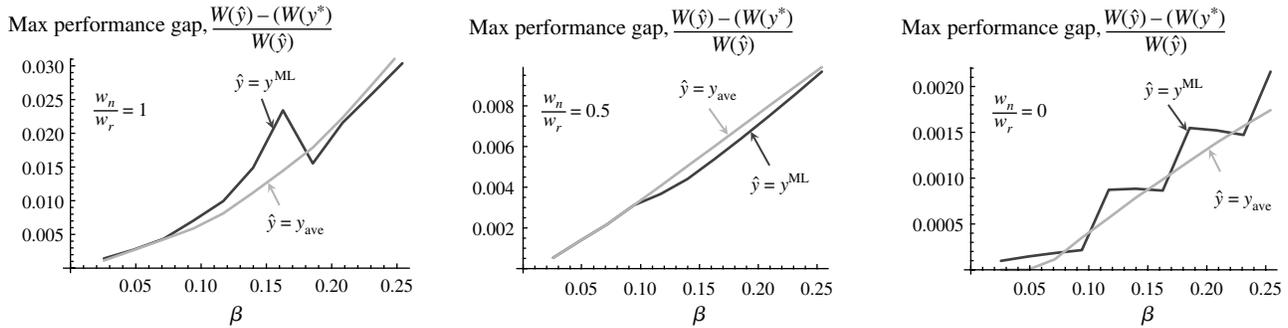
where $y_{min} = 6$ and $y_{max} = 16$ reflect the smallest and largest possible targeted nursing level realizations, and $p_T(y)$ reflects the empirical distribution for the

required number of nurses. In our study, we have compared the performance of three staffing policies: the optimal policy described in Proposition 3, the ML policy described by (24), and the exogenous policy that assumes that the absenteeism rate does not depend on the staffing level and is given by (26). Figure 3 shows the worst-case performance gaps (calculated over the range of cost ratios $w_n/w_e \in [0.5, 0.9]$) of the ML and the exogenous policies as functions of γ_{ave} for three ratio levels $w_n/w_r = 1, 0.5$, and 0 . Our numerical results indicate that in the settings where the average absenteeism rate γ_{ave} is small, ML and exogenous policies represent good approximations for the optimal staffing policies. For example, in the ED we studied, the average absenteeism rate was 7.34%, and the corresponding worst-case performances for these two policies were between 2% and 3% for $w_n = w_r$. However, as γ_{ave} increases, so does the worst-case performance gap for both policies: In particular, in the same setting, the worst-case performance gaps approximately double to 4% (6%) for the exogenous (ML) policy when the average absenteeism rate reaches 15%. As it turns out, a reduction in the amount of hourly compensation paid to absent nurses significantly closes these performance gaps for both policies: For $w_n = 0$, the maximum performance gaps drop below 1%. Figure 4 displays similarly defined performance gaps obtained by focusing on the average absenteeism rate of 7.34% we have observed, and by varying the value of β within the 95% confidence interval around the estimate $\hat{\beta} = 0.1398$ obtained in Model III, namely, $\beta \in [0.1398 - 1.96 \times 0.583, 0.1398 + 1.96 \times 0.583] = [0.0255, 0.2541]$. The values of the performance gaps in Figure 4 are similar to those in Figure 3, displaying, as expected, a substantial drop as the value of β shifts toward the lower limit of the confidence interval.

4. Discussion

Nurses constitute one of the most important resources in hospitals, both in terms of cost and clinical outcomes. Therefore, any insights into how to deploy

Figure 4 Worst-Case Performance Gaps Between the Optimal Staffing Policy and the Misspecified-with-Learning (ML) and Exogenous Policies as Functions of β for $\gamma_{ave} = 7.34\%$ Under the Empirical Targeted Nursing Level Distribution for $w_n/w_r = 1, 0.5, 0$



nurses more effectively are of great interest to hospital managers. This paper focuses on the issue of nurse absenteeism, a problem that has vexed hospital managers for a long time. Using data from a large urban hospital ED, we find that nurse absenteeism is exacerbated when fewer nurses are scheduled for a particular shift. This finding highlights the need for hospital managers to use better methods to identify nurse staffing levels that are adequate to handle the anticipated workload. Our study relies on aggregate data from a single department and thus does not permit a detailed investigation of the impact of workload at the individual nurse level. We leave such an extension to future research. It is, nevertheless, the first study to demonstrate that staffing decisions have an impact on shift-level worker absenteeism—a fact that seems not to have been examined in prior staffing literature.

We analytically examine the implications of absenteeism, both exogenous and endogenous, on optimal staffing policies. To do this, we develop an extension to the single-period newsvendor model, which explicitly accounts for uncertainty in patient census and in the number of nurses that show up for work. Our work suggests that the presence of endogenous absenteeism gives rise to systematic understaffing, which, in turn, has important practical consequences for hospitals. First, as our model demonstrates, endogenous absenteeism gives rise to noticeable cost increases even in settings with low absenteeism rates, as long as absenteeism exhibits a substantial degree of endogeneity and as long as monetary compensation for absent nurses is comparable to that of nurses who show up for work. For model parameters that represent the hospital we study, we find that the cost of ignoring the endogenous nature of absenteeism can be about 2% to 3%. Second, such chronic understaffing harms patients, especially in the life-and-death setting of an ED. Third, it is likely to be a contributing factor to widely reported nurse job dissatisfaction (Aiken et al. 2002). Our research points to an important opportunity for cash-constrained hospitals to improve quality of patient care as well as nurse working conditions, while reducing operating costs.

Turning to the specific context of our analysis, note that our assumption about the unlimited availability of extra/overtime nursing capacity may not be valid in some clinical environments. In such environments it may be impossible to replace absent nurses at a reasonable cost or in reasonable time, and the endogeneity of absenteeism can lead to significant understaffing with the possibility of serious deterioration of service quality and longer ED delays. In other clinical units the use of agency nurses who may be less familiar with the unit can lead to similar declines in quality of patient care and an increase in the rate of medical errors. Thus, an accurate understanding of the nature of nurse absenteeism and the use of a model that accurately incorporates this phenomenon in determining appropriate staffing levels is imperative to ensuring high-quality patient care.

Other than adjusting staffing levels, endogenous absenteeism may be addressed with a number of complementary initiatives. Of particular value is the use of methodologies that lead to better matching of supply and demand through more effective allocation of nurses across units/shifts (Wang and Gupta 2012) and more accurate forecasting of patient census. Also of value are interventions that target the organizational culture of burnout (Maslach et al. 2012). Future research should investigate and compare the efficacy and cost effectiveness of such interventions.

Acknowledgments

The authors thank Dr. Robert Green, associate director of Emergency Medicine, New York–Presbyterian Hospital, and Germaine Nelson, director of nursing, emergency room administration, New York–Presbyterian Hospital, for their cooperation in obtaining the required data and information to conduct this research.

Appendix

Below we present outlines of the proofs of the analytical results.

PROOF OF PROPOSITION 1. Note that

$$\begin{aligned} W_{ave}(y + 1) - W_{ave}(y) &= w + w_e(L(y + 1, \gamma_{ave}) - L(y, \gamma_{ave})) \\ &= w + w_e \Delta L(y, \gamma_{ave}). \end{aligned}$$

Downloaded from informs.org by [163.119.96.41] on 24 March 2016, at 11:01. For personal use only, all rights reserved.

We establish the result of the proposition by showing that $\Delta L(y, \gamma_{ave}) \leq 0$ and $\Delta^2 L(y, \gamma_{ave}) = L(y + 2, \gamma_{ave}) - 2L(y + 1, \gamma_{ave}) + L(y, \gamma_{ave}) \geq 0$. First, note that $q(k + 1) - q(k) = -\sum_{n=k+1}^{\infty} p_T(n) \leq 0$ and $q(k + 2) - 2q(k + 1) + q(k) = p_T(k + 1) \geq 0$. Then, $\Delta L(y, \gamma_{ave}) = \sum_{k=0}^{y+1} q(k)(p(k; y + 1, \gamma_{ave}) - p(k; y, \gamma_{ave}))$, where we have used $p(y + 1; y, \gamma_{ave}) \equiv 0$. Next, using

$$p(k; y + 1, \gamma_{ave}) = (1 - \gamma_{ave})p(k - 1; y, \gamma_{ave}) + \gamma_{ave}p(k; y, \gamma_{ave}), \quad k = 0, \dots, y, \quad (27)$$

we get

$$\begin{aligned} & \sum_{k=0}^{y+1} q(k)(p(k; y + 1, \gamma_{ave}) - p(k; y, \gamma_{ave})) \\ &= (1 - \gamma_{ave}) \sum_{k=0}^y (q(k + 1) - q(k))p(k; y, \gamma_{ave}) \leq 0, \end{aligned} \quad (28)$$

where we have used

$$\sum_{k=0}^{y+1} q(k)p(k - 1; y, \gamma_{ave}) = \sum_{k=0}^{y+1} q(k + 1)p(k; y, \gamma_{ave}).$$

Note that $p(k - 1; y, \gamma_{ave}) \equiv 0$. From $p(y + 2; y, \gamma_{ave}) \equiv 0$ and (27), we get

$$\begin{aligned} \Delta^2 L(y, \gamma_{ave}) &= (1 - \gamma_{ave})^2 \left(\sum_{k=0}^{y+2} (q(k + 1) - q(k)) \right. \\ &\quad \left. \cdot (p(k - 1; y, \gamma_{ave}) - p(k; y, \gamma_{ave})) \right). \end{aligned} \quad (29)$$

Note that

$$\begin{aligned} & \sum_{k=0}^{y+2} (q(k + 1) - q(k))p(k - 1; y, \gamma_{ave}) \\ &= \sum_{k=0}^{y+2} (q(k + 2) - q(k + 1))p(k; y, \gamma_{ave}), \end{aligned}$$

where we have used $p(-1; y, \gamma_{ave}) \equiv p(y + 1; y, \gamma_{ave}) \equiv p(y + 2; y, \gamma_{ave}) \equiv 0$. Thus, we obtain for

$$\begin{aligned} \Delta^2 L(y, \gamma_{ave}) &= (1 - \gamma_{ave})^2 \left(\sum_{k=0}^y (q(k + 2) - 2q(k + 1) + q(k)) \right. \\ &\quad \left. \cdot p(k; y, \gamma_{ave}) \right) \geq 0. \end{aligned}$$

Furthermore, note that $\Delta L(y, \gamma_{ave}) \geq -w/w_e$ is equivalent to

$$\sum_{k=0}^y F_T(k)p(k; y, \gamma_{ave}) \geq 1 - w/(w_e(1 - \gamma_{ave})).$$

Now, consider $F_T^1(k)$ and $F_T^2(k)$ such that $F_T^1(k) \geq F_T^2(k)$ for any $k \in \mathcal{N}^+$. Then, $\Delta L^1(y, \gamma_{ave}) \geq \Delta L^2(y, \gamma_{ave})$ for any $y \in \mathcal{N}^+$ and, respectively, $y_{ave}^{*,1} = \min(y \in \mathcal{N}^+ \mid \Delta L^1(y, \gamma_{ave}) \geq -w/w_e) \leq \min(y \in \mathcal{N}^+ \mid \Delta L^2(y, \gamma_{ave}) \geq -w/w_e) = y_{ave}^{*,2}$. \square

PROOF OF PROPOSITION 2. First, we introduce $G(y, \gamma) = \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma)$ and note that

$$\begin{aligned} G(y + 1, \gamma) &= (1 - \gamma) \sum_{k=1}^{y+1} (1 - F_T(k))p(k - 1; y, \gamma) \\ &\quad + \gamma \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma), \end{aligned} \quad (30)$$

where we have used (27). Then,

$$\begin{aligned} G(y + 1, \gamma) &\leq (1 - \gamma) \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma) \\ &\quad + \gamma \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma) = G(y, \gamma). \end{aligned} \quad (31)$$

Furthermore,

$$\begin{aligned} \frac{\partial G}{\partial \gamma} &= y \left(\sum_{k=0}^{y-1} (1 - F_T(k)) \frac{(y - 1)!}{k!(y - 1 - k)!} (\gamma^{y-1-k} (1 - \gamma)^k) \right) \\ &\quad - y \left(\sum_{k=1}^{y-1} (1 - F_T(k)) \frac{y!}{(k - 1)!(y - 1 - (k - 1))!} \right. \\ &\quad \left. \cdot (\gamma^{y-1-(k-1)} (1 - \gamma)^{k-1}) \right) \\ &= y \left(\sum_{k=0}^{y-1} (F_T(k + 1) - F_T(k))p(k; y - 1, \gamma) \right) \geq 0. \end{aligned} \quad (32)$$

The optimality condition (10) can be rewritten as

$$y_{ave}^* = \min \left(y \in \mathcal{N}^+ \mid G(y, \gamma_{ave}) \leq \frac{w_r - w_n}{w_e} + \frac{w_n}{w_e(1 - \gamma_{ave})} \right).$$

Since y is a discrete variable and γ_{ave} is a continuous parameter, there exist a set of values $\gamma_{ave}^i \in [0, 1]$, $i = 1, \dots, I_{max}$, with $\gamma_{ave}^1 = 0$ and $\gamma_{ave}^{I_{max}} = 1$, with the optimal $y_{ave}^*(\gamma_{ave})$ remaining constant in each interval $(\gamma_{ave}^i, \gamma_{ave}^{i+1})$, $i = 1, \dots, I_{max} - 1$, and exhibiting finite jumps at γ_{ave}^i , i.e., $y_{ave}^*(\gamma_{ave}^i - 0) \neq y_{ave}^*(\gamma_{ave}^i + 0)$, $i = 2, \dots, I_{max} - 1$. Note that the sign of the difference $y_{ave}^*(\gamma_{ave}^i + 0) - y_{ave}^*(\gamma_{ave}^i - 0)$ is completely determined by the sign of

$$\begin{aligned} H(\gamma_{ave}^i) &= \frac{\partial G(\min(y_{ave}^*(\gamma_{ave}^i - 0), y_{ave}^*(\gamma_{ave}^i + 0)), \gamma_{ave}^i)}{\partial \gamma_{ave}^i} \\ &\quad - \frac{w_n}{w_e(1 - \gamma_{ave}^i)^2}. \end{aligned} \quad (33)$$

Indeed, suppose that $y_{ave}^*(\gamma_{ave}^i + 0) > y_{ave}^*(\gamma_{ave}^i - 0)$. Then, we should have $G(y_{ave}^*(\gamma_{ave}^i - 0), \gamma_{ave}^i - 0) \leq (w_r - w_n)/w_e + w_n/(w_e(1 - (\gamma_{ave}^i - 0)))$ and $G(y_{ave}^*(\gamma_{ave}^i - 0), \gamma_{ave}^i + 0) > (w_r - w_n)/w_e + w_n/(w_e(1 - (\gamma_{ave}^i + 0)))$. Combining these two expressions, we get

$$\begin{aligned} & G(y_{ave}^*(\gamma_{ave}^i - 0), \gamma_{ave}^i + 0) - G(y_{ave}^*(\gamma_{ave}^i - 0), \gamma_{ave}^i - 0) \\ & > \frac{w_n}{w_e(1 - (\gamma_{ave}^i + 0))} - \frac{w_n}{w_e(1 - (\gamma_{ave}^i - 0))} \end{aligned} \quad (34)$$

or

$$\frac{\partial G(y_{ave}^*(\gamma_{ave}^i - 0), \gamma_{ave}^i - 0)}{\partial \gamma_{ave}^i} > \frac{w_n}{w_e(1 - (\gamma_{ave}^i - 0))^2}.$$

Similarly, $y_{ave}^*(\gamma_{ave}^i + 0) < y_{ave}^*(\gamma_{ave}^i - 0)$ implies that

$$\frac{\partial G(y_{ave}^*(\gamma_{ave}^i + 0), \gamma_{ave}^i - 0)}{\partial \gamma_{ave}^i} < \frac{w_n}{w_e(1 - (\gamma_{ave}^i - 0))^2}.$$

Note that according to (32),

$$\frac{\partial G(y_{ave}^*, \gamma_{ave}^i)}{\partial \gamma_{ave}^i} = y_{ave}^* \left(\sum_{k=0}^{y_{ave}^*-1} p_T(k + 1)p(k; y_{ave}^* - 1, \gamma_{ave}^i) \right),$$

so that (33) can be expressed as

$$H(\gamma_{ave}^i) = \hat{y} \left(\sum_{k=0}^{\hat{y}-1} p_T(k+1)p(k; \hat{y}-1, \gamma_{ave}^i) \right) - \frac{w_n}{w_e(1-\gamma_{ave}^i)^2},$$

where $\hat{y} = \min(y_{ave}^*(\gamma_{ave}^i - 0), y_{ave}^*(\gamma_{ave}^i + 0))$. Note that for $\gamma_{ave}^i \rightarrow 0, y_{ave}^*(\gamma_{ave}^i) \rightarrow y_0^*$, as expressed in (7), so that $y_0^* = [F_T^{-1}(1 - w_r/w_e)]$. Then,

$$H(\gamma_{ave}^i \rightarrow 0) = \left([F_T^{-1}(1 - \frac{w_r}{w_e})] \right) p_T \left([F_T^{-1}(1 - \frac{w_r}{w_e})] \right) - \frac{w_n}{w_e}.$$

Thus,

$$\left([F_T^{-1}(1 - \frac{w_r}{w_e})] \right) p_T \left([F_T^{-1}(1 - \frac{w_r}{w_e})] \right) \geq \frac{w_n}{w_e}$$

implies that there exist γ_{ave}^l such that y_{ave}^* is a nondecreasing function of γ_{ave} for $\gamma_{ave} \leq \gamma_{ave}^l$. On the other hand, $\gamma_{ave}^i \geq (w_e - w_r)/(w_n + w_e - w_r)$ implies that $y_{ave}^*(\gamma_{ave}^i) = 0$, and for $\gamma_{ave}^i \rightarrow (w_e - w_r)/(w_n + w_e - w_r), y_{ave}^*(\gamma_{ave}^i) \rightarrow 0$, and $H(\gamma_{ave}^i \rightarrow (w_e - w_r)/(w_n + w_e - w_r)) = -(w_n + w_e - w_r)^2/(w_e w_n) < 0$. Thus, there exist γ_{ave}^u such that y_{ave}^* is a nonincreasing function of γ_{ave} for $\gamma_{ave} \geq \gamma_{ave}^u$. □

PROOF OF COROLLARY 1. Under the discrete uniform demand distribution specified by (12), the sum in the expression for the optimal staffing level (10) becomes

$$\sum_{k=0}^y F_T(k)p(k; y, \gamma_{ave}) = \begin{cases} \frac{y(1-\gamma_{ave})+1}{T_{max}+1} & \text{for } y \leq T_{max}, \\ \frac{1}{T_{max}+1} \sum_{k=0}^{T_{max}} (k+1)p(k; y, \gamma_{ave}) & \\ + \sum_{k=T_{max}+1}^y p(k; y, \gamma_{ave}) & \text{for } y \geq T_{max} + 1. \end{cases} \quad (35)$$

Note that for $y = T_{max}$, (35) becomes

$$\frac{T_{max}(1-\gamma_{ave})+1}{T_{max}+1} = 1 - \frac{\gamma_{ave} T_{max}}{T_{max}+1} \geq 1 - \frac{w}{w_e(1-\gamma_{ave})},$$

as long as $w/w_e \geq \gamma_{ave}(1-\gamma_{ave})T_{max}/(T_{max}+1)$. The supremum of the right-hand side of this expression is $\frac{1}{4}$ (for $\gamma_{ave} = 0.5$ and $T_{max} \rightarrow \infty$), so that this expression is implied by $w/w_e \geq \frac{1}{4}$. Thus, under this condition, the optimal staffing level does not exceed T_{max} and, consequently,

$$y_{ave}^* = \min \left(y \in \mathcal{N}^+ \mid \frac{y(1-\gamma_{ave})+1}{T_{max}+1} \geq 1 - \frac{w}{w_e(1-\gamma_{ave})} \right) = \left\lceil \left(\frac{T_{max}}{1-\gamma_{ave}} - \frac{T_{max}+1}{(1-\gamma_{ave})^2} \frac{w}{w_e} \right) \right\rceil.$$

Furthermore, differentiating the expression under the "ceiling" function on the right-hand side with respect to γ_{ave} , we get

$$\frac{1}{(1-\gamma_{ave})^3} \left(\left(T_{max} - (T_{max}+1) \frac{w_r - w_n}{w_e} \right) (1-\gamma_{ave}) - 2(T_{max}+1) \frac{w_r}{w_e} \right),$$

which is nonnegative (nonpositive) if and only if $\gamma_{ave} \leq \gamma_{ave}^u$ ($\gamma_{ave} \geq \gamma_{ave}^l$). □

PROOF OF PROPOSITION 3. Using $L(y, \gamma(y)) = \sum_{k=0}^y q(k) \cdot p(k; y, \gamma(y))$, we have $W(y+1) - W(y) = w_r - (w_r - w_n) \cdot \Delta a(y) + w_e \Delta L(y, \gamma(y))$, where $\Delta L(y, \gamma(y)) = L(y+1, \gamma(y+1)) - L(y, \gamma(y))$, and $\Delta a(y) = a(y+1) - a(y)$. We proceed by identifying sufficient conditions for $\Delta L(y, \gamma(y)) \leq 0, \Delta^2 L(y, \gamma(y)) = \Delta L(y+1, \gamma(y+1)) - \Delta L(y, \gamma(y)) \geq 0$, and $(w_r - w_n) \Delta^2 a(y) = (w_r - w_n)(\Delta a(y+1) - \Delta a(y)) \leq 0$. Since $\gamma(y)$ is continuous and twice differentiable, it follows immediately that $a(y)$ is also continuous and twice differentiable, and therefore a sufficient condition for $(w_r - w_n)(\Delta a(y+1) - \Delta a(y)) \leq 0$ is given by $(w_r - w_n)(d^2/dy^2)a(y) \leq 0$. Now, $\Delta L(y, \gamma(y))$ can be written as

$$\sum_{k=0}^{y+1} q(k)p(k; y+1, \gamma(y+1)) - \sum_{k=0}^y q(k)p(k; y, \gamma(y+1)) + \sum_{k=0}^y q(k)p(k; y, \gamma(y+1)) - \sum_{k=0}^y q(k)p(k; y, \gamma(y)). \quad (36)$$

We are now going to consider separately the first two and the last two terms in Equation (36) and show that both are nonpositive. The first two terms can be expressed as $-(1 - \gamma(y+1)) \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma(y+1))$, which is nonpositive. Next, we examine the second two terms, which can be expressed as

$$\sum_{k=0}^y q(k) \int_y^{y+1} \frac{\partial p(k; y, \gamma(s))}{\partial s} ds = \int_y^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^y q(k) \frac{\partial p(k; y, \gamma(s))}{\partial \gamma(s)}. \quad (37)$$

Note that

$$\begin{aligned} & \sum_{k=0}^y q(k) \frac{\partial p(k; y, \gamma(s))}{\partial \gamma(s)} \\ &= \sum_{k=0}^y q(k) \left(\frac{y!}{k!(y-k)!} (y-k)\gamma(s)^{y-k-1} (1-\gamma(s))^k - k\gamma(s)^{y-k} (1-\gamma(s))^{k-1} \right) \\ &= y \sum_{k=0}^{y-1} q(k)p(k; y-1, \gamma(s)) \\ & \quad - y \sum_{k=1}^y q(k)p(k-1; y-1, \gamma(s)) \\ &= y \sum_{k=0}^{y-1} (1 - F_T(k))p(k; y-1, \gamma(s)), \end{aligned} \quad (38)$$

so that (37) becomes

$$\begin{aligned} & \sum_{k=0}^y q(k)(p(k; y, \gamma(y+1)) - p(k; y, \gamma(y))) \\ &= y \int_y^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1 - F_T(k))p(k; y-1, \gamma(s)). \end{aligned}$$

A sufficient condition for this last expression to be negative is $d\gamma(s)/ds \leq 0$. Thus, $\Delta L(y, \gamma(y)) \leq 0$. Furthermore, consider $\Delta^2 L(y, \gamma(y))$, expressed as

$$\begin{aligned} & -(1 - \gamma(y + 2)) \sum_{k=0}^{y+1} (1 - F_T(k)) p(k; y + 1, \gamma(y + 2)) \\ & + (1 - \gamma(y + 1)) \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(y + 1)) \\ & + (y + 1) \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(s)) \\ & - y \int_y^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1 - F_T(k)) p(k; y - 1, \gamma(s)). \end{aligned} \quad (39)$$

Focusing on the first two terms in (39), we get

$$\begin{aligned} & -(1 - \gamma(y + 2)) \sum_{k=0}^{y+1} (1 - F_T(k)) p(k; y + 1, \gamma(y + 2)) \\ & + (1 - \gamma(y + 1)) \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(y + 1)) \\ & = (1 - \gamma(y + 2)) \left(\sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(y + 2)) \right. \\ & \quad \left. - \sum_{k=0}^{y+1} (1 - F_T(k)) p(k; y + 1, \gamma(y + 2)) \right) \\ & + \sum_{k=0}^y (1 - F_T(k)) ((1 - \gamma(y + 1)) p(k; y, \gamma(y + 1)) \\ & \quad - (1 - \gamma(y + 2)) p(k; y, \gamma(y + 2))). \end{aligned} \quad (40)$$

The first term in (40) is equivalent to

$$(1 - \gamma(y + 2))^2 \left(\sum_{k=0}^y p_T(k + 1) p(k; y, \gamma(y + 2)) \right)$$

and is nonnegative. The second term in (40) is equal to

$$- \sum_{k=0}^y (1 - F_T(k)) \int_{y+1}^{y+2} \frac{\partial((1 - \gamma(s)) p(k; y, \gamma(s)))}{\partial s} ds$$

or

$$- \sum_{k=0}^y (1 - F_T(k)) \int_{y+1}^{y+2} \frac{\partial((1 - \gamma(s)) p(k; y, \gamma(s)))}{\partial \gamma(s)} \frac{d\gamma(s)}{ds} ds.$$

Focusing on the last two terms in (39), we obtain

$$\begin{aligned} & (y + 1) \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(s)) \\ & - y \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1 - F_T(k)) p(k; y - 1, \gamma(s)) \\ & + y \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1 - F_T(k)) p(k; y - 1, \gamma(s)) \\ & - y \int_y^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1 - F_T(k)) p(k; y - 1, \gamma(s)) \end{aligned}$$

$$\begin{aligned} & = \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^y (1 - F_T(k)) \\ & \quad \cdot ((y + 1) p(k; y, \gamma(s)) - y p(k; y - 1, \gamma(s))) \\ & \quad + y \int_{y+1}^{y+2} ds \int_{s-1}^s d\xi \left[\sum_{k=0}^{y-1} (1 - F_T(k)) \frac{\partial}{\partial \xi} \right. \\ & \quad \left. \cdot \left[\frac{d\gamma(\xi)}{d\xi} p(k; y - 1, \gamma(\xi)) \right] \right]. \end{aligned} \quad (41)$$

Thus,

$$\begin{aligned} & \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^y (1 - F_T(k)) \left(- \frac{\partial(1 - \gamma(s)) p(k; y, \gamma(s))}{\partial \gamma(s)} \right. \\ & \quad \left. + (y + 1) p(k; y, \gamma(s)) - y p(k; y - 1, \gamma(s)) \right) \\ & \quad + y \int_{y+1}^{y+2} ds \int_{s-1}^s d\xi \left[\sum_{k=0}^{y-1} (1 - F_T(k)) \right. \\ & \quad \left. \cdot \frac{\partial}{\partial \xi} \left[\frac{d\gamma(\xi)}{d\xi} p(k; y - 1, \gamma(\xi)) \right] \right] \\ & = -2 \int_{y+1}^{y+2} \frac{ds}{\gamma(s)} \frac{d\gamma(s)}{ds} \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(s)) \\ & \quad \cdot (y(1 - \gamma(s)) - k - \gamma(s)) \\ & \quad + y \int_{y+1}^{y+2} ds \int_{s-1}^s d\xi \sum_{k=0}^{y-1} (1 - F_T(k)) \frac{d^2 \gamma(\xi)}{d\xi^2} \\ & \quad + y \int_{y+1}^{y+2} ds \int_{s-1}^s d\xi \times \left(\frac{d\gamma}{d\xi} \right)^2 \frac{1}{\gamma(\xi)(1 - \gamma(\xi))} \\ & \quad \cdot \left[\sum_{k=0}^{y-1} (1 - F_T(k)) p(k; y - 1, \gamma(\xi)) \right. \\ & \quad \left. \cdot ((y - 1)(1 - \gamma(\xi)) - k) \right]. \end{aligned} \quad (42)$$

A sufficient condition for the second line of (42) to be nonnegative is $d^2 \gamma(\xi)/d\xi^2 \geq 0$. Since $d\gamma(s)/ds \leq 0$, a sufficient condition for the first line of (42) to be nonnegative is that, for given y ,

$$\sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma) (y(1 - \gamma) - k - \gamma) \geq 0, \quad (43)$$

for any $\gamma \in [\gamma(y + 2), \gamma(y + 1)]$, and for the third line of (42) is that for given y , $\sum_{k=0}^{y-1} (1 - F_T(k)) p(k; y - 1, \gamma) \cdot ((y - 1)(1 - \gamma) - k) \geq 0$, for any $\gamma \in [\gamma(y), \gamma(y + 1)]$. This last condition is equivalent to the condition, for given y , $\sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma) (y(1 - \gamma) - k) \geq 0$, for any $\gamma \in [\gamma(y + 1), \gamma(y + 2)]$. We next derive sufficient a condition for (43). Note that

$$\begin{aligned} & \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma) (y(1 - \gamma) - k - \gamma) \\ & = \sum_{k=0}^{y-2} (1 - F_T(k)) p(k; y, \gamma) (y(1 - \gamma) - k - \gamma) \\ & \quad + (1 - F_T(y - 1)) \gamma y (1 - \gamma)^{y-1} (1 - (y + 1) \gamma) \\ & \quad + (1 - F_T(y)) (1 - \gamma)^y (-(y + 1) \gamma), \end{aligned} \quad (44)$$

so that, for $(y + 1)\gamma \leq 2$, $y(1 - \gamma) - k - \gamma \geq 0$ for all $k = 0, \dots, y - 2$, and

$$\begin{aligned} & \sum_{k=0}^{y-2} (1 - F_T(k))p(k; y, \gamma)(y(1 - \gamma) - k - \gamma) \\ & \geq (1 - F_T(y - 2)) \sum_{k=0}^{y-2} p(k; y, \gamma)(y(1 - \gamma) - k - \gamma) \\ & = (1 - F_T(y - 2)) \left(\sum_{k=0}^y p(k; y, \gamma)(y(1 - \gamma) - k - \gamma) \right. \\ & \quad \left. - \gamma y(1 - \gamma)^{y-1}(1 - (y + 1)\gamma) \right. \\ & \quad \left. - (1 - \gamma)^y(-(y + 1)\gamma) \right) \\ & = (1 - F_T(y - 2))(-\gamma - \gamma y(1 - \gamma)^{y-1}(1 - (y + 1)\gamma) \\ & \quad - (1 - \gamma)^y(-(y + 1)\gamma)). \quad (45) \end{aligned}$$

Thus, the expression in (44) is nonnegative if $(F_T(y) - F_T(y - 2))(1 - \gamma)^y(y + 1) + (F_T(y - 1) - F_T(y - 2))y(1 - \gamma)^{y-1} \cdot ((y + 1)\gamma - 1) \geq 1 - F_T(y - 2)$. The left-hand side of this expression can be rearranged as

$$\begin{aligned} & (F_T(y) - F_T(y - 2))(1 - \gamma)^y(y + 1) \\ & \quad + (F_T(y - 1) - F_T(y - 2))y(1 - \gamma)^{y-1}((y + 1)\gamma - 1) \\ & = (p_T(y) + p_T(y - 1))(1 - \gamma)^y(y + 1) \\ & \quad + p_T(y - 1)y(1 - \gamma)^{y-1}((y + 1)\gamma - 1) \\ & = p_T(y)(1 - \gamma)^y(y + 1) + p_T(y - 1)(1 - \gamma)^y \left(1 + \frac{y^2\gamma}{1 - \gamma} \right) \\ & \geq (1 - \gamma)^y(p_T(y)(y + 1) + p_T(y - 1)). \quad (46) \end{aligned}$$

Thus, $(F_T(y) - F_T(y - 2))(1 - \gamma)^y(y + 1) + (F_T(y - 1) - F_T(y - 2))y(1 - \gamma)^{y-1}((y + 1)\gamma - 1) \geq 1 - F_T(y - 2)$ as long as $(1 - \gamma)^y(p_T(y)(y + 1) + p_T(y - 1)) \geq 1 - F_T(y - 2)$ or

$$\gamma \leq 1 - \left(\frac{1 - F_T(y - 2)}{(y + 1)p_T(y) + p_T(y - 1)} \right)^{1/y}.$$

Combining this expression with $(y + 1)\gamma \leq 2$ and $\gamma \in [\gamma(y + 2), \gamma(y + 1)]$, and noting that for $\gamma = 0$ the monotonicity of $\Delta L(y, \gamma(y))$ is assured, we obtain the final sufficient condition

$$\begin{aligned} \gamma(y + 1) & \leq \min \left(\frac{2}{y + 1}, \right. \\ & \quad \left. 1 - \min \left(1, \left(\frac{1 - F_T(y - 2)}{(y + 1)p_T(y) + p_T(y - 1)} \right)^{1/y} \right) \right) \\ & = \min \left(\frac{2}{y + 1}, \gamma_T(y + 1) \right). \quad (47) \end{aligned}$$

Given that $\Delta L(y, \gamma(y))$ is a monotone function of y if $\gamma(y)$ is a nonincreasing convex function of y and if (47) is satisfied, we establish the monotonicity of the optimal staffing level y^* with respect to changes in w_r/w_e and $F_T(k)$, following the same arguments used in the proof of Proposition 1. \square

PROOF OF LEMMA 1. Note that for $\alpha, \beta \geq 0$, the function $a(x) = x\gamma(x) = x/(1 + e^{\alpha + \beta x})$ defined on continuous set $x \geq 0$ has a unique global maximum x^* , which satisfies the first-order optimality condition $e^{-\alpha} = e^{\beta x^*}(\beta x^* - 1)$. Thus, the maximum value for this function can be expressed as $x^*\gamma(x^*) = (1/\beta)(\beta x^*/(1 + e^{\alpha + \beta x^*})) = (\beta x^* - 1)/\beta$. This last expression does not exceed 2 if and only if $\beta x^* \leq 2\beta + 1$, which is equivalent to $e^{-\alpha} \leq 2\beta e^{2\beta + 1} \Leftrightarrow 2\beta e^{\alpha + 2\beta + 1} \geq 1$. This condition ensures that the maximum of $y\gamma(y)$ cannot exceed 2 for all integer values of y as well. Furthermore, $d^2a/dy^2 = \beta^2\gamma(y)(1 - \gamma(y))(\beta(1 - 2\gamma(y)) - 2)$, which is always nonpositive for $\beta \leq 2$. \square

PROOF OF PROPOSITION 4. Under (20), the optimal staffing level y^* satisfies (19), which can be re-expressed as $y^* = \min(y \in \mathcal{N}^+ \mid w_r(1 - \gamma(y)) + w_n\gamma(y) - (w_r - w_n)y(\gamma(y + 1) - \gamma(y)) + w_e\Delta L(y, \gamma(y)) \geq 0)$, where the last two terms inside the bracket are

$$\begin{aligned} & -(w_r - w_n)y(\gamma(y + 1) - \gamma(y)) - w_e(1 - \gamma(y + 1)) \\ & \quad \cdot \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma(y + 1)) \\ & \quad + w_e y \int_y^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1 - F_T(k))p(k; y - 1, \gamma(s)) \\ & = -w_e(1 - \gamma(y + 1)) \sum_{k=0}^y (1 - F_T(k))p(k; y, \gamma(y + 1)) \\ & \quad + y \int_y^{y+1} ds \frac{d\gamma(s)}{ds} \left[w_e \left(\sum_{k=0}^{y-1} (1 - F_T(k))p(k; y - 1, \gamma(s)) \right) \right. \\ & \quad \left. - (w_r - w_n) \right]. \quad (48) \end{aligned}$$

Thus,

$$\begin{aligned} & w_r(1 - \gamma(y^*)) + w_n\gamma(y^*) - w_e(1 - \gamma(y^* + 1)) \\ & \quad \cdot \sum_{k=0}^{y^*} (1 - F_T(k))p(k; y^*, \gamma(y^* + 1)) \\ & \quad + y^* \int_{y^*}^{y^*+1} ds \frac{d\gamma(s)}{ds} \left[w_e \left(\sum_{k=0}^{y^*-1} (1 - F_T(k))p(k; y^* - 1, \gamma(s)) \right) \right. \\ & \quad \left. - (w_r - w_n) \right] \geq 0 \end{aligned}$$

or

$$\begin{aligned} & \sum_{k=0}^{y^*} (1 - F_T(k))p(k; y^*, \gamma(y^* + 1)) \\ & \leq \left(w_r - (w_r - w_n)\gamma(y^*) \right. \\ & \quad \left. + y^* \int_{y^*}^{y^*+1} ds \frac{d\gamma(s)}{ds} \right. \\ & \quad \left. \cdot \left[w_e \left(\sum_{k=0}^{y^*-1} (1 - F_T(k))p(k; y^* - 1, \gamma(s)) \right) \right. \right. \\ & \quad \left. \left. - (w_r - w_n) \right] \right) (w_e(1 - \gamma(y^* + 1)))^{-1}. \quad (49) \end{aligned}$$

Now, consider an element of E , y^{ML} , which satisfies

$$y^{ML} = \min \left(y \in \mathcal{N}^+ \mid \sum_{k=0}^y (1 - F_T(k)) p(k; y, \gamma(y^{ML})) \leq \frac{w_r - (w_r - w_n) \gamma(y^{ML})}{w_e (1 - \gamma(y^{ML}))} \right),$$

so that

$$\sum_{k=0}^{y^{ML}} (1 - F_T(k)) p(k; y^{ML}, \gamma(y^{ML})) \leq \frac{w_r - (w_r - w_n) \gamma(y^{ML})}{w_e (1 - \gamma(y^{ML}))}.$$

Below we show by contradiction that, if $a(y^* + 1) < a(y^*)$, then $y^{ML} \leq y^*$. Suppose that $y^{ML} > y^* \Leftrightarrow y^{ML} \geq y^* + 1$. This, in turn, implies that $\gamma(y^{ML}) \leq \gamma(y^* + 1)$ and

$$\begin{aligned} & \sum_{k=0}^{y^*} (1 - F_T(k)) p(k; y^*, \gamma(y^{ML})) \\ & > \frac{w_r - (w_r - w_n) \gamma(y^{ML})}{w_e (1 - \gamma(y^{ML}))}. \end{aligned} \quad (50)$$

Now, since for $\forall s \in [y^*, y^* + 1]$, it follows that $s \leq y^{ML}$ and $\gamma(y^{ML}) \leq \gamma(s)$, we can use (31), (32), and (50) to get

$$\begin{aligned} & \sum_{k=0}^{y^*-1} (1 - F_T(k)) p(k; y^* - 1, \gamma(s)) \\ & \geq \sum_{k=0}^{y^*} (1 - F_T(k)) p(k; y^*, \gamma(s)) \\ & \geq \sum_{k=0}^{y^*} (1 - F_T(k)) p(k; y^*, \gamma(y^{ML})) \\ & > \frac{w_r - (w_r - w_n) \gamma(y^{ML})}{w_e (1 - \gamma(y^{ML}))}. \end{aligned} \quad (51)$$

Furthermore, since $d\gamma(y)/dy \leq 0$, (49) implies that

$$\begin{aligned} & \frac{w_r - (w_r - w_n) \gamma(y^{ML})}{w_e (1 - \gamma(y^{ML}))} \\ & < \sum_{k=0}^{y^*} (1 - F_T(k)) p(k; y^*, \gamma(y^* + 1)) \\ & \leq \left(w_r - (w_r - w_n) \gamma(y^*) \right. \\ & \quad \left. + y^* \int_{y^*}^{y^*+1} ds \frac{d\gamma(s)}{ds} \right. \\ & \quad \left. \cdot \left[w_e \left(\sum_{k=0}^{y^*-1} (1 - F_T(k)) p(k; y^* - 1, \gamma(s)) \right) \right. \right. \\ & \quad \left. \left. - (w_r - w_n) \right] \right) (w_e (1 - \gamma(y^* + 1)))^{-1} \\ & \leq \frac{w_r - (w_r - w_n) \gamma(y^*) + y^* (\gamma(y^* + 1) - \gamma(y^*)) (w_n / (1 - \gamma(y^{ML})))}{w_e (1 - \gamma(y^* + 1))} \\ & \leq (w_r - (w_r - w_n) \gamma(y^* + 1) + y^* (\gamma(y^* + 1) - \gamma(y^*))) \\ & \quad \cdot (w_n / (1 - \gamma(y^{ML}))) \cdot (w_e (1 - \gamma(y^* + 1)))^{-1}. \end{aligned} \quad (52)$$

Equation (52) is equivalent to

$$\begin{aligned} \frac{\gamma(y^{ML})}{1 - \gamma(y^{ML})} & \leq \frac{\gamma(y^* + 1)}{1 - \gamma(y^* + 1)} \\ & \quad + \frac{y^* (\gamma(y^* + 1) - \gamma(y^*)) (1 / (1 - \gamma(y^{ML})))}{(1 - \gamma(y^* + 1))}, \end{aligned}$$

and

$$\begin{aligned} \gamma(y^{ML}) (1 - \gamma(y^* + 1)) & \leq \gamma(y^* + 1) (1 - \gamma(y^{ML})) \\ & \quad + y^* (\gamma(y^* + 1) - \gamma(y^*)), \end{aligned}$$

so that

$$\gamma(y^{ML}) \leq \gamma(y^* + 1) + y^* (\gamma(y^* + 1) - \gamma(y^*)).$$

Note that this contradicts $\gamma(y^{ML}) \geq 0$. \square

References

- Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH (2002) Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *J. Amer. Medical Assoc.* 288(16):1987–1993.
- Allenby GM, Rossi PE (1991) There is no aggregation bias: Why macro logit models work. *J. Bus. Econom. Statist.* 9(1):1–14.
- Bassamboo A, Randhawa RS, Zeevi A (2010) Capacity sizing under parameter uncertainty: Safety staffing principles revisited. *Management Sci.* 56(10):1668–1686.
- Bollapragada S, Morton TE (1999) Myopic heuristics for the random yield problem. *Oper. Res.* 47(5):713–722.
- Bryant C, Fairbrother G, Fenton P (2000) Professional issues: The relative influence of personal and workplace descriptors on stress. *British J. Nursing* 9(13):876–880.
- Cachon G, Kök AG (2007) Implementation of the newsvendor model with clearance pricing: How to (and how not to) estimate a salvage value. *Manufacturing Service Oper. Management* 9(3):276–290.
- Cho SH, Ketefian S, Barkauskas VH, Smith DG (2003) The effects of nurse staffing on adverse outcomes, 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.
- Cooper WL, Homem de Mello T, Kleywegt AJ (2006) Models of the spiral down effect in revenue management. *Oper. Res.* 54(5):968–987.
- Easton FF, Goodale JC (2005) Schedule recovery: Unplanned absences in service operations. *Decision Sci.* 36(3):459–488.
- Erickson RJ, Nichols L, Ritter C (2000) Family influences on absenteeism: Testing an expanded process model. *J. Vocational Behav.* 57(2):246–272.
- Fadiloglu M, Berk E, Gurbuz MC (2008) Supplier diversification under binomial yield. *Oper. Res. Lett.* 36(5):539–542.
- Fry MJ, Magazine MJ, Rao US (2006) Firefighter staffing including temporary absences and wastage. *Oper. Res.* 54(2):353–365.
- Gordon S, Buchanan J, Bretherton T (2008) *Safety in Numbers: Nurse-to-Patient Ratios and the Future of Health Care* (Cornell University Press, Ithaca, New York).
- Greene WH (2005) *Econometric Analysis* (Pearson Education, Harrow Essex, UK).
- 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.
- Gupta D, Cooper WL (2005) Stochastic comparisons in production yield management. *Oper. Res.* 53(2):377–384.

- Harrison JM, Zeevi A (2005) A method for staffing large call centers based on stochastic fluid models. *Manufacturing Service Operations Management* 7(1):20–36.
- He B, Dexter F, Macario A, Zenios S (2012) The timing of staffing decisions in hospital operating rooms: Incorporating workload heterogeneity into the newsvendor problem. *Manufacturing Service Oper. Management* 14(1):99–114.
- Henig M, Gerchak Y (1990) The structure of periodic review policies in the presence of random yield. *Oper. Res.* 38(4):634–643.
- Hill JM, Trist EL (1955) Changes in accidents and other absences with length of service: A further study of their incidence and relation to each other in an iron and steel works. *Human Relations* 8(2):121–152.
- Hobfoll SE (1989) Conservation of resources. *Amer. Psychologist* 44(3):513–524.
- KC D, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Sci.* 55(9):1486–1498.
- Lee S, Homem-de Mello T, Kleywegt AJ (2012) Newsvendor-type models with decision-dependent uncertainty. *Math. Methods Oper. Res.* 76(2):189–221.
- Maslach C, Leiter MP, Jackson SE (2012) Making a significant difference with burnout interventions: Researcher and practitioner collaboration. *J. Organ. Behav.* 33(2):296–300.
- McVicar A (2003) Workplace stress in nursing: A literature review. *J. Adv. Nursing* 44(6):633–642.
- Mersereau AJ (2012) Demand estimation from censored observations with inventory record inaccuracy. Working paper, University of North Carolina, Chapel Hill.
- Miller HE, Pierskalla WP, Rath GJ (1976) Nurse scheduling using mathematical programming. *Oper. Res.* 24(5):857–870.
- 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.
- Nevidjon B, Erickson JI (2001) The nursing shortage: Solutions for the short and long term. *Online J. Issues in Nursing* 6(1):461–462.
- Parkes KR (1982) Occupational stress among student nurses: A natural experiment. *J. Appl. Psych.* 67(6):784–796.
- Powell A, Savin S, Savva N (2012) Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing Service Oper. Management* 14(4):512–528.
- Powell SC, Schultz KL (2004) Throughput in serial lines with state-dependent behavior. *Management Sci.* 50(8):1095–1105.
- Rauhala A, Kivimäki M, Fagerström L, Elovainio M, Virtanen M, Vahtera J, Rainio A-K, Ojaniemi K, Kinnunen J (2007) What degree of work overload is likely to cause increased sickness absenteeism among nurses? Evidence from the RAFAELA patient classification system. *J. Adv. Nursing* 57(3):286–295.
- Schultz KL, Juran DC, Boudreau JW (1999) The effects of low inventory on the development of productivity norms. *Management Sci.* 45(12):1664–1678.
- Seago JA (2001) Nurse staffing, models of care delivery and interventions. Shojania K, Duncan B, McDonald K, Wachter R, eds. *Making Health Care Safer: A Critical Analysis of Patient Safety Practices* (Agency for Healthcare Research and Quality, Rockville, MD).
- Shamian J, O'Brien-Pallas L, Thomson D, Alksnis C, Kerr MS (2003) Nurse absenteeism, stress and workplace injury: What are the contributing factors and what can/should be done about it? *Internat. J. Sociol. Soc. Policy* 23(8/9):81–103.
- Steers RM, Rhodes SR (1978) Major influences on employee attendance: A process model. *J. Appl. Psych.* 63(4):391–407.
- Tummers GER, Janssen PPM, Landeweerd A, Houkes I (2001) A comparative study of work characteristics and reactions between general and mental health nurses: A multi-sample analysis. *J. Adv. Nursing* 36(1):151–162.
- Unruh L, Joseph L, Strickland M (2007) Nurse absenteeism and workload: Negative effect on restraint use, incident reports and mortality. *J. Adv. Nursing* 60(6):673–681.
- U.S. Bureau of Labor Statistics (2008) Industry injury and illness data. Last accessed January 23, 2011, <http://www.bls.gov/iif/oshwc/osh/os/osnr0032.pdf>.
- Wang W-Y, Gupta D (2012) Nurse absenteeism and staffing strategies for hospital inpatient units. Working paper, San Jose State University, San Jose, CA.
- Whitt W (2006) Staffing a call center with uncertain arrival rate and absenteeism. *Production Oper. Management* 15(1):88–102.
- Wright PD, Bretthauer KM, Côté MJ (2006) Reexamining the nurse scheduling problem: Staffing ratios and nursing shortages. *Decision Sci.* 37(1):39–70.
- Yang S, Yang J, Abdel-Malek L (2007) Sourcing with random yields and stochastic demand: A newsvendor approach. *Comput. Oper. Res.* 34(12):3682–3690.
- Yankovic N, Green LV (2011) Identifying good nursing levels: A queueing approach. *Oper. Res.* 59(4):942–955.
- Yano CA, Lee HA (1995) Lot sizing with random yields: A review. *Oper. Res.* 43(2):311–334.