Health Care Manage Sci (2006) 9: 71–85
DOI 10.1007/s10729-006-6281-y

Incremental changes in the workforce to accommodate changes in demand

Jonathan F. Bard · Hadi W. Purnomo

Received: June 2005 / Accepted: October 2005 after two revisions

Abstract In many service organizations, rosters must be constructed weekly or monthly as demand and available personnel change. Once the permanent workforce is fixed, it may not be possible to alter its composition easily, implying that expensive contract labor may be the only option to cover shortages. With respect to nursing resources, this means calling in part-timers, casuals, or agency nurses on a daily basis, or hiring travelers for up to several months at a time.

This paper addresses the latter option and presents two models that can be used to solve what we call the *nurse addition problem*. The first was originally developed to solve the midterm preference scheduling problem and is based on a pattern-view formulation. The second is derived from a shift-view formulation and is solved with a branch-and-price algorithm. In either case, the objective is to hire up to some predetermined number of nurses and assign them midterm schedules that minimize the maximum amount of uncovered shifts per day in the planning horizon. Each roster selected for a new nurse must satisfy a set of hard constraints related to the total working hours, workstretches, time between shifts, and weekend requirements, and a set of soft constraints related to days-on and days-off patterns and transitions from one shift type to another. Extensive testing with data provided by a 400-bed hospital indicated that most instances could be solved in a matter of minutes.

Keywords Workforce planning · Rotational scheduling · Integer programming · Computational complexity · Branch and price

1. Introduction

Structuring a permanent workforce depends on many factors including the nature of demand, contractual and union restrictions, weekly operating hours, and skill requirements. Many service organizations that operate outside a normal 8-hour day often fix the size and composition of their workforce at suboptimal levels because of the difficulty they have in accounting for uncertain or aperiodic demand [1, 2]. Poor personnel planning can lead to an oversupply of workers with too much idle time, or an undersupply with an attendant loss of business. Whether it is a support hotline, airline reservation desks, a brokerage firm, or a hospital, management's goal is to find the best mix of employees so that demand is satisfied at minimum cost (e.g., see [3, 4]).

Rather than taking advantage of worker flexibility and cross-training, management often errs on the side of excess. While large inventories in manufacturing are used to buffer uncertain demand and poor planning, overstaffing in service is a common defense for dealing with surges in demand, customer complaints, and a need to meet productivity goals. Many studies have shown, however, that cross-training and the use of higher skilled workers to fill in for lower skilled workers when idle time exists in their schedules can yield substantial cost savings (e.g., see [5–7]). Different days-off

This work was supported in part by the National Science Foundation under grant DMI-0218701.

J.F. Bard (✉)
Graduate Program in Operations Research
& Industrial Engineering, 1 University Station C2200
The University of Texas,
Austin, Texas 78712-0292
e-mail: jbard@mail.utexas.edu

H.W. Purnomo
American Airlines, AMR Corp Headquarter HDQ1
Mail Drop 5358, Fort Worth, Texas 76155
e-mail:hadiwaskito@yahoo.com

policies, variable start times, and the use of part-time flexible workers are other management options that can yield tangible benefits. Out of habit, though, management typically follows overly restrictive practices, such as providing weekly schedules with uniform start times, without fully examining the cost consequences.

For most organizations, hiring and firing at will to better match a changing demand is not a realistic option. In industries in which there is ample labor, periodic downsizing and rehiring can lead to poor morale, high absenteeism, and shoddy workmanship. In industries in which there is a shortage of skilled labor, one of the primary goals is to hold on to those difficult-to-replace workers by putting them on overhead when demand drops off. From a practical point of view, employers are constantly faced with the problem of making marginal adjustments in their workforce as demand shifts. The purpose of this paper is to present a new model for supporting the adjustment process.

Our interest in this problem grew out a long-term project that was aimed at developing a nurse management system that is now in use at several hospitals in the United States and Great Britain. Because of the erratic nature of demand and a growing shortage of nurses [8, 9], hospitals rarely examine their permanent staffing needs, relying instead on overtime, floaters, and the use of outside resources such as agencies and casuals, to fill in from shift-to-shift when demand outstrips supply. In the nursing area, the problem is complicated by management's desire to accommodate individual preferences while balancing a variety of legally mandated coverage requirements and labor costs [10]. Once a schedule for a nurse is negotiated, it becomes fixed by contract and difficult to change, regardless of need.

In the next section, we give a more detailed description of the problem in terms of nursing services. Many organizations, though, face exactly the same issues. In Section 3, the preference scheduling approach is presented, beginning with our pattern-view formulation for midterm scheduling. This is followed in Section 4 with a cyclic model that decomposes by nurse. We show that the nurse addition problem is NP-hard and then summarize of our branch-and-price algorithm for finding solutions. Extensive test results are reported in Section 5 and conclusions are offered in Section 6. Over the last year, the algorithm has been incorporated in a commercial system that includes a midterm scheduling module and a daily adjustment module.

2. Background and problem statement

Although there has been abundant research on personnel scheduling, almost all efforts have focused on either shift scheduling or days-off scheduling with the objective of minimizing total cost (e.g., see [11–13]). These problems are the central components of midterm or tour scheduling (e.g., see [14]). In the last two decades, most of the research on nurse scheduling has concentrated on rostering with the aim of accommodating individual preferences, which take the form of requests to work specific shifts or to be given specific days off. Each can be defined by various rules related to the number of working hours, shift sequence patterns, or even nurse to patient ratios (see [15] and [16] for surveys). In what is termed *preference scheduling*, nurses are typically asked to sign up for shifts prior to the beginning of the planning horizon and then adjustments are made by the nurse manager to ensure that as much demand as possible is covered with the permanent staff. The outcome of the process is a midterm schedule for each nurse in the hospital.

Midterm scheduling fixes the work assignments for the permanent nursing staff for up to six weeks at a time. Each unit or ward generates its own rosters (schedules) independently using some measure of average demand as input. The model that we developed to solve this problem has the competing objectives of minimizing the total number of uncovered shifts (gaps) over the planning horizon and minimizing some measure of preference violations [17]. The solution algorithm, based on column generation, achieves this end with the help of an intelligent swapping heuristic that identifies good candidate rosters for each nurse.

An important feature of the algorithm is that it also tries to minimize the maximum gap between demand and supply over the planning horizon. By recognizing that there is an inherent tradeoff between the maximum gap and the number of preference violations, it is possible shrink the gap at the expense of solution quality. When personnel shortages exist, however, labor laws, hospital policies and contractual agreements do not in general permit an exact match of supply and demand. This means that there will always be days in which excess coverage exists on some shifts and shortages on others.

In an effort to reduce the gap, at least marginally, we would like to give the nurse manager several options that involve a combination of (1) altering the schedules of a subset of nurses and (2) hiring temporary staff, called *travelers*, for the upcoming scheduling period. Two approaches will be presented. The first is based on our model that emphasizes individual preferences and combines the two options; the second considers adding travelers only and is very similar to cyclic scheduling (e.g., see [18]).

The remaining issue concerns the nature of the daily assignments. In general, nurses are hired to work a specific shift or to rotate between two shifts. Full-timers contract for 72 or 80 hours in a two-week period while part-timers work some reduced percentage, say $\alpha\%$. Three rotational profiles are possible when only 8-hour shifts are permitted: day/night, day/evening, evening/night (a night shift is typically 11:00 p.m. to 7 a.m. the next day). A profile is defined by the "day ratio"; that is, the ratio of the number of day shifts to the total shifts in the first two cases, and the number of evening shifts to total shifts in the third. For example, if a full-time nurse (denoted by 1 FTE) has a day/night profile and a 50% day ratio, then she would ordinarily be assigned a total of (80 hours)/(8 hours/shift) = 10 shifts over two weeks which would be split evenly between day and night (5 each). If an 80% nurse (denoted by 0.8 FTE) had an evening/night profile and a 25% day ratio, then she would be assigned $0.8 \times 80/8 = 8$ shifts in a two-week period of which 25% or 2 would be evening and the remaining 6 would be night. A day ratio of 100% simply means that the nurse does not rotate.

In some hospitals, the rotational profile is a hard constraint; however, a one-shift discrepancy in either direction may be permitted with appropriate penalty. For example, a day/evening nurse with a 50% day ratio might actually be assigned 4 day shifts and 6 evening shifts in a two-week period. Seniority and individual negotiations often determine the degree to which profiles can be violated. The preference scheduling approach discussed in the next section is flexible enough to allow for any level of deviation.

Other hospitals, however, only require that the day ratio be met by either shift in the rotation. For example, if a full-time nurse has a day/evening profile and a 40% day ratio, then she must be assigned at least 4 day shifts and 4 evening shifts over a two-week period. The remaining two shifts could be of either type. We adopt this view in the cycle scheduling approach.

3. Preference scheduling approach

Preference scheduling has become the norm in the majority of hospitals that face a chronic nursing shortage (e.g., see [15], [19]). Our column generation methodology for solving this scheduling problem can be readily adapted as a post-processor to solve the incremental staff adjustment problem. The new model allows the midterm schedule to be perturbed within a set of guidelines provided by the nurse manager and has the objective of minimizing the maximum amount of undercoverage in any one shift. This will be achieved by adjusting existing rosters and adding up to n^{Temp} travelers to the staff.

We begin by outlining the original model and then describe the necessary modifications. The following notation is used in the developments.

Indices and sets

- i index for nurses; $i \in N$
- j index for candidate rosters; $j \in S_i$
- t index for a shift (portion of a day); $t \in T$
- d index for day of the week; $d \in D$
- N set of nurses to be scheduled
- S_i set of rosters considered for nurse i
- D set of days in the planning horizon (typically 28)
- T set of shifts in a day (5 in our case)

Parameters

- c_{ij} penalty "cost" of assigning roster j to nurse i (this value is a function of the number and severity of preference violations in the roster)
- LD_{dt} lower bound on demand for nurses on day d during shift t
- UD_{dt} upper bound on demand for nurses on day d during shift t
- a_{ijdt} 1 if roster j for nurse i contains shift t on day d, 0 otherwise
- M large number representing the cost of an outside nurse (undercoverage) for a shift

Decision variables

- x_{ij} (binary) 1 if nurse i is assigned to roster j, 0 otherwise
- y_{dt} number of outside nurses used on day d during shift t
- s_{dt} excess number of nurses working on day d during shift t

Original Model (single skill)

$$\text{Minimize } z = \sum_{i\in N}\sum_{j\in S_i} c_{ij}x_{ij} + M\sum_{d\in D}\sum_{t\in T} y_{dt} \tag{1a}$$

$$\text{subjectto } -s_{dt} + \sum_{i\in N}\sum_{j\in S_i} a_{ijdt}x_{ij} + y_{dt} = LD_{dt}, \quad \text{for all } d \in D,\ t \in T \tag{1b}$$

$$\sum_{j\in S_i} x_{ij} = 1, \quad \text{for all } i \in N \tag{1c}$$

$$x_{ij} \in \{0, 1\}, \quad \text{for all } i \in N,\ j \in S_i \tag{1d}$$

$$0 \le s_{dt} \le UD_{dt} - LD_{dt},\ y_{dt} \ge 0, \quad \text{for all } d \in D, t \in T \tag{1e}$$

The objective function (1a) sums the "cost" associated with each roster and the number of gaps weighted by a penalty coefficient M. The value of the cost coefficients, c_{ij}, is based on the number and severity of the preference violations and increases exponentially with these two factors. Constraint (1b) along with the variable bounds in (1e) corresponds to the demand requirements for each shift and represents a transformation from a two-sided inequality constraint into a single equality with an upper bound on the slack variable s_{dt}. With some algebra, it can be seen that the number of nurses scheduled during shift t must be at least LD_{dt}. and no more than UD_{dt}. Constraint (1c) ensures that each nurse is given a roster.

In model (1), the following five common shift types are used: day (D), evening (E), night (N), AM, and PM. The first three are 8 hours in length and the last two are 12 hours in length.

To develop an adjustment model, we must consider the set of rosters produced by the midterm model and the extent to which modifications will be permitted. We shall also consider hiring a number of travelers on a temporary basis. In the latter case, it will be necessary to either specify the rotational profile of each temporary nurse or let the model choose the best.

Assuming that (1a) – (1e) has been solved, let

x_{ij^*} = solution for each nurse $i \in N$ where $j^* \in S_i$ is the index of the optimal roster

N^* = set of nurses whose rosters satisfy their rotational profile exactly; $N^* \subseteq N$

For the new model, let us define

n^{Viol} total number of permissible rotational violations in adjusted rosters

n^{Temp} maximum number of temporary nurses to be hired for the upcoming planning horizon

O set of temporary nurses to be hired; $n^{\text{Temp}} = |O|$

S_i set of candidate rosters for nurse $i \in N \cup O$

S_i^* set of candidate rosters for nurse $i \in N^*$ that violate the rotational profile; $S_i^* \subset S_i$

w_{ij} (decision variable) 1 if temporary nurse $i \in O$ is assigned to roster $j \in S_i$, 0 otherwise

The rules that will be used to make the adjustments are as follows.

1. For each nurse $i \in N^*$ whose roster satisfies her rotational profile, an imbalance of up to one shift is permitted.
2. For each nurse $i \in N \backslash N^*$ whose roster does not satisfy her rotational profile, no further deviations are permitted but rosters that adhere to the profile may be assigned.
3. The day ratio for the temporary nurses must be within some upper and lower bound.

The first two rules determine the legality of candidate rosters for each nurse $i \in N$ and hence limit the sets S_i and S_i^*. The third rule addresses a practical concern; namely, that it may not be possible to hire someone if the roster to be assigned is too erratic. With this in mind, we propose the following model.

$$\text{Minimize} \quad \max\{y_{dt} : d \in D, t \in T\} \tag{2a}$$

$$\text{subject to} \quad -s_{dt} + \sum_{i \in N} \sum_{j \in S_i} a_{ijdt} x_{ij} + \sum_{i \in O} \sum_{j \in S_i} a_{ijdt} w_{ij} + y_{dt} = LD_{dt}, \quad \forall d \in D, t \in T \tag{2b}$$

$$\sum_{j \in S_i} x_{ij} = 1, \quad \forall i \in N \tag{2c}$$

$$\sum_{j \in S_i} w_{ij} \le 1, \quad \forall i \in O \tag{2d}$$

$$\sum_{i \in N*} \sum_{j \in S_i^*} x_{ij} \le n^{\text{Viol}} \tag{2e}$$

$$x_{ij} \in \{0,1\}, \forall i \in N, j \in S_i; \quad w_{ij} \in \{0,1\}, \forall i \in O, j \in S_i \tag{2f}$$

$$0 \le s_{dt} \le UD_{dt} - LD_{dt}, y_{dt} \ge 0, \quad \forall d \in D, t \in T \tag{2g}$$

The objective in (2a) is to minimize the maximum gap in any shift over the planning horizon. Although the objective function is nonlinear, the problem can be converted to a mixed-integer linear program with the well-known transformation given below. Constraint (2b) ensures that all demand is met, either with regular nurses (x_{ij}), temporary nurses (w_{ij}), or outside nurses who are contracted on a daily basis (y_{dt}). Constraint (2c) requires that each regular nurse be assigned a roster, while (2d) says that each temporary nurse may be given a roster. The implication is that fewer than n^{Temp} temporary nurses may be included in a solution. For those nurses N^* whose midterm schedules match their rotational profile, constraint (2e) limits the total number of new schedules that violate their profiles to no more than n^{Viol}. While it is possible to change the shifts of these nurses, up to n^{Viol} of them can be given new schedules with day ratios that differ from their contract. Finally, (2f) and (2g) define the variables and their bounds.

To linearize model (2), we replace the objective function in (2a) with a single variable, z, and restrict its value to be no less than the largest gap. The transformed model is

$$\text{Minimize}\, \theta \tag{3a}$$

$$\text{subject to (2b)} - \text{(2g)}$$

$$\theta \ge y_{dt}, \forall d \in D, t \in T \tag{3b}$$

Column generation can now be used to solve model (3). The remaining issue is how to modify the existing procedure to generate candidate rosters that satisfy the adjustment rules. For the original problem, the concern was preference violations; for the new problem, the most important consideration is ensuring that the hard constraints are satisfied, at least for the regular nurses. For the temporary nurses, additional logic based on over- and undercoverage during each shift would be needed to generate good rosters.

4. Cyclic scheduling approach

The basis for our second model for the nurse addition problem is cyclic scheduling in which the major goal is to generate rosters that adhere to the rotational profiles of the nursing staff [20]. It is most appropriate when the same patterns can be used from one planning period to the next, as in health care facilities where individual preferences are overshadowed by the need for consistency. Model input includes the number of nurses contracted for each rotational profile, the demand per shift, and the scheduling rules. As mentioned, a profile specifies the number and types of shifts to be assigned and the number of hours that are to be worked over the planning horizon, which is typically a multiple of two weeks.

Throughout the discussion, it is assumed that some nurses do not have a rotational profile but simply work straight shifts. Once again, five single-shift profiles are considered: three 8-hours shifts (i.e., day (D), evening (E), and night (N)) and two 12-hours shifts (i.e., AM and PM). For a nurse with a rotational profile, there are many possible assignments depending on the shift lengths, the FTE value, and whether or not a combination of 8- and 12-hour shifts is permitted. We consider D/E, D/N, D/AM, D/PM, E/N, E/AM, and N/AM.

The rules that limit the flexibility and legality of a roster constitute the final input. Because they affect the perception of "fairness" of the results, it is essential that they be quantified in a way that ensures an even distribution of preference violations in the final rosters. In general, the rules can be translated into either hard or soft constraints reflecting institutional policies, contracts, individual agreements, and legal restrictions. Hard constraints are rules that must be satisfied, while soft constraints are those that can be violated under certain circumstances but at a cost. The fewer the number of soft constraint violations, the better the schedule.

The following hard constraints are included in our cyclic scheduling model

a. All full-time nurses must be assigned either 72 or 80 hours within a two-week planning period, depending on their contract. When a nurse is assigned fewer hours than specified in the contract, she is still paid her full weekly salary. Cancellations and overtime are taken into account on a daily basis and are not part of our model (see [21] for a discussion of the daily adjustment problem).
b. A nurse can only be assigned to the shifts that define her rotational profile.
c. The number of consecutive working days, also commonly called the *workstretch*, cannot exceed D^{maxon}. In most cases, this parameter is set to 5.
d. A nurse can work for at most 12 hours in a day, which means that at most one shift can be assigned in a day. Also, there needs to be at least an 8-hour break between consecutive assignments. Compliance is generally automatic because of the 12-hour rule; however, additional restrictions are required for those profiles in which back-to-back shifts are possible. In particular, the following sequences are not permitted: N/D, PM/D, N/AM and PM/AM. If the time between consecutive assignments is required to be more than 8 hours, then it would be necessary to expand the definition of back-to-back shifts.
e. Nurses must work two weekend shifts in the same weekend every two weeks. For our purposes, the first weekend shift starts at 7:00 p.m. on Friday and the last weekend shifts starts at 3 p.m. on Sunday.

Two soft constraints considered in the problem are:

f. Days-on and days-off patterns. There are two undesirable working patterns. The first is evidenced by one day off between two working days and is denoted by on-off-on or 1-0-1 (e.g. see [22]). The second is evidenced by a day on between two days off and is denoted by off-on-off or 0-1-0. It is more desirable to have at least two consecutive days off. In our implementation, nurses who only work 12-hour shifts (AM, PM or both) are not subject to the 1-0-1 and 0-1-0 soft constraints because most hospitals view them as too restrictive.
g. Different shift assignments on consecutive working days. This situation may occur when a rotational nurse is assigned to work a sequence such as D/E/D without an intervening day off. This type of pattern is highly undesirable because it disrupts the body's circadian rhythm.

4.1. Model for adding nurses

To solve the cyclic scheduling problem, we developed a branch-and-price (B&P) algorithm based on Dantzig-Wolfe decomposition [23]. In this approach, a set-covering-type master problem is created from the demand and assignment constraints, such that each column represents a feasible roster for a rotational profile. To find promising columns, one subproblem consisting of the remaining hard and soft constraints is set up and solved for each profile. The objective is to minimize a generic representation of the reduced cost of a master problem variable. When a subproblem solution is negative, the associated roster is added to the master problem; when all subproblem solutions are zero, the linear programming relaxation of the original integer program (IP) has been solved. Branch and bound is then used to find the optimal solution to the original problem. At each node in the search tree, column generation is applied to ensure that the optimal rosters are not overlooked.

Rather than present the original pattern-view formulation, we give the modified formulation that can be used for the incremental analysis. Some additional notation follows, all of which refers to the nurse addition problem.

Indices and sets

d	index for days; $d \in D = \{1, 2, \ldots, D_{\max}\}$
m	index for weekends in the 14-day planning period; $m \in W$
j	index for rotational profiles; $j \in N^P$
k	index for rosters
t	index for shifts; $t \in T$
$t_1(t_2)$	first (second) shift in T_i for rotational nurse (actually, a function of $i \in N_R$)
T_j	set of shift types that a nurse assigned to profile j is hired to work
T	set of all possible shift types considered, $T = \bigcup_{j \in N^P} T_j = \{\text{D, E, N, AM, PM}\}$
D_W	set of weekend days in the planning period
W	set of weekends under consideration
$K(j)$	set of alternative rosters for rotational profile j
N^P	set of all rotational profiles; $n^P = \lvert N^P \rvert$

Parameters

h_t	length of shift t (hours)
$H_L(H_U)$	minimum (maximum) number of hours that a nurse assigned to profile j is contracted to work every two weeks ($H_L = 72,\ H_U = 80$)
Δ_{dt}	demand – supply for shift t on day d; $\Delta_{dt} = y_{dt}^{\text{Mid}} - s_{dt}^{\text{Mid}}$, where y_{dt}^{Mid} and s_{dt}^{Mid} are the gap and shortage, respectively, associated with the midterm schedule for shift t on day d and represent input data (can be positive, zero, or negative)
D_j^{maxon}	maximum number of consecutive days (workstretch) that a nurse assigned to profile j is permitted to work
$W_j^{\max}$	number of weekend shifts that a nurse assigned to profile j must work every two weeks
$TR_{\max}$	maximum number of consecutive transitions allowed
$P_{\max}^{010}$	maximum number of 0-1-0 and 1-0-1 pattern violations allowed
$O_{dt}^{\max}$	maximum number of outside nurses that can be assigned to shift t on day d
n^{Temp}	maximum number of temporary nurses to be hired
n_j^{Temp}	maximum number of temporary nurses to be hired with profile j
ε	positive arbitrarily small number
X_{jdt}^k	mapping parameter, 1 if roster k associated with rotational profile j covers shift t on day d, 0 otherwise

Decision variables

x_{jdt}	(binary) 1 if a temporary nurse assigned to profile j works shift t on day d, 0 otherwise
w_{jm}	(binary) 1 if a temporary nurse assigned to profile j works on weekend m, 0 otherwise
b_{jd}	(accounting) 1 if temporary nurse assigned to profile $j \in N^P$ works shift t_1 on day d and shift t_2 on day $d+1$, 0 otherwise; $t_1 \neq t_2$
p_{jd}	(accounting) 1 when a temporary nurse assigned to profile j has a 0-1-0 pattern that starts on day d, 0 otherwise
q_{jd}	(accounting) 1 when a temporary nurse assigned to profile j has a 1-0-1 pattern that starts on day d; 0 otherwise
y_{dt}	number of outside nurses assigned to shift t on day d
s_{dt}	excess number of nurses assigned to shift t on day d
z_{jk}	(decision variable) number of temporary nurses with profile j that are assigned roster k
α_j	number of temporary nurses hired with profile j
θ	maximum number of outside nurses required during any one shift in the planning horizon
μ_{jd}	(unrestricted) dual variable for demand constraint (4b) for shift t on day d
σ_j	(nonnegative) dual variable for upper bound constraint (4c) for profile j
τ	(nonnegative) dual variable for constraint (4d)

Master problem, $\mathcal{MP}$

$$\theta_{\text{IP}} = \text{Minimize}\ \theta + \varepsilon \sum_{d \in D} \sum_{t \in T} y_{dt} \tag{4a}$$

$$\text{subject to} \sum_{j \in N^P} \sum_{k \in K(j)} X_{jdt}^k z_{jk} - s_{dt} + y_{dt} = \Delta_{dt},$$

$$d \in D,\ t \in T \tag{4b}$$

$$\sum_{k \in K(j)} z_{jk} \le n_j^{\text{Temp}},\ j \in N^P \tag{4c}$$

$$\sum_{j \in N^P} \sum_{k \in K(j)} z_{jk} \le n^{\text{Temp}} \tag{4d}$$

$$\theta \ge y_{dt},\ d \in D, t \in T \tag{4e}$$

$$0 \le s_{dt} \le 1,\ 0 \le y_{dt} \le O_{dt}^{\max},\ \forall d, t;$$

$$z_{jk} \ge 0 \text{ and integer},\ \forall j, k \tag{4f}$$

Subproblem j, $\mathcal{SP}_j$

$$\theta_j^{\text{SP}} = \text{Minimize} - \sum_{d \in D} \sum_{t \in T} \mu_{dt} x_{jdt} + \sigma_j + \tau \tag{5a}$$

subject to

$$H_L \leq \sum_{d \in D} \sum_{t \in T_j} h_t x_{jdt} \leq H_U \tag{5b}$$

$$\sum_{t \in T_j} x_{jdt} \leq 1,\ d \in D \tag{5c}$$

$$x_{jdt_2} + x_{j,d+1,t_1} \leq 1,\ d \in D \tag{5d}$$

$$\sum_{\alpha=d}^{d+D_j^{\text{maxon}}} \sum_{t \in T_j} x_{jdt} \leq D_j^{\text{maxon}},\ d \in D \tag{5e}$$

$$\sum_{d \in D_W} \sum_{t \in T_j} x_{jdt} \geq W_j^{\max} w_{jm},\ m \in W \tag{5f}$$

$$\sum_{m \in W} w_{jm} = 1 \tag{5g}$$

$$\sum_{t \in T_j} x_{jdt} + \left(1 - \sum_{t \in T_j} x_{j,d+1,t}\right) + \sum_{t \in T_j} x_{j,d+2,t} + p_{jd} \geq 1,\ d \in D \tag{5h}$$

$$\left(1 - \sum_{t \in T_j} x_{jdt}\right) + \sum_{t \in T_j} x_{j,d+1,t} + \left(1 - \sum_{t \in T_j} x_{j,d+2,t}\right) + q_{jd} \geq 1,\ d \in D \tag{5i}$$

$$1 - x_{jdt_\alpha} + 1 - x_{j,d+1,t_\beta} + b_{jd} \geq 1,\quad d \in D, \alpha \neq \beta \in \{1, 2\} \tag{5j}$$

$$\sum_{d \in D} b_{jd} \leq TR_{\max} \tag{5k}$$

$$\sum_{d \in D} p_{jd} + \sum_{d \in D} q_{jd} \leq P_{\max}^{010} \tag{5l}$$

$$b_{jd}, p_{jd}, q_{jd} \geq 0 \forall j,\ d,\ t;\quad w_{jm} \in \{0, 1\},\ \forall j,\ m \tag{5m}$$

$$x_{jdt} \in \{0, 1\}, \forall j,\ t,\ d,\quad \text{where } x_{j,14+l,t} \equiv x_{jlt},\quad l = 1, \ldots, D_j^{\text{maxon}} \tag{5n}$$

The objective function (4a) in $\mathcal{MP}$, along with constraint (4e), is designed to minimize the maximum number of uncovered shifts during any day of the planning horizon. When there are multiple optima, the second term in (4a) ensures that the solution chosen is the one that minimizes the total number of uncovered shifts. Constraint (4b) corresponds to the demand requirement for each shift t on day d and again represents a transformation from a two-sided inequality into a single equality constraint with an upper bound on the slack variable s_{dt}, as indicated in (4f). Constraints (4c) and (4d) determine the number of additional nurses to be hired for each profile $j \in N^P$. No profile violations are permitted. To complete the formulation, bounds on the slack, gap and profile variables are introduced in (4f). Note that if there is no bound on the number of temporary nurses that can be hired with a particular profile; i.e., $n_j^{\text{Temp}} = n^{\text{Temp}}$ for all j, then (4c) and σ_j can be removed.

The objective function (5a) in $\mathcal{SP}_j$ corresponds to the reduced cost of a column in $\mathcal{MP}$ with respect to the z_{jk} variables. Constraint (5b) states that the total number of hours assigned to a nurse with profile j must be between some lower and upper bound during the two-week planning horizon. Constraint (5c) restricts a nurse to at most one shift assignment within 24 hours. Because the length of a shift is either 8 or 12 hours, constraints (5b) – (5c) automatically ensure an 8-hour break between shifts except for the back-to-back cases defined by the set $N_{BB} = \{$N/D, PM/D, N/AM and PM/AM$\}$ mentioned in rule (d). For $j \in N_{BB}$, these cases are handled by constraint (5d) which permits only one assignment of either an N or PM shift (t_2) on day d, or a D or AM shift (t_1) on day $d + 1$.

Constraint (5e) limits the workstretch of a nurse who is assigned to profile j to no more than D_j^{maxon} days in any time window of $D_j^{\text{maxon}} + 1$ consecutive days. This corresponds to rule (c). The parameter D_j^{maxon} is set to 5 for nurses who work for 8-hour shifts only and to 4 for nurses who work both 8- and 12-hour shifts. Because the problem is cyclic, day 14 is followed by day 1. This is indicated in (5n). The weekend rule (e) is modeled by constraints (5f)–(5g). Weekends are defined by N and PM shifts for Friday and Saturday, and by D, E and AM shifts for Saturday and Sunday. Together, these constraints require that a nurse with profile j work exactly $W_j^{\max}$ weekend days every two weeks. Although the days must fall on the same weekend, it is an easy matter to allow split weekends. Note that the value of $W_j^{\max}$ is a function of the rotational profile. In our implementation, if a nurse with profile j works only 12-hour shifts, then she will be assigned just one weekend day ($W_j^{\max} = 1$) every two weeks; otherwise, $W_j^{\max} = 2$.

Constraints (5h) – (5i) determine the quality of the rosters. The undesirable patterns are counted in the model by the variables p_{jd}, q_{jd} and b_{jd}. A 0-1-0 pattern starting on day d implies that $\sum_{t \in T_j} x_{jdt} = 0$, $\sum_{t \in T_j} x_{j,d+1,t} = 1$ and $\sum_{t \in T_j} x_{j,d+2,t} = 0$. Constraint (5h) is an implication constraint that sets p_{jd} to 1 when such a pattern exists. Because all of the other variables in the constraint are binary and all of the data are integral, p_{jd} will always be integral in an optimal solution so it can be treated as a continuous variable. Constraint (5i) is the corresponding implication constraint for 1-0-1 patterns, which detect the existence of $\sum_{t \in T_j} x_{jdt} = 1$, $\sum_{t \in T_j} x_{j,d+1,t} = 0$ and $\sum_{t \in T_j} x_{j,d+2,t} = 1$ starting on day d. The total number of these patterns for profile j is given by the summation $\sum_{d \in D}(p_{jd}+q_{jd})$. Implicit in the formulation is that a roster is a circulation so that in (5h) and (5i), day $14+d=$ day d. This eliminates the need to introduce initial conditions.

Constraint (5j) detects a shift transition during consecutive days, and must be included for every possible combination of shift transitions that profile j may have. The maximum number permitted is given by the parameter $TR_{\max}$, as indicated in (5k). The next constraint (5l) limits the number of the one-day on and off patterns. Variable bounds are specified in (5m) and (5n).

4.2. Complexity issues

One common question in sensitivity analysis concerns the difficulty of finding a new solution to a problem just solved when the new problem is only marginally different than the original. The nurse addition model (4) – (5) has the same complexity as the original cyclic scheduling model but because we are only considering a handful of new nurses, perhaps it is easier. In fact, when only one nurse is to be added to a unit, the problem is trivial; when two or more are to be added, the problem is difficult. We formally establish this claim for $n^{\text{Temp}}=2$ using a variant of the directed two-commodity integral flow (D2CIF) problem for the reduction (proofs are in Appendix A). We then rely on the principle that any problem in NP that is more general than D2CIF is also NP-complete.

Proposition 1. The feasibility version of the nurse addition problem (NAP) represented by the mixed-integer linear program in model (4) – (5) is NP-complete.

When only one nurse is to be added, NAP is trivial: if there is one day in the planning horizon in which the maximum gap is $\Delta_{\max}$ for Day, Evening and Night shifts, no reduction is possible even if the days are different unless the nurse is allowed to work a D/E/N rotation; in all other cases, either a greedy assignment of shifts provides a reduction or one does not exist. Nevertheless, if the objective of NAP is to first minimize the number of shifts with a gap of $\Delta_{\max}$, and then, subject to that result, minimize the number of shifts with gap $\Delta_{\max}-1$, and so on, this problem can be modeled as a pure network.

Corollary 1. When only one nurse is to be added and the objective is to hierarchically minimize the maximum gap, NAP can be solved as a min-cost network flow problem.

4.3. Solution methodology

The branch-and-price (B&P) algorithm that we developed [23] to solve the midterm cyclic preference scheduling problem was modified to solve NAP. There are three major components to the algorithm: Dantzig-Wolfe decomposition to solve the LP relaxation of the master problem (4), a heuristic for finding feasible solutions, and a branching strategy to create the search tree.

In applying Dantzig-Wolfe, the master problem is solved as an LP and the subproblems as IPs. Starting with only artificial columns in (4), the corresponding dual variables $(\mu_{jd}, \sigma_j, \tau)$ are passed to the subproblems (5) to construct the objective function (5a) at each iteration. Whenever $\theta_j^{\text{SP}} < 0$, the solution (x_{jdt}) is converted to column (X_{jdt}^k), which is then added to $\mathcal{MP}$. In the implementation, all columns with negative objective function values (reduced costs) are added to $\mathcal{MP}$. The process terminates when $\theta_j^{\text{SP}}=0$ for all $j \in n^P$, indicating that the LP relaxation of (4) has been solved. If all the master problem variables $(s_{dt}, y_{dt}, z_{jk}, \alpha_j)$ are integral, then the original problem has also been solved; if not, a heuristic is called to find a feasible solution.

One simple heuristic is to solve the restricted master problem as an IP, disregarding the subproblems. If convergence is quickly achieved, this would be the best course of action, but our experience has shown that several hours may be required just to find a good feasible solution (with a 1% optimality gap) to (4) using all the columns generated during Dantzig-Wolfe decomposition. As an alternative, we have developed a rounding heuristic based on the fractional values of the column variables; call them $\hat{z}_{jk}$. A subset of these variables is fixed to the integer value closest to the fractional solution value found. For the parameter γ where $0<\gamma<0.5$, the following rule is used.

$$\text{If } (\hat{z}_{jk}-\lfloor \hat{z}_{jk} \rfloor) \leq \gamma, \text{ then fix } z_{jk}=\lfloor \hat{z}_{jk} \rfloor \text{ in } \mathcal{MP}$$
$$\text{If } (\lceil \hat{z}_{jk} \rceil-\hat{z}_{jk}) \leq \gamma, \text{ then fix } z_{jk}=\lceil \hat{z}_{jk} \rceil \text{ in } \mathcal{MP}$$
$$\text{If } \hat{z}_{jk}>1, \text{ then fix } z_{jk}=\lfloor \hat{z}_{jk} \rfloor \text{ in } \mathcal{MP}$$

Next, the reduced version of $\mathcal{MP}$ is solved as an IP with the remaining non-fixed variables. These include all z_{jk} variables in the range $\gamma<\hat{z}_{jk} \leq 1-\gamma$ and all z_{jk} variables whose values were 0 in the LP solution of $\mathcal{MP}$.

The final component of the algorithm involves branch and bound to close the optimality gap between the LP and IP solutions. Branching partitions the feasible region by imposing constraints that cut off fractional solutions associated with the relaxed problem. For a fractional solution $\hat{z}_{jk}$ in the master problem, we use the common depth-first strategy of creating two nodes at each iteration defined by the constraints $z_{jk} \leq \lfloor \hat{z}_{jk} \rfloor$ and $z_{jk} \geq \lceil \hat{z}_{jk} \rceil$, respectively. With B&P, it is necessary to take into account the affect of these constraints on SP_j to ensure that only new columns are generated. For the subproblem associated with the $z_{jk} \leq \lfloor \hat{z}_{jk} \rfloor$ node, the inequality $\sum_{(d,t) \in S(k)} x_{jdt} \leqslant \alpha_{jk}-1$, where

$\alpha_{jk} = \sum_{d\in D}\sum_{t\in T} X^k_{jdt}$, must be added to the corresponding subproblem. No additional constraints are required for the complementary node. See Purnomo and Bard [23] for more discussion.

The steps of the algorithm can be summarized as follows. Starting at the root node, $\mathcal{MP}$ is initialized with the artificial columns and $\mathcal{SP}_j$ is set up for all $j \in N^P$. Next, Dantzig-Wolfe decomposition is used to solve the LP relaxation of the master problem. Upon termination, the rounding heuristic is called to find an integer feasible solution to the original problem. Whenever the integrality gap is greater than the predetermined parameter (typically set at 1%), branching is initiated. At each node in the search tree, two descendent nodes are created by adding the appropriate constraints to $\mathcal{MP}$ and $\mathcal{SP}_j$. The modified master problem is then solved as an LP with Dantzig-Wolfe, starting with the previous basis and calling the dual simplex algorithm. Nodes are fathomed when the lower bound provided by the LP solution is within the predetermined percentage of optimality or integrality has been achieved; otherwise, the tree is extended. The procedure terminates when no unexplored nodes remain, a time limit is reached, or a predetermined node threshold is reached.

5. Computational experience

Both the preference scheduling and cyclic scheduling algorithms were implemented in Visual C^{++}. We report results for only the latter, which is more interesting from a computational point of view. During B&P, CPLEX 7.5 was used to solve the LP master problem (4) and the IP subproblems (5). Overall performance was measured on 15 problem instances that were generated from data derived from staffing levels and operations of an intensive care unit in a 400-bed U.S. hospital. The unit has 80 full-time nurses. The experimental design was aimed at determining the effectiveness of the algorithm and the robustness of solution quality. All computations were performed on a Dell PC with a 1.1 GHz processor.

Table 1 summarizes the four input characteristics of the problem sets. The "Δ demand" in the second column measures the total hours of uncovered shifts in the unit under consideration. Assuming that the maximum number of hours a full-time nurse can work in a two-week period is 80, the minimum number of nurses needed to cover the demand is shown under the 'Min nurses' column. The maximum number of nurses, n^{Temp}, that can be added to the unit is given in the 'Max nurses' column for each instance. Early testing showed that there are two components of a problem that make it difficult. The first is the maximum number of new hires allowed and the second is the maximum number of violations permitted for each new nurse. In general, it is more difficult to solve problems when the maximum number of nurses allowed is close to the minimum number required.

Table 1 Input Characteristics of Problem Sets

Problem no.	Δ demand (hours)	Min nurses	Max nurses	Max violations
1	612	8	8	4
2	756	10	10	3
3	984	13	12	3
4	812	11	10	3
5	652	8	9	2
6	392	5	4	2
7	308	4	5	4
8	472	6	6	4
9	560	7	7	2
10	836	11	11	5
11	488	6	7	2

With respect to the 'Max violations' parameter, $P_{\max} = P^{010}_{\max} + TR_{\max}$, listed in the last column, the total number of feasible rosters increases dramatically as more violations are permitted in a solution. In the computations, we set $P^{010}_{\max} = \lceil P_{\max}/2 \rceil$ and $TR_{\max} = \lfloor P_{\max}/2 \rfloor$ in constraints (5k) and (5l), respectively. For AM/PM rotational profiles, we do not impose constraint (5l) as they are commonly accepted in practice.

5.1. Results for basic data sets

The results obtained for the 11 problem instances are summarized in Table 2. Columns 2–4 report the solution quality, as measured by the average violations per nurse (b_{jd}, p_{jd}, q_{jd}), the total surplus hours (s_{dt}), and the gap hours (y_{dt}). The surplus hours are the total hours of overcoverage associated with the new schedules, while the gap hours indicate the undercoverage. The remaining columns highlight the performance of the B&P algorithm. The statistics in the 'Problem size' column indicate the number of rows and columns in the final LP master problem and column 6 gives the number of new columns generated during branch and bound. The next three columns are algorithm-specific. The initial LP solution obtained from solving MP with Dantzig-Wolfe is given first. The difference between the initial IP solution provided by our rounding heuristic at the root node and the initial LP solution is given in the "Initial gap" column. The 'IP soln' column indicates the best integer solution found during the enumeration process. The computations were limited to 30 seconds at this step, which was usually more than enough time for CPLEX to find the optimum. Fractional values are due to the second term in (4a); in the implementation, we set $\varepsilon = 0.01$. The remaining two columns show the computation time in seconds required to solve each instance, and the

Table 2 Output characteristics of problem sets

Problem no.	Avg violations	Surplus hours	Gap hours	Problem size ($m \times n$)	New columns	Initial LP soln	Initial gap (%)	Best IP soln	Total time (sec)	Total nodes
1	3.6	56	56	85 × 572	160	1.04	2	1.06	31	25
2	2.1	40	52	85 × 586	191	1.05	2	1.07	170	40
3	2.7	48	108	85 × 750	350	1.10	1.2	1.11	588	200
4	2.7	88	100	85 × 507	125	1.12	1.1	1.12	259	110
5	1.7	40	60	85 × 579	189	1.05	2.8	1.05	160	40
6	2.2	28	92	85 × 371	45	1.09	1.8	1.1	73	5
7	2.5	12	36	85 × 284	0	1.04	0	1.04	25	1
8	2.7	32	44	85 × 445	63	1.02	3	1.04	69	15
9	1.7	32	48	85 × 390	0	1.03	1	1.04	51	1
10	2.5	104	72	85 × 472	70	1.05	1.5	1.07	136	15
11	1.6	16	68	85 × 473	82	1.00	2	1.02	126	20

total number of nodes created, respectively. All computations were done with a 200-node limit and a 0.5% optimality gap as the stopping criteria.

Because all eleven instances were solved to optimality, the quality of the solutions found were a function of the input demand data and the accompanying parameter settings. In all cases, the number of nurses hired was equal to the maximum allowed. The implication of this result is that the maximum number of permitted violations per nurse was not a constraining factor. To ensure that all nurses hired have a sufficient amount of work, it would be necessary to include constraint on surplus hours in model (4). Limiting the surplus hours would naturally increase the gap hours so a tradeoff exists between these two measures.

Comparing the maximum number of violations permitted per nurse in Table 1 with the average number reported in Table 2, it can be seen that in most instances, the difference is small. For problem no. 10, however, the average is 2.5 violations while the maximum is 5 so setting $P_{\max}$ higher would not provide any advantage. Also, the number of uncovered hours is noticeably low suggesting that all 11 nurses are being used efficiently. In a second set of runs when the max nurse parameter n^{Temp} was set at the minimum number of nurse needed in Table 2, the total gap hours was always below 80. For problems 3, 4 and 6, the total gap hours were more than 80 indicating that adding one more nurse would be beneficial.

A typical solution is enumerated in Table 3 for problem no. 11 with parameter values $n^{\text{Temp}} = 7$ and $P_{\max} = 4$ (see Appendix B for the corresponding demand data). The rotational profiles of each nurse can be inferred from the daily shift assignments. Although the demand on several shifts was sufficient to permit a nurse to be assigned a 100% day ratio, no such rosters were selected. The next to last column identifies the total number of violations per roster (recall that the average is 1.6), while the last column indicates the number of hours that each nurse is to work over the two-week planning horizon.

With regard to problem size and computational performance, the largest number of columns generated when solving model (4) was 750 for problem no. 3. The total number of rows is constant at 156 (=70 demand constraints +14 profile constraints +1 assignment constraint +70 bound constraints on θ). During the B&P, the search trees grew to no more than 40 nodes for most instances and solution times averaged 153 seconds. Each subproblem had approximately the same dimensions and solved within 5 to 6 seconds. Minor differences is size were due to the number of shifts in a rotational profile. The largest subproblem consisted of 86 variables and 93 constraints.

Table 3 Results for problem no. 11

Nurse	Schedule for two weeks														No. viol's	Total hours
	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun		
N1	Day	Day	Day	Day	Day	Off	Off	Eve	Eve	Eve	Off	Off	Eve	Off	2	72
N2	Eve	Eve	Eve	Off	Eve	Eve	Eve	Eve	Off	Off	AM	AM	Off	Off	1	80
N3	Night	Eve	Eve	Off	Night	Night	Night	Night	Off	Off	Eve	Eve	Off	Off	2	72
N4	Eve	Off	Eve	Eve	Off	Off	PM	PM	PM	Off	Off	Off	PM	Eve	2	80
N5	PM	PM	Off	Off	AM	Off	AM	Off	PM	Off	Off	Off	Off	PM	0	72
N6	Day	Off	Off	Day	Day	Off	Off	Day	Day	Off	AM	Day	Day	AM	4	80
N7	Night	Off	PM	PM	Night	Off	Off	Night	Night	Off	Off	Night	Night	Night	2	80

Table 4 Parametric analysis for total nurse hired using data set 11

	Maximum number of nurses permitted to be hired					
Measure	8	7	6	5	4	3
Pattern violations	18	14	9	7	11	7
Transition violations	12	12	8	5	7	3
Average violations	3.7	3.6	2.8	2.4	4.5	3.3
Total gap hours	0	20	28	108	188	252
Total slack hours	128	36	16	8	8	0
Computation time	129	107	64	14	11	9
Total nurses hired	8	7	6	5	4	3
	Profiles chosen					
	D (72)	D/E (72)	D/E (72)	D/E (80)	E (80)	D/E (80)
	E (72)	E/N (72)	AM/N (80)	AM/N (80)	AM/N (76)	E/N (80)
	AM/E (76)	E/N (72)	E/PM (80)	E/PM (80)	E/N (80)	AM/PM (76)
	AM/N (80)	N/AM (72)	E/PM (80)	D/AM (76)	D/PM (80)	
	E/N (80)	D/AM (72)	D/AM (76)	N/PM (80)		
	D/AM (76)	N/PM (72)	N/PM (80)			
	E/PM (80)	N/PM (72)				
	N/PM (80)					

The next to last column in Table 2 indicates that a bit less than 10 minutes was the most time needed to solve a problem. This level of efficiency was primarily due to the small 'Initial gap,' which was never greater than 3%, between the initial LP solution and the IP solution obtained with our heuristic. In two instances (7 and 9), the optimal solution was found at the root node.

5.2. Parametric analysis

Two parametric studies were conducted to determine how solution quality and problem difficulty varied with the maximum number of nurses hired (n^{Temp}) and the maximum number of permitted violations per nurse ($P_{\max}$). The same demand data contained in Appendix B were used as input. Table 4 gives the results for $P_{\max} = 5$ and $n^{\text{Temp}} \in [3, 8]$. As expected, the total gap hours grew significantly and the total slack hours decreased in response as the maximum number of nurses went from 8 down to 3. Because no violation penalties are included in the model, the average number swung widely from 2.4 to 4.5 over the six scenarios.

The results in Table 4 also show that computation times decreased sharply as n^{Temp} decreased. Problems with 3, 4 and 5 nurses solved at the root node within 15 seconds, while the remaining problems required up to 101 seconds and several nodes. Not surprisingly, the instance with 8 nurses took the most time. Although the corresponding solution had 0 gap hours, the 128 slack hours translate into 8 hours of slack per nurse per week.

The profiles chosen for each scenario are shown in the bottom portion of the table. The numbers in parentheses indicate the total working hours during the two-week planning horizon. When the term $\varepsilon \sum_{d\in D} \sum_{t\in T} s_{dt}$ was added to (4a) to act as a secondary objective along with $\varepsilon \sum_{d\in D} \sum_{t\in T} y_{dt}$, only 7 nurses were hired for the case in which n^{Temp} was set to 8.

Table 5 reports algorithmic performance and solution quality when $n^{\text{Temp}} = 7$ and $P_{\max} \in [1, 6]$. One immediate observation is that there appears to exist a threshold value of 5 for the maximum number of permitted violations. This can be seen from the 'Total gap hours' row, where a value of 0 is obtained for $P_{\max} \geq 5$. Setting $P_{\max}$ higher that this value will not lead to a better solution. With respect to the other performance measures, no discernable trends are apparent. The computation times ranged from 90 to 279 seconds while the search trees averaged 10 node, the number reported in Table 3 for $n^{\text{Temp}} = 7$.

6. Conclusions

In the face of rising healthcare costs and a growing nursing shortage, hospitals are constantly reexamining their staffing needs to determine the optimal mix of full-timers, part-timers, and temporary personnel. As demand fluctuates over the year, it often makes better economic sense to hire non-permanent staff for weeks or months at a time, rather than rely on expensive agency nurses to handle the peaks on a daily basis.

In this paper, we proposed two models to help nurse managers make temporary hiring decisions. Extensive testing was done on the second model using a modified version of a branch-and-price algorithm originally developed to solve a related cyclic scheduling problem. The results showed that high quality solution could be obtained quickly for the most

Table 5 Parametric analysis for maximum violations using data set 11

Measures	Maximum number of permitted violations per nurse					
	6	5	4	3	2	1
Pattern violations	12	14	13	12	7	7
Transition violations	13	12	9	6	7	0
Average violations	3.6	3.6	3.1	2.6	2	1
Total gap hours	0	20	0	12	68	36
Total slack hours	60	36	48	68	16	64
Computational time	99	107	141	90	126	274
Total nurse hired	7	7	7	7	7	7
	Profiles chosen					
	D/E (80)	D/E (72)	E (80)	D (80)	D/E (72)	PM (72)
	E/AM (80)	E/N (72)	D/E (72)	D/E (72)	D/AM (80)	D/E (72)
	AM/N (76)	E/N (72)	AM/N (76)	E/N (72)	E/N (72)	D/N (72)
	E/PM (80)	N/AM (72)	E/PM (76)	E/PM (80)	E/PM (80)	E/PM (80)
	D/AM (72)	D/AM (72)	D/AM (72)	D/AM (80)	D/AM (76)	E/N (72)
	N/PM (80)	N/PM (72)	N/PM (80)	N/PM (76)	N/PM (80)	E/PM (76)
	N/PM (80)	N/PM (72)	N/PM (80)	N/PM (80)	N/PM (80)	D/AM (72)

difficult instances. This is especially important in an environment in which scenario analysis is an integral part of the planning process. Our current research is aimed at extending the approaches discussed in the paper to scheduling emergency room physicians.

Appendix A

Proof of Proposition 1. We start with an instance of D2CIF (directed two-commodity integral flow), which is strongly NP-hard [24], and show that it can be polynomially transformed into a restricted instance of the nurse addition problem denoted by RNAP. The recognition version of D2CIF is defined on a directed graph $G = (V, A)$, specific vertices s_1, s_2, e_1, and e_2, capacity $c(a) \in Z^+$ for each $a \in A$, and requirements $R_1, R_2 \in Z^+$. The following question is asked: Are there two flow functions $f_1, f_2 : A \to Z_0^+$ such that

(a) for each $a \in A$, $f_1(a) + f_2(a) \le c(a)$,
(b) for each $v \in V\backslash\{s, e\}$ and $i \in \{1, 2\}$, flow f_i is conserved at v, and
(c) for $i \in \{1, 2\}$, the net flow into e_i under flow f_i is at least R_i?

Gary and Johnson [24] state that this problem is NP-complete and remains so even if $c(a) = 1$ for all $a \in A$ an $R_1 = 1$. Variants in which $s_1 = s_2$, $e_1 = e_2$, and arcs are restricted to carry only one specified commodity are also NP-complete. We are interested in a subset of these conditions for a network with gains, denoted by D2CIF-G.

To simplify the presentation a bit, we consider a two-week instance of RNAP in which the two nurses each work 10 out of 14 days, cannot work more than 5 days in a row, have a distinct rotational profile of either D/E, D/N or E/N, and work 8-hour shifts only. It is not necessary to include straight shifts because they are dominated by rotational profiles. (Although it is possible to include the case in which both nurses work the same profile, the notation is denser and so is omitted.) Let $\Delta_{\max} = \max\{\Delta_{dt} : d \in D, t \in T\}$ be the largest gap over the planning horizon. The feasibility version of RNAP asks the question: Do rosters exist for the two additional nurses that result in a maximum gap of either $\Delta_{\max} - 1$ or $\Delta_{\max} - 2$?

Now consider an instance of D2CIF-G on the graph partially depicted in Figure 1. There is one source node labeled s that is connected to node s' by a single arc with capacity 2. Node s' is connected to each of three nodes labeled D/E, D/N and E/N by arcs with a gain (and capacity) of 14. The commodity flow along exiting arcs is restricted to the corresponding type of nurse. Immediately following each of these three nodes is an array of 14 sets of nodes, one for each day in the planning horizon. Each set consists of 10 nodes labeled D_{1d}, D_{2d}, D_{3d}, E_{1d}, E_{2d}, E_{3d}, N_{1d}, N_{2d}, N_{3d}, O_d for $d = 1, \ldots, 14$, where D_{1d} represents a Day shift with gap $\Delta_{\max}$, D_{2d} represents a Day shift with gap $\Delta_{\max} - 1$, D_{3d} represents a Day shift with a gap $\Delta_{\max} - 2$ or less, and so on. The node labeled O_d corresponds to a day off on day d. Not shown in Figure 1 is the bundle constraint that limits the flow leaving nodes D/E, D/N and E/N to one commodity per day. This restriction can be easily modeled by replacing D/E, D/N and E/N with 14 nodes each. Unless otherwise stated, all arcs have a capacity of 1.

To account for the consecutive days-off restriction, 9 sets of three nodes each are included at the next level in the graph. These nodes are labeled DE_k, DN_k and EN_k, $k = 1, \ldots, 9$.

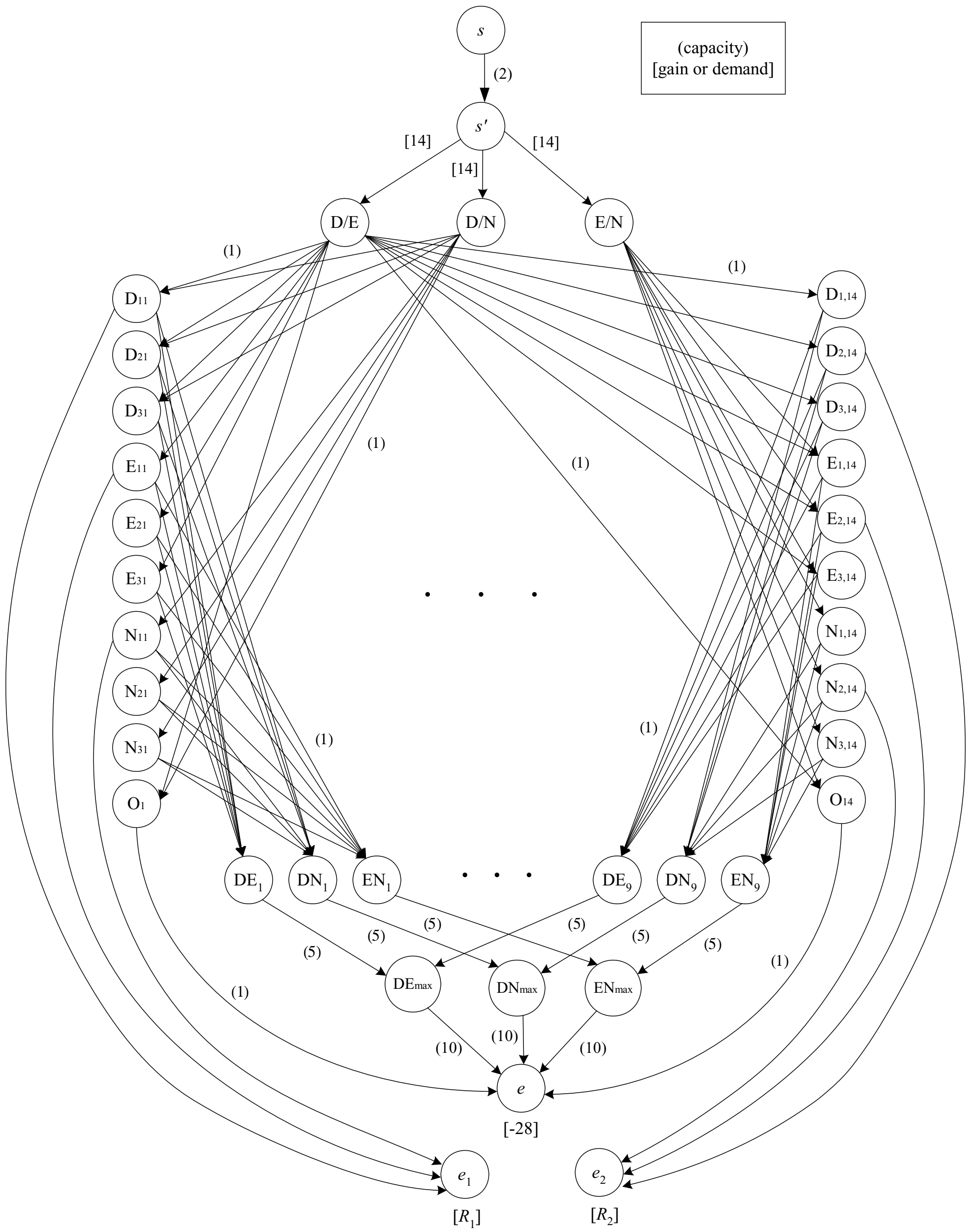

Fig. 1 Network diagram used to show equivalence of RNAP and D2CIF-G

The first node, DE_6, has predecessors $D_{1d}, D_{2d}, D_{3d}, E_{1d}, E_{2d}, E_{3d}$ for $d = 1, \ldots, 6$. In general, node D/E_k has predecessors $D_{1d}, D_{2d}, D_{3d}, E_{1d}, E_{2d}, E_{3d}$ for $d = k, \ldots, 5 + k$ and the only "commodity" that is permitted to flow on the arcs into DE_k corresponds to nurses who work a D/E rotation. The same arrangement exists for nodes DN_k and EN_k, $k = 1, \ldots, 9$.

The flow from nodes DE_k, $k = 1, \ldots, 9$, is channeled into a filtering node, DE_{max}, which accumulates the number of days that a D/E nurse can work. The capacity of each arc from D/E_k to $DE_{\max}$ is set at 5 —- the maximum number of consecutive working days permitted, and the arc cost is set to 0. $DE_{\max}$ is connected directly to the terminal node labeled e by a single arc with a capacity of 10. Each node O_d is also connected to e by single uncapacitated arc. Not shown in Figure 1 is the constraint that exactly 4 of the arcs originating at $O_d (d = 1, \ldots, 14)$ and terminating at e must have positive flow. The required flow into node e is 28. To account for the requirements R_1 and R_2, each node D_{1d}, E_{1d}, N_{1d}, D_{2d}, E_{2d}, N_{2d}, $d = 1, \ldots, 14$, is actually split into two nodes and a gain of 2 is placed on the arc connecting them (this arrangement is not shown in Figure 1). Call the pairs (D^1_{1d}, D^2_{1d}), (E^1_{1d}, E^2_{1d}), (N^1_{1d}, N^2_{1d}), (D^1_{2d}, D^2_{2d}),

(E^1_{2d}, E^2_{2d}), (N^1_{2d}, N^2_{2d}). In each case, the corresponding commodities flow into the first node of each pair and out of the second. The flow out is split into two parts (this is shown). The first part has already been described as the flow into the nodes DE_k, DN_k and EN_k, $k = 1, \ldots, 9$. The second part corresponds to the flow into nodes e_1 and e_2 that have demand R_1 and R_2, respectively. The problem then is to determine whether or not there exists flow functions $f_1(a)$ and $f_2(a)$ for all $a \in A$ such that the all restrictions on the network are satisfied and the respective flow into e_1 and e_2 is R_1 and R_2.

We claim that a feasible solution to RNAP is also a feasible solution to D2CIF-G and vice versa. To see this, note that only two units of flow can exit the source node s. Depending on the path they take at node s', these flows are converted to 14 units of the commodity associated with a D/E nurse, 14 units of the commodity associated with a D/N nurse, or 14 units of the commodity associated with an E/N nurse. Of course, only two of the three arcs can have positive flow.

The restrictions placed on flow into and out of nodes DE_k, DN_k and EN_k, $k = 1, \ldots, 9$, guarantee that at most 5 consecutive shifts associated with the respective rotations will be in a solution. This design corresponds to a realization of constraint (5e). The demand at node e guarantees that 14 units of each of the two commodities selected at node s' will arrive at that node. In a feasible solution, there will be at least R_1 and R_2 units of the two commodities flowing into nodes e_1 and e_2, respectively. With respect to RNAP, this means that the number of shifts that had gaps of at least $\Delta_{\max}$ and $\Delta_{\max} - 1$ in the original schedule will be reduced by at least these amounts.

Now let $R_1 =$ the number of shifts in the original schedule that have a gap of $\Delta_{\max}$ and $R_2 =$ the number of shifts that have a gap of $\Delta_{\max} - 1$ (by definition $R_2 \geq R_1$). Without loss of generality, suppose that only D and E shifts have a gap of $\Delta_{\max}$. To finish the proof, assume that there exists a feasible solution to RNAP that has a maximum gap of $\Delta_{\max} - 1$ but that D2CIF-G is infeasible for the above values of R_1 and R_2. This means that there is no flow through nodes D_{1d} and E_{1d} for all $d = 1, \ldots, 14$, that does not violate the 5-day rule or the days off constraint. But this contradicts the feasibility of the solution to RNAP. The fact that the network can be constructed in polynomial time and that any proposed solution to RNAP can be checked for feasibility in polynomial time, implies that RNAP and hence NAP is NP-complete. □

Proof of Corollary 1. Without loss of generality, assume that $\Delta_{\max}$ occurs on a D shift and on an E shift so the nurse must work a D/E rotation if any reduction is to be realized. Beginning with the network in Figure 1, remove nodes s, s', D/N, E/N, all nodes with an "N," and all arcs incident to them. Merge the split nodes (D^1_{1d}, D^2_{1d}), (E^1_{1d}, E^2_{1d}), (N^1_{1d}, N^2_{1d}), (D^1_{2d}, D^2_{2d}), (E^1_{2d}, E^2_{2d}), (N^1_{2d}, N^2_{2d}), removing the arcs (and gain on those arcs) between them. Also remove nodes e_1 and e_2 and the arcs entering them. Finally, place a supply of 14 units at node D/E, an upper and lower bound of 1 on the 14 arcs leading from D/E to $\mathrm{D/E}_d$ (recall that these nodes are not shown in Figure 1), $d = 1, \ldots, 14$, and a demand of –14 at node e.

Now define the following costs for all arcs leading from node D/E to nodes D_{1d} and E_{1d} for all $d \in D, t \in T$:

$$c_{dt} = \begin{cases} 1, & \text{if } \Delta_{\mathrm{dt}} = \Delta_{\max} \\ 11, & \text{if } \Delta_{\mathrm{dt}} = \Delta_{\max} - 1 \\ 122, & \text{if } \Delta_{\mathrm{dt}} < \Delta_{\max} - 1 \end{cases}$$

This structure implies that it is preferable for a nurse that is going to be added to a unit to work one period in which the gap is maximum, then 10 shifts in which the gap is one less than the maximum. A similar interpretation exists for all shifts in which the gap is two or more below the maximum. Finally, let the cost on the arc entering node O_d be "infinity." The resultant network has no gains and only a single commodity, and thus represents a pure min-cost flow problem. □

Appendix B (Table 6)

References

1. Bard JF (2004) Selecting the Appropriate Input Data Set when Configuring a Permanent Workforce. Computers & Industrial Engineering 47(4): 371–389

Table 6 Incremental demand for data set 11

	Day of planning horizon													
Shift	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Day	1	1	0	0	1	0	0	1	1	0	0	1	1	1
Eve	2	1	2	1	1	1	1	2	1	1	1	1	1	1
Night	2	1	0	0	1	1	1	2	1	0	0	1	1	1
AM	0	0	0	1	1	0	1	0	0	0	1	1	0	1
PM	1	1	1	1	0	0	1	1	1	0	0	1	1	1

2. Bard JF, Morton DP, Wang YM (2006) Workforce Planning at USPS Mail Processing & Distribution Centers Using Stochastic Optimization, to appear in Annals of Operations Research.
3. Ernst AT, Jiang H, Krishnamoorthy M, Sier D (2004) Staff Scheduling and Rostering: A Review. European Journal of Operational Research 153(1): 3–27
4. Lin CKY, Lai KF, Hung SL (2000) Development of a Workforce Management System for a Customer Hotline Service. Computers & Operations Research 27(10): 987–1004
5. Dawid H, Konig J, Strauss C (2001) An Enhanced Rostering Model for Airline Crews. Computers & Operations Research 28(7): 671–688
6. Misra S, Pinker EJ, Shumsky RA (2004) Salesforce Design with Experience-based Learning. IIE Transactions on Scheduling & Logistics 36(10): 941–952
7. Molleman E, Slomp J (1999) Functional Flexibility and Team Performance. International Journal of Production Research 37(8): 1837–1858
8. Kimball B, O'Neil E (2002) The American Nursing Shortage. (The Robert Wood Johnson Foundation, Princeton, NJ)
9. Spratley E, Johnson A, Sochalski J, Fritz M, Spencer W (2000) The Registered Nurse Population. Findings from the National Sample Survey of Registered Nurses, (U.S. Department of Health and Human Services, Washington, DC). Available at http://www.ruralnursing.org/rnsurvey00-1.pdf
10. O'Neil E, Seago JA (2002) Meeting the Challenge of Nursing and the Nation's Health. Journal of the American Medical Association 288(16): 2040–2041
11. Beaulieu H, Ferland JA, Bernard G, Michelon PA (2000) Mathematical Programming Approach for Scheduling Physicians in the Emergency Room. Health Care Management Science 3: 193–200
12. Bechtold SE, Jacobs LW (1996) The Equivalence of General Set-Covering and Implicit Integer Programming Formulations for Shift Scheduling. Naval Research Logistics 43(2): 233–249
13. Burns RN, Carter MW (1985) Work Force Size and Single Shift Schedules with Variable Demands, Management Science 31(5): 599–607
14. Brusco MJ (1998) Solving Personnel Tour Scheduling Problems Using the Dual All-Integer Cutting Plane. IIE Transactions on Operations Engineering 30(9): 835–844
15. Burke EK, Causmaecker PD, Vanden Berghe G, Van Landeghem H (2004) The State of the Art of Nurse Rostering. Journal of Scheduling 7(6): 441–499
16. Cheang B, Li H, Lim A, Rodrigues B (2003) Nurse Rostering Problems – A Bibliographic Survey. European Journal of Operational Research 151(1): 447–460
17. Bard JF, Purnomo HW (2005) Preference Scheduling for Nurses Using Column Generation. European Journal of Operational Research 164(2): 510–534
18. Millar H, Kiragu M (1998) Cyclic and Non-cyclic Scheduling of 12-Hour Shift Nurses by Network Programming. European Journal of Operational Research, 104: 582–592
19. Jaumard B, Semet F, Vovor T (1998) A Generalized Linear Programming Model for Nurse Scheduling. European Journal of Operational Research 107: 1–18
20. Emmons H (1985) Work-force Scheduling with Cyclic Requirements and Constraints on Days Off, Weekends Off, and Work Stretch. IIE Transactions 17(1): 8–15
21. Bard JF, Purnomo HW (2005) Short-Term Nurse Scheduling in Response to Daily Fluctuations in Supply and Demand. Health Care Management Science, 8(4): 315–324
22. Warner DM (1976) Scheduling Nursing Personnel According to Nursing Preference: A Mathematical Programming Approach. Operations Research 24(5): 842–856
23. Purnomo HW, Bard JF (2005) Cyclic Preference Scheduling For Nurses Using Branch and Price, Working paper, (Graduate Program in Operations Research & Industrial Engineering. The University of Texas, Austin)
24. Garey MR, Johnson DS (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness. (W.H. Freeman: New York)