Electronic Companion for Nurse Workload Balancing in a Surgical Ward using Data-Driven Inverse Optimization

Pooya Hoseinpour, Beste Kucukyazici, and Vedat Verter

A1. Proofs

Proof of Proposition 1 $\mathbb{X}^s$ is the set of all assignment matrices; each of its members shows the assignment of nurses to the patients. Because we assume that the nurses are anonymous, the cardinal of $\mathbb{X}^s$ shows the number of ways to distribute a set of $|\hat{\mathbf{J}}^s|$ labeled objects (i.e., patients) into $|\hat{\mathbf{I}}^s|$ nonempty identical subsets (i.e., nurses). This equals $S\left(|\hat{\mathbf{J}}^s|, |\hat{\mathbf{I}}^s|\right)$ and can be computed using recursive formula $S(a+1,b) = bS(a,b) + S(a,b-1)$ with an initial condition of $S(0,0) = 1, S(c,0) = S(0,c) = 0$ for $c > 0$. See Riordan (2012) for more detail.

Proof of Theorem 1 $\tilde{\boldsymbol{w}}^s$ and $\hat{\boldsymbol{w}}^s$ are the optimal solution. This means that by considering each of them $\mathbf{X}^s$ is a weakly nondominated solution. From (4), for any $\mathbf{X}^{s,k} \in \mathbb{X}^s$, we have $\max_{i\in\hat{\mathbf{I}}^s}\{\kappa(\mathbf{X}_i^{s,k};\tilde{\boldsymbol{w}}^s) - \kappa(\mathbf{X}_i^s;\tilde{\boldsymbol{w}}^s)\} \geq 0$, and $\max_{i\in\hat{\mathbf{I}}^s}\{\kappa(\mathbf{X}_i^{s,k};\hat{\boldsymbol{w}}^s) - \kappa(\mathbf{X}_i^s;\hat{\boldsymbol{w}}^s)\} \geq 0$. Therefore,

$$
\begin{aligned}
&\max_{i\in\hat{\mathbf{I}}^s}\{\kappa(\mathbf{X}_i^{s,k};\theta\tilde{\boldsymbol{w}}^s + (1-\theta)\hat{\boldsymbol{w}}^s) - \kappa(\mathbf{X}_i^s;\tilde{\boldsymbol{w}}^s + (1-\theta)\hat{\boldsymbol{w}}^s)\}\\
&= \max_{i\in\hat{\mathbf{I}}^s}\{\theta\kappa(\mathbf{X}_i^{s,k};\tilde{\boldsymbol{w}}^s) + (1-\theta)\kappa(\mathbf{X}_i^{s,k};\hat{\boldsymbol{w}}^s) - \theta\kappa(\mathbf{X}_i^s;\tilde{\boldsymbol{w}}^s) - (1-\theta)\kappa(\mathbf{X}_i^s;\hat{\boldsymbol{w}}^s)\}\\
&\geq \max_{i\in\hat{\mathbf{I}}^s}\{\theta\kappa(\mathbf{X}_i^{s,k};\tilde{\boldsymbol{w}}^s) - \theta\kappa(\mathbf{X}_i^s;\tilde{\boldsymbol{w}}^s)\} + \max_{i\in\hat{\mathbf{I}}^s}\{(1-\theta)\kappa(\mathbf{X}_i^{s,k};\hat{\boldsymbol{w}}^s) - (1-\theta)\kappa(\mathbf{X}_i^s;\hat{\boldsymbol{w}}^s)\} \geq 0.
\end{aligned}
$$

This holds because the workload function $\kappa(\mathbf{X}_i^s;\tilde{\boldsymbol{w}})$ is assumed to be linear with respect to $\boldsymbol{w}$ and the concavity of max function.

Proof of Proposition 2 In each iteration of Step 2, two-by-two checking are required, which equals

$$
\binom{|\mathbb{W}|}{2} + \binom{|\mathbb{W}|-1}{2} + \cdots + \binom{2}{2} = \frac{|\mathbb{W}|\left(|\mathbb{W}|^2-1\right)}{6}.
$$

This completes the proof.

Proof of Proposition 3 The algorithm's main step is Step 3, which is a sorting and the well-known sorting algorithm is the *merge sort* with complexity of $\mathrm{O}(n\log(n))$ for n numbers. This completes the proof.

Proof of Theorem 2 For each $i \in I^s$, the robust constraint is

$$
\sum_{r\in R} w_r f_r(X_i^s) \leq z \qquad \forall \boldsymbol{w}:\ \|\boldsymbol{w} - \tilde{\boldsymbol{w}}^0\|_2 \leq \theta.
$$

This is equivalent to

$$\sup_{\|\boldsymbol{w}-\tilde{\boldsymbol{w}}^0\|_2\le\theta} \sum_{r\in R} w_r f_r(X_i^s) \le z.$$

Let $\boldsymbol{u} = \boldsymbol{w} - \tilde{\boldsymbol{w}}^0$. Then $\|\boldsymbol{u}\|_2 \le \theta$, and thus

$$\sup_{\|\boldsymbol{u}\|_2\le\theta} \sum_{r\in R} (\tilde{w}_r^0 + u_r) f_r(X_i^s) \le z.$$

Hence,

$$\sum_{r\in R} \tilde{w}_r^0 f_r(X_i^s) + \sup_{\|\boldsymbol{u}\|_2\le\theta} \sum_{r\in R} u_r f_r(X_i^s) \le z.$$

It follows that the robust constraint is equivalent to as follows Ben-Tal and Nemirovski (1998):

$$\sum_{r\in R} \tilde{w}_r^0 f_r(X_i^s) + \theta \sqrt{\sum_{r\in R} \left(f_r(X_i^s)\right)^2} \le z.$$

A2. Generalization: Multi-Linear Workload Function

When workload is a multi-linear function, the resulting NWB$^s(\boldsymbol{w})$ model can be formulated as follows:

$$\begin{aligned}
\text{minimize} \quad & \max_{i\in\mathbf{I}^s} \left\{ \sum_{k\in K} \left(w_k \prod_{r\in\mathbf{R}^{(k)}} f_r(\mathbf{X}_i^s) \right) \right\} \\
\text{subject to} \quad & \mathbf{X}^s \mathbf{1}_{|\mathbf{I}^s|} = \mathbf{1}_{|\mathbf{J}^s|} \\
& \mathbf{X}^s \in \{0,1\}^{|\mathbf{J}^s|\times|\mathbf{I}^s|}.
\end{aligned}$$

The objective of the model above is a multi-linear function and, in general, its convexity is not guaranteed. We can reformulate the problem as follows:

$$\begin{aligned}
\text{minimize} \quad & z \\
\text{subject to} \quad & \sum_{k\in K} \left(w_k \prod_{r\in\mathbf{R}^{(k)}} q_{ir} \right) \le z \qquad \forall i \in \mathbf{I}^s \\
& f_r(\mathbf{X}_i^s) \le q_{ir} \qquad \forall r \in \mathbf{R}, \forall i \in \mathbf{I}^s \\
& \mathbf{X}^s \mathbf{1}_{|\mathbf{I}^s|} = \mathbf{1}_{|\mathbf{J}^s|} \\
& q_{ir} \ge 0 \qquad \forall r \in \mathbf{R}, \forall i \in \mathbf{I}^s \\
& \mathbf{X}^s \in \{0,1\}^{|\mathbf{J}^s|\times|\mathbf{I}^s|},
\end{aligned}$$

The model above can be transformed into a mixed-integer convex program by introducing $q_{ir} = e^{q'_{ir}}$, $z = e^{z'}$ (for more detail see (Boyd and Vandenberghe 2004)).

$$
\begin{aligned}
\text{minimize} \quad & e^{z\prime} \\
\text{subject to} \quad & \log\left(\sum_{k\in K} w_k e^{\sum_{r\in\mathbf{R}^{(k)}} q_{ir} - z}\right) \le 0 \qquad \forall i \in \mathbf{I}^s \\
& f_r(\mathbf{X}_i^s) \le \hbar(q_{ir}) \qquad \forall r \in \mathbf{R}, \forall i \in \mathbf{I}^s \\
& \mathbf{X}^s \mathbf{1}_{|\mathbf{I}^s|} = \mathbf{1}_{|\mathbf{J}^s|} \\
& q_{ir} \sim urs \qquad \forall r \in \mathbf{R}, \forall i \in \mathbf{I}^s \\
& \mathbf{X}^s \in \{0,1\}^{|\mathbf{J}^s|\times|\mathbf{I}^s|},
\end{aligned}
$$

Above, q_{ir} for $r \in \mathbf{R}$ and $i \in \mathbf{I}^s$ are auxiliary variables that are used for reconfiguring the model and $\hbar(x)$ is a piecewise linear function that approximates the exponential function. When the parameters are normalized to less than one, $\hbar(x)$ can be approximated as $\hbar(x) \sim x + 1$. This Taylor approximation is necessary to maintain the convexity of the reformulation and it has been commonly used. This model is in the form of a mixed-integer convex program if each attribute function $f_r(\mathbf{X}_i^s)$ is presented as a convex function.

The multi-linear workload function, defined over the power set of attributes, has $2^R - 1$ weights. The weights of pairwise interaction terms represent how the workload burden of one attribute depends on the level of another. For example, a positive weight for the interaction between time spent for direct care and travel distance indicates that the marginal disutility of non-nursing activities increases when care intensity is higher — a pattern consistent with the feedback we received from nurse managers. Higher-order interaction weights capture compounded interaction effects, although they might be harder to interpret. In general, attribute weights are interpretable as proxies to the intensity of the trade-offs between workload attributes. However, from a practical perspective, not all interaction terms would be active, and we expect the weight vector to be sparse.

A3. Modifying the Uncertainty Set

An ellipsoid can be defined as

$$\xi(\tilde{\boldsymbol{w}}^0, \Sigma) := \left\{\tilde{\boldsymbol{w}}^0 + \Sigma^{1/2}\gamma \,|\, \|\gamma\| \le 1\right\},$$

where the vector $\tilde{\boldsymbol{w}}^0$ is the ellipsoid center and $\Sigma = \Sigma^T \succ 0$ is symmetric and positive definite, which determines its shape. This ellipsoid volume is propositional to $(\det\Sigma)^{1/2}$. In the case that $\Sigma = \theta^2 I$, the ellipsoid is the ball, previously defined and used in Section 5.2.

The following mathematical program can be used to find an ellipsoid, centered in $\tilde{\boldsymbol{w}}^0$, with minimum volume that contains a $(1-\epsilon)$ portion of the points belonging to cluster $g \in G$:

$$\begin{aligned}
\text{minimize} \quad & \log(\det \Sigma) \\
\text{subject to} \quad & (\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^0)^T \Sigma^{-1} (\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^0) \le y_k + (1-y_k)M \qquad \forall k = 1, \dots, |\mathbf{W}_g| \\
& \sum_{k=1,\dots,|\mathbf{W}_g|} y_k \ge (1-\epsilon) \times |\mathbf{W}_g| \\
& y_k \in \{0,1\} \qquad \forall k = 1, \dots, |\mathbf{W}_g| \\
& \Sigma \succ 0.
\end{aligned}$$

Here, $y_k \in \{0,1\}$ is a binary variable, and for $k = 1, \dots, |\mathbf{W}_g|$, equals zero if the kth point is outside the ellipsoid and equals one, otherwise. The log function measures minimizing the volume, as it is extremely well-behaved and easy to optimize for practical purposes, e.g., in branch-and-bound algorithms. In Section 5.2, we use Algorithm 3 to simply solve the aforementioned problem for the case where $\Sigma = \theta^2 I$, i.e., the ellipsoid is a ball.

The robust counterpart of

$$\kappa(\mathbf{X}^s, \boldsymbol{w}) \le 0 \quad \forall \boldsymbol{w} \in \xi(\tilde{\boldsymbol{w}}^0, \Sigma),$$

will be as follows:

$$\left\| \Sigma^{1/2} \boldsymbol{u}(\mathbf{X}^{\mathbf{s}}) \right\| + \tilde{\boldsymbol{w}}^{0^T} \boldsymbol{u}(\mathbf{X}^{\mathbf{s}}) \le 0,$$

where $\kappa(\mathbf{X}^s, \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{u}(\mathbf{X}^s)$.

Moreover, to consider the dependency in the components of an uncertain weight vector, we might be interested in considering the uncertainty set as an intersection of the ellipsoid, i.e., $\xi(\tilde{\boldsymbol{w}}^0, \Sigma)$, and polyhedral $\omega = \{\boldsymbol{w} | \boldsymbol{w}^T \mathbf{1} = 1, \mathbf{0} \le \boldsymbol{w} \le \mathbf{1}\}$, which comes from our assumptions. In other words, for $\boldsymbol{w} \in \xi \cap \omega$, the robust counterpart will be

$$\begin{aligned}
& \delta^*(\hat{\boldsymbol{v}} | \xi) + \delta^*(\tilde{\boldsymbol{v}} | \omega) - \kappa_*(\mathbf{X}^s, \boldsymbol{v}) \le 0 \\
& \hat{\boldsymbol{v}} + \tilde{\boldsymbol{v}} = \boldsymbol{v},
\end{aligned}$$

where $\delta^*(\boldsymbol{v}|\psi) := \sup_{\boldsymbol{w} \in \psi} \{\boldsymbol{w}^T \boldsymbol{v}\}$ and $\kappa_*(\mathbf{X}^s, \boldsymbol{v})$ is the concave conjugate function of $\kappa(\mathbf{X}^s, \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{u}(\mathbf{X}^s)$ which results $\kappa_*(\mathbf{X}^s, \boldsymbol{v}) = 0$ and $\boldsymbol{v} = \boldsymbol{u}(\mathbf{X}^s)$. Therefore, it equals

$$\begin{aligned}
& \left\| \Sigma^{1/2} \hat{\boldsymbol{v}} \right\| + \tilde{\boldsymbol{w}}^{0^T} \hat{\boldsymbol{v}} + \max_{k=1,\dots,|\boldsymbol{u}|} \{\tilde{v}_k\} \le 0 \\
& \hat{\boldsymbol{v}} + \tilde{\boldsymbol{v}} = \boldsymbol{v} \\
& v_k = u_k(\mathbf{X}^s),
\end{aligned}$$

which can be transformed to a convex program for the model with a linear additive workload function (for more detail, see Ben-Tal et al. (2015)).

A4. Description of the Data

We draw on three sources of data: (i) RTLS installed in the surgical services department of the hospital, which allows us to track the nurses and patients dynamically; (ii) electronic medical records that include detailed information on patient characteristics; and (iii) department floor plans, which enable us to calculate physical distances between the rooms of units. Our focal dataset is the RTLS data, and we incorporated additional key variables from others.

For collecting data through RTLS, an active tag is placed on the ID bracelet of each patient upon admission to surgical services. Similarly, an RTLS tag is attached to the ID badge of each staff member, including nurses. Sensors have been installed in patients' rooms and hallways of 5NW. The RTLS hardware tracks the tags movement. Consequently, the system continuously tracks patients and nurses along with their interactions within the department. Thus, it collects and stores detailed time-stamped location data. Although each tagged patient is identified by her/his name, the nurses' identifications are kept anonymous, in keeping with the agreements with the labor union. Therefore, the nurse station is the only unobserved area.

Using initial raw data produced by RTLS, two files are created—patient-file and nurse-file. Both data files are preanalyzed and sifted for data cleaning, integration, and reduction. Figure 8 summarizes these steps. In data cleaning, some discrepancies are found and solved. For example, a patient in bed A of a room might also be seen as in bed B. This kind of inconsistency in data is decoded. Also, repetition and missing values are detected and mostly remedied by creating useful rules. The iterative trial-and-error method has been applied for finding the suitable rules. Then, the outlier detecting and dealing processes are performed on the data. To detect outliers in each data frame, Tukey's method Tukey (1977)–a well-known technique in data mining–is applied. After the detection process, different approaches from data mining literature are used to contend with outliers Han et al. (2011).

In data integration, the observations with the same values for all variables except the time variable—but with continuous values for time—are merged. Therefore, the time column at the merged observation now shows the patient's or nurse's duration of time in a specific location. Finally, in data reduction, we eliminate redundant variables that are not needed in our analysis. As a result, patient-file includes the patient name, identification code, diagnosis type, location, and

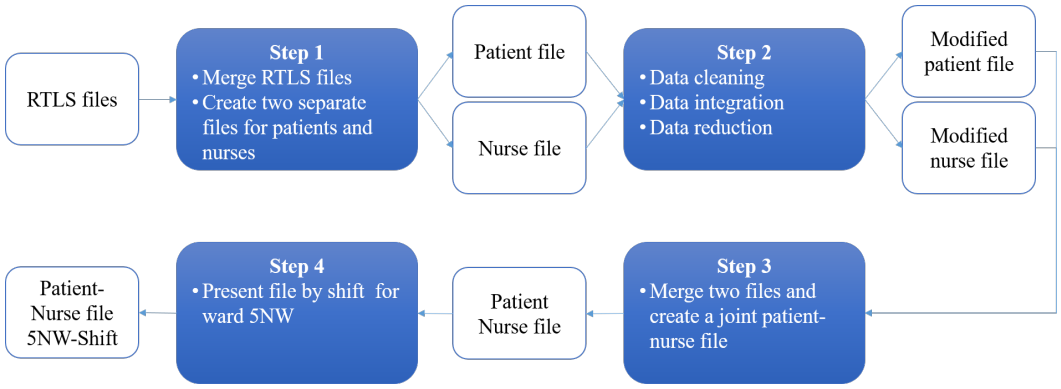


Figure 8 Preprocessing procedure for RTLS data.

periods. Nurse-file contains variables such as identification code, location, and periods. Patient diagnosis is categorized as colorectal, gynecology, gyneoncology, urology, and general surgery. Also, the nurses' identification code is not identical for each nurse; it can be a different code for the same nurse from one shift to another.

At the next step, both files are considered simultaneously to determine each patient's assigned nurse and the time of nursing services that each patient received in each shift. The nurse-patient assignment matrices are produced shift-by-shift. So, if more than one patient is in a specific location during a shift, we checked the data and decided considering which one is correct—e.g., the patient who was only a short-term visitor at a location, or the one who had determined as outlier data is ignored. If there is still more than one patient, we choose the one with a longer length of stay in the location. If a patient had more than one nurse during a shift, we select the nurse who provided the care for the maximum duration as the assigned nurse.

Our final dataset includes the patient's name, identification code, diagnosis type, location, and duration of stay in each location, as well as the nurse's identification code, location, and duration of stay in each location. Surgeries are categorized based on their specialty as colorectal, gynecology, gyneoncology, urology, and general surgery. This merged and organized dataset lets us identify which nurse was assigned to each patient and how much direct care was provided to each patient in each shift.

A5. Case Study Specification

Consider a case where the nurse workload function is in linear additive of the following attributes: (i) total direct care needed by the assigned patients, (ii) distances among the assigned patients, and (iii) total number of the assigned patients- these are attributes listed by the nurse managers

and nurses in our partner hospital. Indirect care refers to patient-specific service provided without interaction with the patient, such as chart/report writing, consulting or communicating with the families. Since indirect care is independent from patient specific characteristics and might be assumed to be a constant time per patient (Kakushi and Evora 2014). Therefore, (iii) can serve as a means to incorporate indirect care in workload estimations. Unit-related activities required of nurses and non-nursing activities also play a key role in workload estimations. Our discussions with the nurses show that the time allocated to unit related activities is quite homogeneous among nurses, which is also confirmed by a recent study (Yen et al. 2018); thus, assignment decisions are not expected to have a significant impact on the additional workload due to unit-related activities. Finally, walking in the hallways to go to patient beds or to supply room for medication or medical equipment is the leading factor in the context of non-nursing activities (Yen et al. 2018).

To define the multiattribute workload function of each nurse, we first define each of these attributes mathematically, and then we define the workload function.

Total direct care provided by the nurse: Assume that the total time of direct nursing service for each patient measures the amount of care needed by the patient. For each shift $s \in \mathbf{S}$, let the care needed by each patient $j \in \hat{\mathbf{J}}^s$ be τ_j^s, which is calculated from historical data. Therefore, the total amount of care required from nurse $i \in \hat{\mathbf{I}}^s$ in shift s is

$$c_i^s := \sum_{j \in \hat{\mathbf{J}}^s} \tau_j^s x_{ji}^s. \tag{10}$$

Distance traveled by the nurse: Assume that the maximum of distances among the locations of the assigned patients measures this attribute. Suppose l_{jk} shows the distance between the locations of patient $j \in \hat{\mathbf{J}}^s$ and $k \in \hat{\mathbf{J}}^s$. The maximum distance among the assigned patients to nurse i is measured by

$$t_i^s := \max_{j,k \in \hat{\mathbf{J}}^s} \{l_{jk} x_{ji}^s x_{ki}^s\}. \tag{11}$$

Number of patients assigned to the nurse: The total number of assigned patients is calculated by

$$u_i^s := \sum_{j \in \hat{\mathbf{J}}^s} x_{ji}^s. \tag{12}$$

Nurse workload function: In the remainder of this paper, we use a linear additive version of the general workload function in (3) for nurse i. We have

$$\kappa(\mathbf{X}_i^s) = w_1 c_i^s + w_2 t_i^s + w_3 u_i^s = w_1 \sum_{j \in \hat{\mathbf{J}}^s} \tau_j^s x_{ji}^s + w_2 \max_{j,k \in \hat{\mathbf{J}}^s} \{l_{jk} x_{ji}^s x_{ki}^s\} + w_3 \sum_{j \in \hat{\mathbf{J}}^s} x_{ji}^s. \tag{13}$$

Using the historical data available for previous shifts, we have the values of parameters τ_j^s, and l_{jk} in (10) and (11), respectively. Therefore, we calculate the values of attributes c_i^s, t_i^s and u_i^s by (10)-(12). Assuming a linear additive utility function (13) for each nurse $i \in \hat{\mathbf{I}}$, we would like to find w_1, w_2, w_3 such that assignment $\mathbf{X}^s$ is a weakly nondominated solution for almost all shifts $s \in \mathbf{S}$. Problem $\text{ARE}^s(\mathbf{X}^s)$ for the linear additive function is

$$\begin{aligned} \text{maximize} \quad & \frac{1}{\alpha_s} \sum_k Z_k^s && \text{(14a)} \\ \text{subject to} \quad & w_1 \Delta c_i^{s,k} + w_2 \Delta t_i^{s,k} + w_3 \Delta u_i^{s,k} \geq M(y_{ik} - 1) \qquad \forall i \in \hat{\mathbf{I}}^s, \forall \mathbf{X}^{s,k} \in \mathbb{X}^s && \text{(14b)} \\ & \text{(5c)} - \text{(5f)}, \end{aligned}$$

where $\Delta c_i^{s,k} = c_i^{s,k} - c_i^s$, $\Delta t_i^{s,k} = t_i^{s,k} - t_i^s$, and $\Delta u_i^{s,k} = u_i^{s,k} - u_i^s$ for shift $s \in \mathbf{S}$ and are normalized to have the value of all three components of the workload function between zero and one. Then we calculate $c_i^{s,k}$, $t_i^{s,k}$, $u_i^{s,k}$ using (10)-(12), respectively, considering $\mathbf{X}^{s,k}$ instead of $\mathbf{X}^s$.

When the workload function defined by (13) is used in problem NWB^s we have

$$\begin{aligned} \text{minimize} \quad & \max_{i \in \mathbf{I}^s} \left\{ w_1 \sum_{j \in \mathbf{J}^s} \tau_j^s x_{ji}^s + w_2 \max_{j,k \in \mathbf{J}^s} \{ l_{jk} x_{ji}^s x_{ki}^s \} + w_3 \sum_{j \in \mathbf{J}^s} x_{ji}^s \right\} \\ \text{subject to} \quad & \sum_{i \in \mathbf{I}^s} x_{ji}^s = 1 \qquad \forall j \in \mathbf{J}^s \\ & x_{ji}^s \in \{0,1\} \qquad \forall i \in \mathbf{I}^s, \forall j \in \mathbf{J}^s. \end{aligned}$$

For the deterministic model, as discussed, we transform this to an MILP as

$$\begin{aligned} \text{minimize} \quad & z \\ \text{subject to} \quad & w_1 \sum_{j \in \mathbf{J}^s} \tau_j^s x_{ji}^s + w_2 g_i^s + w_3 \sum_{j \in \mathbf{J}^s} x_{ji}^s \leq z \qquad \forall i \in \mathbf{I}^s \\ & l_{jk} y_{jki}^s \leq g_i^s \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\ & y_{jki}^s \leq x_{ji}^s \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\ & y_{jki}^s \leq x_{ki}^s \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\ & x_{ji}^s + x_{ki}^s \leq y_{jki}^s + 1 \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\ & \sum_{i \in \mathbf{I}^s} x_{ji}^s = 1 \qquad \forall j \in \mathbf{J}^s \\ & g_i^s \geq 0 \qquad \forall i \in \mathbf{I}^s \\ & y_{jki}^s, x_{ji}^s \in \{0,1\} \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \hat{\mathbf{I}}^s. \end{aligned}$$

In addition, we develop the following robust mathematical program for each uncertainty set:

$$
\begin{aligned}
\text{minimize} \quad & z \\
\text{subject to} \quad & w_1 \sum_{j\in\mathbf{J}^s} \tau_j^s x_{ji}^s + w_2 g_i^s + w_3 \sum_{j\in\mathbf{J}^s} x_{ji}^s \le z \qquad \forall i \in \mathbf{I}^s, \forall \boldsymbol{w} \in \mathbf{B}(\tilde{\boldsymbol{w}}^0, \theta) \\
& l_{jk} y_{jki}^s \le g_i^s \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\
& y_{jki}^s \le x_{ji}^s \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\
& y_{jki}^s \le x_{ki}^s \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\
& x_{ji}^s + x_{ki}^s \le y_{jki}^s + 1 \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\
& \sum_{i\in\mathbf{I}^s} x_{ji}^s = 1 \qquad \forall j \in \mathbf{J}^s \\
& g_{jki}^s \ge 0 \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s \\
& y_{jki}^s, x_{ji}^s \in \{0,1\} \qquad \forall j,k \in \mathbf{J}^s, \forall i \in \mathbf{I}^s.
\end{aligned}
$$

We can rewrite the robust counterpart of this program by replacing the constraints

$$
\theta \sqrt{\left(\sum_{j\in\mathbf{J}^s} \tau_j^s x_{ji}^s\right)^2 + \left(g_i^s\right)^2 + \left(\sum_{j\in\mathbf{J}^s} x_{ji}^s\right)^2} + \tilde{w}_1^0 \sum_{j\in\mathbf{J}^s} \tau_j^s x_{ji}^s + \tilde{w}_2^0 g_i^s + \tilde{w}_3^0 \sum_{j\in\mathbf{J}^s} x_{ji}^s \le z \qquad \forall i \in \mathbf{I}^s,
$$

instead of the first set of constraints in the model, which is finally in MISOCP format.

A6. Sensitivity Analysis on the Design Parameters of Algorithm 1

Sensitivity to the Random Generation of Feasible Assignments Because Algorithm 1 approximates the feasible assignment space using randomly generated feasible assignment matrices, we examined whether the results depend on the particular random sample used. To this end, Algorithm 1 was run with 10 different random seeds while all other model and algorithmic settings were held fixed. In particular, all runs used the same sequence of α values and the same stopping criterion based on the quality measure β.

Table 9 summarizes the results. For each run, we report the final value of β and the number of behavioral rules obtained by applying hierarchical clustering to the shift-specific weight vectors. Each rule is labeled by its dominant workload attribute, defined as the attribute with the largest weight.

The results suggest that the framework is robust to the random assignments used to approximate the feasible assignment space. The final value of β varies moderately across seeds, ranging from

Table 9 Sensitivity of the inverse optimization framework to the random seed.

Seed	Final β	No. of Rules	Rule 1	Rule 2	Rule 3
1	78.1%	3	Patients	Distance	Direct care
2	76.4%	3	Patients	Distance	Direct care
3	79.1%	3	Patients	Distance	Direct care
4	81.6%	3	Patients	Distance	Direct care
5	69.8%	3	Patients	Distance	Direct care
6	78.5%	3	Patients	Distance	Direct care
7	77.7%	3	Patients	Distance	Direct care
8	74.8%	3	Patients	Distance	Direct care
9	79.0%	3	Patients	Distance	Direct care
10	83.3%	3	Patients	Distance	Direct care
Mean ± SD	**77.8% ± 3.7%**	**3.0 ± 0.0**	–	–	–

69.8% to 83.3%. However, every run identifies three rules with the same qualitative structure: one rule dominated by the number of patients, one by maximum travel distance, and one by total direct care. Hence, the random sample affects the degree to which observed assignments are explained, but not the behavioral rules that are recovered.

Sensitivity to the Number of Clusters Algorithm 1 produces shift-specific weight vectors, which are then aggregated through hierarchical clustering into a practical set of behavioral assignment rules. Since the number of clusters determines the level of aggregation, we examined how this choice affects the resulting behavioral representation. We varied the number of clusters from $K = 2$ to $K = 5$ while keeping the set of shift-specific weight vectors unchanged. For each value of K, Table 10 reports the dominant workload attribute of each cluster.

Table 10 Sensitivity of the identified behavioral rules to the number of clusters.

K	Dominant Attribute of Each Cluster	Interpretation
2	Patients; Direct care	Broad behavioral segmentation
3	Patients; Distance; Direct care	Distinct and interpretable behavioral rules
4	Patients; Distance; Direct care; Mixed	More granular behavioral segmentation
5	Patients; Distance; Direct care; Mixed; Mixed	Further subdivision of existing behaviors

The number of clusters mainly controls the granularity of the behavioral representation. With $K = 2$, distinct workload preferences are merged into broader groups, which yields a coarse representation. Larger values of K separate some of these preferences into additional groups, but the added clusters are either mixed or subdivisions of existing behaviors and therefore offer little additional behavioral insight. The $K = 3$ solution provides the most parsimonious and interpretable representation, with three clusters corresponding to three distinct workload-oriented behaviors: one dominated by the number of patients, one by maximum travel distance, and one by total direct care. We therefore use the three-cluster solution as the primary behavioral representation in the remainder of the analysis.