Optimizing Family Medicine Residency Schedules under the Clinic First Principles

Mohammad Al Syouf* Ankit Bansal† Bjorn P. Berg‡ Jean-Philippe Richard§

Abstract

Family medicine residency programs must balance educational and operational requirements while providing residents with consistent exposure to continuity clinics, a central principle of the `Clinic First Model`. We study the Family Medicine Residency Scheduling problem and develop a binary integer programming (BIP) framework that incorporates Clinic First principles through two criteria: Clinic Time Consistency (`CTC`), which promotes consistent clinic involvement within resident schedules, and Level Loading (`LL`), which promotes a balanced distribution of total resident clinic sessions across time periods. To provide a simple and interpretable schedule structure, we integrate the proposed BIP framework with the *staggered scheduling framework*, in which resident schedules are generated by successive temporal shifts of a common `base rotation schedule`. We establish a theoretical bound on the loss in `CTC` resulting from the *staggered scheduling framework* and derive valid inequalities that strengthen the resulting BIP formulation. Computational experiments based on data from the University of Minnesota Family Medicine Residency program demonstrate substantial computational benefits of the valid inequalities and show that the *staggered scheduling framework* incurs only a small loss in `CTC`. Compared with a schedule used in current practice, the proposed approach substantially improves `CTC` while producing a more balanced distribution of clinic sessions. Comparisons with an existing scheduling objective further demonstrate the benefits of explicitly incorporating `CTC` and `LL`.

Keywords: Family Medicine Residency Scheduling, Clinic First Model, Staggered Scheduling, Valid Inequalities, Convex Hull

Acknowledgment: This material is based upon work supported by the National Science Foundation under Grant Numbers 2427201, 2427202. Any opinions, findings, and conclusions or recom-

*School of Systems Science and Industrial Engineering, State University of New York-Binghamton, Binghamton, NY, United States

†School of Systems Science and Industrial Engineering, State University of New York-Binghamton, Binghamton, NY, United States, abansal@binghamton.edu (*Corresponding author*)

‡Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN, USA

§Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN, USA

mendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

1 Introduction

Primary care is a cornerstone of the U.S. healthcare system, promoting population health, preventing chronic disease, and serving as the primary point of contact for patients seeking medical care (Manolakis and Skelton 2010, Stange et al. 2023, Starfield et al. 2005). Strong primary care systems are associated with improved health outcomes, more equitable access to care, and lower healthcare costs (Starfield 2012, Bodenheimer and Sinsky 2014, Brown et al. 2019). Despite its importance, the United States faces a persistent shortage of primary care physicians (PCPs). The American Association of Medical Colleges projects a shortage of 21,100 to 55,200 PCPs by 2032 (Dall et al. 2017), highlighting the need to strengthen the primary care workforce.

Family medicine physicians constitute the largest segment of the primary care workforce, accounting for approximately 40% of all PCPs. Consequently, improving family medicine residency training is widely recognized as a critical strategy for addressing the projected shortage of PCPs (Adam et al. 2022, Zeller et al. 2019). A major challenge, however, is that many residency programs emphasize inpatient training at the expense of outpatient `continuity clinic` experiences, leading to reduced patient continuity, irregular clinic schedules, resident dissatisfaction, and ultimately fewer graduates pursuing careers in primary care (Gupta et al. 2016, Adam et al. 2022, Zeller et al. 2019, Hofkamp et al. 2020).

The `Clinic First Model` has emerged as a promising educational framework for addressing this imbalance by prioritizing `continuity clinic` experiences while maintaining required inpatient and specialty training (Gupta et al. 2016). Of its six guiding principles, we focus on the following two: *"(1) design resident schedules that prioritize continuity of care and eliminate tension between inpatient and outpatient duties, and (2) create operationally excellent clinics"*. Recent implementations of the `Clinic First Model` have demonstrated improvements in resident satisfaction, continuity of care, clinic operations, and preparation for careers in primary care (Paul et al. 2021, Hofkamp et al. 2020, Hersch et al. 2023, Adam et al. 2022). Realizing these benefits requires incorporating `Clinic First` principles into annual rotation scheduling. At the beginning of each academic year, every resident receives a schedule specifying rotation assignments throughout the year. These schedules must satisfy accreditation requirements, educational objectives, staffing needs across clinical services, and `continuity clinic` obligations (Bard et al. 2016b, 2017, Proano and Agarwal 2018, Guo et al.

2023). Incorporating the `Clinic First Model` while satisfying these educational and operational requirements substantially increases the complexity of determining annual rotation schedules.

In current practice, annual rotation schedules in family medicine residency programs are developed primarily through manual processes, with implementation supported by commercial scheduling systems such as Amion and QGenda (Day et al. 2006, Proano and Agarwal 2018). Manual scheduling is time-consuming, difficult to adapt when unexpected changes occur, and limits the ability of program administrators to systematically evaluate alternative scheduling policies. As a result, opportunities to implement the `Clinic First` principles are often missed, adversely affecting resident experiences and the operational performance of the `continuity clinics`.

We use the term *Family Medicine Residency Scheduling* (`FMRS`) to denote the problem of determining annual rotation schedules for all residents in a family medicine residency program. Each academic year is divided into equal-length time periods, or *blocks*, and an annual rotation schedule specifies the rotation assigned to each resident in every block of the academic year. Family medicine residency programs typically span three years, with cohorts of approximately 6-10 residents in each program year (PGY). When a resident is assigned to a rotation for a time period, their time is allocated between rotation-specific activities, `continuity clinic`, and other programmatic requirements. Each rotation has its own template schedule of how often and when residents in that rotation are in the `continuity clinic`. As a result, the rotation schedule directly determines residents' availability in the `continuity clinic` throughout the year. Consequently, rotation scheduling has a profound impact on both residents' educational experiences and the operational performance of the `continuity clinic`.

In this paper, we develop a binary integer programming (BIP) framework for `FMRS` that explicitly incorporates the guiding principles of the `Clinic First Model` through both the objective function and the model constraints. Since the `Clinic First Model` prioritizes residents' `continuity clinic` experiences, we must account for clinic schedules when determining rotation assignments. Specifically, the proposed framework is designed around the following two criteria.

- Clinic Time Consistency (`CTC`): The number of clinic sessions varies across rotations because of specialty-specific training requirements and the availability of supervising faculty. Consequently, assigning several rotations with similar clinic-session requirements consecutively can lead to prolonged periods of either high or low clinic involvement, adversely affecting both residents' clinic experiences and continuity of care. To promote a more consistent clinic experience throughout the year for each resident, we introduce the *Clinic Time Consistency* (`CTC`) metric, defined as the sum of the absolute differences in the numbers of clinic sessions between consecutive rotations. Maximizing `CTC` encourages schedules that alternate between rotations with relatively high and low clinic-session requirements, thereby distributing clinic involvement more evenly over time for each resident. This criterion is directly responsive to the guiding principle (1) of the `Clinic First Model`.

- Level Loading (`LL`): The rotation schedule also determines the total number of resident clinic sessions available during each time period and, consequently, the staffing levels of the `continuity clinic`. Large fluctuations in available clinic capacity across time periods may reduce patient access, create operational inefficiencies, and adversely affect residents' clinical experiences. To address this issue, we introduce the *Level Loading* (`LL`) criterion, which measures the uniformity of total resident clinic sessions throughout the planning horizon. Our model enforces a minimum level of `LL` through a set of constraints, thereby ensuring a sufficiently balanced distribution of clinic capacity over time. This criterion is motivated by guiding principle (1) and (2) of the `Clinic First Model`.

To better align with current scheduling practice, we integrate our BIP framework with a *staggered scheduling framework*, in which the schedules of all residents are generated by successively right-shifting a single `base rotation schedule` across the planning horizon. This yields a simple and interpretable schedule structure in which every resident completes the same sequence of rotations in a different temporal order, thereby promoting equitable training experiences while producing schedules that are straightforward to construct, communicate, and administer. We establish a theoretical bound on the loss in `CTC` due to the *staggered scheduling framework*.

The resulting BIP formulation is computationally challenging. To enhance its computational tractability, we derive valid inequalities, show that they provide the convex hull of a subset of the constraints in the formulation. Computational experiments based on data from a family medicine residency program at the University of Minnesota (UMN) demonstrate that the proposed valid inequalities substantially improve the computational efficiency of the branch-and-bound algorithm for solving the BIP formulation. The experiments further show that the proposed approach produces schedules with substantially higher `CTC` and a more balanced distribution of clinic sessions than the schedule currently implemented in practice. Additionally, we demonstrate the benefits of incorporating the `CTC` and `LL` criteria by comparing the schedules generated by our proposed approach with those obtained using an alternative objective from the existing literature.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the `FMRS` problem, integrating the two `Clinic First` criteria and the *staggered scheduling framework*. Sections 4 and 5 present the proposed BIP formulation and the valid inequalities used to strengthen it, respectively. Section 6 presents the computational results, and Section 7 concludes the paper.

2 Literature Review

Physician scheduling has received considerable attention in the literature over the past several decades (Erhard et al. 2018). Resident rotation scheduling constitutes an important class of physician scheduling problems and involves assigning residents to rotations over a sequence of time periods while satisfying program-specific requirements and optimizing relevant objectives. Among the early contributions, Franz and Miller (1993) developed an integer programming (IP) formulation for resident rotation scheduling that focuses on optimizing resident preferences. Subsequently, Guo et al. (2014) studied a generic rotation scheduling problem incorporating constraints commonly encountered in residency programs and established that such a problem is *NP-hard*.

There is no universally applicable model for resident rotation scheduling, as program requirements and scheduling priorities vary across medical specialties and countries (Li et al. 2025, Kraul 2020, Guo et al. 2023, Hong et al. 2019). Consequently, the literature has considered a variety of objectives, including resident preferences (Akbarzadeh and Maenhout 2021a,b, Lemay et al. 2017, Smalley and Keskinocak 2016), workload management (Kraul 2020), and equity among residents (Li et al. 2025, Smalley and Keskinocak 2016). Despite these advances, relatively limited attention has been devoted to rotation scheduling in specialties where experience in the `continuity clinic` is an integral component of residency training.

Family medicine and internal medicine are two specialties in which residents' experiences in the `continuity clinic` play an important role in their decision to continue practicing in the specialty after completing residency. From a rotation scheduling perspective, an important distinction between the two specialties lies in how `continuity clinic` assignments are incorporated into rotations. In family medicine, rotations are typically associated with predefined templates that specify the `continuity clinic` sessions to be completed during each week of the rotation. In contrast, internal medicine rotation scheduling generally does not rely on such rotation-specific clinic templates (Bard et al. 2017).

Within the internal medicine setting, Bard et al. (2016a, 2017) develop IP-based approaches for annual resident rotation scheduling that explicitly account for clinic assignments. Their models seek, among other objectives, to balance the total number of clinic sessions across residents over the year and the number of patients seen by each clinic across months. Bard et al. (2016a) consider a 4+1 scheduling structure and obtain optimal solutions by decomposing the underlying model into two subproblems. In contrast, Bard et al. (2017) consider a more general scheduling setting that results in a substantially more complex optimization problem and, consequently, which they address using a heuristic. Proano and Agarwal (2018) develop a multi-objective integer programming framework for annual rotation scheduling of internal medicine residents. They employ the Analytical Hierarchy Process (AHP) to incorporate multiple scheduling criteria and promote equity among residents. Their objectives include, among others, maximizing resident satisfaction with vacation assignments and increasing assignments to elective rotations. Wang et al. (2022) also develop an IP formulation for annual block rotation scheduling of internal medicine residents. Their formulation seeks to avoid assigning residents to heavy-workload rotations in consecutive blocks and to distribute rotations evenly across the planning horizon, with the goal of maintaining the quality of patient care; they employ a heuristic to solve the resulting model.

Within the family medicine setting, Bard et al. (2016b) develop an IP-based approach for determining annual rotation schedules for family medicine residents. Their objective is to minimize deviations from predetermined rotation templates while balancing the number of patients seen per clinic session across the planning horizon. Given the computational complexity of the resulting formulation, they develop a heuristic solution approach.

Despite these contributions, the aforementioned studies in both internal medicine and family medicine do not explicitly focus on improving residents' experiences in the `continuity clinic`, a central objective of the `Clinic First Model`. Moreover, although some of these approaches promote equitable training experiences among residents, the resulting schedules may lack a simple underlying structure, making them more difficult to communicate and administer in practice. Our work addresses these gaps by explicitly incorporating `Clinic First` objectives into the family medicine rotation scheduling problem and integrating them within a structured scheduling framework that promotes equitable training experiences while yielding schedules that are straightforward to construct, communicate, and administer. The main contributions of this paper are summarized as follows:

1. We introduce the *staggered scheduling framework* for `FMRS`, which imposes a simple and interpretable schedule structure in which all residents complete the same sequence of rotations in different temporal orders. This structure promotes equitable training experiences while yielding

schedules that are straightforward to construct, communicate, and administer. Additionally, we establish a theoretical bound on the loss in `CTC` due to the *staggered scheduling framework*.

2. We develop an BIP formulation for `FMRS` under the *staggered scheduling framework* that explicitly incorporates the guiding principles of the `Clinic First Model` through its objective function and constraints. Specifically, the formulation maximizes `CTC` while enforcing a minimum level of `LL`.

3. We derive valid inequalities to strengthen the BIP formulation and show that these inequalities provide the convex hull of a set defined by a subset of the constraints in the formulation.

4. Using computational experiments based on data from a family medicine program at UMN, we demonstrate the computational benefits of the proposed valid inequalities and the *staggered scheduling framework*. We further compare the schedules produced under the *staggered scheduling framework* with those implemented in current practice and analyze structural properties of the resulting schedules to provide insights into how the proposed approach improves `CTC` and `LL`. Finally, we compare the schedules obtained using the `CTC` and `LL` criteria with those generated using the objective from Bard et al. (2016b).

3 Problem Description and Staggered Scheduling Framework

Let Y denote the set of PGYs. For each PGY $y \in Y$, let R_y and S_y represent the sets of rotations and residents, respectively. Define $\gamma_y := |S_y|$ as the number of residents in PGY y. For each rotation $r \in R_y$, let $\bar{c}^y_r$ denote the number of clinic sessions associated with that rotation. Finally, let $M = \{1, \ldots, |M|\}$ denote the set of equal-length discrete time periods within an academic year, and let $r^y_{im} \in R_y$ denote the rotation assigned to resident $i \in S_y$ during time period $m \in M$. Since each resident is assigned to exactly one rotation in each time period and each rotation is assigned exactly once over the calendar year, it follows that $|R_y| = |M|$ for every $y \in Y$. This is without loss of generality, as multiple rotations within the same specialty during an academic year can be represented as distinct rotations in the model. Accordingly, the schedule of resident $i \in S_y$ is represented by the sequence $\omega^y_i := (r^y_{im})_{m \in M}$, where ω^y_i is a permutation of the rotations in R_y. For ease of reference, Table 1 summarizes the notation used throughout the paper.

In traditional residency schedules, outpatient clinic responsibilities are often subordinated to inpatient requirements, leading to irregular and fragmented clinic exposure. The `Clinic First Model`

Notation	Description
Sets and Indices	
Y	Set of PGYs
R_y	Set of rotations in PGY y
S_y	Set of residents in PGY y
M	Set of time periods
P	Set of positions in the `base rotation schedule`
I_y	Set of unordered pairs of distinct rotations in R_y, i.e., $\{(r_1, r_2) \in R_y \times R_y : r_1 < r_2\}$
$\bar{P}_m^y$	Set of positions in the `base rotation schedule` of PGY y whose rotations are assigned to residents in time period m
Parameters	
γ_y	Number of residents in PGY y
$\bar{c}_r^y$	Number of clinic sessions associated with rotation r in PGY y
$\bar{L}$	Lower bound on `LL`
$\hat{c}_{r_1 r_2 p}^y$	Contribution to the `CTC` from assigning rotations r_1 and r_2 consecutively starting at position p in PGY y
Decision Variables	
x_{rp}^y	Binary variable equal to 1 if rotation r is assigned to position p in the `base rotation schedule` of PGY y, and 0 otherwise
$\eta_{r_1 r_2 p}^y$	Binary variable equal to 1 if rotations r_1 and r_2 are scheduled consecutively starting at position p (in either order) in the `base rotation schedule` of PGY y, and 0 otherwise

Table 1: Description of the notations.

addresses this imbalance by placing greater emphasis on `continuity clinic` while better managing the competing demands of outpatient clinic responsibilities and inpatient training. To assess the quality of a resident schedule, we consider two performance measures that capture the primary goals of the `Clinic First Model`. The first measure, *Clinic Time Consistency* (`CTC`), quantifies how evenly clinic sessions are distributed over the course of the academic year. In our optimization model, `CTC` is the objective to be maximized. A higher value of `CTC` corresponds to a more balanced distribution of clinic sessions across time periods for each resident, preventing extended intervals with either too few or too many clinic sessions. Consequently, maximizing `CTC` promotes regular and consistent outpatient clinic participation throughout the year. For a given schedule, `CTC` is defined as $\texttt{CTC} := \sum_{y \in Y} \sum_{i \in S_y} \sum_{m=2}^{|M|} \left| \bar{c}_{r_{im}}^y - \bar{c}_{r_{i(m-1)}}^y \right|$.

The second performance measure, *Level Loading* (`LL`), evaluates the distribution of clinic capacity across time periods. Specifically, `LL` is defined as the minimum total number of clinic sessions scheduled in any time period: $\texttt{LL} := \min_{m \in M} \left(\sum_{y \in Y} \sum_{i \in S_y} \bar{c}_{r_{im}}^y \right)$. Because the number of clinic sessions in each rotation varies, the total number clinic sessions per time period across all residents could be inconsistent across the year if individual resident schedules are not coordinated at the clinic level. Too few clinic sessions in a time period results in poor access to care for patients, while too many clinic sessions in a time period can result in congestion in the clinic and over-utilization of shared staff and resources. In our proposed model, we require `LL` to exceed a prescribed threshold $\bar{L}$, thereby guaranteeing a minimum level of clinic capacity in every time period. Enforcing a lower bound, $\bar{L}$ on monthly resident clinic sessions forces the model to balance annual capacity, naturally suppressing extreme monthly spikes by redistributing sessions across the year. This requirement is enforced through the constraints $\sum_{y\in Y}\sum_{i\in S_y}\bar{c}^y_{rim} \geq \bar{L},\ \forall m \in M$. These constraints ensure that no time period is underserved by maintaining a minimum number of scheduled clinic sessions throughout the planning horizon. Consequently, they improve the availability of outpatient care while preventing highly uneven allocations of clinic workload across time periods.

The two performance measures play distinct but complementary roles. `CTC` is the primary optimization objective because it directly reflects the educational goal of providing residents with a consistent outpatient experience by distributing clinic sessions evenly throughout the academic year. In contrast, `LL` is imposed as a lower-bound constraint to guarantee a minimum level of clinic capacity in every time period. This allows the model to optimize the quality of individual resident schedules while simultaneously satisfying an essential operational requirement for clinic coverage. Taken together, the `CTC` objective and the `LL` constraints provide a mathematical formulation of the *Clinic First* model by promoting consistent outpatient exposure for individual residents while ensuring balanced clinic capacity across all time periods.

To construct resident schedules, we adopt a *staggered scheduling framework*, in which each PGY is represented by a single `base rotation schedule` and the schedules of all residents are obtained by successively right-shifting this schedule. This approach yields a simple, interpretable schedule structure in which every resident completes the same sequence of rotations in a different temporal order. Consequently, it promotes equitable training experiences while producing schedules that are straightforward to understand, communicate, and administer. As an illustration, consider a PGY y with $|R_y| = |M| = 5$ and $\gamma_y = 4$. Let the `base rotation schedule` be $(\hat{r}^y_1, \hat{r}^y_2, \hat{r}^y_3, \hat{r}^y_4, \hat{r}^y_5)$, where $\hat{r}^y_j \in R_y\ \forall j$ and $\hat{r}^y_i \neq \hat{r}^y_j$ whenever $i \neq j$. Figure 1 illustrates how the schedules of all four residents are generated from this common sequence. The `base rotation schedule` is assigned to Resident 1, while the schedule of Resident 2 is obtained by shifting Resident 1's schedule one time period to the right. Thus, if Resident 1 follows $(\hat{r}^y_1, \hat{r}^y_2, \hat{r}^y_3, \hat{r}^y_4, \hat{r}^y_5)$, then Resident 2 follows $(\hat{r}^y_5, \hat{r}^y_1, \hat{r}^y_2, \hat{r}^y_3, \hat{r}^y_4)$. The schedules of the remaining residents are generated analogously by repeatedly applying a one-period right shift to the preceding resident's schedule. Building on the *staggered scheduling framework*, the next section presents a BIP formulation for determining an optimal `base rotation schedule` for each PGY, while maximizing `CTC` and satisfying the `LL` constraints.

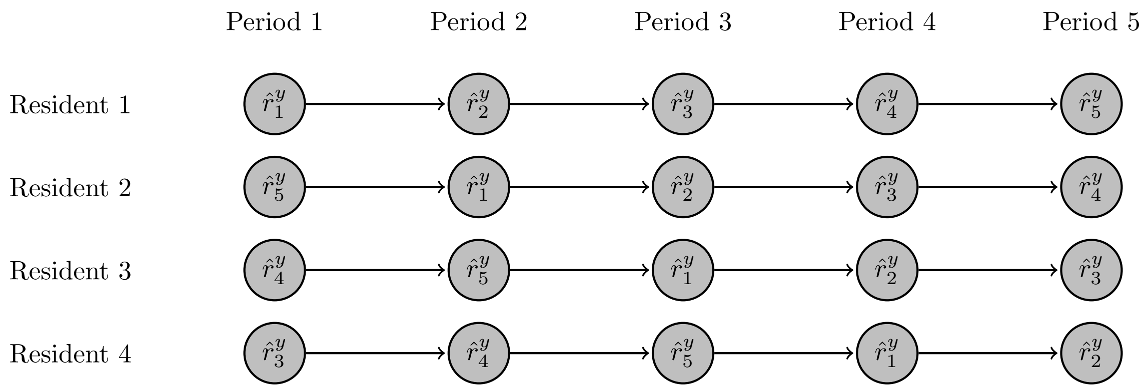


Figure 1: Example of the *staggered scheduling framework*.

4 BIP Formulation for Staggered Scheduling

Let $P := M$ denote the set of positions in a `base rotation schedule`. For each PGY $y \in Y$, let x^y_{rp} be a binary variable that equals 1 if rotation $r \in R_y$ is assigned to position $p \in P$ in the `base rotation schedule`, and 0 otherwise. To model consecutive rotations, define $I_y := \{(r_1, r_2) \in R_y \times R_y : r_1 < r_2\}$, the set of all unordered pairs of distinct rotations in PGY y. For each $(r_1, r_2) \in I_y$ and $p \in P$, let $\eta^y_{r_1 r_2 p}$ be a binary variable that equals 1 if rotations r_1 and r_2 occupy consecutive positions in the `base rotation schedule`, with the first rotation appearing in position p, and 0 otherwise. Here position 1 is treated as the successor of position $|P|$. To relate the `base rotation schedule` to the staggered resident schedules, for each PGY $y \in Y$ and $m \in M$, let $\bar{P}^y_m \subseteq P$ denote the set of positions in the `base rotation schedule` whose assigned rotations appear in time period m in the resident schedules generated under the right-shifting *staggered scheduling framework*. For example, consider the staggered schedule in Figure 1. During the first time period, the assigned rotations are $\hat{r}^y_1$, $\hat{r}^y_5$, $\hat{r}^y_4$, and $\hat{r}^y_3$. These rotations occupy positions 1, 5, 4, and 3 in the `base rotation schedule`, respectively, yielding $\bar{P}^y_1 = 1, 5, 4, 3$. Thus, under the *staggered scheduling framework*,

$$\bar{P}^y_m = \begin{cases} m - i + 1, & 1 \le i \le m, \\ |M| - (i - m - 1), & m + 1 \le i \le \gamma_y \end{cases} : i = 1, \ldots, \gamma_y. \tag{1}$$

Let $\hat{c}^y_{r_1 r_2 p}$ denote the contribution to `CTC` from assigning rotations r_1 and r_2 to consecutive positions p and $p + 1$, respectively, in the `base rotation schedule` of PGY y. Thus, for each PGY $y \in Y$,

$(r_1, r_2) \in I_y$, and $p \in P$,

$$\hat{c}^y_{r_1 r_2 p} = \begin{cases} \gamma_y \, |\bar{c}^y_{r_1} - \bar{c}^y_{r_2}|, & \text{if } p \leq |P| - \gamma_y, \\ (\gamma_y - 1) \, |\bar{c}^y_{r_1} - \bar{c}^y_{r_2}|, & \text{if } p > |P| - \gamma_y. \end{cases} \tag{2}$$

The weighting by γ_y in (2) accounts for the number of resident schedules to which each consecutive rotation pair contributes under the *staggered scheduling framework*. If the pair begins at a position $p \leq |P| - \gamma_y$ in `base rotation schedule`, it appears consecutively in the schedules of all γ_y residents. Otherwise, it appears consecutively in only $\gamma_y - 1$ resident schedules. Using these notations, we formulate the following BIP model to determine an optimal `base rotation schedule` for each PGY.

$$(\mathcal{H}) \quad \max \sum_{y \in Y} \sum_{(r_1, r_2) \in I_y} \sum_{p \in P} \hat{c}^y_{r_1 r_2 p} \, \eta^y_{r_1 r_2 p} \tag{3a}$$

$$\sum_{r \in R_y} x^y_{rp} = 1 \qquad \forall \, y \in Y, \; p \in P \tag{3b}$$

$$\sum_{p \in P} x^y_{rp} = 1 \qquad \forall \, y \in Y, \; r \in R_y \tag{3c}$$

$$\eta^y_{r_1 r_2 p} \leq x^y_{r_1 p} + x^y_{r_1 \, p+1} \qquad \forall \, p \in P, \; (r_1, r_2) \in I_y, \; y \in Y \tag{3d}$$

$$\eta^y_{r_1 r_2 p} \leq x^y_{r_2 p} + x^y_{r_2 \, p+1} \qquad \forall \, p \in P, \; (r_1, r_2) \in I_y, \; y \in Y \tag{3e}$$

$$\sum_{y \in Y} \sum_{p \in \bar{P}^y_m} \sum_{r \in R_y} \bar{c}^y_r \, x^y_{rp} \geq \bar{L} \qquad \forall \, m \in M \tag{3f}$$

$$\eta^y_{r_1 r_2 p} \in \{0, 1\} \qquad \forall \, p \in P, \; (r_1, r_2) \in I_y, \; y \in Y \tag{3g}$$

$$x^y_{rp} \in \{0, 1\} \qquad \forall \, r \in R_y, \; p \in P, \; y \in Y. \tag{3h}$$

The objective function (3a) maximizes the `CTC`. For each PGY, Constraints (3b) and (3c) ensure that each position is assigned exactly one rotation and each rotation is assigned to exactly one position. Constraint (3d) permits $\eta^y_{r_1 r_2 p}$ to equal 1 only if rotation r_1 is assigned to either position p or $p+1$, while Constraint (3e) imposes the analogous condition for rotation r_2. Together, these constraints ensure that $\eta^y_{r_1 r_2 p}$ is 1 only when rotations r_1 and r_2 occupy consecutive positions in the `base rotation schedule` starting at position p. Finally, Constraint (3f) enforces the `LL` requirement by ensuring that the total number of clinic sessions scheduled across all residents in every time period are at-least the prescribed lower bound $\bar{L}$.

While the model does not explicitly include program-specific constraints due to their idiosyncratic nature, it is generalizable in its design and allows these rules to be easily integrated. Such program-specific constraints typically fall into three categories: contractual coverage requirements (e.g., staffing Emergency Medicine or OB/GYN shifts), spacing guidelines that prevent repetitive rotations from occurring back-to-back, and educational sequencing rules that dictate when specific training blocks must occur during the academic year.

In the following proposition, we show that the optimal rotation schedules generated by $\mathcal{H}$ under the right-shifting *staggered scheduling framework* are also optimal under its left-shifting counterpart. Therefore, without loss of generality, we restrict our attention to the right-shifting *staggered scheduling framework* throughout this paper.

Proposition 1. *Let $[\bar{\omega}_i^y]_{i \in R_y, y \in Y}$ denote the rotation schedules of all residents obtained from an optimal solution of $\mathcal{H}$. Then, there exists an optimal* `base rotation schedule` *for the analogous formulation under the left-shifting counterpart that generates the same set of rotation schedules.*

Proof. Let $\mathcal{H}^L$ denote the formulation analogous to $\mathcal{H}$ under the left-shifting counterpart of the *staggered scheduling framework* and fix some PGY $y \in Y$. Without loss of generality, let $\bar{\omega}_1^y$ denote an optimal `base rotation schedule` obtained from $\mathcal{H}$, and let $\bar{\omega}_{\gamma_y}^y$ denote the last rotation schedule generated by successively right-shifting $\bar{\omega}_1^y$. Then, $\bar{\omega}_{\gamma_y}^y$ constitutes a feasible `base rotation schedule` for $\mathcal{H}^L$. Successively left-shifting $\bar{\omega}_{\gamma_y}^y$ generates the same set of rotation schedules as those obtained from $\mathcal{H}$ and, therefore, yields the same `CTC`.

Suppose, by contradiction, that $\bar{\omega}_{\gamma_y}^y$ is not an optimal `base rotation schedule` for $\mathcal{H}^L$. Then, $\mathcal{H}^L$ admits an optimal `base rotation schedule` $\tilde{\omega}_{\gamma_y}^y$ that yields a strictly higher `CTC`. Let $\tilde{\omega}_1^y$ denote the last rotation schedule obtained by successively left-shifting $\tilde{\omega}_{\gamma_y}^y$. By construction, $\tilde{\omega}_1^y$ is a feasible `base rotation schedule` for $\mathcal{H}$, and successively right-shifting it generates the same set of rotation schedules as those generated from $\tilde{\omega}_{\gamma_y}^y$. Hence, $\tilde{\omega}_1^y$ yields the same `CTC` as $\tilde{\omega}_{\gamma_y}^y$ and, therefore, a strictly higher `CTC` than $\bar{\omega}_1^y$. This contradicts the optimality of $\bar{\omega}_1^y$ for $\mathcal{H}$. $\square$

Next, we establish a theoretical bound on the loss in `CTC` due to the *staggered scheduling framework*. Let $\mathcal{P}$ (presented in Appendix A) denote the BIP formulation for `FMRS` without the *staggered scheduling framework*, which maximizes `CTC` subject to the `LL` constraints. Since $\mathcal{P}$ is not restricted to schedules satisfying the *staggered scheduling framework*, we refer to it as the *unrestricted formulation*.

Proposition 2. *For each PGY $y \in Y$, let $\underline{\nu}^y = \min_{\substack{r_1, r_2 \in R_y \\ r_1 \neq r_2}} |\bar{c}_{r_1}^y - \bar{c}_{r_2}^y|$ and $\bar{\nu}^y = \max_{\substack{r_1, r_2 \in R_y \\ r_1 \neq r_2}} |\bar{c}_{r_1}^y - \bar{c}_{r_2}^y|$. Furthermore, let $z_{\mathcal{P}}^*$ and $z_{\mathcal{H}}^*$ denote the optimal objective values of $\mathcal{P}$ and $\mathcal{H}$, respectively. Then,$z_{\mathcal{P}}^* - z_{\mathcal{H}}^* \leq \sum_{y \in Y} \left((\gamma_y |R_y| - 1)\bar{\nu}^y - (\gamma_y - 1)\underline{\nu}^y\right)$.*

Proof. Let $z^*_{\mathcal{H}_{NL}}$ denote the optimal objective value of $\mathcal{H}$ without the `LL` constraints (3f) and $\tilde{\omega}^y_1$ be the rotation schedule for PGY y with the maximum single resident `CTC` value, $\bar{C}^y$. We have the following inequalities:

$$z^*_{\mathcal{P}} \leq \sum_{y \in Y} \gamma_y \bar{C}^y, \tag{4a}$$

$$z^*_{\mathcal{H}_{NL}} \geq \sum_{y \in Y} \gamma_y \bar{C}^y - \sum_{y \in Y} (\gamma_y - 1) \left(\overline{\nu}^y - \underline{\nu}^y \right), \tag{4b}$$

$$z^*_{\mathcal{H}_{NL}} \leq \sum_{y \in Y} \gamma_y (|R_y| - 1) \overline{\nu}^y. \tag{4c}$$

Inequality (4a) follows immediately from the definition of $\bar{C}^y$. To establish inequality (4b), we exploit the structure of the *staggered scheduling framework*. $\tilde{\omega}^y_1$ is a feasible `base rotation schedule` for $\mathcal{H}$ without the `LL` constraints (3f) where the `base rotation schedule` attains `CTC` of $\bar{C}^y$. Each subsequent schedule is obtained by a one-period right shift of the preceding schedule and therefore differs from the `base rotation schedule` by the removal of one consecutive rotation pair and the addition of another. Consequently, the `CTC` value of each shifted schedule can decrease by at most $\overline{\nu}^y - \underline{\nu}^y$, implying that every such schedule has clinic time consistency of at least $\bar{C}^y - (\overline{\nu}^y - \underline{\nu}^y)$. Since there are $\gamma_y - 1$ shifted schedules in PGY y, summing the `CTC` contribution of the `base rotation schedule` and these lower bounds over all other schedules establishes inequality (4b).

Inequality (4c) follows from the observation that each resident schedule in PGY y contains exactly $|R_y| - 1$ consecutive rotation pairs. Since the contribution of any consecutive rotation pair to the `CTC` is at most $\overline{\nu}^y$, `CTC` of an individual resident schedule is bounded above by $(|R_y| - 1)\overline{\nu}^y$. Summing this bound over all γ_y residents and then over all PGYs establishes inequality (4c).

From (4a) and (4b), we obtain $z^*_{\mathcal{P}} - z^*_{\mathcal{H}_{NL}} \leq \sum_{y \in Y} (\gamma_y - 1)(\overline{\nu}^y - \underline{\nu}^y)$. From (4c) and $z^*_{\mathcal{H}} \geq 0$ we obtain $z^*_{\mathcal{H}_{NL}} - z^*_{\mathcal{H}} \leq \sum_{y \in Y} \gamma_y (|R_y| - 1)\overline{\nu}^y$. Adding these two inequalities, we get $z^*_{\mathcal{P}} - z^*_{\mathcal{H}} \leq \sum_{y \in Y} \left((\gamma_y |R_y| - 1)\overline{\nu}^y - (\gamma_y - 1)\underline{\nu}^y \right)$. □

Proposition 2 provides a bound on the loss in `CTC` due to the *staggered scheduling framework*. This bound decreases as the number of residents, the number of rotations, and $\overline{\nu}^y$ decrease, and as $\underline{\nu}^y$ increases. Furthermore, as expected, the bound on the loss in `CTC` under the *staggered scheduling framework* approaches zero as $\overline{\nu}^y$ approaches zero. When $\overline{\nu}^y$ is small, the number of clinic sessions differs only slightly across rotations. Consequently, replacing one pair of consecutive rotations with another can result in only a small change in the corresponding `CTC` contribution, thereby limiting the loss associated with the *staggered scheduling framework*.

5 Valid Inequalities for $\mathcal{H}$

Our computational experiments indicate that the linear programming (LP) relaxation of $\mathcal{H}$ is weak, leading to poor computational performance. Thus, in this section, we derive two families of Valid Inequalities to strengthen formulation $\mathcal{H}$, namely VI-a and VI-b. We further show that VI-b provides a complete description of the convex hull of the discrete set defined by VI-a and a subset of the constraints in $\mathcal{H}$. Computational experiments (Section 6) demonstrate that incorporating VI-b substantially reduces the solution time required to solve $\mathcal{H}$.

Proposition 3. *The following inequalities are valid for $\mathcal{H}$:*

$$\text{(VI-a)} \qquad \sum_{r \in R_y : r < k} \eta^y_{rkp} + \sum_{r \in R_y : r > k} \eta^y_{krp} \leq 1 \qquad \forall p \in P, k \in R_y, y \in Y. \tag{5}$$

Proof. Let $(\mathbf{x}^*, \boldsymbol{\eta}^*)$ be a feasible solution to $\mathcal{H}$ such that $\sum_{r \in R_y : r < k} \eta^{y*}_{rkp} + \sum_{r \in R_y : r > k} \eta^{y*}_{krp} > 1$ for some $y \in Y$, $k \in R_y$, and $p \in P$. By (3g), there exist distinct rotations $r_1, r_2 \in R_y \setminus k$ satisfying one of the following:

- $r_1 < r_2 < k$, with $\eta^{y*}_{r_1 kp} = 1$ and $\eta^{y*}_{r_2 kp} = 1$;
- $r_1 < k < r_2$, with $\eta^{y*}_{r_1 kp} = 1$ and $\eta^{y*}_{kr_2 p} = 1$;
- $k < r_1 < r_2$, with $\eta^{y*}_{kr_1 p} = 1$ and $\eta^{y*}_{kr_2 p} = 1$.

In the first case where $r_1 < r_2 < k$, from (3d) and (3e) imply $x^{y*}_{r_1 p} + x^{y*}_{r_1 p+1} \geq 1$, $x^{y*}_{r_2 p} + x^{y*}_{r_2 p+1} \geq 1$, $x^{y*}_{kp} + x^{y*}_{kp+1} \geq 1$. Together with (3c) and (3h), this implies that rotations r_1, r_2, and k must all be assigned to positions p and $p+1$, contradicting (3b). The remaining two cases lead to analogous contradictions. □

The inequalities VI-a enforce that each rotation $k \in R_y$ can be paired with at most one other rotation $r \neq k$ to form a consecutive pair in the `base rotation schedule`, with the earlier rotation of the pair assigned to position p. In the following proposition, we strengthen VI-a.

Proposition 4. *The following inequalities are valid for $\mathcal{H}$:*

$$\text{(VI-b)} \qquad \sum_{r \in R_y : r > k} \eta^y_{krp} + \sum_{r \in R_y : r < k} \eta^y_{rkp} \leq x^y_{kp} + x^y_{kp+1} \qquad \forall p \in P, k \in R_y, y \in Y. \tag{6}$$

Proof. Let $(\mathbf{x}^*, \boldsymbol{\eta}^*)$ be a feasible solution to $\mathcal{H}$ such that $\sum_{r\in R_y:r>k} \eta^{y*}_{krp} + \sum_{r\in R_y:r<k} \eta^{y*}_{rkp} > x^{y*}_{kp} + x^{y*}_{kp+1}$ for some $y \in Y$, $k \in R_y$, and $p \in P$. By VI-a and (3c), it follows that $\sum_{r\in R_y:r>k} \eta^{y*}_{krp} + \sum_{r\in R_y:r<k} \eta^{y*}_{rkp} = 1$, and $x^{y*}_{kp} = x^{y*}_{kp+1} = 0$. By (3d), (3e), and (3g), this implies $\eta^{y*}_{krp} = 0, \quad \forall r \in R_y : r > k$, and $\eta^{y*}_{rkp} = 0, \quad \forall r \in R_y : r < k$, contradicting $\sum_{r\in R_y:r>k} \eta^{y*}_{krp} + \sum_{r\in R_y:r<k} \eta^{y*}_{rkp} = 1$. □

VI-b ensure that each rotation $k \in R_y$ can be paired with at most one other rotation $r \neq k$ to form a consecutive pair in the `base rotation schedule`, with the earlier rotation of the pair assigned to position p only if k is assigned to position p or $p+1$. VI-b weakly dominates VI-a, since both families share the same left-hand side, while the right-hand side of VI-b satisfies $x^y_{kp} + x^y_{k,p+1} \leq 1$ due to (3c).

Next, we establish the strength of VI-b by proving that they describe the convex hull of the discrete set defined by VI-a and a subset of the constraints in $\mathcal{H}$. Specifically, let $\mathcal{Z}^y_k$ denote the set defined by the following constraints:

$$\sum_{p\in P} x^y_{kp} = 1 \tag{7a}$$

$$\eta^y_{krp} \leq x^y_{kp} + x^y_{kp+1} \qquad \forall p \in P, r \in R_y : r > k \tag{7b}$$

$$\eta^y_{rkp} \leq x^y_{kp} + x^y_{kp+1} \qquad \forall p \in P, r \in R_y : r < k \tag{7c}$$

$$\sum_{r\in R_y:r<k} \eta^y_{rkp} + \sum_{r\in R_y:r>k} \eta^y_{krp} \leq 1 \quad \forall p \in P \tag{7d}$$

$$\eta^y_{r_1 r_2 p} \in \{0,1\} \qquad \forall p \in P, (r_1, r_2) \in I_y : k \in \{r_1, r_2\} \tag{7e}$$

$$x^y_{kp} \in \{0,1\} \qquad \forall p \in P. \tag{7f}$$

The following proposition shows that replacing constraints (7b)-(7d) in (7) with VI-b and relaxing the integer restriction yields $\operatorname{conv}(\mathcal{Z}^y_k)$.

Proposition 5. $\operatorname{conv}(\mathcal{Z}^y_k)$ *is given by the following set of constraints:*

$$\sum_{p\in P} x^y_{kp} = 1 \tag{8a}$$

$$\sum_{r\in R_y:r<k} \eta^y_{rkp} + \sum_{r\in R_y:r>k} \eta^y_{krp} \leq x^y_{kp} + x^y_{kp+1} \quad \forall p \in P \tag{8b}$$

$$\eta^y_{r_1 r_2 p} \geq 0 \qquad \forall p \in P, (r_1, r_2) \in I_y : k \in \{r_1, r_2\} \tag{8c}$$

$$x^y_{kp} \geq 0 \qquad \forall p \in P. \tag{8d}$$

Proof. Let $\mathcal{Q}^y_k$ denote the polyhedron defined by (8), and let $\bar{\mathcal{Q}}^y_k$ be the set of binary points satisfying

(8). We show that (i) $\mathcal{Z}_k^y = \bar{\mathcal{Q}}_k^y$ and (ii) *Vertices of* $\mathcal{Q}_k^y$ *are binary.* Together, these imply $\mathcal{Q}_k^y = \operatorname{conv}(\bar{\mathcal{Q}}_k^y) = \operatorname{conv}(\mathcal{Z}_k^y)$.

(i) $\mathcal{Z}_k^y = \bar{\mathcal{Q}}_k^y$*:* The inclusion $\mathcal{Z}_k^y \subseteq \bar{\mathcal{Q}}_k^y$ follows because (8b) is valid for $\mathcal{Z}_k^y$, as shown in Proposition 4. Conversely, $\bar{\mathcal{Q}}_k^y \subseteq \mathcal{Z}_k^y$ because every feasible solution in $\bar{\mathcal{Q}}_k^y$ satisfies the following inequalities derived from (8b):

$$\eta_{krp}^y \le \sum_{r\in R_y:r<k} \eta_{rkp}^y + \sum_{r\in R_y:r>k} \eta_{krp}^y \le x_{kp}^y + x_{kp+1}^y \quad \forall p \in P, r \in R_y : r > k \tag{9a}$$

$$\eta_{rkp}^y \le \sum_{r\in R_y:r<k} \eta_{rkp}^y + \sum_{r\in R_y:r>k} \eta_{krp}^y \le x_{kp}^y + x_{kp+1}^y \quad \forall p \in P, r \in R_y : k < r \tag{9b}$$

$$\sum_{r\in R_y:r<k} \eta_{rkp}^y + \sum_{r\in R_y:r>k} \eta_{krp}^y \le x_{kp}^y + x_{kp+1}^y \le 1 \qquad \forall p \in P. \tag{9c}$$

The first inequalities in (9a) and (9b) follow from the non-negativity of $\boldsymbol{\eta}_k^y$, while the second inequality in (9c) follows from (8a) and the non-negativity of $\mathbf{x}_k^y$.

(ii) *Vertices of* $\mathcal{Q}_k^y$ *are binary:* Consider the following linear program that maximizes an arbitrary linear objective over $\mathcal{Q}_k^y$:

$$(\mathcal{W}) \quad \max \sum_{p\in P} \tilde{f}_{kp}^y x_{kp}^y + \sum_{r\in R_y:r>k} \sum_{p\in P} \hat{f}_{krp}^y \eta_{krp}^y + \sum_{r\in R_y:r<k} \sum_{p\in P} \hat{f}_{rkp}^y \eta_{rkp}^y \tag{10a}$$

$$[\alpha] \quad \sum_{p\in P} x_{kp}^y = 1 \tag{10b}$$

$$[\beta_p] \quad \sum_{r\in R_y:r>k} \eta_{krp}^y + \sum_{r\in R_y:r<k} \eta_{rkp}^y \le x_{kp}^y + x_{kp+1}^y \quad \forall p \in P \tag{10c}$$

$$\eta_{r_1r_2p}^y \ge 0 \quad \forall p \in P, (r_1, r_2) \in I_y : k \in \{r_1, r_2\} \tag{10d}$$

$$x_{kp}^y \ge 0 \quad \forall p \in P. \tag{10e}$$

The corresponding dual variables are shown beside the constraints. The dual of $\mathcal{W}$ is given by:

$$(\mathcal{W}_D) \quad \min \ \alpha \tag{11a}$$

$$\alpha - \beta_p - \beta_{p-1} \ge \tilde{f}_{kp}^y \quad \forall p \in P \tag{11b}$$

$$\beta_p \ge \hat{f}_{krp}^y \quad \forall r \in R_y : r > k, p \in P \tag{11c}$$

$$\beta_p \ge \hat{f}_{rkp}^y \quad \forall r \in R_y : r < k, p \in P \tag{11d}$$

$$\beta_p \ge 0 \quad \forall p \in P, \tag{11e}$$

where $\beta_0 = \beta_{|P|}$.

$\bar{\beta}_p = \max\Big(0, \max_{r \in R_y : r > k} \hat{f}^y_{krp}, \max_{r \in R_y : r < k} \hat{f}^y_{rkp}\Big) \ \forall p \in P$, $\bar{\alpha} = \max_{p \in P} \left\{ \tilde{f}^y_{kp} + \bar{\beta}_p + \bar{\beta}_{p-1} \right\}$ is a feasible solution of $\mathcal{W}_D$. By the weak duality theorem, the optimal value of $\mathcal{W}$ is at most $\bar{\alpha}$. Let $\bar{p} = \mathrm{argmax}_{p \in P} \left\{ \tilde{f}^y_{kp} + \bar{\beta}_p + \bar{\beta}_{p-1} \right\}$, and without the loss of generality, assume that $(k, \bar{r}_{\bar{p}}) = \mathrm{argmax}_{(r_1,r_2) \in I_y : k \in \{r_1,r_2\}} \hat{f}^y_{r_1 r_2 \bar{p}}$, $(k, \bar{r}_{\bar{p}-1}) = \mathrm{argmax}_{(r_1,r_2) \in I_y : k \in \{r_1,r_2\}} \hat{f}^y_{r_1 r_2 \bar{p}-1}$. Let $(\tilde{\mathbf{x}}^y_k, \tilde{\boldsymbol{\eta}}^y_k)$, be such that

- $\tilde{x}^y_{k\bar{p}} = 1$, $x^y_{kp} = 0 \ \forall p \in P, p \neq \bar{p}$,
- $\tilde{\eta}^y_{r_1 r_2 p} = 0 \ \forall p \in P \setminus \{\bar{p}, \bar{p}-1\}, (r_1, r_2) \in I_y : k \in \{r_1, r_2\}$,
- $\tilde{\eta}^y_{k\bar{r}_{\bar{p}}\bar{p}} = 1$ if $\hat{f}^y_{k\bar{r}_{\bar{p}}\bar{p}} > 0$ and 0, otherwise; $\tilde{\eta}^y_{r_1,r_2\bar{p}} = 0 \ \forall (r_1, r_2) \in \{I_y | k \in \{r_1, r_2\}, \bar{r}_{\bar{p}} \notin \{r_1, r_2\}\}$,
- $\tilde{\eta}^y_{k\bar{r}_{\bar{p}-1}\bar{p}-1} = 1$ if $\hat{f}^y_{k\bar{r}_{\bar{p}-1}\bar{p}-1} > 0$ and 0, otherwise; $\tilde{\eta}^y_{r_1,r_2\bar{p}-1} = 0 \ \forall (r_1, r_2) \in \{I_y | k \in \{r_1, r_2\}, \bar{r}_{\bar{p}-1} \notin \{r_1, r_2\}\}$.

The solution $(\tilde{\mathbf{x}}^y_k, \tilde{\boldsymbol{\eta}}^y_k)$ is feasible for $\mathcal{W}$ and attains the objective function value $\bar{\alpha}$. Hence, it is optimal for $\mathcal{W}$. Moreover, since $(\tilde{\mathbf{x}}^y_k, \tilde{\boldsymbol{\eta}}^y_k)$ is integral and the objective function coefficients of $\mathcal{W}$ are arbitrary, all vertices of $\mathcal{Q}^y_k$ are binary.

□

Based on Proposition 5, we strengthen $\mathcal{H}$ by replacing constraints (3d) and (3e) with VI-b. The resulting formulation is stated below:

$$(\bar{\mathcal{H}}) \quad \max \sum_{y \in Y} \sum_{(r_1,r_2) \in I_y} \sum_{p \in P} \hat{c}^y_{r_1 r_2 p} \eta^y_{r_1 r_2 p} \ \text{ s.t. } \quad \text{(3b), (3c), VI-b, (3f)} - \text{(3h)}. \tag{12}$$

6 Computational Experiments

The computational experiments are based on historical data from the UMN Family Medicine program. Each rotation spans multiple weeks. For each rotation type, we estimate the average weekly number of clinic sessions from the historical data (see Table 7 in Appendix D) and use this estimate as the parameter of a Poisson distribution to generate random instances of weekly clinic-session counts for each rotation. The total number of clinic sessions assigned to a rotation is then obtained by multiplying the realized weekly count by the rotation duration (in weeks). We consider two discretizations of the academic year, consisting of $|M| \in \{12, 24\}$ time periods. Consequently, each PGY is partitioned into

Case	Runtime (s)		Opt-Gap (%)		LB-I	UB-I
$\gamma \mid \phi$	$\mathcal{H}$	$\bar{\mathcal{H}}$	$\mathcal{H}$	$\bar{\mathcal{H}}$		
9 | 0.90	3600 (3600)	2 (5)	1313.14 (1344.32)	0 (0)	0.54 (0.15)	92.25 (91.89)
9 | 0.95	3600 (3600)	2 (4)	1258.43 (1354.30)	0 (0)	0.65 (0.31)	92.20 (91.86)
9 | 0.98	3600 (3600)	8 (10)	1028.58 (1393.45)	0 (0)	1.89 (0.76)	92.22 (91.92)
12 | 0.90	3600 (3600)	2 (5)	833.73 (1072.72)	0 (0)	0.28 (0.00)	92.41 (92.16)
12 | 0.95	3600 (3600)	2 (3)	1049.90 (1083.11)	0 (0)	0.65 (0.15)	92.32 (92.10)
12 | 0.98	3600 (3600)	3 (6)	1056.01 (1094.19)	0 (0)	0.80 (0.12)	92.37 (91.95)
15 | 0.90	3600 (3600)	2 (4)	1047.04 (1382.49)	0 (0)	0.41 (0.00)	92.35 (92.11)
15 | 0.95	3600 (3600)	3 (5)	1024.25 (1375.22)	0 (0)	0.41 (0.00)	92.35 (92.09)
15 | 0.98	3600 (3600)	4 (7)	1295.04 (1401.72)	0 (0)	0.68 (0.22)	92.40 (91.90)

Table 2: Comparison of $\bar{\mathcal{H}}$ with $\mathcal{H}$ for $|M|= 24$.

either 12 four-week rotations or 24 two-week rotations. For $|M|= 12$, we consider instances with $\gamma_y \in \{4, 6, 8\}$ for all $y \in Y$, whereas for $|M|= 24$, we consider $\gamma_y \in \{9, 12, 15\}$ for all $y \in Y$. Since we consider the same number of residents in each PGY, we omit the y index from γ throughout this section. We set $\bar{L} = \lfloor \phi \tilde{\texttt{LL}} \rfloor$, where $\phi \in \{0.90, 0.95, 0.98\}$ is a scaling factor and $\tilde{\texttt{LL}}$ is an upper bound on LL. The computation of $\tilde{\texttt{LL}}$ is described in Appendix C. We conduct a full-factorial experimental design over $|M|$, γ, and ϕ. For each parameter combination, we generate five random instances, resulting in a total of 90 test instances. All computational experiments are performed in Python using Gurobi 11.0 on a workstation equipped with an Intel Xeon Gold 6132 processor and 64 GB of RAM.

The remainder of this section is organized as follows. Section 6.1 evaluates the computational performance of $\bar{\mathcal{H}}$. Section 6.2 compares the schedules produced by $\bar{\mathcal{H}}$ with those implemented in current practice. Section 6.3 analyzes the structural properties of the schedules produced by $\bar{\mathcal{H}}$. Finally, Section 6.4 compares the schedules produced by $\bar{\mathcal{H}}$ with those obtained from its analogous formulation using the objective function proposed by Bard et al. (2016b).

6.1 Computational Performance Analysis

In this section, we compare the computational performance of $\mathcal{H}$ and $\bar{\mathcal{H}}$. We also evaluate the computational benefits of warm-starting $\mathcal{P}$ with the optimal solution of $\bar{\mathcal{H}}$. Finally, we quantify the loss in CTC due to the *staggered scheduling framework*. All BIP formulations are solved using Gurobi's branch-and-bound algorithm with a time limit of 3,600 seconds.

For each of the nine combinations of γ and ϕ, Tables 2 and 3 compare the computational performance of $\mathcal{H}$ and $\bar{\mathcal{H}}$ for $|M|= 24$ and $|M|= 12$, respectively. Columns 2–3 report the average (maximum)

Case	Runtime (s)		Opt-Gap (%)		LB-I	UB-I
$\gamma \mid \phi$	$\mathcal{H}$	$\bar{\mathcal{H}}$	$\mathcal{H}$	$\bar{\mathcal{H}}$		
4 | 0.90	3600 (3600)	0.11 (0.17)	234.38 (276.47)	0 (0)	0.16 (0.00)	69.60 (63.52)
4 | 0.95	3600 (3600)	0.34 (0.46)	260.53 (278.46)	0 (0)	0.34 (0.00)	72.13 (70.88)
4 | 0.98	3600 (3600)	0.69 (1.42)	243.76 (277.25)	0 (0)	1.29 (0.20)	70.40 (67.05)
6 | 0.90	3600 (3600)	0.12 (0.27)	233.17 (281.00)	0 (0)	0.03 (0.00)	69.22 (61.48)
6 | 0.95	3600 (3600)	0.33 (0.49)	253.43 (297.01)	0 (0)	0.23 (0.00)	71.41 (66.89)
6 | 0.98	3600 (3600)	0.51 (0.70)	252.90 (294.13)	0 (0)	0.20 (0.00)	71.51 (70.06)
8 | 0.90	3600 (3600)	0.30 (1.23)	205.73 (302.68)	0 (0)	0.06 (0.00)	66.56 (62.96)
8 | 0.95	3600 (3600)	0.29 (0.68)	250.50 (336.91)	0 (0)	0.34 (0.10)	70.85 (67.45)
8 | 0.98	3600 (3600)	0.40 (0.53)	260.16 (332.47)	0 (0)	0.33 (0.20)	71.88 (69.89)

Table 3: Comparison of $\bar{\mathcal{H}}$ and $\mathcal{H}$ for $|M|= 12$.

solution times, in seconds, while Columns 4–5 report the average (maximum) optimality gaps (%) for $\mathcal{H}$ and $\bar{\mathcal{H}}$, respectively. Columns 6 and 7 report the average (minimum) percentage improvements in the best lower bound (LB-I) and best upper bound (UB-I) obtained by $\bar{\mathcal{H}}$ relative to $\mathcal{H}$, computed as $\texttt{LB-I} = \frac{LB^{\bar{\mathcal{H}}}-LB^{\mathcal{H}}}{LB^{\mathcal{H}}} \times 100$, $\texttt{UB-I} = \frac{UB^{\mathcal{H}}-UB^{\bar{\mathcal{H}}}}{UB^{\mathcal{H}}} \times 100$. Here, $LB^{\mathcal{H}}$ and $LB^{\bar{\mathcal{H}}}$ denote the best lower bounds obtained by $\mathcal{H}$ and $\bar{\mathcal{H}}$, respectively, whereas $UB^{\mathcal{H}}$ and $UB^{\bar{\mathcal{H}}}$ denote the corresponding best upper bounds.

As shown in Tables 2 and 3, replacing constraints (3d) and (3e) in $\mathcal{H}$ with VI-b yields a significant improvement in computational performance. For every test instance and for both values of $|M|$, $\mathcal{H}$ reaches the 3600-second time limit with large optimality gaps, whereas $\bar{\mathcal{H}}$ solves every instance to optimality within 10 seconds. The poor performance of $\mathcal{H}$ is primarily due to the weakness of its LP relaxation. Although the solver is able to identify high-quality feasible solutions, it is unable to sufficiently tighten the upper bound within the allotted time, resulting in large optimality gaps. By contrast, incorporating VI-b significantly strengthens the formulation. This observation is supported by Columns 6-7 of Tables 2 and 3. Specifically, the improvement in the best lower bound is consistently small, indicating that $\mathcal{H}$ already produces near-optimal feasible solutions. In contrast, the improvement in the best upper bound exceeds 66% on average for all test instances, demonstrating that the primary benefit of VI-b is a substantially tighter relaxation, which enables the solver to close the optimality gap rapidly.

For each of the nine combinations of γ and ϕ, Tables 4 and 5 compare the computational performance of the *unrestricted formulation*, $\mathcal{P}$, with its warm-started counterpart, denoted by $\mathcal{P}^{\bar{\mathcal{H}}}$, in which $\mathcal{P}$ is initialized with the optimal solution of $\bar{\mathcal{H}}$. The two tables report results for $|M|= 24$

Case	%-Inst-F		%-Inst-S		Runtime (seconds)		δ (%)
$\gamma \mid \phi$	$\mathcal{P}$	$\mathcal{P}^{\bar{\mathcal{H}}}$	$\mathcal{P}$	$\mathcal{P}^{\bar{\mathcal{H}}}$	$\mathcal{P}$	$\mathcal{P}^{\bar{\mathcal{H}}}$	
9 | 0.90	100	100	100	100	2659 (3232)	17 (17)	0.69 (0.99)
9 | 0.95	100	100	40	100	3516 (3602)	17 (17)	0.85 (0.99)
9 | 0.98	0	100	0	100	3601 (3601)	78 (320)	0.89 (1.17)
12 | 0.90	20	100	0	100	3603 (3610)	24 (25)	0.81 (0.99)
12 | 0.95	0	100	0	100	3600 (3601)	24 (25)	0.91 (1.00)
12 | 0.98	0	100	0	100	3600 (3601)	24 (24)	0.92 (0.99)
15 | 0.90	0	100	0	100	3600 (3600)	28 (29)	0.99 (1.00)
15 | 0.95	0	100	0	100	3600 (3600)	29 (29)	0.99 (1.00)
15 | 0.98	0	100	0	100	3600 (3600)	29 (29)	0.99 (1.00)

Table 4: Comparison of $\mathcal{P}^{\bar{\mathcal{H}}}$ with $\mathcal{P}$ for $|M|$=24.

Case	%-Inst-F		%-Inst-S		Runtime (seconds)		δ (%)
$\gamma \mid \phi$	$\mathcal{P}$	$\mathcal{P}^{\bar{\mathcal{H}}}$	$\mathcal{P}$	$\mathcal{P}^{\bar{\mathcal{H}}}$	$\mathcal{P}$	$\mathcal{P}^{\bar{\mathcal{H}}}$	
4 | 0.90	100	100	100	100	4 (7)	1 (2)	0.36 (1.17)
4 | 0.95	100	100	100	100	6 (11)	1 (2)	0.63 (1.57)
4 | 0.98	100	100	100	100	22 (53)	4 (15)	0.69 (1.59)
6 | 0.90	100	100	100	100	7 (8)	3 (6)	0.60 (1.58)
6 | 0.95	100	100	100	100	19 (28)	5 (13)	1.03 (2.39)
6 | 0.98	100	100	100	100	25 (45)	14 (37)	1.35 (2.67)
8 | 0.90	100	100	100	100	31 (46)	6 (8)	0.61 (0.90)
8 | 0.95	100	100	100	100	47 (83)	14 (40)	0.68 (1.10)
8 | 0.98	100	100	100	100	83 (115)	20 (48)	0.96 (1.61)

Table 5: Comparison of $\mathcal{P}^{\bar{\mathcal{H}}}$ with $\mathcal{P}$ for $|M|$=12.

and $|M|$= 12, respectively. Columns 2 and 3 report the percentage of instances for which the commercial solver finds a feasible solution within the 3600-second time limit (`%-Inst-F`) for $\mathcal{P}$ and $\mathcal{P}^{\bar{\mathcal{H}}}$, respectively. Columns 4 and 5 similarly report the percentage of instances solved to 1% optimality. Columns 6 and 7 report the average (maximum) solution time for $\mathcal{P}$ and $\mathcal{P}^{\bar{\mathcal{H}}}$, respectively. Finally, Column 8 reports the average (maximum) value of $\delta = \frac{UB^{\mathcal{P}^{\bar{\mathcal{H}}}} - LB^{\bar{\mathcal{H}}}}{LB^{\bar{\mathcal{H}}}} \times 100$. Here, $UB^{\mathcal{P}^{\bar{\mathcal{H}}}}$ denotes the best upper bound obtained by $\mathcal{P}^{\bar{\mathcal{H}}}$, whereas $LB^{\bar{\mathcal{H}}}$ denotes the optimal objective value of $\bar{\mathcal{H}}$. Thus, δ measures the percentage gap between the optimal objective value under the *staggered scheduling framework* and the best upper bound obtained for $\mathcal{P}^{\bar{\mathcal{H}}}$. A small value of δ therefore indicates that the *staggered scheduling framework* achieves a `CTC` value close to the best achievable value under the *unrestricted formulation*, $\mathcal{P}$.

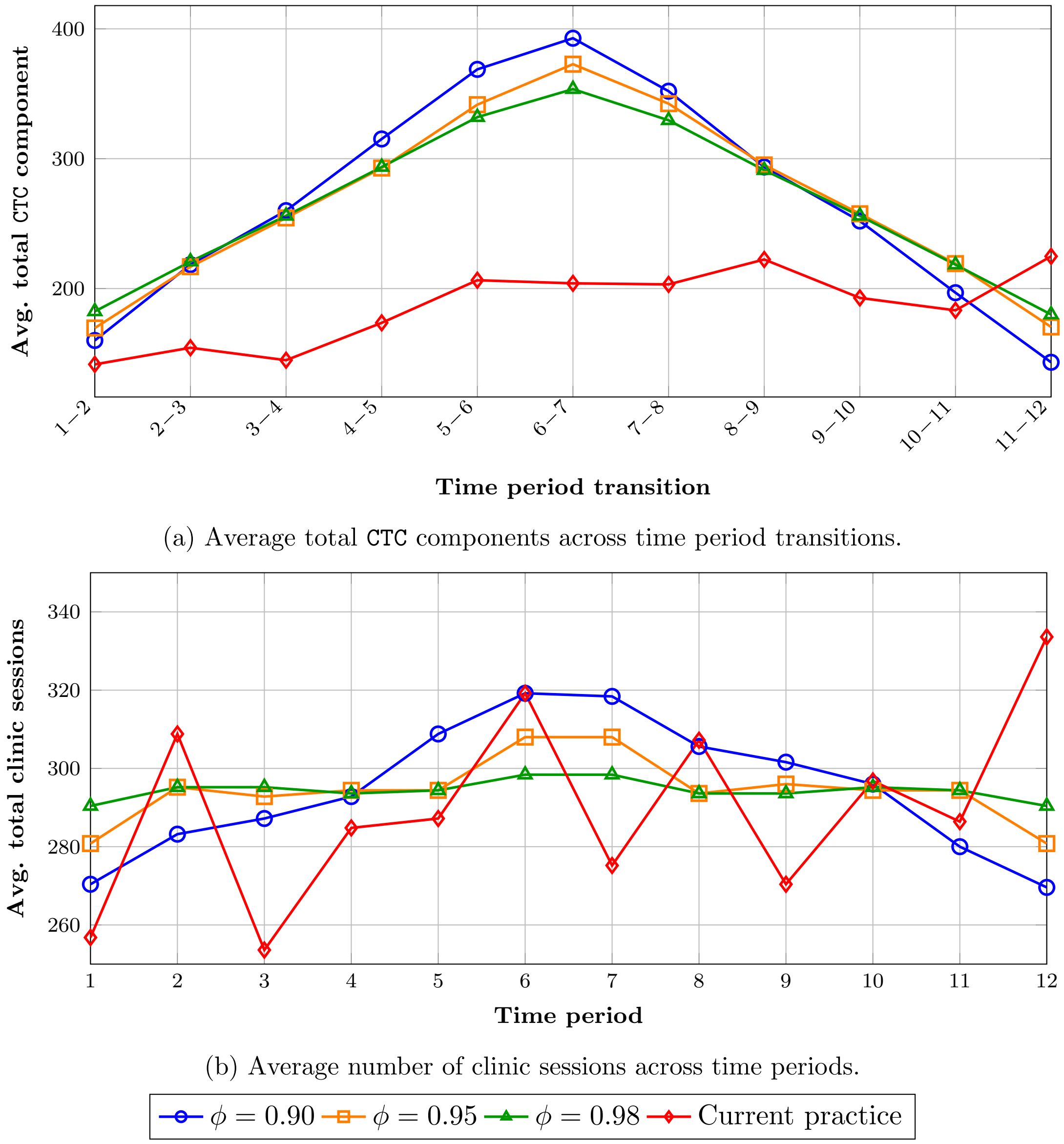
As shown in Tables 4 and 5, the computational difficulty of $\mathcal{P}$ increases substantially with instance size. For $|M|= 12$, the commercial solver solves all instances to within 1% optimality in less than 115 seconds. In contrast, for $|M|= 24$, even instances with smaller values of γ require more than 2,600 seconds to reach the same optimality tolerance. For larger values of γ, the solver frequently fails to even identify a feasible solution within the 3,600-second time limit. Specifically, among the 45 instances with $|M|= 24$, $\mathcal{P}$ yields a feasible solution for only 11 instances, all but one has $\gamma = 9$. Furthermore, for a given instance size, the computational difficulty of $\mathcal{P}$ increases with ϕ. For example, when $|M|= 24$ and $\gamma = 9$, the percentage of instances solved to within 1% optimality within 3,600 seconds decreases from 100% to 0% as ϕ increases from 0.90 to 0.98.

Warm-starting $\mathcal{P}$ with an optimal solution of $\bar{\mathcal{H}}$ substantially improves its computational performance. With the warm start, the solver reaches the 1% optimality tolerance in less than 80 seconds on average, demonstrating the computational benefit of using $\bar{\mathcal{H}}$ to provide a high-quality solution. Moreover, the average values of δ are approximately 1% or lower, indicating that the optimal solutions obtained under the *staggered scheduling framework* achieve `CTC` values close to the best upper bounds obtained for the *unrestricted formulation*. Thus, in addition to accelerating the solution of $\mathcal{P}$, $\bar{\mathcal{H}}$ provides high-quality feasible solutions for `FMRS`.

6.2 Comparison with Current Practice

Figure 2 compares the schedules obtained from $\bar{\mathcal{H}}$ with the schedule implemented in the UMN Family Medicine Residency program. For each value of ϕ, we consider instances with $|M|= 6$ and $\gamma = 6$. Figure 2a compares the average total `CTC` component over five random instances for each pair of consecutive time periods in M between the schedules produced by $\bar{\mathcal{H}}$ and the current practice. Here, the `CTC` component for two consecutive time periods in a rotation schedule is defined as the absolute difference in the number of clinic sessions associated with the rotations assigned to those periods. The total `CTC` component for two consecutive time periods is obtained by summing these individual components across all residents. Similarly, Figure 2b reports the average total number of clinic sessions in each time period for the schedules produced by $\bar{\mathcal{H}}$ and the schedule implemented in current practice.

As shown in Figure 2a, for each value of ϕ, the schedules produced by $\bar{\mathcal{H}}$ achieve substantially higher total `CTC` components than the current practice for 11 of the 12 consecutive time periods, resulting in a significant overall improvement in `CTC`. Figure 2b shows that, as expected, increasing ϕ produces a more uniform distribution of clinic sessions across time periods. Compared with the current practice, the schedules generated by $\bar{\mathcal{H}}$ exhibit substantially lower temporal variability in clinic capacity, leading



(a) Average total `CTC` components across time period transitions.

(b) Average number of clinic sessions across time periods.

Figure 2: Comparison of schedules obtained from $\bar{\mathcal{H}}$ under different values of ϕ with the schedule implemented in the current practice.

to a more balanced allocation of clinic sessions over time. In the following section, we demonstrate that this improvement in the uniformity of clinic capacity, achieved at higher values of ϕ, is attained with little loss in `CTC`.

6.3 Solution Analysis

In this section, we examine the structural properties of the schedules produced by $\bar{\mathcal{H}}$. Specifically, we consider instances with $|M|= 12$ and $\gamma = 6$. Figure 3 reports the average `CTC`, computed over five random instances, for the first six and the last six positions of the optimal `base rotation schedule` under each value of ϕ. Table 6 reports the corresponding average percentage of `CTC` retained relative

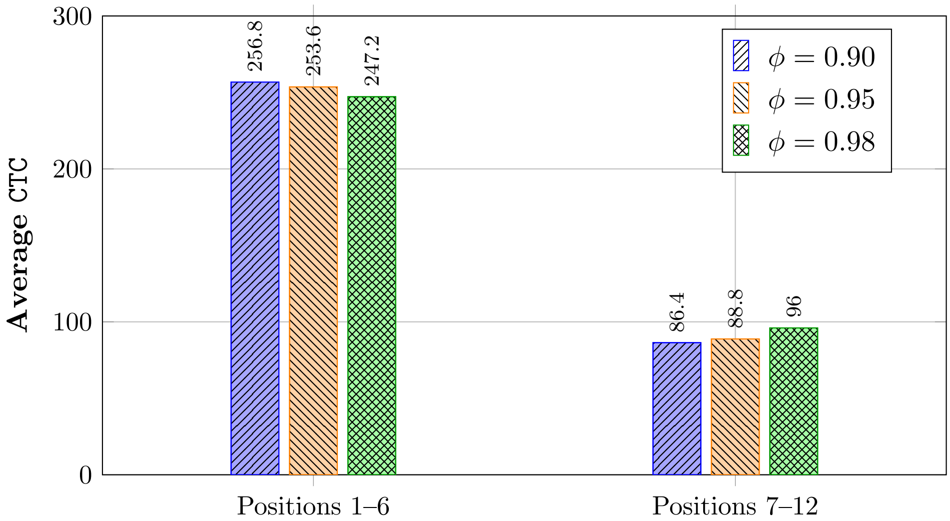


Figure 3: Average CTC over the first six and last six positions in the base rotation schedule for $|M| = 12$, $\gamma = 6$, and $\phi \in \{0.90, 0.95, 0.98\}$.

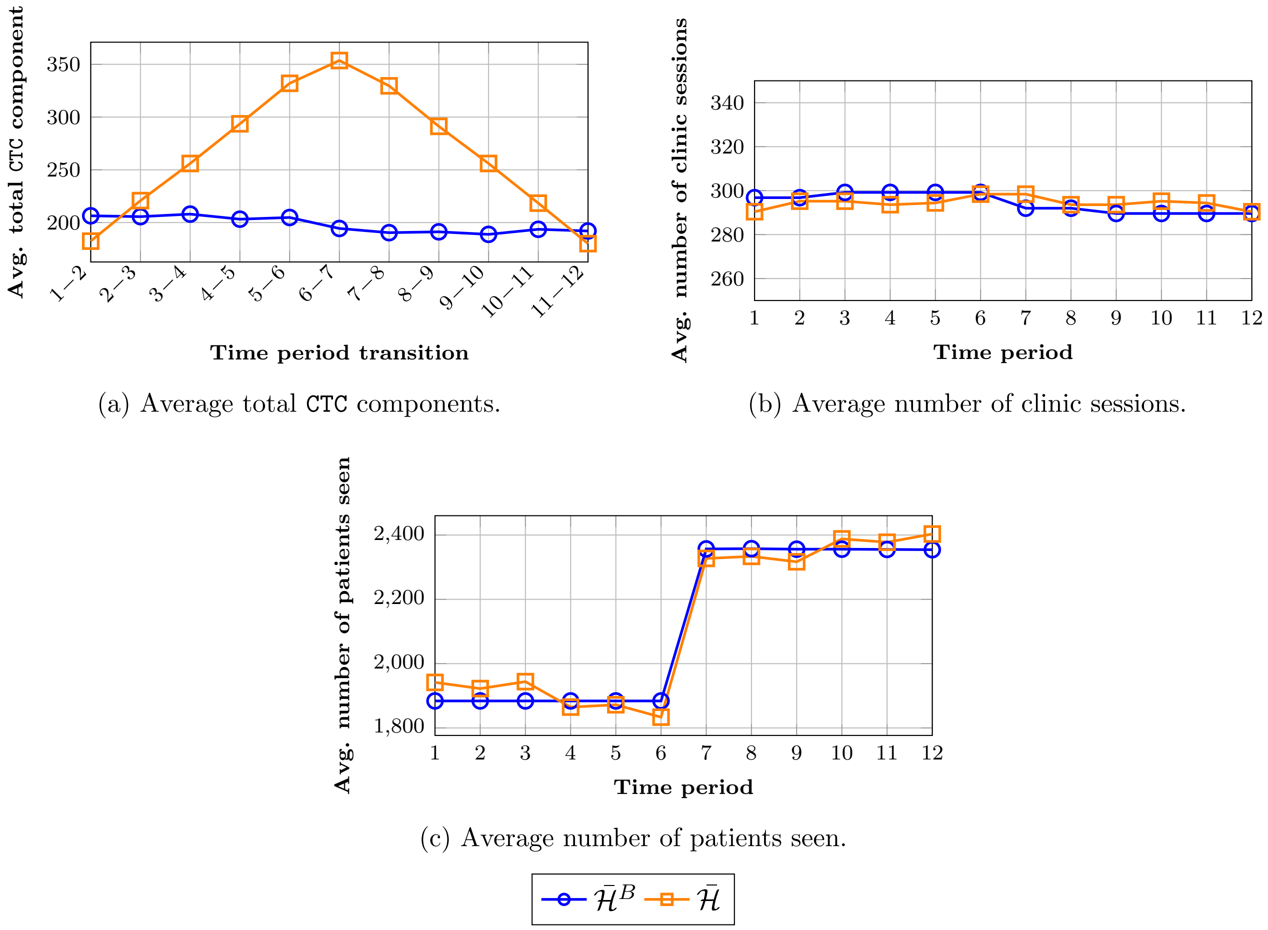
ϕ	CTC retained (%)
0.90	100.0
0.95	99.3
0.98	98.7

Table 6: Average percentage of CTC retained relative to $\phi = 0.90$ for $\gamma = 6$.

to the solution obtained with $\phi = 0.90$ as ϕ increases.

Figure 3 shows that the optimal base rotation schedule consistently assigns rotations with larger CTC components to earlier positions. This behavior is a direct consequence of the objective coefficients of $\bar{\mathcal{H}}$ in (2). Under the *staggered scheduling framework*, CTC components associated with earlier positions contribute to a larger number of resident schedules than those associated with later positions and therefore receive greater weight in the objective function.

Furthermore, Figure 3 shows that increasing ϕ decreases the average CTC associated with the first six positions of the base rotation schedule while increasing that of the last six positions. This redistribution explains the modest reduction in CTC reported in Table 6. Under the *staggered scheduling framework*, CTC components associated with earlier positions of the base rotation schedule primarily affect the middle portions of the schedules of the remaining residents, whereas those associated with later positions primarily affect the beginning and end of their schedules. Consequently, Figure 2a shows that the total CTC components initially increase across time periods before subsequently decreasing. As ϕ increases, however, the total CTC components decrease in the middle time periods while increasing near the beginning and end of the planning horizon. Thus, the reduction in CTC in one portion of the planning horizon is largely offset by an increase elsewhere, resulting in only a modest



(a) Average total CTC components.

(b) Average number of clinic sessions.

(c) Average number of patients seen.

Figure 4: Comparison of schedules obtained from $\bar{\mathcal{H}}^B$ and $\bar{\mathcal{H}}$ for $\phi = 0.98$, $|M| = 12$, and $\gamma = 6$.

reduction in the overall CTC. This balance is achieved by assigning rotations with larger number of clinic sessions to later positions in the `base rotation schedule` to satisfy the increasingly restrictive lower bound on LL, resulting in a more uniform distribution of clinic sessions across time periods, as shown in Figure 2b.

6.4 Comparison with the objective of Bard et al. (2016b)

In this section, we compare the schedules generated under the *staggered scheduling framework* using two different objectives. The first maximizes CTC subject to a lower bound on LL, while the second uses the objective considered in Bard et al. (2016b), which minimizes the maximum deviation in the number of patients seen across half-years. To facilitate this comparison, Appendix C presents a variant of $\bar{\mathcal{H}}$, denoted by $\bar{\mathcal{H}}^B$, in which we replace the CTC objective with the objective of Bard et al. (2016b) and remove the lower-bound constraint on LL. The objective function in Bard et al. (2016b) also includes a term that penalizes modifications to the predefined rotation templates. Because our setting does not permit modifications to these templates, this term is omitted from $\bar{\mathcal{H}}^B$. For the comparison, we use the same data as Bard et al. (2016b) on the number of patients that can be seen by residents in each PGY. This data is reported in Table 8 in Appendix D.

Figure 4 compares the schedules obtained from $\bar{\mathcal{H}}$ and $\bar{\mathcal{H}}^B$ for $\phi = 0.98$, $|M| = 12$, and $\gamma = 6$. Specifically, Figure 4a reports the average total `CTC` component, computed across five random instances, for each pair of consecutive time periods. Figures 4b and 4c report the corresponding average numbers of clinic sessions and patients seen, respectively, in each time period.

As shown in Figures 4b and 4c, the schedules produced by $\bar{\mathcal{H}}$ and $\bar{\mathcal{H}}^B$ result in similar distributions of clinic sessions and patients seen across the planning horizon. However, Figure 4a shows that $\bar{\mathcal{H}}$ yields substantially higher `CTC` components for most consecutive time period transitions. This result highlights an important benefit of the proposed formulation. In $\bar{\mathcal{H}}$, the `LL` constraint, particularly at the relatively high value of $\phi = 0.98$, promotes a balanced distribution of clinic sessions across time periods, which in turn leads to a balanced distribution of patients seen. At the same time, the objective function directly maximizes `CTC`. Consequently, $\bar{\mathcal{H}}$ achieves a level of balance in number of patients seen comparable to that obtained by $\bar{\mathcal{H}}^B$ while producing schedules with substantially higher `CTC`. In contrast, $\bar{\mathcal{H}}^B$ focuses directly on balancing the number of patients seen and does not explicitly account for `CTC` in its objective.

7 Conclusion

In this paper, we developed an BIP framework for `FMRS` that explicitly incorporates the guiding principles of the `Clinic First Model` through its objective function and constraints. Specifically, the proposed formulation maximizes `CTC` to promote a more consistent distribution of clinic sessions within individual resident schedules, thereby avoiding extended periods with either too few or too many clinic sessions. The formulation also enforces a lower bound on `LL` to promote a more uniform distribution of total resident clinic sessions across the planning horizon. By incorporating these criteria into `FMRS`, the proposed framework seeks to improve residents' experiences in the `continuity clinic` and, more broadly, support the goals of the `Clinic First Model`.

We integrated the proposed BIP framework with the *staggered scheduling framework*, under which resident schedules are generated by successively right-shifting a single `base rotation schedule` across the planning horizon. This structure promotes equitable training experiences while yielding schedules that are simple to construct, communicate, and administer. We also established a theoretical bound on the loss in `CTC` resulting from the *staggered scheduling framework*. To improve the computational performance of the resulting BIP formulation, we derived valid inequalities and showed that they describe the convex hull of a set defined by a subset of the formulation's constraints.

Using computational experiments based on data from the UMN Family Medicine Residency program, we demonstrated that the proposed valid inequalities substantially improve the computational performance of the BIP formulation. Our results also showed that the *staggered scheduling framework* results in only a small loss in `CTC`. Moreover, using its optimal solution to warm-start the *unrestricted formulation* substantially improves computational performance. Compared with a schedule implemented in current practice, the schedules generated by our approach achieve substantially higher `CTC` while providing a more balanced distribution of clinic sessions across time periods. Finally, our analysis of the structural properties of the resulting schedules provides further insight into how the proposed approach improves `CTC` while balancing the number of clinic sessions across time periods. A comparison with the objective function considered in Bard et al. (2016b) further demonstrates the benefits of explicitly incorporating `CTC` and `LL` into `FMRS`.

Several directions for future research arise from this work. First, the proposed framework could be extended to account for uncertainty in the number of clinic sessions associated with each rotation. Second, resident team formation could be integrated with `FMRS` so that team composition and annual rotation schedules are determined jointly. Such an integrated approach could account for interactions among residents' clinic schedules when forming teams and potentially further improve continuity of care for patients.

References

Adam P, Hersch D, Peek CJ (2022) Implementing clinic first guiding actions across 4 family medicine residency clinics. *Academic Medicine* 97(2):233–238.

Akbarzadeh B, Maenhout B (2021a) A decomposition-based heuristic procedure for the medical student scheduling problem. *European Journal of Operational Research* 288(1):63–79.

Akbarzadeh B, Maenhout B (2021b) An exact branch-and-price approach for the medical student scheduling problem. *Computers & Operations Research* 129:105209.

Bard JF, Shu Z, Morrice DJ, Leykum LK (2016a) Annual block scheduling for internal medicine residents with 4+ 1 templates. *Journal of the Operational Research Society* 67(7):911–927.

Bard JF, Shu Z, Morrice DJ, Leykum LK (2017) Constructing block schedules for internal medicine residents. *IISE Transactions on Healthcare Systems Engineering* 7(1):1–14.

Bard JF, Shu Z, Morrice DJ, Leykum LK, Poursani R (2016b) Annual block scheduling for family medicine residency programs with continuity clinic considerations. *IIE Transactions* 48(9):797–811.

Bodenheimer T, Sinsky C (2014) From triple to quadruple aim: care of the patient requires care of the provider. *The Annals of Family Medicine* 12(6):573–576.

Brown SR, Bodenheimer T, Kong M (2019) High-performing primary care residency clinics: a collaboration. *Annals of Family Medicine* 17(5):470.

Dall T, Chakrabarti R, Iacobucci W, Hansari A, West T (2017) The complexities of physician supply and demand: projections from 2015 to 2030. *Association of American Medical Colleges* .

Day TE, Napoli JT, Kuo PC (2006) Scheduling the resident 80-hour work week: an operations research algorithm. *Current Surgery* 63(2):136–141.

Erhard M, Schoenfelder J, Fügener A, Brunner JO (2018) State of the art in physician scheduling. *European Journal of Operational Research* 265(1):1–18.

Franz LS, Miller JL (1993) Scheduling medical residents to rotations: solving the large-scale multiperiod staff assignment problem. *Operations Research* 41(2):269–279.

Guo J, Morrison DR, Jacobson SH, Jokela JA (2014) Complexity results for the basic residency scheduling problem. *Journal of Scheduling* 17:211–223.

Guo J, Pozehl W, Cohn A (2023) A two-stage partial fixing approach for solving the residency block scheduling problem. *Health Care Management Science* 26(2):363–393.

Gupta R, Barnes K, Bodenheimer T (2016) Clinic first: 6 actions to transform ambulatory residency training. *Journal of Graduate Medical Education* 8(4):500–503.

Hersch D, Klemenhagen K, Martin C, Berg B, Adam P (2023) Impact of set-day clinic on physician continuity in a family medicine residency clinic. *Family Medicine* 55(9):612–615.

Hofkamp H, Verdieck A, Cheng A, Ramalingam N, Rdesinski R, Garvin R (2020) Effects of clinic first educational interventions on resident wellness and engagement. *Family Medicine* 52(6):422–426.

Hong YC, Cohn A, Gorga S, O'Brien E, Pozehl W, Zank J (2019) Using optimization techniques and multidisciplinary collaboration to solve a challenging real-world residency scheduling problem. *INFORMS Journal on Applied Analytics* 49(3):201–212.

Kraul S (2020) Annual scheduling for anesthesiology medicine residents in task-related programs with a focus on continuity of care: S. kraul. *Flexible Services and Manufacturing Journal* 32(1):181–212.

Lemay B, Cohn A, Epelman M, Gorga S (2017) New methods for resolving conflicting requests with examples from medical residency scheduling. *Production and Operations Management* 26(9):1778–1793.

Li S, Shehadeh KS, Curtis FE, Hochman BR (2025) Equity-promoting integer programming approaches for medical resident rotation scheduling. *Health Care Management Science* 28(4):890–929.

Manolakis PG, Skelton JB (2010) Pharmacists' contributions to primary care in the united states collaborating to address unmet patient care needs: the emerging role for pharmacists to address the shortage of primary care providers. *American Journal of Pharmaceutical Education* 74(10).

Paul KJ, Hidaka BH, Ford P, Morris C (2021) Measuring greater patient-provider continuity in a clinic-first family medicine residency curriculum. *The Permanente Journal* 25.

Proano RA, Agarwal A (2018) Scheduling internal medicine resident rotations to ensure fairness and facilitate continuity of care. *Health Care Management Science* 21:461–474.

Smalley HK, Keskinocak P (2016) Automated medical resident rotation and shift scheduling to ensure quality resident education and patient care. *Health Care Management Science* 19(1):66–88.

Stange KC, Miller WL, Etz RS (2023) The role of primary care in improving population health. *The Milbank Quarterly* 101(S1):795–840.

Starfield B (2012) Primary care: an increasingly important contributor to effectiveness, equity, and efficiency of health services. sespas report 2012. *Gaceta Sanitaria* 26:20–26.

Starfield B, Shi L, Macinko J (2005) Contribution of primary care to health systems and health. *The Milbank Quarterly* 83(3):457–502.

Wang F, Huang YL, Ju F, Grega JC (2022) Resident rotation scheduling for categorical internal medicine residency program. *IEEE Robotics and Automation Letters* 7(2):4142–4148.

Zeller T, Ewing J, Asif I (2019) Prevalence of clinic first curricula: a survey of afmrd members. *Family Medicine* 51(4):338–343.

Appendices

Appendix A: *Unrestricted Formulation* for FMRS

Let $\kappa^y_{ir_1r_2m}$ be a binary variable that equals 1 if resident $i \in S_y$ in PGY $y \in Y$ is assigned to rotation $r_1 \in R_y$ in time period $m-1$ and to rotation $r_2 \in R_y$ in time period m, and 0 otherwise. Define $c^y_{r_1r_2} = |\bar{c}^y_{r_1} - \bar{c}^y_{r_2}|$ as the contribution to CTC obtained by scheduling rotations $r_1, r_2 \in R_y$ consecutively, $[a] = \{1, 2, 3, \ldots, a\}$, and $\bar{M} = M \setminus \{1\}$. With this notation, *unrestricted formulation* for FMRS is as follows:

$$(\mathcal{P}) \quad \max \sum_{y\in Y} \sum_{i\in S_y} \sum_{r_1\in R_y} \sum_{r_2\in R_y\setminus\{r_1\}} \sum_{m\in\bar{M}} c^y_{r_1r_2}\, \kappa^y_{ir_1r_2m} \tag{13a}$$

$$\sum_{i\in S_y} \sum_{r_2\in R_y\setminus\{r_1\}} \kappa^y_{ir_1r_22} \le 1, \quad \forall r_1 \in R_y, y \in Y \tag{13b}$$

$$\sum_{i\in S_y} \sum_{r_1\in R_y\setminus\{r_2\}} \kappa^y_{ir_1r_2m} \le 1, \quad \forall r_2 \in R_y,\ m \in \bar{M},\ y \in Y \tag{13c}$$

$$\sum_{r_1\in R_y\setminus\{r_2\}} \kappa^y_{ir_2r_12} + \sum_{r_1\in R_y\setminus\{r_2\}} \sum_{m\in\bar{M}} \kappa^y_{ir_1r_2m} = 1, \quad \forall r_2 \in R_y,\ i \in S_y,\ y \in Y \tag{13d}$$

$$\sum_{r_1\in R_y\setminus\{r_2\}} \kappa^y_{ir_1r_2m} = \sum_{r_3\in R_y\setminus\{r_2\}} \kappa^y_{ir_2r_3(m+1)}, \quad \forall m \in \bar{M} \setminus \{|M|\},\ r_2 \in R_y,\ i \in S_y,\ y \in Y \tag{13e}$$

$$\sum_{y\in Y} \sum_{r_2\in R_y} \sum_{r_1\in R_y\setminus\{r_2\}} \sum_{i\in S_y} \bar{c}^y_{r_1}\, \kappa^y_{ir_1r_22} \ge \bar{L}, \tag{13f}$$

$$\sum_{y\in Y} \sum_{r_2\in R_y} \sum_{r_1\in R_y\setminus\{r_2\}} \sum_{i\in S_y} \bar{c}^y_{r_2}\, \kappa^y_{ir_1r_2m} \ge \bar{L}, \quad \forall m \in \bar{M} \tag{13g}$$

$$\sum_{r_2\in R_y\setminus\{r_1\}} \kappa^y_{kr_1r_22} = 0, \quad \forall r_1 \in [k-1],\ k \in S_y,\ y \in Y \tag{13h}$$

$$\sum_{r_2\in R_y\setminus\{r_1\}} \kappa^y_{kr_1r_22} \le \sum_{r_3=1}^{r_1-1} \sum_{r_2\in R_y\setminus\{r_3\}} \kappa^y_{lr_3r_22}, \quad \forall l \in [k-1], r_1 \in R_y \setminus [k-1],\ k \in S_y,\ y \in Y \tag{13i}$$

$$\kappa^y_{ir_1r_2m} \in \{0,1\}, \quad \forall i \in S_y,\ r_1 \in R_y,\ r_2 \in R_y \setminus \{r_1\},\ m \in M,\ y \in Y \tag{13j}$$

Constraints (13b) and (13c) ensure that each rotation in R_y is assigned to at most one resident in each time period, whereas Constraints (13d) ensure that each resident is assigned to every rotation in R_y exactly once over the planning horizon. Constraints (13e) enforce consistency in rotation assignments across consecutive time periods, while Constraints (13f) and (13g) impose the lower bound on LL. Constraints (13h) and (13i) are symmetry-breaking constraints. Specifically, Constraints (13h) prevent resident k from being assigned to any rotation $r_1 < k$ in the first time period. Constraints (13i) ensure that if resident k is assigned to rotation r_1 in the first time period, then every resident $l < k$ is assigned to a rotation $r_3 < r_1$ in that period. The objective function (13a) maximizes CTC.

Appendix B: Upper Bound on the Maximum Level Loading

The following proposition gives an upper bound on $\texttt{LL}$.

Proposition 6. $\texttt{LL} \leq \tilde{\texttt{LL}} = \left\lfloor \frac{\sum_{y\in Y}\gamma_y\left(\sum_{r\in R_y}\bar{c}_r^y\right)}{|M|} \right\rfloor$

Proof. Let $\texttt{TC} = \sum_{y\in Y}\gamma_y\left(\sum_{r\in R_y}\bar{c}_r^y\right)$ denote the total number of clinic sessions across the planning horizon. Indeed, for each PGY $y \in Y$, $\gamma_y\left(\sum_{r\in R_y}\bar{c}_r^y\right)$ is the total number of clinic sessions assigned to all residents in that PGY. Suppose, by contradiction, that $\texttt{LL} > \left\lfloor \frac{\texttt{TC}}{|M|} \right\rfloor$. Since $\texttt{LL}$ is integer-valued, it follows that there exists some $\zeta > 0$ such that $\texttt{LL} = \frac{\texttt{TC}}{|M|} + \zeta$. By definition, $\texttt{LL}$ is the minimum total number of clinic sessions assigned in any time period. Therefore, every time period must contain at least $\texttt{LL}$ clinic sessions, implying $\texttt{TC} \geq |M|\texttt{LL} = |M|\left(\frac{\texttt{TC}}{|M|} + \zeta\right) = \texttt{TC} + |M|\zeta$. This implies $|M|\zeta \leq 0$, which is a contradiction. $\square$

Appendix C: *Staggered Scheduling* formulation with the objective from Bard et al. (2016b)

Let $L = \{1, 2\}$ denote the set of half-years, and let $\mathcal{M}_l$ denote the set of time periods in half-year $l \in L$. For each PGY y and half-year l, let ρ_l^y denote the number of patients seen per clinic session, and let μ_l denote the average number of patients seen per time period in half-year l. Finally, let Δ denote the maximum absolute deviation in the number of patients seen in any time period from the corresponding half-year average μ_l. Using these notations, the following formulation determines a `base rotation schedule` under the *staggered scheduling framework* that minimizes Δ.

$$(\bar{\mathcal{H}}^B) \quad \min \quad \Delta \tag{14a}$$

$$(3b), (3c), (3h) \tag{14b}$$

$$\mu_l = \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}_l}\sum_{y\in Y}\sum_{p\in\bar{P}_m^y}\sum_{r\in R_y}\rho_l^y\,\bar{c}_r^y\,x_{rp}^y \quad \forall\, l \in L \tag{14c}$$

$$\Delta \geq \sum_{y\in Y}\sum_{p\in\bar{P}_m^y}\sum_{r\in R_y}\rho_l^y\,\bar{c}_r^y\,x_{rp}^y - \mu_l \quad \forall\, l \in L,\ m \in \mathcal{M}_l \tag{14d}$$

$$\Delta \geq \mu_l - \sum_{y\in Y}\sum_{p\in\bar{P}_m^y}\sum_{r\in R_y}\rho_l^y\,\bar{c}_r^y\,x_{rp}^y \quad \forall\, l \in L,\ m \in \mathcal{M}_l. \tag{14e}$$

$\bar{\mathcal{H}}^B$ includes a subset of the constraints from $\bar{\mathcal{H}}$, as specified in (14b). Constraints (14c) determine the average number of patients seen per time period for each half-year. Finally, Constraints (14d)–(14e), together with the objective function (14a), minimize the maximum absolute deviation in the number of patients seen in each time period from the corresponding half-year average, μ_l.

Appendix D: Supplementary Tables

Rotation	PGY-1	PGY-2	PGY-3	Avg. Weekly Clinic Sessions
Inpt1	2	0	0	1
ICU	1	0	0	1
Emed	1	0	0	10
Clinic1	1	0	0	8
OB	2	0	0	3
Card1	1	0	0	2
NICU	1	0	0	1
Ped	1	0	0	1
Surg	1	0	0	3
Neur1	1	0	0	3
AO	0	1	0	4
Bmed	0	1	0	4
ELEC2	0	2	0	4
Derm	0	1	0	4
CHlth	0	1	0	4
Ger	0	1	0	4
Emed2	0	1	0	4
Ortho	0	1	0	6
PedSv	0	1	0	3
Inpt2	0	2	0	3
Clinic3	0	0	1	5
Card3	0	0	1	4
ELEC3	0	0	3	4
Inpt3	0	0	2	4
CoUro	0	0	1	5
Gyn	0	0	1	6
Nephro	0	0	1	4
PedER	0	0	1	4
PMed	0	0	1	5

Table 7: Family medicine rotations by PGYs and average weekly clinic sessions in each rotation.

PGY	Periods 1–6	Periods 7–12
PGY-1	3	5
PGY-2	6	8
PGY-3	9	10

Table 8: Number of patients seen per clinic session by PGY and time period (Bard et al. 2016b).