Industrial and
Systems Engineering

The Inmate Assignment and Scheduling Problem and its Application in the PA Department of Correction

Mohammad Shahabsafa , Tamás Terlaky, Chaitanya Gudapati, Anshul Sharma, Louis Plebani, and George Wilson

Department of Industrial and Systems Engineering,
Lehigh University, USA

Kristofer B. Bucklen

Pennsylvania Department of Corrections, USA

ISE Technical Report 17T-013

The Inmate Assignment and Scheduling Problem and its Application in the PA Department of Correction

Mohammad Shahabsafa [1], Tamás Terlaky[1], Chaitanya Gudapati[1], Anshul Sharma[1], Louis Plebani[1], George Wilson[1], and Kristofer B. Bucklen[2]

[1] Department of Industrial and Systems Engineering,, Lehigh University, USA ,
[2] Pennsylvania Department of Corrections, USA

October 20, 2017

Abstract

The inmate assignment project, in close collaboration with the Pennsylvania Department of Corrections (PADoC), took five years from start to successful implementation. In this project, we developed the Inmate Assignment Decision Support System (IADSS), where the primary goal is simultaneous and system-wide optimal assignment of inmates to correctional institutions (CIs). We develop a novel hierarchical, multi-objective Mixed Integer Linear Optimization (MILO) model, which accurately describes the inmate assignment problem (IAP). The IAP is the mathematical optimization formulation of the problem every correctional system faces which is to assign inmates to CIs and schedule their programs, while all legal restrictions and best practice constraints are considered. By using real inmate data sets from the PADoC, we also demonstrate that the MILO model can be solved efficiently. IADSS enables PADoC to significantly reduce the population management costs, and enhance public safety and security of the CIs.

To the best of our knowledge, this is the first time that Operations Research (OR) methodologies have been built directly into the routine business practice of a correctional system, and used to optimize its operations. This successful project opens a rich and untouched area for the application of OR and optimization methodology. The new model and methodology can be utilized for the assignment of inmates in any correctional system.

Introduction

According to the International Centre for Prison Studies, the U.S. incarcerates 698 people for every 100,000 of its population. Despite accounting for approximately 4.5% of the world's population, the U.S. has 21.4 % of the world's incarcerated population (Walmsley 2017). In 2010, all levels of government in the U.S. spent more than $80 billion on corrections (Kyckelhahn and Martin 2010), implying $260 tax burden for each U.S. resident. Adjusted to inflation, the expenditures on corrections in 2010 are more than three times of that in 1979 (Schanzenbach et al. 2016).

Due to insufficient capacity of the correctional institutions (CIs), there is a growing problem of *overcrowding* in the CIs. Population management of the inmates is one of the most critical operations within a correctional system, and requires significant monetary and human resources. Efficiently managing the inmate population results in huge savings. Appropriate assignment of the inmates to the CIs is a key element of population management, which can lead to significant savings, as well as enhancing public safety and security of the CIs.

When a court delivers a sentence, the inmate often receives a list of treatment programs based on the various assessments, including the crime committed. Research shows that inmates who complete the programs, offered by the CIs, have lower recidivism rate (Davis et al. 2013); hence, programs have the capability of saving CI capacity and promoting a safe and healthy society. Inmates usually are given a *minimum sentence length* in "indeterminate sentencing states" like Pennsylvania. Having served the minimum sentence length, they are eligible to be conditionally released, also known as *parole*, if they satisfy all of the parole requirements. *One of the parole requirements is to complete all the required treatment programs.* Overcrowding of CIs adversely affects the way inmates receive their treatment programming and delays scheduling as the resources for the programs are limited. Inmates who receive timely programming have a better chance of becoming eligible for parole and leaving the correctional institution earlier, thereby reducing the population of the CIs.

In 2015, the PADoC had a staggering $2.15 billion in expenditures to house 50,366 inmates (Mai and Subramanian 2017). All inmates, who enter the correctional system, have their own programming needs and special requirements. Often, a CI can offer only certain programs as it has only limited personnel and infrastructure resources and so might not be able to meet the needs of all inmates. We briefly describe the inmate assignment process before this project started. Each new inmate would be assigned to CIs, *manually*, by a staff member of the Office of Population Management (OPM). Numerous factors, i.e., rules and criteria, are considered in assigning inmates to CIs, including but not limited to, security concerns, mental and medical conditions, program needs, separation from other inmates, capacities of the CIs, and home county of the inmates. Having to consider all the factors for the assignment of each inmate, individually, is time-consuming and prone to human errors. Additionally, when inmate assignment is done individually and *sequentially*, inmates assigned later are not considered in the current assignment. This greedy sequential assignment of inmates to CIs makes the process highly inefficient, and results in numerous violations of the factors, or the capacity constraints, or both.

The optimal inmate assignment project in collaboration with the PADoC, spanned five years from idea to successful implementation. The main goal of the project is to develop an Inmate Assignment Decision Support System (IADSS) for the PADoC, which *simultaneously* assigns the inmates to CIs and schedules the treatment programs for the inmates, while all the factors and criteria of the assignment are considered. The IADSS is comprised of a user-friendly web based interface, which is linked to the PADoC databases, and an optimization engine which does the assignment of the inmates to CIs.

The goal of the IAP is to optimize inmate assignments, transfers, and program scheduling, while numerous restrictions and constraints are considered to advance the following objectives:

- reduce the total population of inmates at the CIs,
- minimize inmate movements during prison terms,
- reduce treatment services waiting lists.

Literature Review

The IAP is a novel class of the *assignment problem* (Flood 1953, Votaw and Orden 1952) with several side constraints. The classic assignment problem and algorithms to solve it have been extensively studied in the 50s (Dantzig 1951, Orden 1951). Kuhn (1955) suggests the well-known Hungarian method for solving the assignment problem. Assignment models have been used in a large variety of applications of optimization. For instance, crew scheduling is a broadly-used problem class using generalized assignment models. Airline crew scheduling is one of the most important crew scheduling problems that received attention within the optimization community in the 60s (Arabeyre et al. 1969) and it has been extensively studied since then. Furthermore, Caprara et al. (1998) have used the assignment model for crew scheduling in the railway industry. To the best of our knowledge, the only OR paper for the IAP is by Li et al. (2014), who studied the inmate assignment process and developed a decision tree representing all the factors of the inmate assignment to CIs.

Contributions: Novel Modeling and Solution Methodology

The IAP mainly revolves around the assignment of inmates to the CIs and scheduling of programs for the inmates at the CIs. In order to develop a mathematical optimization model, all the processes of the inmate assignment were mapped and formalized, which in fact was a challenging process, because there are no OR experts at the PADoC, nor to the best of our knowledge at DoCs elsewhere today. Due to scarce resources and often conflicting rules, the IAP is inherently an infeasible problem. In order to address the need for simultaneous system-wide optimization of inmate assignments, while considering all the conflicting factors, we developed and fine-tuned a *hierarchically weighted multi-objective MILO model*. In conjunction with model development, data collection and preparation procedures, which interface with the DoC database systems, have been developed. Ultimately, the web-based IADSS was developed which enables the user to make optimal decisions in a fraction of the time needed before. Since September 2016, the integrated IADSS has been in daily use by PADoC. The IADSS makes the assignment process efficient, while significantly improving the quality and consistency of the assignments. These goals are achieved by advanced optimization modeling of system-wide assignment and scheduling needs, and the use of state of the art optimization methodology.

Impact

IADSS enables the PADoC to have high-quality consistent assignments, which also increases security and reduces violence. IADSS has resulted in cost savings by reducing the population of the inmates and the number of transfers between the CIs. It has also enabled the PADoC to reduce the staff needed for making assignments, and it has led to a smaller number of assaults in the CIs. As a result of using the IADSS for the assignment of inmates, the PADoC has saved \$2.9 million in the first year, and it is expected to reduce the cost by \$19.8 million over 5 years.

The broader impact of this project, and the highly successful development of the IADSS is that it can be adapted and used in the correctional systems of other states and countries. Thus, this project, and the developed solution methodology, is opening a new, high impact area for the application of OR and analytics methodologies.

This paper is structured as follows. The *Preliminaries and Problem Description* section presents the IAP and the numerous factors and program scheduling requirements which define the IAP. Modeling and solution methodology details are presented in the *Modeling and the Solution Methodology* section. The *Implementation at the PADoC* section presents the development of the IADSS, and the implementation at the PADoC. We list and quantify the benefits of using IADSS in the *Benefits and Impact of the IADSS* section, and the *Summary* section presents the summary of the paper. The multi-objective MILO model for simultaneously assigning the inmates to the CIs is presented in the *Hierarchical Multi-Objective Mathematical Model* appendix.

Preliminaries and Problem Description

In this section, we discuss the preliminary developments at the PADoC and we formally define the IAP and elaborate the rules and criteria used for inmate assignment to the CIs.

Preliminary Development

We now discuss the developmental evolution of the model and the web based IADSS at the PADoC. When the project started, we discussed with PADoC the need for a decision support system to assist the OPM in assigning the inmates to the best possible CI, considering both the needs and limitations of the inmates and the available limited resources of the PADoC. This is a complex problem where ideal assignment of all inmates is not possible. Inmate-specific factors are a combination of several categories such as medical, psychological, educational, custody level, and sentence conditions. On the other hand, CIs have numerous limitations, such as security level, treatment programs availability, and capacity.

Conventionally, the assignment process has been manual and subjective, where a staff member with the provided information of the inmate and the CIs from the PADoC database assigns the inmates one-by-one to the CIs. While the general guidelines for the assignment are known, the large number of factors, the daily changing capacities of the CIs, and the subjective nature of this sequential ad-hoc assignment made the efficiency and quality of the assignment heavily dependent on the experience and judgment of the staff. In order to remove the subjective component of the assignment, initially we developed a decision-tree based decision support system (DTDSS) to reduce bias and variability in assignments, while improving adherence to the guidelines. The DTDSS provided the DoC with a ranked order of the CIs for a particular inmate from which the staff member can choose the assignment. This eliminated much of the tedious work of evaluating various combinations of factors, thus, freeing staff to use their experience to choose from a smaller subset of the most suitable CIs.

Figure 1 illustrates the decision tree of the DTDSS. The development and use of the decision tree in the DTDSS was critical in classifying and refining all the relevant factors and their importance level in inmate assignment. After discussing with PADoC personnel the factors which influenced the inmate assignment in detail, we identified and incorporated 60 of the most important factors used in the manual assignment procedure. The DTDSS uses these factors and rules to evaluate and, subsequently, rank the CIs with respect to their suitability for the inmate being assigned. DTDSS assigns weights and penalties for each factor, and the accumulated penalties are used to rank the CIs for the inmates.

This approach could conceivably have been deemed sufficient, while clearly not optimal, if inmates were arriving to the system in a sequence (one by one). The greedy assignment strategy embodied in the sequential application of DTDSS cannot adequately anticipate the bottlenecks in the CIs, several assignments into the future. When a batch of inmates need assignment, there is an opportunity to make resource tradeoffs performing the batch assignment that is not present in the sequential approach. In a sequential assignment, the sequence of the inmates is critical and significantly affects the succeeding assignments. The need for system-wide, simultaneous assignment made clear the need for a multiple-objective optimization model which treats all the inmates needing assignment and considers the current state of all the CIs, simultaneously, from a system's perspective.

Assignment Criteria

In this section, we present the essential elements of the inmate assignment problem. First, we give a brief description of the inmate assignment process. Inmates are evaluated and classified at "intake CIs". Each period, the accumulated inmates have to be assigned to CIs, while all restrictions and constraints need to be taken into account. This is the basic inmate initial assignment problem. The map of PA with the 25 currently running CIs of the PADoC are shown in Figure 2. A crucial feature of the inmate initial assignment problem is that inmates need to go through individually specified programs, which are scheduled according to specific rules and requirements. Furthermore, there are a variety of reasons leading to inmate transfers from their initially assigned CI to another one. The need for this transfer of inmates further complicates the problem. Next we explain the criteria that need to be taken into account at the initial assignment of inmates to the CIs.

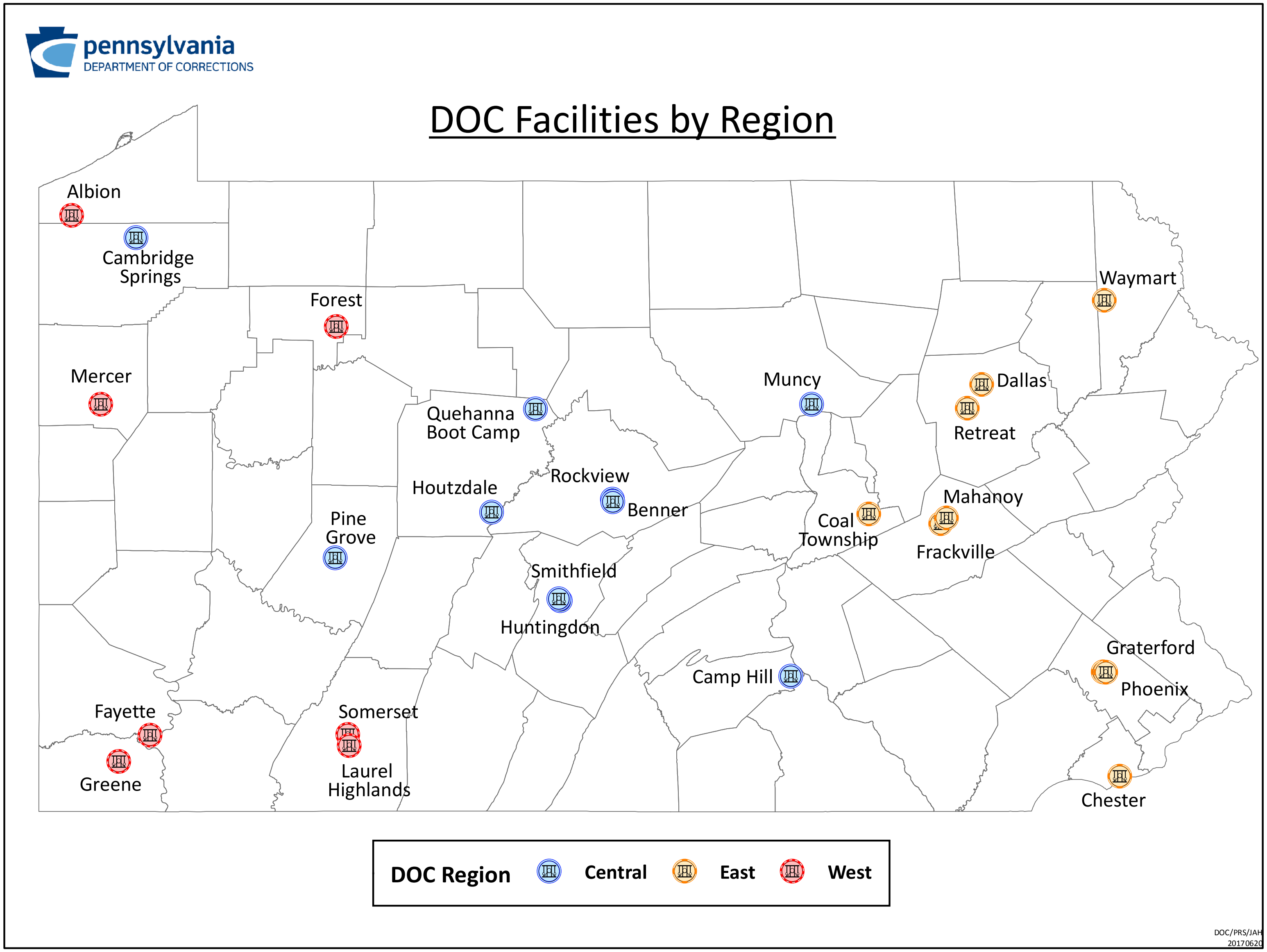


Figure 2: The 25 state CIs of PADoC and their placement in one of the three main regions of the state

General factors: There are a variety of factors, that have to be satisfied at initial inmate assignment, including but not limited to,

- *High risk inmates* have to be assigned to a predefined set of CIs.
- Inmates who are *mentally unstable* should be assigned to a predefined set of CIs.
- *Young adult offenders* should be assigned to a predefined set of CIs.
- Inmates serving a *life sentence* have to be assigned to a predefined set of CIs.
- CIs are gender specific; thus, inmates have to be assigned accordingly.

Available beds: The number of available beds for each CI is determined prior to assigning the inmates. At least a minimum number of inmates, which is a function of the available beds, should be assigned to each CI in order to properly and proportionally utilize bed spaces. Additionally, for each CI, the maximum number of inmates, which is again a function of the available beds, is specified to avoid creating long lists of inmates waiting for beds to become available at the CIs. Furthermore, the number of inmates assigned to the CIs should be *proportional to the available beds* when it is in the minimum and maximum range.
Home county: Inmates need to be assigned to a *CI near their home county.*
Separations: Considering previous inmate-inmate and inmate-staff conflicts, some inmates cannot be assigned to certain CIs. Additionally, there might be pairs or groups of inmates, waiting to be assigned, that *cannot be assigned to the same CI.*

Treatment Programs

Inmates usually are given minimum sentence length, i.e., the minimum time they have to stay in CIs, and they have a scheduled parole board interview before their minimum sentence date. To be eligible for parole, they need to satisfy all of the requirements of their sentences. One of the requirements is to complete all of their treatment programs before the parole board interview. Treatment programs are prescribed by the court, or by the correctional system.

Ideally, inmates should be assigned to a CI which can offer their program(s) before their parole board meeting. However, due to limited capacity of the programs at CIs, not all the inmates are able to finish their program(s) before their parole board meeting. This results in creating *inmate waiting lists* for the programs at the CIs, which provides one of the most important criteria in the IAP. Furthermore, inmates can start their programs only within the 24-month window before their minimum sentence date.

Programs can either be *open-enrollment* or *closed-enrollment.* In an open enrollment program, enrollments can happen any time. If an inmate completes an open program, the next inmate can start that program immediately. However, in a closed-enrollment program, a group is identified and they all start and complete the program at the same time.

The number of inmates that start an open-enrollment program at time t is driven by the number of open spots of that program at time t. However, the number of inmates that can start a closed-enrollment program at time t is driven by the number of groups of that program that can start at time t. There is a minimum and maximum for the number of inmates that can be enrolled in a group for each of the closed-enrollment programs.

Another concept which is important in handling the program waiting lists is *clusters.* A cluster is a group of closed-enrollment programs that have *common instructors*, i.e., an instructor can handle all the programs in a cluster and it needs to be determined which program(s) the given instructor runs at a given time. Notice that clusters are only defined for closed-enrollment programs.

One of the main goals of the IAP is to ensure that inmates start their programs as soon as possible. This goal is formalized as minimizing the maximum waiting time of the inmates for starting their required program(s). To reach this goal we schedule the programs for the incoming inmates, while considering the limited available resources of the CIs and the inmates that are already in the CIs.

Transfer Constraints

After the initial assignment, some of the inmates need to be transferred. Some of the reasons for transfers after the initial assignment are as follows

Parole violator: Inmates who are released on parole and have violated their parole terms are brought back to a "parole intake facility" and need to be assigned to a CI afterwards.

Program placement: It may happen that the CI, to which an inmate is initially assigned, does not have all the inmate's required programs. Additionally, treatment programs may be prescribed after the initial assignment and some programs might not be available in the current CI. In these cases, the inmate should be moved to a CI where all the required programs are offered.

Incentive based transfers: Satisfying specific predefined requirements, inmates can request to be moved to other CIs.

Separation: Separation of an inmate from other inmates or from DoC staff can lead to a transfer request.

Constraints and restrictions for transfer placements are the same as the ones explained in the *Assignment Criteria* section for the initial assignment of the inmates. However, the importance of the factors for a transfer placement might differ from those of an initial assignment.

Modeling and the Solution Methodology

As it was explained in the *Preliminaries and Problem Description* section, one of the main goals of the IAP is to assign the inmates to CIs. However, it is not a basic assignment problem, since there are a variety of factors that need to be considered in the assignment of each inmate. General factors, elaborated in the *Assignment Criteria* section, need to be satisfied in the assignment of each inmate. We not only need to satisfy the bound constraints on the number of inmates that can be assigned to each CI, but we also need to assign the inmates in proportion to the capacities of the CIs. Another criterion is that inmates should be assigned to CIs that are nearest to their home county. Furthermore, we need to schedule the required programs for the inmates, which brings a scheduling component to the IAP. Due to limited availability of resources and the conflicting rules of the assignment, it is impossible to make an ideal assignment and perfectly satisfy all the factors and program scheduling needs in the assignment of a batch of inmates.

In order to address all the conflicting factors of the assignment, we developed a *hierarchically weighted multi-objective MILO model*. As the problem is inherently infeasible, we allow the violation of the factors, and penalize the violations according to their importance. To do so, we define a weight for each factor of the assignment, which represents the importance of the factor in the assignment process. The violations of the factors are hierarchically weighted according to their importance, and the sum of the hierarchically penalized violations serve as the objective function of the MILO model. The mathematical model is presented in detail in the appendix.

The optimization software package Gurobi (2016) was used to solve the MILO models. Having developed the MILO model, it was extensively tested with various real data sets from PADoC with the goal of specifying and fine-tuning the weights of each of the factors, and ensuring the robustness of the model in recommending appropriate simultaneous assignments and program scheduling.

It is worth mentioning that, any time we solve the MILO model and schedule the programs, not everybody who is going to start the programs in the given time horizon is currently in the system. For instance, inmates with short sentence times who need immediate program enrollment enter the correctional system every week. Thus, there is a lot of freedom in scheduling the programs for periods towards the end of the time horizon. As a result, the MILO model has many equally good solutions. This in turn increases the solution time, since a significant amount of time need to be spent to prove optimality. Knowing that proving optimality requires excessive amount of time, we stop the MILO solver when the absolute optimality gap reaches a predefined threshold.

Implementation at the PADoC

The project from idea to successful implementation took five years. Before this project started, inmates were assigned to CIs manually by a staff member of the OPM. This manual process had three main drawbacks:

- A variety of factors need to be considered in assigning each inmate to a CI, including security concerns, mental and medical conditions, program needs, separation from other inmates, capacities of the CIs, home county of the inmates, etc. Having all the factors of the assignment and characteristics and capacities of CIs in mind, and considering them for each individual is time-consuming and prone to human errors. As a result there were numerous inappropriate assignment of the inmates.

- If the inmate assignment is done sequentially, then the inmates that are assigned later, are not considered in the earlier assignments. This makes the process inefficient and suboptimal. In fact, if the assignment is done manually, it is hardly possible to consider the following inmate assignments appropriately in the assignment of the current inmate.

- Scheduling of treatment programs was not considered in the manual inmate assignment. This resulted in inmates having longer waiting times to get their programming, thus postponed their eligibility to go on parole, and so increased the population of the CIs.

The DTDSS, which was initially developed, enabled the PADoC to address the first drawback of the manual inmate assignment and consider the rules and criteria of the assignment in assigning each individual inmate to the CIs. However, DTDSS lacks the ability to simultaneously assign a batch of inmates to the CIs, and it does not consider the treatment program scheduling in the assignment. This stressed the need to develop the multi-objective optimization model, which became the heart of the IADSS. The rigorous optimization model enables OPM to consistently account for all the the factors of the assignment. It also enables OPM to simultaneously assign the inmates to CIs, as well as schedule programs optimally to minimize the waiting time of each individual inmate in starting their program(s).

Development of IADSS

The development of the IADSS took three years. First, a mathematical optimization model was developed as a proof of concept to optimize the simultaneous initial assignment of the inmates to the CIs. It demonstrated to OPM personnel that mathematical optimization provides a powerful tool to optimally assign inmates to the CIs. In conjunction with model development, data had to be harvested from the PADoC databases; thus, data collection and clean up procedures were set up and implemented to link the model to the live databases. The workflow of IADSS is presented in Figure 3.

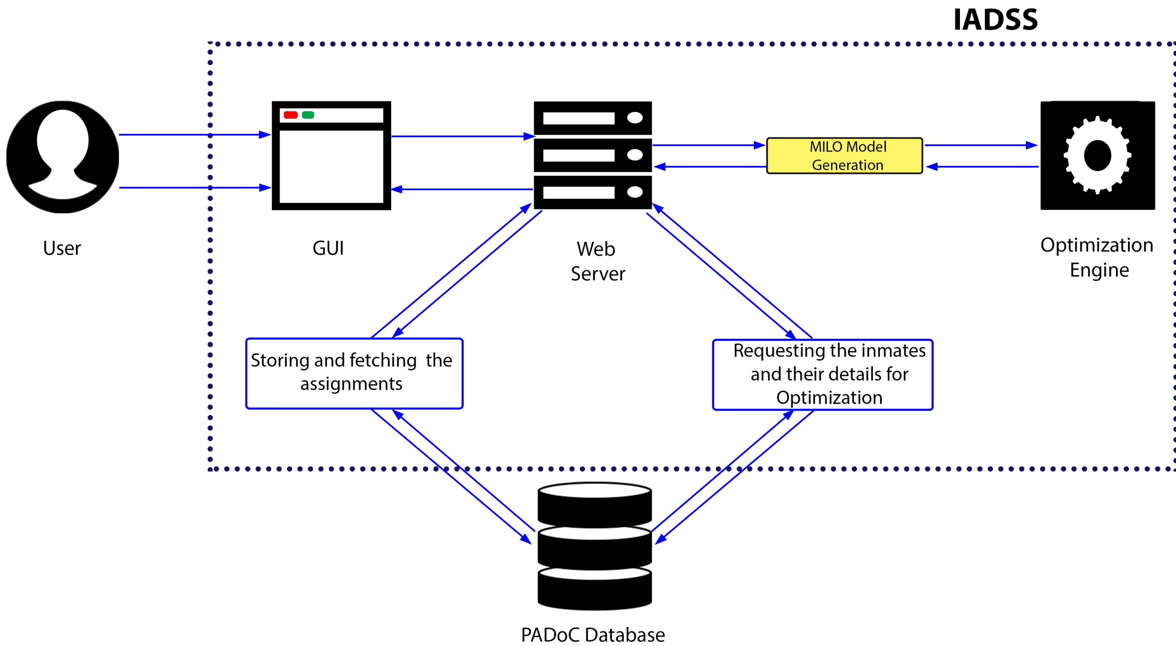


Figure 3: Workflow of the IADSS

The heart of the IADSS is the *optimization module* which generates the mathematical optimization model of the IAP using the data extracted from the PADoC databases, and solves the model. As the inmate assignment to CIs is a multi-objective process, we propose a hierarchical multi-objective optimization model. We consider the weighted sum method (Sawaragi et al. 1985) to combine the objectives. The choice of the weighted sum of the objectives is validated by solving real data instances from the PADoC.

The time sequence of the development phases followed the anticipated increasing mathematical sophistication and complexity of the modules. The violations of the inmate assignment factors were interpreted as the penalty objectives of the assignment and were added one-by-one to the optimization model. As explained in the *Preliminaries and Problem Description* section, we need to make two main decisions: assignment of inmates to the CIs and scheduling the start of their program(s). We initially developed a model which only did the assignment of the inmates to the CIs, and tested the model with real data from PADoC to validate the assignment recommendations. Then we extended the model to include the scheduling of the programs for the inmates. Executing the project in this sequence brought meaningful capability online in a judicious manner, while demonstrating to OPM what was possible with an optimization model, and how to utilize a decision support system to optimally execute their most critical task. The model which does the assignment of the inmates and schedules the programs has been used for the daily assignment of the inmates since September 2016.

Benefits and Impact of the IADSS

The successful development and implementation of IADSS has both significant financial and non-quantifiable human benefits.

High-Quality, Consistent Assignment

- The assignment of the inmates is done simultaneously for all the inmates with a petition for assignment or transfer. Simultaneous assignment ensures system-wide optimum.

- All the factors of the assignment are considered for each individual. As a result, consistently high-quality assignments are made. Current errors are almost exclusively due to data inconsistency, so undesired assignments help OPM to identify data errors.

- The inmate assignment process was previously fragmented in the sense that assignment was done by OPM and the program waiting list was monitored by the Bureau of Treatment Services (BTS) and reported to OPM on a monthly basis. With the implementation of the IADSS, the process is integrated and all the necessary elements of the assignment are considered in one system.

- Program schedules and wait lists at each CI are generated as an integral part of the inmate assignment output. The integrated IADSS minimizes the wait time of the inmates for their required program(s), thus allowing timely release of inmates, and so reducing the inmate population.

- In addition to simultaneous assignment, *individual assignment* can be done for the inmates. Facilities are sorted in the individual assignment for each inmate considering all the factors of the assignment only for that inmate. In case the simultaneous assignment recommendation, for some reason, is not appropriate for an individual, the individual assignment results can be used to evaluate possible assignment to other CIs . The simultaneous assignment recommendation and individual assignment recommendations from the IADSS interface are demonstrated in the first and third panel of Figure 4, respectively.

- Three geographical regions (west, central, east) are defined in PA. Counties and CIs are placed in each of these regions. In Figure 2, the regions of the CIs are given. Due to the complexity of considering the distance of the home county to the CIs, only assignment of an inmate to his home region was considered before. The IADSS enables DoC to consider the actual distance of the home county to the CIs for each inmate.

- The rate of acceptance of the simultaneous assignments and individual assignments has been measured to validate the MILO model and ensure that the MILO model captures the hierarchy of the factors of the assignment. In Jan 2017, over 90% of the inmates were assigned to the facility that was suggested by the simultaneous assignment. Among the remaining 10 percent of the inmates, more than 6 percent were assigned to one of the first three CIs recommended by the individual assignment. The remaining 4 percent, that were assigned to other CIs, were either because of data inconsistency, or the special conditions of those inmates. In Table 1, results of the IADSS for the first 10 days of the year 2017 are presented.

Table 1: Assignment recommendations

Date	# of inmates	Sim. assignment match	Ind. assignment used and matched	Not ind. nor sim. matched	Sim. assignment match	Ind. or sim. assignment match
3 Jan	15	12	3	0	80%	100 %
4 Jan	54	53	1	0	98.15%	100 %
5 Jan	53	43	5	5	81.13 %	90.57 %
9 Jan	14	12	1	1	85.71%	92.86 %
10 Jan	98	91	5	2	92.86 %	97.96 %
Total	234	211	15	8	90.17%	96.58%

User-Friendly Web Application

- A web-based Graphical User Interface (GUI) is developed to enable the user to have interaction with the IADSS. In Figure 4, a screenshot of the GUI is demonstrated.

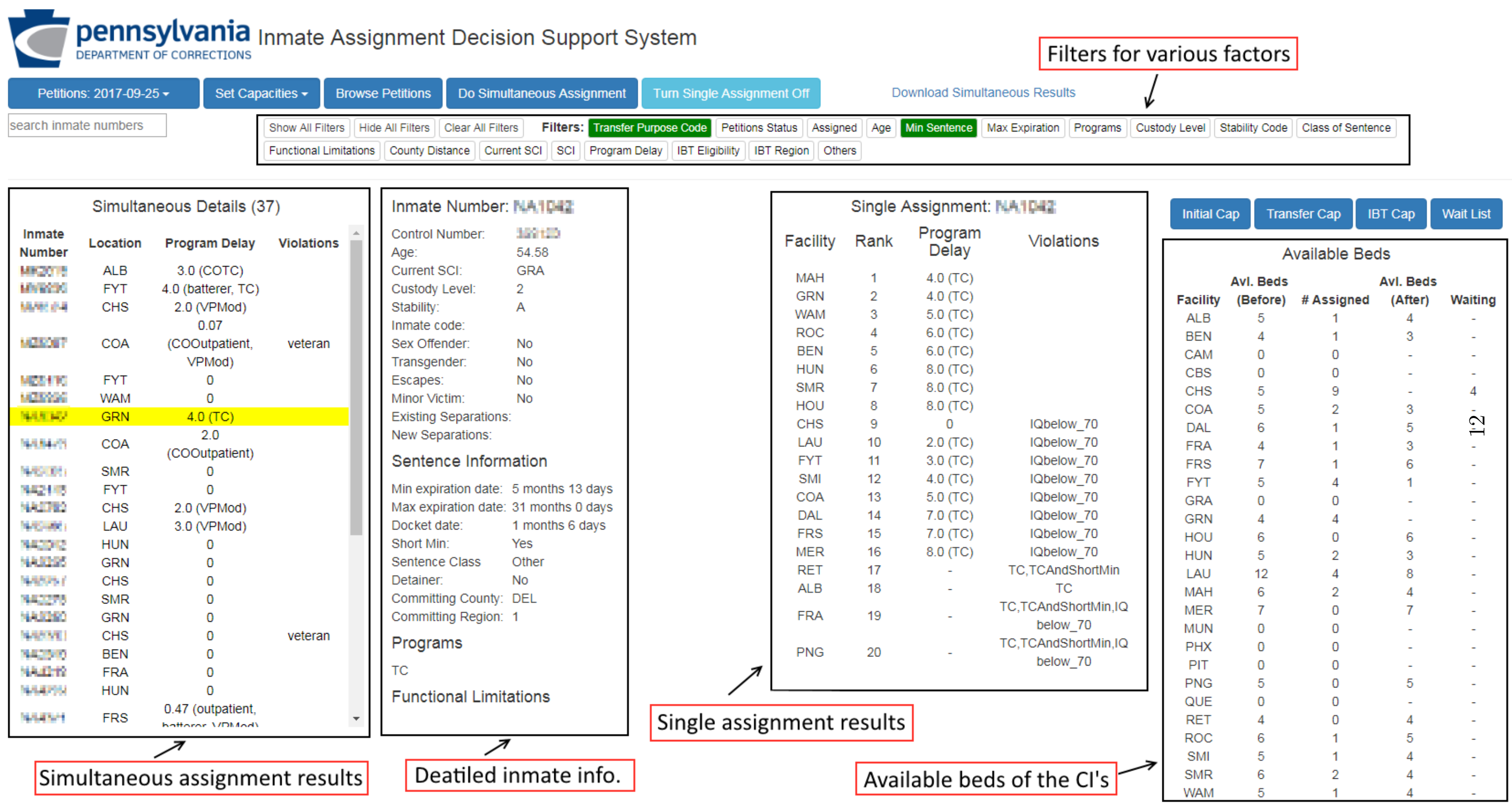


Figure 4: A screen shot of the web-based UI of the IADSS

- All the personal and sentence information needed for the assignment of an inmate is collected and displayed in the GUI to facilitate the review of the assignment. In the second panel of Figure 4, the inmate display page is demonstrated.
- Reporting of the program waiting list alerts BTS for current and future bottlenecks in program schedules and availability.

Security Enhancement

Security enhancement is hard to quantify, however, it was one of the main motivations for initiating this project. The use of IADSS at the PADoC has already resulted in the following identified security enhancements.

- It is stated by the PADoC Secretary, Wetzel, that inmate transportation is one of the riskiest operations at PADoC. By doing proper initial assignment, IADSS has reduced inmate transfers, and so enhancing the security of the CIs and public safety.
- IADSS considers inmates' demographic information and enforces the separations in the assignment, which in turn reduces the number of assaults, thus increasing the security of the CIs.

Quantified Savings

In this section, we present the cost savings resulted from the implementation of IADSS in the first year, and project the benefits to a period of five years. Four areas of significant savings are identified.

- *Reduced waiting time*: IADSS helps to decrease the waiting time for treatment programs, which reduces the length of stay for inmates past their minimum sentence date. We consider the inmates that have less than 9 months to their minimum sentence date at the time of their initial petition, who need at least one treatment program. These inmates must start their programs immediately, since the delay in starting their program(s) directly postpones their parole eligibility. The waiting time of these inmates to start all their programs is calculated with the goal to see how much IADSS has helped to reduce the waiting time for programs. In Figure 5, the cumulative distribution functions of the waiting time of these inmates for the first and second quarters of 2016 (2016-1, 2016-2), and the first and second quarters of 2017 (2017-1, 2017-2) are plotted. Notice that the waiting time in the second quarter of 2017 stops at three months, since we did not have the data for longer waiting times at the time of writing the paper. For both quarters 1 and 2, the cumulative distribution function for 2017 is above and to the left of the one for 2016, showing that the use of IADSS has reduced waiting times substantially. Comparing the waiting time of the inmates with initial petition requests in the first quarter of 2016, when IADSS was not yet used, with the first quarter of 2017, we found out that the average waiting time of the inmates in the first quarter of 2016 is 143 days while the average waiting time of the inmates in the first quarter of 2017 is 89 days. Therefore, the average waiting time decreased by 54 days from 2016-1 to 2017-1.

  In average, PADoC has 10000 initial petitions annually. 12% of those petitions have less than 9 months to their minimum sentence date and need at least one program. The marginal cost of keeping an inmate in a CI is \$16 per day. As a result, the total annual saving of reducing the inmates' waiting time in starting their program(s) is $10000 \times 0.12 \times 54 \times 16 = \$1,036,800$. As we can see in Figure 5, the waiting time is significantly decreasing from 2016 to 2017. Based on already achieved 54 days reduction in the waiting time, 90 days reduction in the waiting time can be projected at the steady state of the system in years 4 and 5. The 90 days reduction in the waiting time of programs enables the PADoC to close a full CI unit. Closing a CI unit allows for more savings than the marginal cost of keeping an inmate in a CI. If a CI unit is closed, the savings per day for each inmate is \$30. Thus, the saving in years 4 and 5 will be $10000 \times 0.12 \times 90 \times 30 = \$3,240,000$.

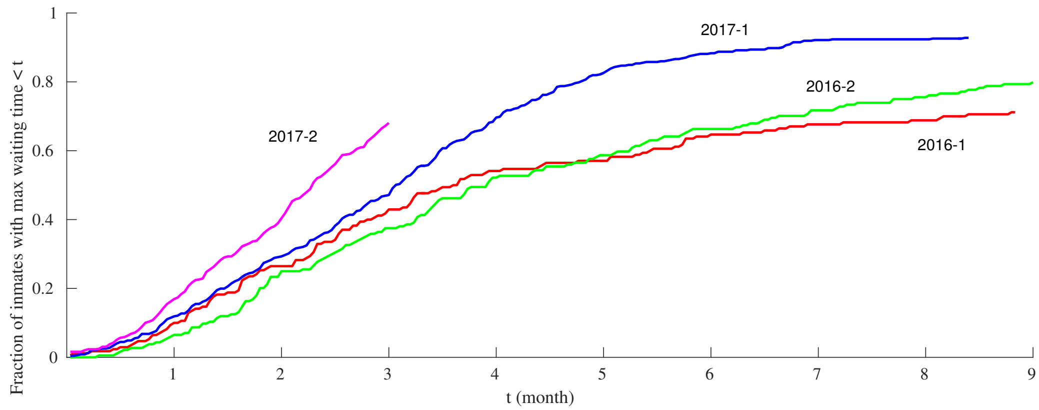


Figure 5: Program waiting time for inmates with less than 9 months to their minimum sentence date

- *Fewer assaults*: There are fewer assaults, due to assigning the right combination of inmates to the most appropriate CIs. We compared the number of assaults in the period January-July of 2017 to the same period in 2016; 95 fewer assaults were reported. If we project this result to the full year, 163 fewer assaults are expected in 2017. The PADoC estimates that approximately 10% to 15% of this reduction is due to the introduction of IADSS, thus IADSS results in 20 fewer assaults in 2017. The criminal justice literature (Cohen 2005) documents that an assault on average costs $70,000. Thus IADSS has resulted in $20 \times 70,000 = \$1,400,000$ saving by reducing the number of assaults.

- *Reduced staff*: Fewer staff are required in OPM to oversee inmate assignments and transfers. As a result of using IADSS, one less Captain position is needed to do the inmate assignments at PADoC. The salary and benefits of a captain is $\$134,742$ annually.

- *Fewer transfers*: Due to initially assigning inmates to the correct CI, fewer transfers are later required. By making better assignments with IADSS, $4,672$ fewer transfers were needed in 2017. The cost of each transfer is on average $82.85 at the PADoC. Hence, the total annual transportation saving is equal to $4,672 \times 82.85 = \$387,075$.

Considering the four main saving points, IADSS has decreased the annual cost at PADoC by $\$2,958,617$, and the projected saving over five years is $\$19,199,485$.

Summary

The IAP is a problem every correctional system faces on a daily basis. Various constraints need to be satisfied in the assignment of inmates, including general assignment factors, CI capacity constraints, scheduling of treatment programs for the inmates, assigning the inmates near their home county, etc. It is impossible to make an ideal assignment, i.e., satisfy all the constraints of the assignment; thus, the IAP is inherently an infeasible problem.

In this paper, we develop a novel hierarchical multi-objective MILO model for the IAP. We penalize the sum of the weighted violation of the constraints of the assignment in the multi-objective MILO model, which is in fact the heart of the IADSS. The IADSS enables PADoC to simultaneously and optimally assign the inmates to the CIs in the PA correctional system, and schedule the treatment programs for them as well, while all the rules and criteria of the assignment are considered. The IADSS minimizes the waiting time of the inmates for getting the required program(s); hence, it facilitates the timely eligibility of the inmates for parole, which ultimately reduces the population of the inmates in the correctional system. The IADSS has been used with proven success at the PADoC for the daily assignment of the inmates to CIs since September 2016.

To our knowledge, this has been the first time that OR methodology was built directly into the routine business operations of a correctional system. The success of this project opens new avenues to: a) adapt and introduce the IADSS methodology to optimize the operations of correctional systems of other states and countries and, b) explore other applications of OR methodology in the complex operations of correctional systems. Correctional systems in this nation and elsewhere have numerous problems that cry out for the use of OR methodologies. This highly successful application of OR in a large correctional system will open a rich application area of OR, just as the first crew scheduling application did in the airline industry.

Acknowledgement

Research is partially supported by a grant from the Pennsylvania Department of Corrections. We gratefully acknowledge the contributions of PADoC personnel for their valuable support of this project. We specifically thank William F. Nicklow, director of the Office of the Population Management, for his leadership and invaluable insight for this project, and Jessica Campbell from the PADoC Bureau of Planning, Research, and Statistics for her substantial, strong support. We also thank the Bureau of Treatment Services and Bureau of Information Technology for their technical assistance. Last but not least, we acknowledge Jennifer Hendricks and Tanya Brandt, from the Office of Population Management, for their tireless efforts in testing the IADSS and patiently providing invaluable and insightful feedback.

Hierarchical Multi-Objective Mathematical Model

In this section, we present a MILO model for the IAP. We first explain the assignment and the treatment program constraints, and finally the objectives of the problem.

Assignment Criteria Constraints

Let $\mathcal{I}$ be the set of inmates waiting to be assigned and let $\mathcal{J}$ be the set of the available CIs for the assignment. Each inmate should be assigned to one facility, i.e.

$$\sum_{j\in\mathcal{J}} x_{ij} = 1 \quad \forall i \in \mathcal{I},$$

where x_{ij}, for all $i \in \mathcal{I}$ and $j \in \mathcal{J}$ is a binary variable and is equal to 1 if inmate i is assigned to facility j. Let $\mathcal{K}$ be the set of general factors, and let coefficient κ_{ik} for $i \in \mathcal{I}$, $k \in \mathcal{K}$ be equal to 1 if factor k applies to inmate i; and equal to 0 otherwise. Additionally, for all $j \in \mathcal{J}$, $k \in \mathcal{K}$ let ρ_{jk} be equal to 1 if facility j can accommodate inmates with factor k; otherwise, $\rho_{jk} = 0$. The following constraints describe the general-factors violations of inmates.

$$\kappa_{ik}\big(1 - \sum_{j\in\mathcal{J}} \rho_{jk} x_{ij}\big) = v_{ik} \quad \forall\ i \in \mathcal{I}, \forall\ k \in \mathcal{K},$$

where v_{ik} indicates the violation of factor k by inmate i and is equal to one if inmate i violates factor k; otherwise, v_{ik} is equal to zero. Furthermore, we have capacity related constraints:

$$\sum_{i\in\mathcal{I}} x_{ij} = s_j \quad \forall j \in \mathcal{J},$$

where s_j, $j \in \mathcal{J}$ denotes the number of the inmates that are assigned to facility j. Let c_j be the capacity of facility j. Ideally, for each pair j_1 and j_2 of CIs, we want to assign inmates proportional to their capacities, i.e., ideally we would have

$$c_{j_1}/s_{j_1} = c_{j_2}/s_{j_2}.$$

Variables $\delta^+_{j_1j_2}, \delta^-_{j_1j_2}$ are the decision variables representing the deviation from assigning inmates proportional to the capacities of CIs j_1 and j_2 and are defined as

$$c_{j_2}s_{j_1} - c_{j_1}s_{j_2} = \delta^+_{j_1j_2} - \delta^-_{j_1j_2} \quad \forall j_1, j_2 \in \mathcal{J}, j_1 \neq j_2. \tag{1}$$

We aim to minimize $\delta^+_{j_1j_2}, \delta^-_{j_1j_2}$ by penalizing them in the objective function.
Additionally, we define upper and lower bounds on the number of inmates that can be assigned to each facility. Let $c^{\min}_j$ and $c^{\max}_j$ be, respectively, the minimum required and maximum allowed capacity of facility j, which are functions of the capacity c_j of facility j. For instance, $c^{\min}_j = \zeta^-_j c_j$ and $c^{\max}_j = \zeta^+_j c_j$ for appropriately chosen constants $\zeta^-_j \leq 1 \leq \zeta^+_j$. Let o_j be the number of inmates assigned over the maximum capacity of facility j and let u_j be the number of inmates needed to reach the minimum capacity of facility j. We have

$$\begin{aligned} s_j &\leq c^{\max}_j + o_j \qquad \forall j \in \mathcal{J}, \\ s_j &\geq c^{\min}_j - u_j \qquad \forall j \in \mathcal{J}. \end{aligned}$$

We aim to minimize o_j and u_j by penalizing them in the objective function.

Another important criterion for the inmate assignment is the *separations*. Considering the history of inmates, there might be pairs of inmates that can not be assigned to the same facility. Let $\mathcal{I}^s$ be the set of inmate pairs that should be separated from each other. Additionally, an inmate might have already on his/her file that he/she has to be separated from certain staff or inmates that are already in a facility. Let $\mathcal{J}^s_i$ be the set of CIs that inmate i should be separated from. We have

$$\begin{aligned} &\sum_{j \in \mathcal{J}^s_i} x_{ij} = 0 && \forall i \in \mathcal{I}, \\ &x_{i_1j} + x_{i_2j} \leq 1 && \forall (i_1, i_2) \in \mathcal{I}^s. \end{aligned}$$

Treatment Program Constraints

Next, we explain the constraints needed to describe the waiting lists of the programs at the CIs. Let $\mathcal{P}^o, \mathcal{P}^c$ be, respectively, the set of open-enrollment and closed-enrollment programs, and let $\mathcal{C}$ be the set of program clusters. Let $\hat{t}_{ip}$ and $\bar{t}_{ip}$ be, respectively, the latest time and earliest time that inmate i is supposed to start program p, and let $\alpha_{ipt} = 1$ if $t \geq \hat{t}_{ip}$, otherwise $\alpha_{ipt} = 0$, i.e., inmate i should not start program p later than t if $\alpha_{ipt} = 1$. Similarly, $\beta_{ipt} = 1$ if $t \geq \hat{t}_{ip}$, otherwise $\beta_{ipt} = 0$, i.e., inmate i can start program p at time t if $\beta_{ipt} = 1$.

We would like to minimize the number of inmates that can not start their programs earlier than their latest start time $\hat{t}_{ip}$. The decision variable y_{jpt} represents the number of inmates at facility j that are prescribed program p and have to start it by time t but can not do so. We aim to minimize y_{jpt} by penalizing it in the objective function.

Let $\mathcal{T} = \{1, 2, \ldots, t'\}$ be the set of the time periods in our decision horizon. Parameter t' is the last time period in the decision horizon, and let ψ_{jpt}, $\underline{q}_{jpt}$, and $\overline{q}_{jpt}$ be defined as

ψ_{jpt}: The number of inmates starting program p at t in facility j.

$\underline{q}_{jpt}$: The number of inmates, already in facility j, that should start program p at time t or earlier, i.e., the number of inmates with $\hat{t}_{ip} \leq t$

$\overline{q}_{jpt}$: The maximum number of inmates, already in facility j, that can start program p at time t, i.e., the number of inmates with $\bar{t}_{ip} \leq t$.

The following two sets of constraints compute the lower and upper bound on the number of inmates that can start the programs at each time period in the CIs

$$\begin{aligned} &\sum_{i \in \mathcal{I}} \alpha_{ipt} x_{ij} + \underline{q}_{jpt} \leq y_{jpt} + \sum_{\tau=0}^{t} \psi_{jp\tau} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\ &\sum_{i \in \mathcal{I}} \beta_{ipt} x_{ij} + \overline{q}_{jpt} \geq y_{jpt} \sum_{\tau=0}^{t} \psi_{jp\tau} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}. \end{aligned}$$

Let R_{jpt} be the number of available spots for open-enrollment program p at time t in facility j. The following constraints assure that the number of inmates starting an open-enrollment program does not exceed the number of spots available for that program at the CIs

$$\sum_{\tau=\max(0,t-d_p)}^{t} \psi_{jp\tau} \leq R_{jpt} \quad \forall j \in \mathcal{J}, \forall p \in \mathcal{P}^o,\ \forall t \in \mathcal{T},$$

where d_p is the duration of program p.
Next, we explain the constraints related to the closed-enrollment programs. As mentioned previously, closed-enrollment programs are categorized in clusters. All the programs in a cluster can be facilitated by one instructor, i.e., programs in a cluster use common instructors. Let R'_{jct} and ψ'_{jpt} be defined as

ψ'_{jpt}: The number of groups of the closed program p that start at time t in facility j.

R'_{jct}: The number of available groups of cluster c that can start at time t in facility j.

Then we have

$$\sum_{p\in\mathcal{P}_c} \sum_{\tau=\max(0,d_p)}^{t} \psi'_{jp\tau} \leq R'_{jct} \qquad \forall j \in \mathcal{J}, \forall c \in \mathcal{C},\ \forall t \in \mathcal{T},$$

where $\mathcal{P}_c$ is the set of the programs of the cluster c. Let $\underline{G}_p$ and $\overline{G}_p$ be, respectively, the minimum and maximum number of inmates that can be enrolled in closed-enrollment program p. The following set of constraints enforce these capacity bounds for the closed-enrollment programs.

$$\underline{G}_p\psi'_{jpt} \leq \psi_{jpt} \leq \overline{G}_p\psi'_{jpt} \qquad \forall j \in \mathcal{J},\ \forall p \in \mathcal{P}^c,\ \forall t \in \mathcal{T}.$$

Scheduling of the Programs for the Inmates

One of the main objectives of the IAP is to minimize the maximum waiting time of inmates to start their program(s). In this section, we present the constraints needed to minimize the maximum waiting time of inmates to start their program(s).

Let $\mathcal{T}' = \mathcal{T} \cup \{\infty\}$, and let $\mathcal{P}_i$ be the set of the programs prescribed for inmate i. The new decision variable z_{ijpt}, for $i \in \mathcal{I},\ j \in \mathcal{J}, p \in \mathcal{P}_i, t \in \mathcal{T}'$, is equal to one if inmate i is assigned to facility j, starting program p at time t; otherwise, it is equal to zero. If $z_{ijp\infty} = 1$, it implies that inmate i is not going to start program p in the decision horizon, i.e. later than the last time period of the decision horizon. Following is the set of constraints that define the relationship between z_{ijpt} and x_{ij}

$$\sum_{t\in\mathcal{T}'} z_{ijpt} = x_{ij} \qquad \forall i \in \mathcal{I},\ \forall j \in \mathcal{J},\ \forall p \in \mathcal{P}_i.$$

Let y^a_{jpt} and ψ^a_{jpt}, for $j \in \mathcal{J},\ p \in \mathcal{P},\ t \in \mathcal{T}$, be defined as follows,

y^a_{jpt}: The number of inmates already in facility j, who are prescribed program p and have to start it by time t but can not do so.

ψ^a_{jpt}: The number of inmates already in facility j, starting program p at time t.

We have

$$\underline{q}_{jpt} \leq \sum_{\tau=0}^{t} \psi^a_{jp\tau} + y^a_{jpt} \leq \bar{q}_{jpt} \quad \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}.$$

Additionally, let y^n_{jpt} and ψ^n_{jpt}, for $j \in \mathcal{J},\ p \in \mathcal{P},\ t \in \mathcal{T}$, be defined as follows

y^n_{jpt}: The number of inmates assigned to facility j, who are prescribed program p and have to start it by time t, but can not do so.

ψ^n_{jpt}: The number of inmates assigned to facility j, starting program p at time t.

We have

$$\psi^n_{jpt} = \sum_{i \in \mathcal{I}_p} z_{ijpt} \qquad \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T},$$

$$\sum_{i \in \mathcal{I}} \alpha_{ipt} x_{ij} \le \sum_{\tau=1}^{t} \psi^n_{jp\tau} + y^n_{jpt} \le \sum_{i \in \mathcal{I}} \beta_{ipt} x_{ij} \quad \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T},$$

where $\mathcal{I}_p$ is the set of the inmates who need program p.

Suppose the number of inmates, already in facility j that should start program p at time t is more than the available spots for program p at time t. Then we have $y^a_{jpt} > 0$. In this case, ψ^n_{jpt} should be equal to zero. In other words, if there are not enough spots of program p for the inmates that are already in facility j, then the number of inmates, assigned to facility j through the model, that are going to start program p at time t should be zero. In order to satisfy this constraint, the indicator variable ϕ_{jpt}, for $j \in \mathcal{J},\ p \in \mathcal{P},\ t \in \mathcal{T}$ is equal to one if $y^a_{jpt} > 0$; otherwise, it is equal to zero. Then we have

$$\begin{aligned} &y^a_{jpt} \le M\phi_{jpt} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\ &\psi^n_{jpt} \le M(1 - \phi_{jpt}) && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \end{aligned}$$

where M is a big number.

Additionally, we have the following set of constraints, which defines the relationship between the decision variables of the problem

$$\begin{aligned} &\psi^a_{jpt} + \psi^n_{jpt} = \psi_{jpt} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\ &y^a_{jpt} + y^n_{jpt} = y_{jpt} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \end{aligned}$$

Let w_{ip}, for $i \in \mathcal{I},\ p \in \mathcal{P}_i$ be the waiting time of inmate i to start program p after his latest possible start time $\hat{t}_{ip}$. We have

$$w_{ip} = \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} \max(0, t - \hat{t}_{ip}) z_{ijpt} \qquad \forall i \in \mathcal{I},\ \forall p \in \mathcal{P}_i.$$

Finally, let w'_i be the maximum waiting time of inmate i to start his/her program(s). Then

$$w'_i \ge w_{ip} \qquad \forall i \in \mathcal{I},\ \forall p \in \mathcal{P}_i.$$

Transfer Constraints

The constraints needed to account for the inmate transfers after the initial assignment are the same kind of constraints as the ones for the initial assignment in the current model. However, as the importance of these constraints are frequently different for transfers, the weights of the factors in the objective function differ from an initial assignment.

The Objective Function

The IAP is a multi-objective problem. There are different approaches in the literature to deal with a multi-objective optimization problem. We consider the weighted sum method (Sawaragi et al. 1985) to combine the objectives and have a one-shot optimization in assigning the inmates. The choice of the weighted sum of the objectives is validated by solving real data instances from the Pennsylvania Department of Corrections. It is worth mentioning that the weights of all the objectives are assumed to be positive. The objectives of the IAP are listed as follows:

- Violation of the general factors should be minimized. The violation is equal to

$$\vartheta = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \lambda^f_{ik} v_{ik},$$

  where λ^f_{ik} is the weight of factor k for inmate i.

- Assignment of inmates under the capacity and over the capacity of the CIs should be minimized. The violations of the capacity constraints are defined as

$$o_j = \sum_{i \in \mathcal{I}} x_{ij} - c_j^{\max} \quad \forall j \in \mathcal{J},$$
$$u_j = c_j^{\min} - \sum_{i \in \mathcal{I}} x_{ij} \quad \forall j \in \mathcal{J}.$$

  Then, the overall capacity violation is equal to

$$\eta = \sum_{j \in \mathcal{J}} \lambda_j^o o_j + \lambda_j^u u_j,$$

  where λ_j^o and λ_j^u for $j \in \mathcal{J}$ are, respectively, the weights of over-assignment and under-assignment to the CIs.

- The difference between the capacities of the CIs should be minimized

$$\delta = \lambda^\delta \sum_{j_1 \in \mathcal{J}} \sum_{j_2 \in \mathcal{J} | j_2 \neq j_1} \left(\delta_{j_1 j_2}^+ + \delta_{j_1 j_2}^- \right),$$

  where λ^δ is the weight of the capacity difference, and $\delta_{j_1 j_2}^+$ and $\delta_{j_1 j_2}^-$ are defined in equation (1).

- Distance to the home county of the CIs should be minimized.

$$\gamma = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \lambda_i^d d_{ij} x_{ij},$$

  where λ_i^d is the weight of the distance for inmate i.

- The number of inmates that can not start their program on time should be minimized.

$$\omega = \sum_{j \in \mathcal{J}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} \lambda_{jpt}^\omega \, y_{jpt},$$

  where λ_{jpt}^ω is the weight of the wait list of program p at facility j in time t.

- The maximum program waiting time of inmates need to be minimized

$$\theta = \sum_{i \in \mathcal{I}} \lambda_i^\theta w_i',$$

  where λ_i^t is the penalty weight of waiting time of inmate i.

The weighted sum of the objectives is defined as

$$\lambda_\vartheta \vartheta + \lambda_\eta \eta + \lambda_\delta \delta + \lambda_\gamma \gamma + \lambda_\omega \omega + \lambda_\theta \theta,$$

where the weights of all the objective elements are positive. Objective hierarchies are being enforced through order of magnitude differences in the weight applied. General factors have the highest priority in assigning inmates to CIs. Minimizing the maximum waiting time for each inmate is second in the hierarchy of objectives. Assigning in the range of the minimum and maximum capacity of each facility has the next highest priority. Additionally, in order to reduce the population of the CIs, program waiting lists have a high priority in the objective function. Assigning inmates to a facility near their home county is less important compared to the other objectives of the problem.

The Multi-Objective MILO Model

Now we present the complete optimization model for the inmate assignment and scheduling problem. The lists of parameters and decision variables of IAP are summarized in Tables 2 and 3, respectively. We utilize the hierarchically weighted sum method to combine the objectives and have a single-objective optimization problem. The MILO model is as follows:

$$
\begin{aligned}
\min \quad & \lambda_\vartheta \vartheta + \lambda_\eta \eta + \lambda_\delta \delta + \lambda_\gamma \gamma + \lambda_\omega \omega + \lambda_\tau \tau \\
\text{s.t.} \quad & \sum_{j \in \mathcal{J}} x_{ij} = 1 && \forall i \in \mathcal{I}, \\
& \sum_{t \in \mathcal{T}'} z_{ijpt} = x_{ij} && \forall i \in \mathcal{I}, \forall j \in \mathcal{J},\ \forall p \in \mathcal{P}_i, \\
& \psi^n_{jpt} = \sum_{i \in \mathcal{I}_p} z_{ijpt}, && \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& \kappa_{ik}\big(1 - \sum_{j \in \mathcal{J}} \rho_{jk} x_{ij}\big) = v_{ik} && \forall\ i \in \mathcal{I}, \forall\ k \in \mathcal{K}, \\
& \sum_{i \in \mathcal{I}} \alpha_{ipt} x_{ij} + \underline{q}_{jpt} \le y_{jpt} + \sum_{\tau=0}^{t} \psi_{jp\tau} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& \sum_{i \in \mathcal{I}} \beta_{ipt} x_{ij} + \overline{q}_{jpt} \ge y_{jpt} + \sum_{\tau=0}^{t} \psi_{jp\tau} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& \sum_{\tau=\max(0, t-\hat{t}_p)}^{t} \psi_{jp\tau} \le R_{jpt} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P}^o,\ \forall t \in \mathcal{T}, \\
& \underline{G}_p \psi'_{jpt} \le \psi_{jpt} \le \overline{G}_p \psi'_{jpt} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P}^c,\ \forall t \in \mathcal{T} \\
& \sum_{p \in \mathcal{P}_c} \sum_{\tau=\max(0,d_p)}^{t} \psi'_{jp\tau} \le R'_{jct} && \forall j \in \mathcal{J}, \forall c \in \mathcal{C},\ \forall t \in \mathcal{T} \\
& \underline{q}_{jpt} \le \sum_{\tau=1}^{t} (\psi^a_{jp\tau}) + y^a_{jpt} \le \overline{q}_{jpt} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& \sum_{i \in \mathcal{I}} \alpha_{ipt} x_{ij} \le \sum_{\tau=1}^{t} (\psi^n_{jpt}) + y^n_{jp\tau} \le \sum_{i \in \mathcal{I}} \beta_{ipt} x_{ij}, && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& y^a_{jpt} \le M \phi_{jpt} && \forall j \in \mathcal{J}, \forall p \in \mathcal{P}, \forall t \in \mathcal{T}, \\
& \psi^n_{jpt} \le M(1 - \phi_{jpt}) && \forall j \in \mathcal{J}, \forall p \in \mathcal{P}, \forall t \in \mathcal{T}, \\
& \psi^a_{jpt} + \psi^n_{jpt} = \psi_{jpt} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& y^a_{jpt} + y^n_{jpt} = y_{jpt} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& w_{ip} = \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} \max(0, t - \hat{t}_{ip}) z_{ijpt} && \forall i \in \mathcal{I},\ \forall p \in \mathcal{P}_i, \\
& w'_i \ge w_{ip} && \forall i \in \mathcal{I},\ \forall p \in \mathcal{P}_i, \\
& \sum_{i \in \mathcal{I}} x_{ij} = s_j && \forall j \in \mathcal{J}, \\
& c_{j_2} s_{j_1} - c_{j_1} s_{j_2} = \delta^+_{j_1 j_2} - \delta^-_{j_1 j_2} && \forall j_1, j_2 \in \mathcal{J}, \\
& s_j \le c^{\max}_j + o_j && \forall j \in \mathcal{J},
\end{aligned}
$$

$$
\begin{aligned}
& s_j \geq c_j^{\min} - u_j && \forall j \in \mathcal{J}, \\
& \sum_{j \in \mathcal{J}_i^s} x_{ij} = 0 && \forall i \in \mathcal{I}, \\
& x_{i_1 j} + x_{i_2 j} \leq 1 && \forall (i_1, i_2) \in \mathcal{I}^s, \\
& z_{ijpt} \in \{0,1\} && \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \forall p \in \mathcal{P}_i, \forall t \in \mathcal{T}', \\
& x_{ij} \in \{0,1\} && \forall i \in \mathcal{I},\ \forall j \in \mathcal{J}, \\
& v_{ik} \in \{0,1\} && \forall i \in \mathcal{I},\ \forall k \in \mathcal{K}, \\
& \phi_{jpt} \in \{0,1\} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& y_{jpt}^a, y_{jpt}^n, y_{jpt} \in \mathbb{N} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& \psi_{jpt}^a, \psi_{jpt}^n, \psi_{jpt} \in \mathbb{N} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P},\ \forall t \in \mathcal{T}, \\
& \psi'_{jpt} \in \mathbb{N} && \forall j \in \mathcal{J},\ \forall p \in \mathcal{P}^c,\ \forall t \in \mathcal{T}, \\
& s_j,\ o_j,\ u_j \in \mathbb{N} && \forall j \in \mathcal{J}, \\
& \delta_{j_1 j_2}^+, \delta_{j_1 j_2}^- \in \mathbb{N} && \forall j_1, j_2 \in \mathcal{J},\ j_1 \neq j_2, \\
& w_{ip} \geq 0 && \forall i \in \mathcal{I},\ \forall p \in \mathcal{P}_i, \\
& w_i \geq 0 && \forall i \in \mathcal{I}.
\end{aligned}
$$

Table 2: The parameters of the IAP

Parameter	Definition
$\mathcal{I}$	The set of inmates that need to be assigned
$\mathcal{J}$	The set of the CIs
$\mathcal{K}$	The set of factors
$\mathcal{P}$	The set of programs
$\mathcal{P}^o$	The set of open-enrollment programs
$\mathcal{P}^c$	The set of closed-enrollment programs
$\mathcal{C}$	The set of program clusters
$\mathcal{P}_i$	The set of the program(s) of inmate i
$\mathcal{P}_c$	The set of closed-enrollment programs of cluster c
$\mathcal{J}_i^s$	Set of CIs that inmate i should be separated from
$\mathcal{I}^s$	Set of inmate pairs that should be separated from each other
$\mathcal{T}$	The set of the time periods in the time horizon
$\mathcal{T}'$	$\mathcal{T} \cup \infty$
κ_{ik}	1 if factor k applies to inmate i 0 otherwise
ρ_{jk}	1 if CI j can accommodate inmates with factor k 0 otherwise
$\hat{t}_ip$	The latest time that inmate i can start program p and finish it before his scheduled board meeting
$\tilde{t}_ip$	The earliest time that inmate i can start program p based on the system regulations
α_{ipt}	1 if inmate i should not start program p later than time t 0 otherwise
β_{ipt}	1 if inmate i can start program p at time t 0 otherwise
$\underline{q}_{jpt}$	The number of inmates, already in facility j, that should have started program p by time t to be able to finish their program before their parole board meeting, i.e., the number of inmates with $\hat{t}_{ip} \leq t$
$\overline{q}_{jpt}$	The maximum number of inmates, already in facility j, that can start program p at time t, i.e., the number of inmates with $\tilde{t}_{ip} \leq t$
R_{jpt}	Number of spots available for open-enrollment program p in CI j at time period t
R'_{jct}	Number of groups available for cluster c in CI j at time period t
$\overline{G}_p$	Maximum number of inmates in a group of program p
$\underline{G}_p$	Minimum number of inmates needed to run program p
d_{ij}	The distance between the home county of inmate i to facility j
c_j	Capacity of facility j
$c_j^{\min}$, $c_j^{\max}$	Minimum and maximum capacity at facility j which are functions of c_j

Table 3: The decision variables of the IAP

Parameter	Definition
x_{ij}	1 if inmate i is assigned to CI j 0 otherwise
z_{ijpt}	1 if inmate i is assigned to CI j, starting program p at time t 0 otherwise
v_{ik}	1 if inmate i violates factor k 0 otherwise
ψ_{jpt} ψ^a_{jpt} ψ^n_{jpt} ψ'_{jpt}	The number of inmates starting program p at t in facility j The number of inmates already in facility j, starting program p at time t The number of inmates assigned to facility j, starting program p at time t The number of groups of the closed program p that start at time t in facility j
y_{jpt} y^a_{jpt} y^n_{jpt}	The number of inmates at facility j that are prescribed program p and have to start it by time t but can not do so The number of inmates already in facility j, who are prescribed program p and have to start it by time t but can not do so The number of inmates assigned to facility j, who are prescribed program p and have to start it by time t, but can not do so
ϕ_{jpt}	1 if $y^a_{jpt} > 0$ 0 otherwise
w_{ip} w'_i	The waiting time of inmate i to start program p from his latest possible start time $\hat{t}_{ip}$ The maximum waiting time of inmate i to start his/her program(s)
s_j o_j u_j	The total number of inmates assigned to facility j The number of inmates assigned over the maximum capacity of facility j The number of inmates assigned under the minimum capacity of facility j
$\delta^+_{j_1 j_2}, \delta^-_{j_1 j_2}$	Variables representing the difference in capacities between the CIs j_1 and j_2

We can strengthen the MILO model formulation by adding a set of constraints for the inmates who have prescribed program(s) as follows

$$\sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}'} z_{ijpt} = 1 \qquad \forall i \in \mathcal{I},\ \forall p \in \mathcal{P}_i.$$

While these constraints are redundant, notably if we add them to the model, the solution time decreases significantly. Further, in order to generate a good solution quickly, we set the MILO solver to perform the highest level of preprocessing before starting the branch & bound algorithm, which further reduces the overall solution time.

References

Arabeyre JP, Fearnley J, Steiger FC, Teather W (1969) The airline crew scheduling problem: a survey. *Transportation Science* 3(2):140–163.

Caprara A, Toth P, Vigo D, Fischetti M (1998) Modeling and solving the crew rostering problem. *Operations Research* 46(6):820–830.

Cohen MA (2005) *The Costs of Crime and Justice* (Routledge).

Dantzig GB (1951) Application of the simplex method to a transportation problem. Koopmans TC, ed., *Activity Analysis of Production and Allocation*, volume 13, 359–373 (John Wiley and Sons).

Davis L, Bozick R, Steele J (2013) *Evaluating the Effectiveness of Correctional Education: A Meta-Analysis of Programs That Provide Education to Incarcerated Adults.* G - Reference, Information and Interdisciplinary Subjects Series (RAND Corporation).

Flood MM (1953) On the Hitchcock distribution problem. *Pacific Journal of Mathematics* 3(2):369–386.

Gurobi Optimization Inc (2016) Gurobi optimizer reference manual, http://www.gurobi.com.

Kuhn HW (1955) The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2(1-2):83–97.

Kyckelhahn T, Martin T (2010) Justice expenditure and employment extracts, 2010 - preliminary. http://www.bjs.gov/index.cfm?ty=pbdetail&iid=4679, accessed: 2017-08-15.

Li D, Plebani L, Terlaky T, Wilson GR, Bucklen KB (2014) Inmate classification: decision support tool gives help to pennsylvania department of corrections. volume 46, 44–48 (Industrial Engineer).

Mai C, Subramanian R (2017) The price of prisons: examining state spending trends, 2010-2015. https://storage.googleapis.com/vera-web-assets/downloads/Publications/price-of-prisons-2015-state-spending-trends/legacy_downloads/the-price-of-prisons-2015-state-spending-trends.pdf, accessed: 2017-08-15.

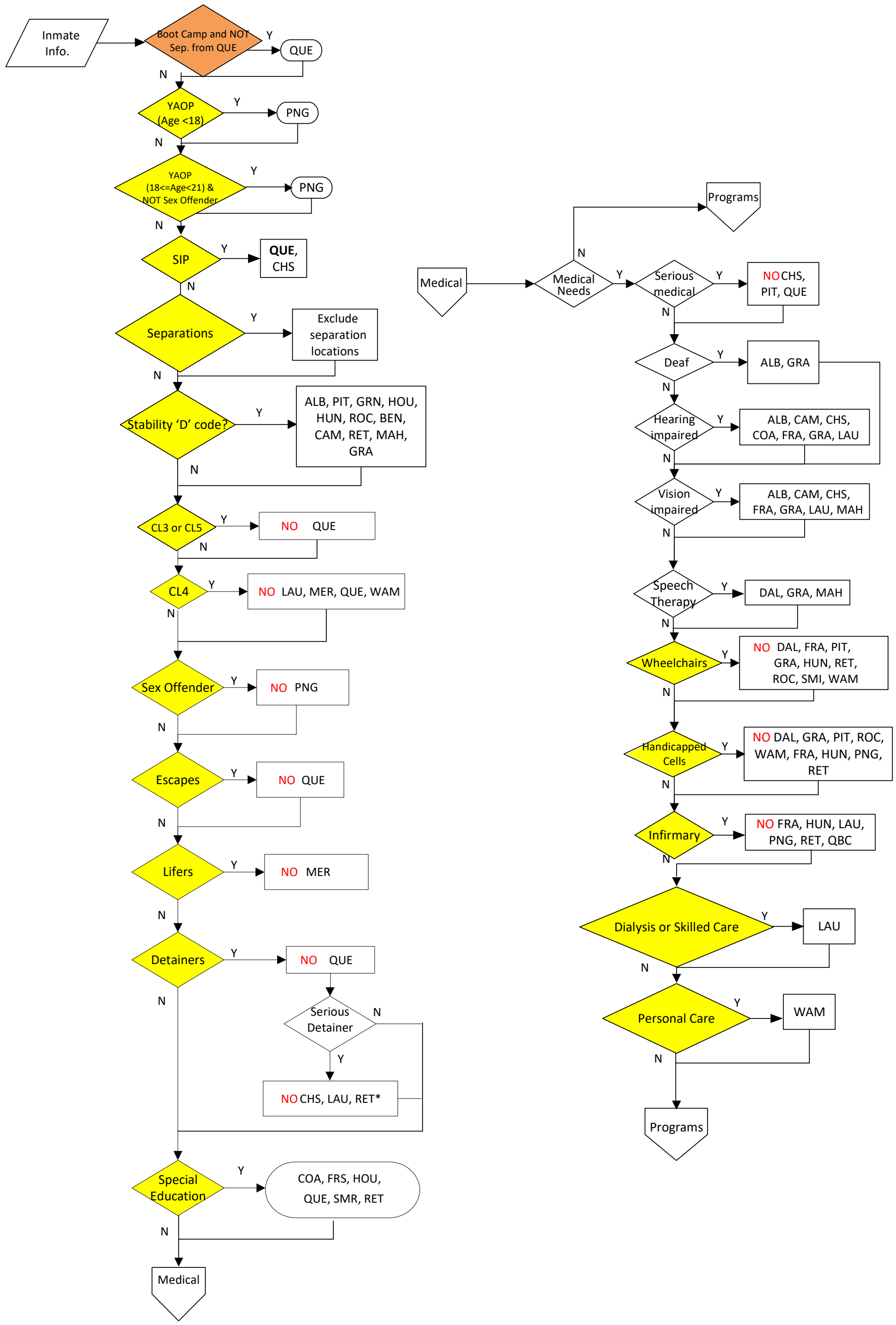
Orden A (1951) A procedure for handling degeneracy in the transportation problem. *DCS/Comptroller, Headquarters US Air Force, Washington, DC* .

Sawaragi Y, Nakayama H, Tanino T (1985) *Theory of Multiobjective Optimization*, volume 176 of *Mathematics in Science and Engineering* (Elsevier).

Schanzenbach DW, Nunn R, Bauer L, Breitwieser A, Mumford M, Nantz G (2016) Twelve facts about incarceration and prisoner reentry. *The Hamilton Project* .

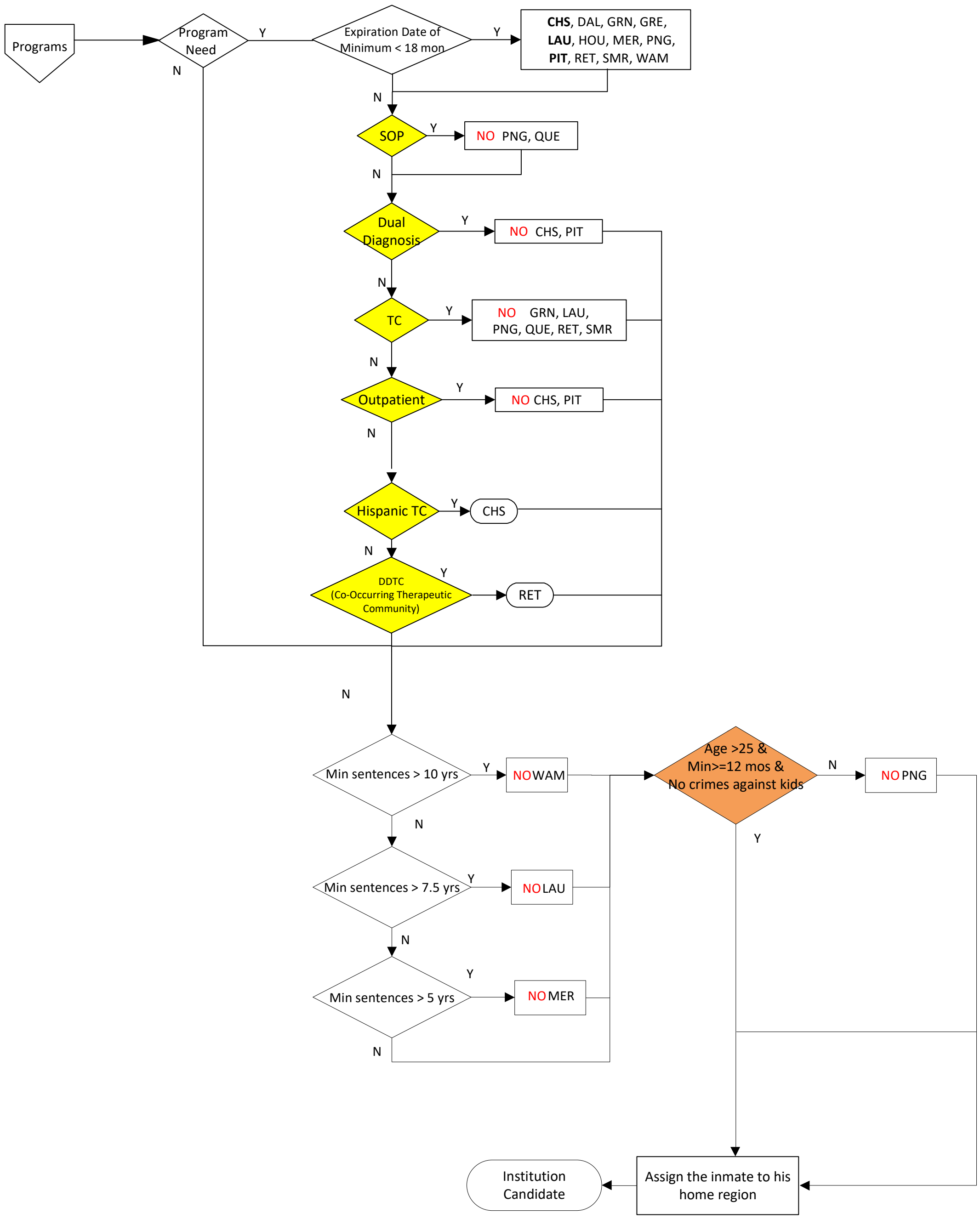
Votaw DF, Orden A (1952) The personnel assignment problem. *Symposium on Linear Inequalities and Programming, SCOOP 10, USAF* 155–163.

Walmsley R (2017) World prison population list. http://www.prisonstudies.org/sites/default/files/resources/downloads/world_prison_population_list_11th_edition_0.pdf, accessed: 2017-08-15.



(a) Section one

Figure 1: The decision tree of the inmate assignment process: There are three major types of nodes in the decision tree: judge nodes (denoted as diamonds), showing different outcomes corresponding to a condition; activity nodes (denoted as rectangular box), presenting the current decision pool; and end nodes (denoted as rectangular box with rounded corners), indicating final decision(s).



(b) Section two

Figure 1: The decision tree of the inmate assignment process