MODELING AND COMPUPATIONAL STRATEGIES FOR MEDICAL DECISION MAKING

A Dissertation
Presented to
The Academic Faculty

by

Fan Yuan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2015

MODELING AND COMPUTATIONAL STRATEGIES FOR MEDICAL DECISION MAKING

Approved by:

Dr. Eva K. Lee, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Andy Sun
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Yajun Mei
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. James C.H. Chu
College of Health Sciences
Rush University

Dr. Renato D.C. Monteiro
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: November 19, 2014

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisor, Dr. Eva Lee, for her supervision, guidance, and financial support for these years. Without her, the accomplishment of this dissertation would have been impossible.

I want to express my appreciation to Dr. Renato Monteiro, Dr. Yajun Mei, Dr. Andy Sun, and Dr. James Chu for serving on my dissertation committee. In particular, I am thankful to their valuable feedback on my work.

I want to acknowledge the National Science Foundation, the National Institutes of Health, and the Centers for Disease Control and Prevention for their funding support.

I would like to acknowledge many ISyE faculty who teach me professional knowledge, ISyE staff who give me administrative support, and current and former fellow students who help me in both of my academic and personal life. All of them play important roles during my doctoral study.

I also want to thank all my friends in Atlanta for their company in these years. They make my life delightful, and I am particularly pleased that I have gained some lifelong friendships.

I really appreciate my parents for the endless support and the sacrificial love.
To my wife, Fei Zhao, thank you for staying with me. It would be impossible for me to complete the Ph.D without your support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Technological advances in prevention, diagnosis, and treatment of diseases help predict disease, prolong life, and promote health. However, with the increase in volume and complexity of data and evidence, medical decision making can be a complex process. Many decisions involve uncertainties and tradeoffs, and can have serious consequences to patients and the clinical practice. For example, to design a radiation therapy treatment plan, physicians must determine the locations of over 50 seeds to deliver precise dosage to the tumor such that the cancer cells are killed while the functionalities of surrounding organs are preserved. To make such complex decisions, providers must balance the potential harm and benefit of medical interventions. Computational methods such as mathematical programming, simulation, and classification have found broad applications in these areas.

In this dissertation, we investigate three topics: predictive models for disease diagnosis and patient behavior, optimization for cancer treatment planning, and public health decision making for infectious disease prevention.

In the first topic, we propose a multi-stage classification framework that incorporates Particle Swarm Optimization (PSO) for feature selection and discriminant analysis via mixed integer programming (DAMIP) for classification. By utilizing the reserved judgment region, it allows the classifier to delay making decisions on 'difficult-to-classify' observations and develop new classification rules in later stage. We apply the framework to four real-life medical problems: 1) Patient readmissions: identifies the patients in emergency department who return within 72 hours using patient's demographic information, complaints, diagnosis, tests, and hospital real-time utility. 2) Flu vaccine responder: predicts high/low responders of flu vaccine on subjects in 5 years using gene signatures. 3) Knee reinjection: predicts whether a patient needs to take a second surgery within 3 years of his/her first knee injection and tackles with missing

data. 4) Alzheimer's disease: distinguishes subjects in normal, mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups using neuropsychological tests.

In the second topic, we first investigate multi-objective optimization approaches to determine the optimal dose configuration and radiation seed locations in brachytherapy treatment planning. Tumor dose escalation and dose-volume constraints on critical organs are incorporated to kill the tumor while preserving the functionality of organs. Based on the optimization framework, we propose a non-linear optimization model that optimizes the tumor control probability (TCP). The model is solved by a solution strategy that incorporates piecewise linear approximation and local search.

In the third topic, we study optimal strategies for public health emergencies under limited resources. First we investigate the vaccination strategies against a pandemic flu to find the optimal strategy when limited vaccines are available by constructing a mathematical model for the course of the pandemic flu and the process of the vaccination. Second, we analyze the cost-effectiveness of emergency response strategies again a large-scale anthrax attack to protect the entire regional population.

CHAPTER I
OVERVIEW OF MEDICAL DECISION MAKING

Medical decisions are intrinsically complex. They affect the health and clinical care of individuals and they can also influence or facilitate health policy development. Recent years have seen numerous technological advances in prevention, diagnosis, and treatment of diseases. These innovations assist in preventing illness, prolonging life, and promoting health. However they also add new dimensions and extra complexity to the medical decision making process. More parameters, variables, and effects have to be taken into account to determine the best course of actions. Since medical decisions have substantial consequences and involve uncertainties and trade-offs, decision makers must balance the potential harm and benefit of their chosen interventions. Computational methods such as mathematical programming, simulation, machine learning and classification have found broad applications in this area to assist in determining the best decision(s), understanding the alternatives, and estimating the impact of each option. In the following sections, we briefly review some emerging challenges and opportunities in applying these computational methods.

Disease diagnosis. Historically, diseases were detected when symptoms manifested themselves. Screening and surveillance methods including imaging, laboratory and blood tests help achieve earlier detection, resulting in improved quality of life and/or reduced morbidity. With the advent in human genome sequencing, microarrays, omics, and discovery of biomarkers, large-scale biological and clinical data are frequently generated. These big data require powerful and sophisticated classification and predictive tools for effective analysis to accelerate the diagnosis process.

Patient behavior. An important and challenging aspect in medical advances entails understanding patient behavior and outcome success. Learning patient behavior may allow hospitals to improve their care to the patients. One example is to predict the

length of stay of arriving patients and the unnecessary return visits in emergency department to avoid potential overcrowding.

Treatment decisions. Treatment options are now readily available to many diseases that can cure or delay their progression. In some cases treatment decisions are one-time decisions, but in other cases treatment decisions recur, often involving the coordination of multiple treatment types to achieve control of one or more risk factors. An important application area is the design of radiation therapy treatment plans for cancer patients. Here, the decision involves three major pieces 1) selectingthe radiation sources, 2) deciding on the amount of radiation dosage to deliver, and 3) optimizing the delivery of the treatment.

Disease prevention. Preventing diseases, especially infectious diseases, requires organized efforts of public health and the society. One key challenge is to deliver mass vaccine/medications to a population. The design of response logistics must consider the impact of epidemic and valuable resource simultaneously. The limited resources should be optimally allocated such that the loss from the disease is minimized or mitigated.

In this dissertation, we investigate three topics in medical decision making: predictive models for disease diagnosis and patient behavior, optimization for cancer treatment planning, and public health decision making for infectious disease prevention.

In the first topic, we propose a multi-stage classification framework that incorporates Particle Swarm Optimization (PSO) for feature selection and discriminant analysis via mixed integer programming (DAMIP) for classification. By utilizing the reserved judgment region, it allows the classifier to delay making decisions on 'difficult-to-classify' observations and develop new classification rules in later stage. We apply the framework to four real-life medical problems: 1) Patient readmissions: identifies the patients in emergency department who return within 72 hours using patient's demographic information, complaints, diagnosis, tests, and hospital real-time utility. 2) Flu vaccine responder: predicts high/low responders of flu vaccine on subjects using gene

signatures. 3) Knee reinjection: predicts whether a patient needs to undergo a second surgery within 3 years of his/her first knee injection and tackles with missing data. 4) Alzheimer's disease: discriminates subjects in normal, mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups using neuropsychological tests and biomarkers.

In the second topic, we first investigate multi-objective optimization approaches to determine the optimal dose configuration and radiation seed locations in the design of high-dose rate brachytherapy treatment plans for cancer patients. Tumor dose escalation and dose-volume constraints on critical organs are incorporated to enhance the dose to tumor for better tumor control while minimizing the damage to healthy organs-at-risks and normal tissues. Based on the optimization framework, we propose a non-linear optimization model that optimizes the tumor control probability (TCP) directly. The model is solved by a solution strategy that incorporates piecewise linear approximation and local search.

In the third topic, we study optimal strategies for public health emergencies under limited resources. First we investigate vaccine prioritization strategies against a pandemic flu to find the optimal strategy when limited vaccines are available by constructing a mathematical model for the course of the pandemic flu and the process of the vaccination. Second, we analyze the cost-effectiveness of emergency response strategies against a large-scale anthrax attack to protect the entire regional population.

CHAPTER II
OPTIMIZATION-BASED PREDICTIVE MODELS

In this chapter we proposed a multi-stage classification framework incorporated with Particle Swarm Optimization (PSO) for feature selection and discriminant analysis via mixed integer programming (DAMIP) for classification to find the feature subsets that can correctly predict the medical outcomes.

2.1 DAMIP

2.1.1 Mathematical Programming Based Classification Models

Classification is a fundamental machine learning problem of identifying the group status of new observations, on the basis of a set of observations of which the group memberships are known. This approach has been applied frequently in disease diagnosis where the disease stage of a patient is detected based on symptoms and lab tests.

Traditional classification methods including linear discriminant analysis for two group classification problems (LDA) [49], quadratic discriminant analysis (QDA) [171], and other discriminant analyses, such as logit and probit, have been well studied over the past several decades. These models are parametric as they incorporate assumptions about the distribution of data. Mathematical-programming-based (MP) classification methods emerge in the 1960's, gain popularity in the 1980's, and have grown drastically ever since. Most MP approaches are nonparametric, and can be considered a valuable alternative to the classical models of discriminant analysis (DA) [178].

2.1.1.1 Linear Programming Classification Models

In a two group classification, the MP formulations define a hyperplane which allows the two groups to be separated [14, 52, 71, 123]. Models including minimizing the sum of deviations (MSD), minimizing the maximum deviation (MMD), and minimizing the sum of interior distances (MSID) were proposed and tested by computational studies by Freed

et al [53]. The objective function is L_1-norm distance and is zero if the two groups can be perfectly separated by a hyperplane. By introducing gap between two regions and normalization constraints, Glover et al [65] proposed the Hybrid Model which includes different combinations of deviations in the objective function. Stam and Ungar [177] introduced a software package RAGNU for solving two-group classification problems using LP methods. Comparisons between the linear programming models and the traditional methods including LDA and QDA were performed by previous studies [45, 81, 159]. To solve the issues with unbounded or trivial solutions [126], variant models were proposed, including normalization constraints [64], reverse of group designations [160], and regularization method [198].

Bennett and Mangasarian [19] proposed a robust linear programming model which minimizes the average of the deviations, and tested it in the diagnosis and prognosis of breast cancer [125]. Effect of introducing second-order terms of the feature values in the LP models were investigated by Silva [168] and Wanarat [190]. Asparoukhov and Stam [10] proposed MSD models to solve the two-group classification problem with binary features, which have been frequently found in medical data.

Freed and Glover [54] extended the two-group LP models onto multi-group classification problems. An artificial misclassification cost is introduced for each group, and the sum of the misclassification cost for all groups is maximized. It uses single discriminant function and requires valuation of the misclassification costs. Bennett and Mangasarian [18] proposed a LP model that generates a piecewise-linear separation for multiple groups. It returns an error-minimizing separation if no feasible separation exists. Gochet et al [66] introduced the goodness and badness of fit for an observation with respect to a certain group, and solved a multi-group LP model that minimizes the badness of fit.

2.1.1.2 Mixed Integer Programming Classification Models

Mixed integer programming generates linear discriminant functions by assigning binary variables associated to each observation. In two-group classification problems, binary variables can be used to represent the exact number of misclassifications. A basic model is to minimize the number of total misclassifications (MM). Variant models including hybrid models and model of minimizing the expected misclassification cost have been proposed in previous studies [1, 14, 15, 161, 167]. Pavur et al. [149] proposed secondary goals which maximize the difference between the discriminant scores of the two groups. Glen [61, 62] proposed IP techniques for normalization by adding normalization constraints into two-group models which maximizes the classification accuracy (MCA). From the MCA model, two-stage approaches that aim to identify the observations which are difficult to classify and solve a second model with these observations were proposed in Glen [60] and Sueyoshi [181]. Glen [63] proposed piecewise-linear programming models to approximate the nonlinearity of discriminant functions of MCA.

To extend the two-group models on multi-group classification problem, Gehrlein [59] proposed two MIP models of maximizing the number of proper classifications: a single function classification scheme (DSFC) and a multiple function classification scheme (DMFC). Stam and Joachimsthaler [175] studied the performance of MIP models compared to traditional methods including MSD, LDF, and QDF. MIP models perform better when the group overlap is higher, but there are no general conclusions yet [81, 175].

2.1.1.3 Non-Linear Programming Classification Models

Nonlinear discriminant functions can be generated from MP methods by transforming the variables [168], by forming dichotomous categorical variables from the original variables [60], or based on piecewise linear function [63]. Silva and Stam [169] proposed a second-order model by introducing external deviations of each observation to the formulation. Stam and Joachimsthaler [176] proposed a class of nonlinear programming methods using L_p-norm distance as objective. Mangasarian et al. [124] proposed a non-convex model that can be solved in polynomial-time and works successfully for diagnosis of breast cancer.

Support vector machine (SVM) has been widely studied in the field of classification. SVM classification problem can be formulated as a convex quadratic programming for two-group classification problems. Mangasarian proposed a general mathematical programming framework for SVM (GSVM) [122]. Variant SVM based methods have been developed for solving large scale classification problems [21, 55, 113], and problems in healthcare including breast cancer [114] and genome prediction [120]. Hsu and Lin [76] compared the SVM methods for multi-group classification that are derived from two-group SVMs.

2.1.2 Anderson's Model and DAMIP

In this section, we first introduce Anderson's model [9], a classification rule which incorporates misclassification limits, and then discuss the discriminant analysis – mixed integer programming (DAMIP) model which provides the optimal classification rule of Anderson's model [57, 58].

We introduce the notations used in this chapter. Let $\mathcal{O}$ be the set of observations, $\mathcal{G}$ be the set of groups, and $\mathcal{F}$ be the set of features. Let n, G, and m denote the number of observations, groups, and features, respectively. The ith observation in $\mathcal{O}$ is denoted as $(y_i, \boldsymbol{x_i}) = (y_i, x_{i1}, x_{i2}, \dots, x_{im})$, where y_i is group membership and the vector $\boldsymbol{x_i}$ is the feature vector. In a classification problem, the discriminant function $f: (\mathcal{F}_1 \times \dots \times \mathcal{F}_m) \rightarrow \mathcal{G}$ is determined, it predicts the group membership of new observations based on their features.

2.1.2.1 Anderson's Model

Assuming that the group density functions and prior probabilities are known, Anderson [9] showed that an optimal rule for the problem of maximizing the probability of correct classification subject to the constraints on the misclassification probabilities must be of a specific form when discriminating among multiple groups with a simplified model. This model is parametric – assuming data of each group follow certain distribution. Let π_g be

the prior probability of group g and $f_g(\boldsymbol{x})$ be the conditional probability density function of group g, $g \in \mathcal{G}$ for the data point $\boldsymbol{x} \in \mathbb{R}^m$. Let $\alpha_{hg} \in (0,\ 1)$, $h, g \in \mathcal{G}$, $h \neq g$ be the predetermined limit on the misclassifications where the observations of group g are classified to group h. The group assignment decisions of observations that are classified into reserved judgment region, denoted by group $g = 0$ are reserved. The proposed model is to seek for a partition $\{R_0, R_1, \ldots, R_G\}$ of $\mathbb{R}^m$, where R_g is the region assigned to group g and R_0 is the reserved judgment region in which the group-assignment decisions of observations are reserved/delayed. The formulation of Anderson's model is described as follows.

$$\max \sum_{g \in \mathcal{G}} \pi_g \int_{R_g} f_g(x)dx$$

$$s.t. \int_{R_h} f_g(x)dx \leq \alpha_{hg} \qquad \forall\ h, g \in \mathcal{G}, h \neq g \qquad (2.1.1)$$

Anderson showed that there exist nonnegative constants λ_{hg}, $h,\ g \in \mathcal{G},\ h \neq g$, such that the optimal decision rule, which is referred to the Anderson's rule, is given by

$$R_g = \left\{x \in \mathbb{R}^m : L_g(x) = \max_{h \in \{0\} \cup \mathcal{G}} L_h(x)\right\},\ g \in \{0\} \cup \mathcal{G}, \qquad (2.1.2)$$

where

$$L_0(\boldsymbol{x}) = 0$$

$$L_g(\boldsymbol{x}) = \pi_g f_g(\boldsymbol{x}) - \sum_{h \in \mathcal{G}, h \neq g} \lambda_{hg} f_h(\boldsymbol{x}),\ g \in \mathcal{G}$$

For $G = 2$ the optimal solution can be modeled rather straightforward. However, finding optimal λ's for the general case, $G \geq 3$, is a difficult problem. The complexity increases as G increases [23, 24, 57, 58].

2.1.2.2 Discriminant Analysis via Mixed Integer Program (DAMIP)

Gallagher et al. [57, 58] first proposed mixed integer programming formulations, named DAMIP, for obtaining the optimal λ's in Anderson's rule. Let $\mathcal{O}_g$ denote the set of

observations in group g, and n_g denote the number of observations in group g. Let u_{hgi} represent the binary variable indicates whether observation i in group g is classified to group h, $h \in \{0\} \cup \mathcal{G}$. Thus $u_{ggi} = 1$ denotes a correct classification for observation i in group g. Gallagher et al. formulation is presented below.

$$\max \quad \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{O}_g} u_{ggj} \tag{2.1.3}$$

s.t.

$$L_{hgj} = \pi_g f_g(\boldsymbol{x}_j) - \sum_{h \in \mathcal{G}, h \neq g} \lambda_{hg} f_h(\boldsymbol{x}_j), \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.4}$$

$$y_{gj} - L_{hgj} \leq M(1 - u_{hgj}), \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.5}$$

$$y_{gj} \leq M(1 - u_{0gj}), \qquad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.6}$$

$$y_{gj} - L_{hgj} \geq \varepsilon(1 - u_{hgj}), \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.7}$$

$$y_{gj} \geq \varepsilon\, u_{hgj,} \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.8}$$

$$\sum_{h \in \{0\} \cup \mathcal{G}} u_{hgj} = 1, \qquad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.9}$$

$$\sum_{j \in \mathcal{O}_g} u_{hgj} \leq \lfloor \alpha_{hg} n_g \rfloor, \qquad \forall\, h, g \in \mathcal{G}, g \neq h \tag{2.1.10}$$

$$u_{hgj} \in \{0,1\}, \qquad \forall\, h \in \{0\} \cup \mathcal{G}, g \in \mathcal{G}, j \in \mathcal{O}_g$$

$$L_{hgj}\ urs, \qquad \forall\ h, g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$y_{gj}\ \geq 0, \qquad \forall\ g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$\lambda_{hg} \geq 0 \qquad \forall\, h, g \in \mathcal{G}, g \neq h$$

Constraints (2.1.4) define the loss functions $L_g(\boldsymbol{x})$ in Anderson's rule, constraints (2.1.5)-(2.1.9) guarantee an observation is uniquely assigned to the group with the maximum value of $L_g(\boldsymbol{x})$ among all groups, and constraints (2.1.10) set the misclassification limits. Brooks [24] and Brooks and Lee [23] showed that DAMIP is polynomial solvable for $G = 2$ but is $\mathcal{NP}$-complete for a general G.

Lee et al. [110] proposed a linear programming approach, named DALP (discriminant analysis - linear program), as a heuristic method to obtain the λ's in Anderson's rule.

$$\max \quad \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{O}_g} (c_1 w_{gj} + c_2 y_{gj}) \tag{2.1.11}$$

$$s.t. \quad L_{hgj} = \pi_g f_g(\boldsymbol{x}_j) - \sum_{h \in \mathcal{G}, h \neq g} \lambda_{hg} f_h(\boldsymbol{x}_j), \qquad \forall\, h, g \in \mathcal{G}, j \in \mathcal{O}_g \tag{2.1.12}$$

$$L_{ggj} - L_{hgj} + w_{gj} \geq 0, \qquad \forall\, h, g \in \mathcal{G}, g \neq h, j \in \mathcal{O}_g \tag{2.1.13}$$

$$L_{ggj} + w_{gj} \geq 0, \qquad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.1.14}$$

$$-L_{hgj} + y_{gj} \geq 0, \qquad \forall\, h, g \in \mathcal{G}, j \in \mathcal{O}_g \tag{2.1.15}$$

$$L_{hgj}\ urs, \qquad \forall\ h, g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$w_{gj},\ y_{gj} \geq 0, \qquad \forall\ g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$\lambda_{hg} \geq 0 \qquad \forall\, h, g \in \mathcal{G}, g \neq h$$

where c_1 and c_2 are constants/weights controlling the emphasis on correctly classifying observations or placing them in the reserved judgment region. Constraints (2.1.13), (2.1.14) and (2.1.15) link the objective-function variables with the L_{hgj} variables in such a way that correct classification of observations, and allocation of observations into the reserved judgment region, are captured by the objective-function variables.

The DAMIP/DALP approaches have been successfully applied to various multi-group disease diagnosis and biological/medical prediction problems [46, 47, 102, 108, 109, 110, 111, 154].

2.2 The PSO/DAMIP Machine Learning Framework

In this section, we review the feature selection algorithms, discuss a heuristic method called particle swarm optimization (PSO), and introduce a machine learning framework based on PSO for feature selection and DAMIP for classification.

2.2.1 Feature Selection Algorithms

Feature selection is the process of selecting a subset of the relevant features for the use of model construction in data mining problems, including regression, classification, and clustering. The basic assumption of feature selection is that data contains redundant or irrelevant features. Using too many features in constructing classification models may result in over-fitting. An example of the use of feature selection is analyzing DNA microarrays, where there are thousands to millions of features, and a few tens to hundreds of subjects. The benefits of adopting feature selection techniques are 1) improving the prediction performance, 2) reducing over-fitting, 3) providing faster predictor, and 4) improving model interpretability.

There are three main categories of feature selection algorithms: wrappers, filters and embedded methods. Wrapper methods use a search algorithm to search through the space of features and evaluate the subsets by running the classification models on them. Filter methods are similar to wrapper methods, but instead of evaluating by the classification models, a simple filter that is independent of the classification models is evaluated. Many filter methods provide a feature ranking rather than best subsets, where top ranking features can be used in classification models. Embedded feature selection algorithms are built in the classifier during the model construction.

Feature subset selection which is to find an optimal subset of features in the space of features can be considered as a combinatorial optimization problem. Dash and Liu [33] described a feature selection algorithm as in four steps: subset generation, subset evaluation, stopping criterion, and result validation. Search strategies including branch

and bound [72, 139], greedy procedure and sequential search [153, 172, 179], and random search [166] have been widely studied.

Tibshirani [185] proposed the Lasso (least absolute shrinkage and selection operator) model that estimates the coefficients w_j by solving the following quadratic program:

$$\min \quad \sum_{i \in \mathcal{O}} (y_i - w_0 - \sum_{j \in \mathcal{F}} x_{ij} w_j)^2$$

$$s.t. \sum_{j \in \mathcal{F}} |w_j| \leq t$$

$$w_j \; urs \qquad \forall \; j \in \{0\} \cup \mathcal{F}$$

where t is a tuning parameter, and x_{ij}'s are standardized. Efron et al. [41] proposed the LARS (least-angle regression) model which estimates the coefficients in a regression model, one w_j at each step, based on the equiangular directions. LARS is computationally efficient, and can be easily modified to produce solutions for other estimators, like the Lasso. Both Lasso and LARS are embedded methods.

Bertolazzi et al. [20] proposed the idea of solving feature selection using mathematical programming. The problem for the two group classification problem with binary features can be formulated as a set cover problem. Glen [62] incorporated an additional constraint into the two group mathematical programming model such that the optimal hyperplane contains no more than p features. Iannarilli and Rubin [79] propose a nonlinear integer program to find the feature subsets in multi-group classification problems.

2.2.2 Particle Swam Optimization

Particle swarm optimization was originally proposed by Kennedy and Eberhart [88]. It solves an optimization problem by iteratively trying to improve a candidate solution in

the population of candidate solutions, named particles. A particle is moved around in the search space based on its position and a randomly generated velocity. Its movement is influenced by its best known position (achieves the best objective value) and the best known position of the particles in its neighborhood. Let $\boldsymbol{x}_i$ denote the position and $\boldsymbol{v}_i$ denote the velocity of particle *i*. Let $\boldsymbol{p}_i$ be the best position of particle *i* so far. Initially, $\boldsymbol{x}_i$ and $\boldsymbol{v}_i$ for each particle *i* are randomly generated within some predetermined ranges. At each iteration, $\boldsymbol{x}_i$ and $\boldsymbol{v}_i$ are updated by

$$\boldsymbol{v}_i \leftarrow \boldsymbol{v}_i\omega + (\boldsymbol{p}_i - \boldsymbol{x}_i)c_1 rand() + (\boldsymbol{p}_{n^*(i)} - \boldsymbol{x}_i)c_2 rand()$$

$$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i + \boldsymbol{v}_i$$

where $\boldsymbol{p}_{n^*(i)}$ is the position in the neighborhood of particle *i* that holds the best objective value so far, $rand()$ is a random number between 0 and 1, and ω, c_1, and c_2 are parameters. The PSO algorithm can be terminate by criteria including number of iterations, target objective value, and the particle movement.

In a feature selection problem, the selection of features can be represented as a binary vector, i.e., 1 means that the feature is selected and 0 means that the feature is not selected. Kennedy and Eberhart [87] modify the PSO algorithm to work on binary variables. Variant PSO based algorithms have been proposed for feature selection [2, 29, 135, 184, 197].

2.2.3 The PSO-DAMIP Machine Learning Framework

Wu proposed a modified PSO algorithm to solve the feature selection algorithm where the number of selected features is determined [197]. We describe a PSO/DAMIP framework which uses the modified PSO algorithm for feature selection and the DAMIP model for classification. For particle *i*, let $\boldsymbol{v}_i$ denote the velocity and $\boldsymbol{x}_i$ represent a binary vector of length *m* where *m* is the number of features. x_{ij} denotes whether the *j*th feature is selected in particle *i*. In each iteration of the modified PSO algorithm, a DAMIP model is solved using the selected features in each particle. Particle *i* records the current selected features $\boldsymbol{x}_i$ and the best achieved objective value of DAMIP thus far, denoted by y_i. And $\boldsymbol{v}_i$and $\boldsymbol{x}_i$ in the next iteration is then determined by a random combination of $\boldsymbol{v}_i$, $\boldsymbol{x}_i$, $\boldsymbol{p}_i$,

and $\boldsymbol{p}_{n(i)}$ in the current iteration where $n(i)$ is the set of particles in the neighborhood of particle *i*. The algorithm is described as follows.

Initialization:

For each particle *i*,

$\boldsymbol{x}_i$ is generated by randomly selecting *k* 1's and *m*-*k* 0's,

$\boldsymbol{v}_i$ is a random number in $(-V_{max}, V_{max})$ where V_{max} is predetermined,

$\boldsymbol{p}_i \leftarrow \boldsymbol{x}_i$,

$y_i \leftarrow 0$.

Update:

For each particle *i* in each iteration,

Run DAMIP model with the selected features by $\boldsymbol{x}_i$.

If the objective value $y' > y_i$, update $\boldsymbol{p}_i \leftarrow \boldsymbol{x}_i$, $y_i \leftarrow y'$.

$\boldsymbol{v}_i \leftarrow \boldsymbol{v}_i\omega + (\boldsymbol{p}_i - \boldsymbol{x}_i)c_1\, rand() + (\boldsymbol{p}_{n(i)} - \boldsymbol{x}_i)c_2\, rand()$.

$\boldsymbol{x}_i$ is determined by selecting k features with the largest values in $\boldsymbol{v}_i$.

Termination:

The PSO algorithm terminates when 1) the maximum number of iterations is achieved, or 2) the percentage of the number of moving particles is less than a threshold.

We adopt the von Neumann neighborhood topology, a two dimensional grid with neighbors to the top, bottom, left and right, to construct the particles. The number of particles can be chosen as 9, 12, 16, 25, 36, or other numbers, depending on the size of the feature set. We set the parameters in the PSO algorithm: $V_{max} = 1$, ω = 0.7298, and $c_1 = c_2$ = 1.49618, which are frequently used in studies.

2.3 Multi-stage Classification Model

In this section, we introduce a multi-stage DAMIP-based classification model incorporated with reserved judgment region. In addition to the multi-stage framework, we propose modified DAMIP models to improve the performance at each stage.

2.3.1 Multi-stage Classification Model

The multi-stage classification model aims to improve the performance of the PSO/DAMIP framework by utilizing the reserved judgment region in DAMIP. A DAMIP model bisects the data set into 'easy–to-classify' subset which it classifies to specific groups, and 'difficult-to-classify' subset which it classifies to reserved judgment region. It delays making group-assignment decision to subjects that are difficult to be classified by the DAMIP with selected features. In the multi-stage model we propose, those subjects are moved to the next stage where new feature set is selected and new DAMIP classifier is developed. In such way the multi-stage framework constructs a chain of successive classifiers using different subsets of features. The classifier at the *i*th stage, denoted by f_i, can be represented by a discriminant function $f(\boldsymbol{x}_i, \boldsymbol{\lambda_i})$, which is determined by the feature subset $\boldsymbol{x}_i$, and the decision variables $\boldsymbol{\lambda_i}$ in DAMIP. More stages do not necessarily produce a better model. At each stage, the framework selects the better of two models: a single-stage model that solves a DAMIP model without reserved judgment region, and a multi-stage model that solves a DAMIP model with reserved judgment region at this stage. The algorithm naturally terminates when there are no observations in the reserved judgment region. As more stages are processed, fewer observations remain for DAMIP and the constructed model consists of too many successive classifiers. This may result in over-fitting. Hence we propose two additional stopping criteria to terminate the process: 1) the number of observations is less than the minimum number of remaining observations, denoted by *n*, and 2) the maximum allowed depth, denoted by *d* is reached. *n* and *d* are pre-determined according to the number of observations and the number of selected features in the given data. Let *S* denote the set of observations to be used in a stage. The algorithm can be described as followed:

Initialization:

$S \leftarrow \mathcal{O}$, $(\boldsymbol{x}_1, \boldsymbol{\lambda}_1)$ is determined by solving the PSO/DAMIP framework with S.

$f_1 \leftarrow f(\boldsymbol{x}_1, \boldsymbol{\lambda}_1)$

Update at stage *i*:

$(\boldsymbol{x}_i, \boldsymbol{\lambda}_i)$ is determined by solving the PSO/DAMIP framework with S.

$f_i \leftarrow f(\boldsymbol{x}_i, \boldsymbol{\lambda}_i)$

$S \leftarrow \mathcal{O}_0$, where $\mathcal{O}_0$ is the set of observations in the reserved judgment region at stage *i-1*.

Termination

The multi-stage model terminates at stage i when 1) $|S| \leq n$, or 2) $i \geq d$.

if $S \neq \emptyset$,

Solve a PSO/DAMIP model without reserved judgment region using S.

Let $f_{I+1} \leftarrow f(\boldsymbol{x}_{I+1}, \boldsymbol{\lambda}_{I+1})$.

End *if*.

Return the chain of classifiers $f_1, f_2, ..., f_I$, (f_{I+1} , if one exists).

2.3.2 Modified DAMIP Models

In the DAMIP model introduced by Gallagher [57, 58] and presented in Section 2.1.2, the size of the reserved judgment region is bounded by the misclassification rates specified in constraint (2.1.9). DAMIP is able to return good classification results through problem fine-tuning of the misclassification rates, especially when the groups are unbalanced. To ease this fine-tuning process, we envision that the classifiers in our multi-stage model to have the ability of balancing misclassifications and 'difficult to classify' observations in order to maximize the prediction accuracy through a multi-stage structure. For group g, let α_g be the misclassification rate, β_g be the proper classification rate, and γ_g be the 'difficult to classify' rate, i.e. the rate of reserved judgment region. These three parameters can be defined in DAMIP as follows:

$$\alpha_g = \frac{1}{n_g} \sum_{h \in \mathcal{G}, h \neq g} \sum_{j \in \mathcal{O}_g} u_{hgj}$$

$$\beta_g = \frac{1}{n_g} \sum_{j \in \mathcal{O}_g} u_{ggj}$$

$$\gamma_g = \frac{1}{n_g} \sum_{j \in \mathcal{O}_g} u_{0gj}$$

Recall u_{hgj} is the 0/1 variable that denotes whether to classify an observation j in group g into group h, N_g is the set of observations of group g, and n_g is the size of N_g (i.e, $n_g = |N_g|$). The three parameters satisfy that $\alpha_g + \beta_g + \gamma_g = 1$ for each group g. We propose three modified DAMIP models to 1) better utilize reserved judgment region and 2) handle imbalanced groups more efficiently.

Variant 1: The base model

$$max \quad \min_{g \in \mathcal{G}} \beta_g \tag{2.3.1}$$

s.t.

$$L_{hgj} = \pi_g f_g(\boldsymbol{x}_j) - \sum_{h \in \mathcal{G}, h \neq g} \lambda_{hg} f_h(\boldsymbol{x}_j), \quad \forall h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.2}$$

$$y_{gj} - L_{hgj} \leq M(1 - u_{hgj}), \quad \forall h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.3}$$

$$y_{gj} \leq M(1 - u_{0gj}), \quad \forall g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.4}$$

$$y_{gj} - L_{hgj} \geq \varepsilon(1 - u_{hgj}), \quad \forall h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.5}$$

$$y_{gj} \geq \varepsilon\, u_{hgj}, \quad \forall h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.6}$$

$$\sum_{h \in \{0\} \cup \mathcal{G}} u_{hgj} = 1, \quad \forall g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.7}$$

$$u_{hgj} \in \{0,1\}, \quad \forall h \in \{0\} \cup \mathcal{G}, g \in \mathcal{G}, j \in \mathcal{O}_g$$

$$L_{hgj}\ urs, \quad \forall\ h, g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$y_{gj} \geq 0, \quad \forall\ g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$\lambda_{hg} \geq 0 \quad \forall h, g \in \mathcal{G}, g \neq h$$

The base model aims to generate the optimal classification rule without using misclassification limits and reserved judgment. The objective is to maximize the

minimum value of correct classification rates β_g among all groups. It ensures that the minority groups are treated equally as the majority groups, and hence it can perfectly deal with imbalanced groups. Compared to Gallagher's DAMIP model mentioned in 2.1.2, it removes the misclassification rate constraints (2.1.10) and hence drops the reserved judgment region. It produces a lower bound of the prediction accuracy of each group, and the optimal values β_g and the associated α_g can be used in the misclassification limits in DAMIP.

Variant 2: The $\beta - \alpha$ model

$$max \quad \min_{g \in \mathcal{G}} \left(\beta_g - \alpha_g\right) \tag{2.3.8}$$

s.t.

$$L_{hgj} = \pi_g f_g(\boldsymbol{x}_j) - \sum_{h \in \mathcal{G}, h \neq g} \lambda_{hg} f_h(\boldsymbol{x}_j), \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.9}$$

$$y_{gj} - L_{hgj} \leq M\left(1 - u_{hgj}\right), \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.10}$$

$$y_{gj} \leq M\left(1 - u_{0gj}\right), \qquad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.11}$$

$$y_{gj} - L_{hgj} \geq \varepsilon\left(1 - u_{hgj}\right), \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.12}$$

$$y_{gj} \geq \varepsilon\, u_{hgj}, \qquad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.13}$$

$$\sum_{h \in \{0\} \cup \mathcal{G}} u_{hgj} = 1, \qquad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.14}$$

$$\sum_{j \in \mathcal{O}_g} u_{hgj} \leq \left\lfloor \alpha_{hg} n_g \right\rfloor, \qquad \forall\, h, g \in \mathcal{G}, g \neq h \tag{2.3.15}$$

$$u_{hgj} \in \{0,1\}, \qquad \forall\, h \in \{0\} \cup \mathcal{G}, g \in \mathcal{G}, j \in \mathcal{O}_g$$

$$L_{hgj}\ urs, \qquad \forall\ h, g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$y_{gj}\ \geq 0, \qquad \forall\ g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$\lambda_{hg} \geq 0 \qquad \forall\, h, g \in \mathcal{G}, g \neq h$$

The $\beta - \alpha$ model maximizes the minimum difference between β_g and α_g by moving a small proportion of observations into reserved judgment region. Instead of using misclassification rate constraints, it incorporates both α and β into the objective function to keep the reserved judgment region from getting too large that weakens the performance of the model.

Variant 3: The γ model

$$max \quad \sum_{g \in \mathcal{G}} \beta_g \tag{2.3.16}$$

$$s.t. \quad L_{hgj} = \pi_g f_g(\boldsymbol{x}_j) - \sum_{h \in \mathcal{G}, h \neq g} \lambda_{hg} f_h(\boldsymbol{x}_j), \quad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.17}$$

$$y_{gj} - L_{hgj} \leq M(1 - u_{hgj}), \quad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.18}$$

$$y_{gj} \leq M(1 - u_{0gj}), \quad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.19}$$

$$y_{gj} - L_{hgj} \geq \varepsilon(1 - u_{hgj}), \quad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.20}$$

$$y_{gj} \geq \varepsilon\, u_{hgj,} \quad \forall\, h, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.21}$$

$$\sum_{h \in \{0\} \cup \mathcal{G}} u_{hgj} = 1, \quad \forall\, g \in \mathcal{G},\ j \in \mathcal{O}_g \tag{2.3.22}$$

$$\sum_{j \in \mathcal{O}_g} u_{0gj} \leq \lfloor \gamma_g n_g \rfloor, \quad \forall g \in \mathcal{G} \tag{2.3.23}$$

$$u_{hgj} \in \{0,1\}, \quad \forall\, h \in \{0\} \cup \mathcal{G}, g \in \mathcal{G}, j \in \mathcal{O}_g$$

$$L_{hgj}\ urs, \quad \forall\ h, g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$y_{gj}\ \geq 0, \quad \forall\ g \in \mathcal{G},\ j \in \mathcal{O}_g$$

$$\lambda_{hg} \geq 0 \quad \forall\, h, g \in \mathcal{G}, g \neq h$$

The γ model maximizes the prediction accuracy while limiting the size of reserved judgment region by adding constraints (2.3.23) on the percentage of reserved judgment γ_g for each group g. It provides accurate control of the reserved judgment region to avoid too many stages in the model. The maximum percentage $\overline{\gamma_g}$ for each group g is predetermined according to the size of the problem. Thus the γ model resembles the original DAMIP model (as described in 2.1.2.2) except it constrains the reserved judgment instead of constraining the misclassification rate.

In two-group, the modified DAMIP models can be solved in polynomial time. The constraints that define $L(x)$ in Anderson's rule can be written as:

$$L_{1i} = \pi_1 f_1(x_i) - \lambda_{21} f_2(x_i) \quad \forall\ i \in\ \mathcal{O},$$

$$L_{2i} = \pi_2 f_2(x_i) - \lambda_{12} f_1(x_i) \quad \forall\ i \in\ \mathcal{O},$$

where optimal λ_{12} and λ_{21} are determined in DAMIP. Wu [197] proved that optimal λ_{12} and λ_{21} in a two group DAMIP model that maximizes the total correct classifications can be found by searching on the sorted array f_2/f_1 where f_1 and f_2 are the density functions in constraint (2.1.4) of group 1 and 2 respectively.

When no reserved judgment region is used in the modified DAMIP model, i.e., the base model, we define a partition p on the sorted array f_2/f_1 such that observations having $f_2(x)/f_1(x) \leq p$ are classified to group 1, and observations having $f_2(x)/f_1(x) > p$ are classified to group 2. By searching on the sorted array f_2/f_1, p^* can be found such that the objective function which is the minimum of the correct classifications of the two group in the base model is maximized. An optimal solution of $(\lambda_{12}, \lambda_{21})$ then can be determined by $\frac{\pi_1+\lambda_{12}}{\pi_2+\lambda_{21}} = p^*$.

When reserved judgment region is used in the DAMIP models, we define two partitions of the sorted array f_2/f_1 p_1 and p_2: observations having $f_2(x)/f_1(x) \leq p_1$ are classified to group 1, observations having $p_1 < f_2(x)/f_1(x) \leq p_2$ are classified to reserved judgment region, and observations having $f_2(x)/f_1(x) > p_2$ are classified to group 2. By searching on the sorted array f_2/f_1 , (p_1^*, p_2^*) can be found such that the objective function is optimized. An optimal solution of $(\lambda_{12}, \lambda_{21})$ then can be determined by $\frac{\pi_1}{\lambda_{21}} = p_1^*$ and $\frac{\lambda_{12}}{\pi_1} = p_2^*$.

The optimal partition may not be unique: any partition $p \in [l_1, l_2)$ results in the same objective value as $p^* \in [l_1, l_2)$ where l_1 is the maximum value of f_2/f_1 of observations that is less than or equal to p^* and l_2 is the minimum value of f_2/f_1 of observations that is greater than p^*. A proper way of determining p^* when searching on the sorted array is to choose the mid-point $p^* = \frac{l_1+l_2}{2}$. The complexity of this algorithm is *O*(*nlogn*): it takes *O*(*nlogn*) to sort the array f_2/f_1, and *O*(*n*) to search through the array to find the partition that reaches the optimal objective.

2.4 Medical Predictive Analysis

We apply the classification framework to real-world medical predictive problems. In this section, the results of four applications are presented: readmissions in emergency department, flu vaccine responders, knee reinjections, and Alzheimer's disease.

2.4.1 Readmissions in Emergency Department

This section contains the paper appeared in American Medical Informatics Association Proceedings 2012, 495-504.

A Clinical Decision Tool for Predicting Patient Care Characteristics: Patients returning within 72 Hours in the Emergency Department

Eva K. Lee, Ph.D[*,1,2,3], Fan Yuan[1,2,3], Daniel A. Hirsh, MD[4,5], Michael D. Mallory[4,6], MD, Harold K. Simon, MD, MBA[4,5]

[1]Center for Operations Research in Medicine and HealthCare; [2]Industrial & Systems Engineering; [3]NSF I/UCRC Center for Health Organization Transformation, Georgia Institute of Technology; [4]Children's HealthCare of Atlanta, [5]Emory University School of Medicine, [6]Pediatric Emergency Medicine Associates, Georgia.

Abstract

The primary purpose of this study was to develop a clinical tool capable of identifying discriminatory characteristics that can predict patients who will return within 72 hours to the Pediatric emergency department (PED). We studied 66,861 patients who were discharged from the EDs during the period from May 1 2009 to December 31 2009. We used a classification model to predict return visits based on factors extracted from patient demographic information, chief complaint, diagnosis, treatment, and hospital real-time ED statistics census. We began with a large pool of potentially important factors, and used particle swarm optimization techniques for feature selection coupled with an optimization-based discriminant analysis model (DAMIP) to identify a classification rule with relatively small subsets of discriminatory factors that can be used to predict — with 80% accuracy or greater — return within 72 hours. The analysis involves using a subset of the patient cohort for training and establishment of the predictive rule, and blind predicting the return of the remaining patients.

Good candidate factors for revisit prediction are obtained where the accuracy of cross validation and blind prediction are over 80%. Among the predictive rules, the most frequent discriminatory factors identified include diagnosis (> 97%), patient complaint (>97%), and provider type (> 57%). There are significant differences in the readmission

characteristics among different acuity levels. For Level 1 patients, critical readmission factors include patient complaint (>57%), time when the patient arrived until he/she got an ED bed (> 64%), and type/number of providers (>50%). For Level 4/5 patients, physician diagnosis (100%), patient complaint (99%), disposition type when patient arrives and leaves the ED (>30%), and if patient has lab test (>33%) appear to be significant. The model was demonstrated to be consistent and predictive across multiple PED sites.

The resulting tool could enable ED staff and administrators to use patient specific values for each of a small number of discriminatory factors, and in return receive a prediction as to whether the patient will return to the ED within 72 hours. Our prediction accuracy can be as high as over 85%. This provides an opportunity for improving care and offering additional care or guidance to reduce ED readmission.

*Corresponding author: eva.lee@gatech.edu

2.4.1.1 Introduction

Among patients who are discharged from the ED, 3%-4% return within 72 hours. Revisits can be related to the nature of the disease, medical errors, and/or care during their initial treatment [86, 115, 151].

Early returns to the ED may involve patients who are in a high-risk population, but other factors, such as an overcrowded ED, which decreases efficiency, can also contribute to the problem [77, 85, 151, 185]. Alessandrini et al analyzed unscheduled revisits and the similarity of return visit rates between pediatric ED and general ED [3]. Previous studies have identified risk factors for the early return to the ED, including diagnosis, complaints, and patient demographic factors [127, 129]. Gordon et al. indicated that initial diagnosis may be a useful predictor of early ED return [69]. McCusker et al. developed a screening tool called the Identification of Senior at Risk (ISAR) to identify elderly patients at high risk of return to the ED [128]. Other efforts have focused on predictors of the return for pediatric mental health care [143], Acute Pulmonary Embolism [11], and chronic obstructive pulmonary disease (COPD) exacerbations [68].

Although these studies have identified factors that appear to be linked to return visits, little is known about actually predicting return visits. Studies have applied prediction and classification methods to a variety of types of healthcare data [92, 96, 146]. In 1997, Gallagher et al. presented a mixed integer programming model (DAMIP) for constrained discriminant analysis, an approach to classification with constraints to control the likelihood of misclassification [58]. Lee et al. subsequently demonstrated the capability of DAMIP on a wide variety of medical problems compared to other classification methods [23, 102, 108, 154]. In this study, we leverage DAMIP along with swarm optimization to develop a clinical tool capable of identifying discriminatory characteristics that can predict patients who will return to the ED within 72 hours. We contrast the DAMIP results against other classification approaches.

2.4.1.2 Methods

This study was conducted in the EDs of two sites of Children's Healthcare of Atlanta (CHOA): CHOA at Hospital 1 and CHOA at Hospital 2. Included in this study are 66,861 patients who were discharged from the EDs during the period from May 1, 2009 to December 31, 2009. Patients were identified from the ED information system, including 2519 patients (3.77%) who returned within 72 hours. The patients were classified into two groups as the input of the classification model: the patients who revisit within 72 hours, and other discharged patients.

The data included 96 factors for each of the patients, including chief and secondary complaint, physician diagnosis, 5 factors related to demographic information, 8 factors related to patient arrivals, 44 factors related to the treatment and procedures received, and 35 factors related to the hospital environment.

Factors of patient information, diagnosis, and treatment have been used in previous studies to analyze the early return patients [3, 69, 127, 128, 129]. In this study, the demographic factors include age, sex, race, and weight; the hospital environment factors include day of week, time of arrival, method of arrival, payor status, triage category (acuity

level), number of patients in the ED, number of patients waiting for triage, number of available physicians, and number of available beds when the patient arrives; and the treatment factors include length of service, waiting time before a physician arrives, number of orders, number of requested resources, and whether they have taken CT scan, lab tests, radiology test, or IV therapy.

The hospital environment data was extracted from the ED electronic medical record and tracking system (Picis ED PulseCheck) into the hospitals enterprise Oracle database. For ED descriptors and available patient level details, this occurred on an hourly basis. During extraction, variables were recorded and calculations for aggregate indicators were written to an Oracle datamart. Final patient data determined after the visit (final icd-9 codes) were written to the datamart when they became available.

Early return of patients is considered a measure of quality of health care [17]. Many studies have indicated that the errors in medical care or patient education may increase the risk of early return. However, studies have not adequately analyzed the effect of the hospital environment on the patient's decision to revisit. Previous studies have used logistic linear regression models to find patients at increased risk of return. However, these models fail to accurately predict a return visit since the association between the Boolean value of return and the risk factors is more complicated than linear association. In order to predict the revisit patients among the discharged patients, we used a classification model as the predictive model. The implemented classifier is discriminant analysis via mixed integer program (DAMIP) which realizes the optimal parameters of the Anderson's classification model [9, 58, 108]. DAMIP aims to maximize the overall prediction accuracy using a set of factors, subject to an upper bound on the misclassification rate. In the next section, we describe the DAMIP-based machine learning framework.

Machine Learning Framework for Establishing Predictive Rules

The computational design of our machine learning framework focuses on the 'wrapper approach', where a feature selection algorithm is coupled to the DAMIP learning/classification module. The feature selection, classification and cross validation

procedures are coupled such that the feature selection algorithm searches through the space of attribute subsets using the cross-validation accuracy from the classification module as a measure of goodness. The attributes selected can be viewed as critical clinical/hospital variables that drive certain diagnosis or early detection. This allows for feedback to clinical decision makers for prioritization/intervention of patients and tasks.

Optimization-Based Classifier: Discriminant Analysis via Mixed Integer Program

Suppose we have n entities from K groups with m features. Let $\boldsymbol{\mathcal{G}} = \{1, 2, \dots, K\}$ be the group index set, $\boldsymbol{\mathcal{O}} = \{1, 2, \dots, n\}$ be the entity index set, and $\boldsymbol{\mathcal{F}} = \{1, 2, \dots, m\}$ be the feature index set. Also, let $\boldsymbol{\mathcal{O}}_k$, $k \in \boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{O}}_k \subseteq \boldsymbol{\mathcal{O}}$, be the entity set which belong to group k. Moreover, let $\boldsymbol{\mathcal{F}}_j$, $j \in \boldsymbol{\mathcal{F}}$, be the domain of feature j, which could be the space of real, integer, or binary values. The ith entity, $i \in \boldsymbol{\mathcal{O}}$, is represented as $(y_i, \boldsymbol{x}_i) = (y_i, x_{i1}, \dots, x_{im}) \in \boldsymbol{\mathcal{G}} \times \boldsymbol{\mathcal{F}}_1 \times \cdots \times \boldsymbol{\mathcal{F}}_m$, where y_i is the group to which entity i belongs, and $(x_{i1}, \dots, x_{im})$ is the feature vector of entity i. The classification model finds a function $f: (\boldsymbol{\mathcal{F}}_1 \times \cdots \times \boldsymbol{\mathcal{F}}_m) \to \boldsymbol{\mathcal{G}}$ to classify entities into groups based on a selected set of features.

Let π_k be the prior probability of group k and $f_k(\boldsymbol{x})$ be the conditional probability density function for the entity $\boldsymbol{x} \in \mathbb{R}^m$ of group k, $k \in \boldsymbol{\mathcal{G}}$. Also let $\alpha_{hk} \in (0,1)$, $h, k \in \boldsymbol{\mathcal{G}}, h \neq k$, be the upperbound for the misclassification percentage that group h entities are misclassified into group k. DAMIP seeks a partition $\{P_0, P_1, \dots, P_K\}$ of $\mathbb{R}^K$, where P_k , $k \in \boldsymbol{\mathcal{G}}$, is the region for group k, and P_0 is the reserved judgement region with entities for which group assignment are reserved (for potential further exploration).

Let u_{ki} be the binary variable to denote if entity i is classified to group k or not. Mathematically, DAMIP can be formulated as

(Nonlinear DAMIP) [58, 108]

$$\max \sum_{i \in \mathcal{O}} u_{y_i i} \tag{1}$$

$$\text{s.t. } L_{ki} = \pi_k f_k(\boldsymbol{x}_i) - \sum_{h \in \mathcal{G}, h \neq k} f_h(\boldsymbol{x}_i)\lambda_{hk} \qquad \forall\, i \in \mathcal{O},\ k \in \mathcal{G} \tag{2}$$

$$u_{ki} = \begin{cases} 1 & \text{if } k = \arg\max\{0, L_{hi}: h \in \mathcal{G}\} \\ 0 & \text{otherwise} \end{cases} \qquad \forall\, i \in \mathcal{O}, k \in \{0\} \cup \mathcal{G} \tag{3}$$

$$\sum_{k \in \{0\} \cup \mathcal{G}} u_{ki} = 1 \qquad \forall\, i \in \mathcal{O} \tag{4}$$

$$\sum_{i:\, i \in \mathcal{O}_h} u_{ki} \leq \lfloor \alpha_{hk} n_h \rfloor \qquad \forall\, h, k \in \mathcal{G}, h \neq k \tag{5}$$

$$u_{ki} \in \{0,1\} \qquad \forall\, i \in \mathcal{O}, k \in \{0\} \cup \mathcal{G}$$

$$L_{ki} \text{ unrestricted in sign} \qquad \forall\, i \in \mathcal{O}, k \in \mathcal{G}$$

$$\lambda_{hk} \geq 0 \qquad \forall\, h, k \in \mathcal{G}, h \neq k$$

DAMIP has many appealing characteristics including: 1) the resulting classification rule is strongly universally consistent, given that the Bayes optimal rule for classification is known [97]; 2) the misclassification rates using the DAMIP method are consistently lower than other classification approaches in both simulated data and real-world data; 3) the classification rules from DAMIP appear to be insensitive to the specification of prior probabilities, yet capable of reducing misclassification rates when the number of training entities from each group is different; 4) the DAMIP model generates stable classification rules regardless of the proportions of training entities from each group [23, 58, 102, 108, 154].

In the ED readmission classification experiments and analysis, there are two groups of patients: Group 1: non-returning, Group 2: return within 72-hour. Each entity is a patient, and each feature is the factor. Patient data from the period May 1, 2009 to December 31, 2009 were randomly divided into two sets: a training set and an independent set for blind prediction. The DAMIP classifier is first applied to the training set to establish the classification rule. The accuracy of the rule is first gauged by performing 10-fold cross validation, and can be further gauged by applying the rule to the independent set of patient data for blind prediction. Blind prediction is performed only when the 10-fold classification for the training set satisfies a pre-set minimum accuracy criteria. To gauge the performance

of our classifier, we compare the results with linear discriminate analysis, Naive Bayesian classifier, support vector machine, logistic regression, decision tree, random forest and nearest shrunken centroid approaches that are implemented in the R© language/environment.

By design, the machine learning process strives to identify the smallest set of discriminatory features that offers reliable prediction. In our application, a data stream can be fed automatically into our machine learning framework. In other applications, e.g., using hand-held device for early diagnostics etc., it is desirable that the final prediction rule depends on relatively few factors so that it is not a burden on the healthcare staff to enter the information. Regardless of the input of data stream, these discriminatory features impact decisions for hospital policy, and thus should contain only the critical factors.

Incorporating the Feature Selection Algorithm

We developed a heuristic algorithm using particle swarm optimization (PSO) to iteratively search among subsets of factors. PSO, originally developed by Kennedy and Eberhart [88, 152], is an evolutionary computation technique for solving optimization problems. Below, we describe the DAMIP/PSO machine learning framework.

Let n be the desired number of factors to be selected. Let m be the size of the particles population. Let x_i and p_i be binary vectors representing sets of chosen factors. Let v_i be a real-value vector representing the velocity of particle i. v_i is randomly assigned during initialization.

Associated with each particle is a current set of factors, x_i, and a record of the best classification accuracy with its corresponding factor set, p_i, reached thus far by this particle. We use a Von Neumann topology with 36 particles (6 × 6 block). Each particle's neighborhood is defined by its top, bottom, left and right.

At each iteration, the factor set for a particle is updated by the following algorithm:

- *Step 1* Perform DAMIP classifier for cross validation on the training data using the selected set of factors x_i;
- *Step 2* If the overall accuracy and the accuracy of *each* group in the 10-fold cross validation are over a pre-set value (e.g., > 70%), perform blind prediction using this rule on the independent set and output results. Otherwise go to Step 3.
- *Step 3* Update the velocity of the particle: The new velocity v_i is obtained from the current velocity, the current factor set, and the best accuracy of this particle and its neighborhood and their corresponding factor set:

$$v_i \leftarrow \omega \cdot v_i + c_1 \cdot r_1 \cdot (p_i - x_i) + c_2 \cdot r_2 \cdot (p_{N(i)} - x_i),$$

  where ω, c_1 and c_2 are fixed positive coefficients, r_1 and r_2 are randomly generated in the range (0,1), $N(i)$ is the neighborhood of particle i.
- *Step 4* The highest n velocity entries of this new v_i form the associated new factor set of this particle.

The algorithm updates the m particles sequentially in each iteration, and terminates when it reaches a pre-determined maximum number of iterations.

We implemented the DAMIP classifier and PSO feature selection algorithm in C++. In this study, the particle population is 36, and the machine learning process consists of 1000 DAMIP/PSO iterations (= one complete learning cycle). Each cycle requires an average of 1,080 CPU seconds. The experiment is repeated 100 times with randomly selected starting subsets of factors to strategize our search space and to avoid local optimum.

The output of the algorithm is a collection of discriminatory subsets of factors that are good candidates for the prediction of return visits within 72 hours. While users can set the desired number of discriminatory factors, the size of factors reported herein ≤ 10) is reflected from our experimental findings (see Figure 2.4.1).

2.4.1.3 Results

Due to the diversity of patients in the two hospital sites, we ran the classification model separately for each site. There were 27,534 ED patients at Hospital 1, 996 (3.62%) of whom returned within 72 hours; and there were 39,327 at Hospital 2, 1523 (3.87%) of whom returned within 72 hours. All patients went home after the first ED visit. In our analysis, the training set is 15,000 and 20,000 respectively, and the blind prediction set consists of the rest of the patients.

Table 2.4.1. Selected characteristics of patient information

	CHOA at Hospital 1			CHOA at Hospital 2		
	Total number	Percent (%)	% of return in 72 hours (%)	Total number	Percent (%)	% of return in 72 hours (%)
Total	27534	100	3.62	39327	100	3.87
Day of week						
Monday	4036	14.66	3.20	5835	14.84	3.75
Tuesday	3910	14.20	3.12	5514	14.02	3.46
Wednesday	3782	13.74	3.64	5481	13.94	3.45
Thursday	3794	13.78	3.61	5334	13.56	3.86
Friday	3683	13.38	4.13	5323	13.54	4.30
Saturday	4061	14.80	4.14	5660	14.39	4.47
Sunday	4268	15.50	3.51	6180	15.71	3.82
Time of arrival						
20:00-08:00	10681	38.79	3.95	16013	40.72	4.05
08:00-12:00	4110	14.93	3.58	5854	14.89	3.79
12:00-16:00	5772	20.96	3.50	8139	20.70	3.70
16:00-20:00	6971	25.32	3.23	9321	23.70	3.78
Acuity Level						
1	1333	4.84	2.32	637	1.62	2.67
2	8514	30.92	2.65	8382	21.31	3.30
3	12060	43.80	3.91	18781	47.76	3.89
4	5583	20.28	4.76	11372	28.92	4.36
5	44	0.16	4.55	155	0.39	1.29

The patient factors were acquired from the patient records and the ED information system. Table 2.4.1.1 shows the selected patient information for the two sites. We categorized free text factors including method of arrival, patient complaint, physician diagnosis, race, payor code, financial class, and disposition type via natural language

processing, and then ranked the categories for each factor based on the corresponding revisit rate.

Good candidates for revisit prediction are obtained by filtering the good results among the 100 complete learning cycles. These results are filtered using the criteria that the accuracy of cross validation and blind prediction for both groups are over 70%. Based on the filtered criteria, we found 7 sets of discriminatory factors for Hospital 1 and 70 sets of discriminatory factors for Hospital 2 with set size less than 10. The most frequent factors appearing among these discriminatory sets are listed in Table 2.4.1.2. The factors patient chief complaint, patient diagnosis, and provider type appear in the list of both sites.

Table 2.4.1.2. Factors most frequently occurring among the 7 sets of discriminatory factors for predicting <72-hour returns at Hospital 1, and those most frequently occurring among the 70 sets of discriminatory factors for predicting <72-hour returns at Hospital 2.

Hospital 1		Hospital 2	
Factor Name	Frequency (%)	Factor Name	Frequency (%)
Patient diagnosis	7 (100%)	Patient diagnosis	68 (97.14%)
Patient chief complaint	7 (100%)	Patient chief complaint	68 (97.14%)
Training Physician: Resident or Fellow.	4 (57.14%)	Physician Extender (i.e., nurse practitioners or others)	51 (72.86%)
If IV antibiotics was ordered.	4 (57.14%)	If the patient received a radiological test	27 (38.57%)
Attending Provider Ratio (The provider ratio (PR) determines the volume of patients that can be evaluated and treated by the physician providers). See http://www.ncbi.nlm.nih.gov/pubmed/11691670 for reference	3 (42.86%)	Expectant Patient: This is a patient is on the way to the ED who was called in by a care provider.	21 (30%)
Patient has been in ED in last 72 hours	3 (42.86%)	Time it took when the first medical doctor arrived until the attending arrived	19 (27.14%)
Primary nurse involved	2 (28.57%)	Patients who arrived ambulance	14 (20%)
Time when the patient got an ED bed to time until first medical doctor arrived	2 (28.57%)	Number of triaged patients at time	13 (18.57%)
Number of nursing resources requested	2 (28.57%)		

Figure 2.4.1 depicts the highest accuracy values achieved in DAMIP/PSO cross validation and blind prediction. The classification accuracy increases as the number of factors selected in the classification rule increases, and the highest accuracy was achieved when 4 to 10 factors were used. Figure 2.4.1 also shows that performance levels off as the number of factors increases. We include both the cross-validation and the blind prediction results to reflect the consistency of predictive power of the developed classification rules.

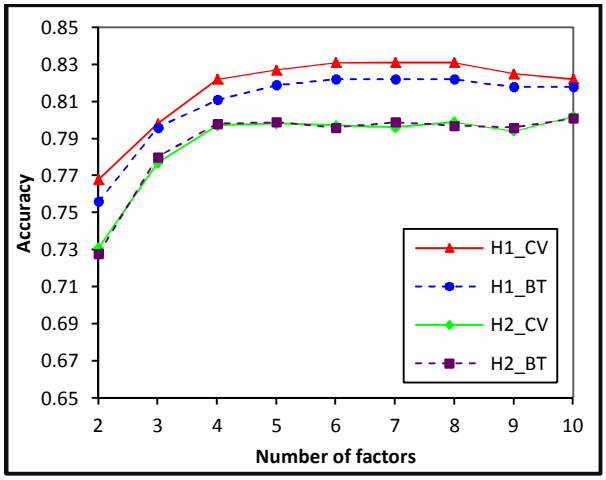

Figure 2.4.1. The highest prediction accuracy obtained via DAMIP/PSO for the two hospital sites. The solid lines represent the accuracy of cross validation, and the dashed lines represent the blind prediction accuracy. H1: Hospital 1, H2: Hospital 2, CV: 10-fold cross validation, BT: blind prediction.

Table 2.4.1.3 contrasts DAMIP/PSO results with other classification methods. Uniformly other classification methods suffer from group imbalance and the classifiers tend to place all entities into the Non-return group. In particular, linear discriminant analysis, support vector machine, logistic regression, classification trees, and random forest placed almost all patients (> 99%) into the "Non-return" group, by sacrificing the very small percentage of "Return" patients. This table also showcases the importance of reporting the classification accuracy for each group, in addition to the overall accuracy.

Acuity level is a crucial indicator of ED patient treatment resource and service needs. To better understand the 72-hour readmission characteristics of patients across different acuity levels, we perform DAMIP/PSO classification on patients with acuity level 1, 2, 3, and 4/5 – level 1 being the highest acuity. We combine Levels 4 and 5 patients in this analysis since there are only 44 and 155 Level 5 patients in each hospital respectively. The classification results, reported in Table 2.4.1.4, show higher classification and predictive accuracy for Level 1 and Level 4/5 patients. This may be explained by the fact that these patients have less diagnosis uncertainty than those in Levels 2 and 3.

Specifically, patients with Level 1 acuity have the lowest re-admission percentage (Table 2.4.1.1). These patients have the highest acuity, and thus require the most urgent rapid service. The prediction accuracy for these patients can be as high as 88%. Patients at Levels 4 and 5 have the highest re-admission percentage (Table 2.4.1.1). These patients have less pain severity, and are more concerned with quality of service. The most frequent factors shown in the discriminatory sets at the two hospitals are listed in Tables 2.4.1.5a, and 2.4.1.5b. We observe similarities between the two hospital sites for Level 1 patients (and Level 4/5 patients) critical readmission factors. For Level 1 patients, among the acquired discriminatory sets with good predictive results, time when the patient arrived until he/she got an ED bed, patient complaint, type/number of providers, and patients receive radiologic/CT scans are common and most frequent factors in both sites. For Level 4/5 patients, patient diagnosis, patient complaint, disposition type when patient arrives and leaves the ED, if the patient has a lab test, and if an IV was ordered are among the most critical readmission factors in both hospitals.

Table 2.4.1.3. Comparison of DAMIP/PSO results against other classification methods.

	10-fold Cross Validation Accuracy			Blind Prediction Accuracy		
Hospital 1	Training Set: 15,000			Blind Prediction Set: 12,534		
Classification Method	Overall	Non-return	Return	Overall	Non-return	Return
Linear Discriminant Analysis	96.3%	99.6%	5.5%	96.1%	99.6%	5.3%
Naïve Bayesian	51.6%	50.3%	87.0%	51.7%	50.2%	89.2%
Support Vector Machine	96.5%	100.0%	0.0%	96.2%	100.0%	0.0%
Logistic Regression	96.5%	99.8%	5.9%	96.3%	99.8%	8.3%
Classification Tree	96.6%	99.9%	4.4%	96.3%	100.0%	3.0%
Random Forest	96.6%	100.0%	1.5%	96.3%	100.0%	1.9%
Nearest Shrunken Centroid	62.7%	62.9%	50.0%	48.7%	48.2%	64.7%
DAMIP/PSO	**83.1%**	**83.9%**	**70.1%**	**82.2%**	**83.1%**	**70.5%**
Hospital 2	Training Set: 20,000			Blind Prediction: 19,327		
	Overall	Non-return	Return	Overall	Non-return	Return
LDA	96.2%	100.0%	0.1%	96.0%	100.0%	0.3%
Naïve Bayesian	53.4%	52.2%	83.9%	54.4%	53.2%	84.2%
SVM	96.3%	100.0%	0.0%	96.0%	100.0%	0.0%
Logistic Regression	96.3%	100.0%	0.0%	96.1%	99.9%	3.3%
Classification Tree	96.2%	100.0%	0.0%	96.0%	100.0%	0.0%
Random Forest	96.2%	100.0%	0.5%	96.1%	100.0%	0.5%
Nearest Shrunken Centroid	60.5%	60.6%	50.1%	45.8%	45.1%	61.2%
DAMIP/PSO	**80.1%**	**81.1%**	**70.1%**	**80.5%**	**81.5%**	**70.0%**

Table 2.4.1.4. DAMIP/PSO classification results for patients in each of the acuity levels.

	Training set				Testing set			
Acuity	Size	10-fold Cross Validation Accuracy			Size	Blind Prediction Accuracy		
Hospital 1		Overall	Non-return	Return		Overall	Non-return	Return
1	700	87.9%	82.9%	92.8%	633	85.1%	85.3%	76.4%
2	5000	76.0%	76.4%	71.4%	3514	73.2%	73.6%	71.6%
3	6000	80.2%	81.0%	70.2%	6060	80.3%	81.1%	70.3%
4 and 5	3000	85.2%	81.1%	89.2%	2627	81.0%	81.0%	81.3%
Hospital 2		Overall	Non-return	Return		Overall	Non-return	Return
1	350	77.2%	75.6%	78.8%	287	76.4%	76.4%	75.0%
2	4500	74.8%	75.3%	70.0%	3882	74.2%	74.7%	70.5%
3	10000	77.5%	78.2%	70.1%	8781	77.5%	78.2%	70.0%
4 and 5	6000	80.1%	83.7%	76.5%	5527	78.3%	78.4%	76.2%

Table 2.4.1.5a. Factors most frequently occurring among the 236 sets of discriminatory factors for predicting <72-hour returns for acuity-level 1 patients at Hospital 1, and those most frequently occurring among the 42 sets of discriminatory factors for predicting <72-hour returns for acuity-level 1 patients at Hospital 2.

Hospital 1		Hospital 2	
Factor Name	Frequency (%)	Factor Name	Frequency (%)
Patient diagnosis	230 (97.46%)	Time when the patient arrived until he/she got a bed	27 (64.29%)
Time when the patient arrived until he/she got a bed	228 (96.61%)	Number of beds reserved at time	24 (57.14%)
Payor type	180 (76.27%)	Patient chief complaint	24 (57.14%)
Patient chief complaint	178 (75.42%)	Physician extender (i.e.. Nurse Practitioners or others).	22 (52.38%)
Number of residents in ED	129 (54.66%)	Month	16 (38.1%)
Number of medical students in ED	106 (44.92%)	If the patient received a radiologic test	12 (28.57%)
If the patient received a CT scan for his/her head	80 (33.9%)	Patient's weight	11 (26.19%)
If the patient had a rapid strep test	77 (32.63%)	If the patient received Chest X-Ray	11 (26.19%)
Number or waiting patients divided by the number of available beds[29]	66 (27.97%)		

Table 2.4.1.5b. Factors most frequently occurring among the 246 sets of discriminatory factors for predicting <72-hour returns for acuity-level 4/5 patients at Hospital 1, and those most frequently occurring among the 491 sets of discriminatory factors for predicting <72-hour returns for acuity-level 4/5 patients at Hospital 2.

Hospital 1		Hospital 2	
Factor Name	Frequency (%)	Factor Name	Frequency (%)
Patient diagnosis	246 (100%)	Patient diagnosis	491 (100%)
Patient chief complaint	246 (100%)	Patient chief complaint	487 (99.19%)
Disposition type when patient first arrives	206 (83.74%)	Disposition type when patient leaves	399 (81.26%)
if the patient comprehensive metabolic panel	122 (49.59%)	Called in, patient is on way	210 (42.77%)
Disposition type when patient leaves	72 (29.27%)	If the patient had any lab tests done	166 (33.81%)
Number of patients in bed waiting to be discharged	69 (28.05%)	Disposition type when patient first arrives	163 (33.2%)
Time when patient arrived until a first medical doctor arrived on scene	44 (17.89%)	Month	145 (29.53%)
Arrival method	43 (17.48%)	Acuity level when the patient leaves	138 (28.11%)
If an IV of ondansetron was ordered	42 (17.07%)	If an IV of fluids was ordered	117 (23.83%)

2.4.1.4 Discussion and Conclusion

In this study, we developed a machine-learning framework combining a PSO feature selection algorithm and a DAMIP classifier to predict patients who will return to the ED within 72 hours. We used this model to select sets of discriminatory factors to establish classification rules, and to develop prediction criteria based on these rules that differentiate the revisit patients from the rest of the patients with predictive accuracy over 80%.

The input factor pool included patient information, patient complaint, physician diagnosis, operations and treatment, and hospital real-time utilization records. For Level 1 patients, among the acquired discriminatory sets with good predictive results, time when the patient arrived until he/she got an ED bed, patient complaint, type/number of providers, and patients receive a radiologic/CT scan are common and most frequent factors in both sites. For Level 4/5 patients, physician diagnosis, patient complaint, disposition type when patient arrives and leaves, if the patient has a lab test, and if an IV was ordered are among the most common factors across the two hospitals. We also note that some key hospital environment factors (e.g., time when the patient arrived until he/she got an ED bed, type/number of providers) appear among the most frequently chosen factors. Besides the common factors, the predictive factors for the two sites are different due to the diversity of the patients and the hospital characteristics. This supports the point indicated by Joynt et al. that the hospital location may affect readmission of the patients [82].

Our classification model was demonstrated to be consistent when the hospital environment varies, and its objective can be extended from short-term revisit to any class of revisit. The DAMIP/PSO machine learning framework is generalizable for predictive analytics across different hospital sites. It can adapt to different feature input and identify the appropriate set of discriminatory features for consistent prediction.

Among the ED patients, about 3-4% are return patients. Their returns may be related to their first visit experience. Being able to anticipate and predict return patterns may facilitate quality of ED service and quality of patient care and allow ED providers to

intervene appropriately. The DAMIP/PSO classifier is able to blind predict with over 80% accuracy, and outperforms other classifiers.

Based on the set of discriminatory factors with high accuracy, we developed a decision support tool for predicting patients returning within 72 hours. When implemented in clinical settings, the tool can potentially acquire data in real-time from the ED database and acquire the current hospital resource status. As the relevant factors for a patient are entered by the ED staff or through automated data-streaming, the system will return readmission prediction status of the patient. Since each discriminatory set of factors corresponds to a delivery or policy change, and requires action from ED staff, we would expect the set of discriminatory factors to be rather small, as discovered in our study.

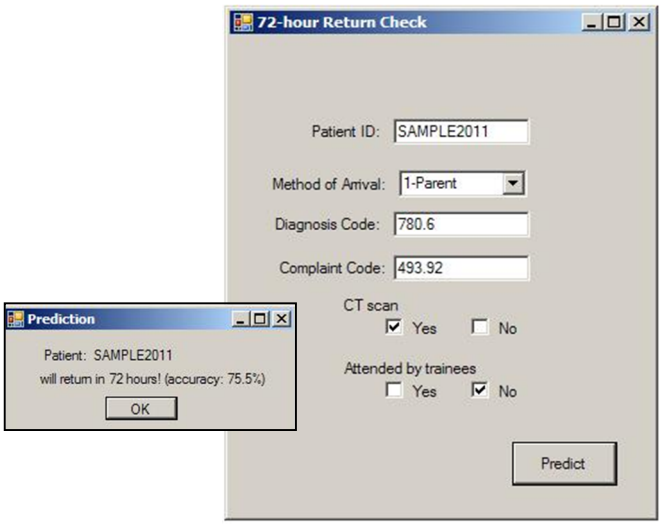

Figure 2.4.2. A sample user interface of a prediction tool for 72-hour return. Key features are typed in or selected.

Figure 2.4.2 shows a simple user interface based on a set of factors that predicted return visits with accuracy over 80%. After the required patient data is entered, and the employee clicks the "predict" button, the tool will retrieve the hospital related factors from the hospital database system, and present the revisit prediction result based on the implemented criteria. Such a computerized system allows real-time decision making, and ongoing learning and retraining of the predictive rule (and thus the discriminatory factors) as the ED data evolves over time.

We caution that these are only preliminary results based on a subset of patients in predicting readmission cases. Currently, we are conducting more detailed analysis where different patient cases will be drawn for training, and consistency among the discriminant features will be analyzed. Although we obtain better predictive accuracy ($\geq$ 85%) when more discriminatory factors are selected, it is important to keep in mind that using too many factors is impractical.

Acknowledgement

The authors acknowledge the AMIA reviewers for providing useful comments to improve the paper. This research is partially supported by a grant from the National Science Foundation (0832390), and from the National Institutes of Health.

2.4.2 Flu Vaccine Responders

In this section, we apply the classification model to predict the flu vaccine responders. The analysis and results of three data sets are presented.

2.4.2.1 High Responders versus Low Responders

HAI stands for "hemagglutination inhibition". Hemagglutination happens when the virus envelope protein called hemagglutinin (HA) binds to the sialic acid receptors on cells causing the formation of a lattice. When there is no virus or very low amount of virus, the hemagglutination will not occur and the red blood cells will sink to the bottom of a well and form a visible button. The amount of antibody in a serum by mixing several dilutions of the serum (in different wells) with a fixed concentration of virus that normally causes hemagglutination can be found through test: If a button (red blood cells that sank to the bottom) can be seen, it means that the antibody level on the serum on a given dilution is enough to prevent the hemagglutination of red blood cells. The objective of HAI assay is to detect the highest dilution of serum (that contains the Influenza-antibodies) that prevents hemagglutination. Therefore, this assay is an indirect measure of antibody level for each virus strain.

Each shot of vaccine contains 3 flu virus strains and each virus strain has its own HAI titer response. We will use the "maximum response" which is the "maximum log2 fold-change among all 3 strains". If a vaccinee has a log2 HAI response (day 28/day0) of 1, 2 and 0 for strain 1, 2 and 3, respectively, the maximum HAI response will be equal to 2, which is a 4-fold increase HAI response.

We operationally classified the vaccinees as 'low HAI responders' or 'high HAI responders' based on whether or not a fourfold increase occurred after vaccination. The high or low responder is measured by Method 1 (4-fold increase on day 28 versus day 0 and 1:40 titers on day 28 for at least one strain out of 3).

It takes 28 days to identify a patient's HAI responses after he/she takes the flu shot vaccine. In this study, gene signatures are collected by the Emory Vaccine Center for each subject before the vaccination and at day 3 and day 7 after the vaccination. We propose a machine-learning model that can predict the maximum HAI responders using the gene signatures. This scheme can correctly predict ≥ 80% of the patients as being high or low responders, and shorten the time of measuring the antibody levels for the flu shot from one month to a few days. This could significantly accelerate the process of flu shot development.

2.4.3.2 Analysis 1

This study involves 213 subjects from 5-consecutive-year trials (2007-2011) in Analysis 1, among which 141 subjects are with high responses and 72 subjects are with low responses. Each subject's maximum HAI response of all 3 strains and responder (High or Low) on day 28 are given. For each subject, we selected 54,613 gene signatures that correlated with the magnitude of HAI response on day 3 or day 7. 170 subjects are collected in day 3 data set and 200 subjects are collected in day 7 data set as shown in Table 2.4.2.1. We aim to use minimum number of gene signatures to predict high responders versus low responders.

Table 2.4.2.1. Number of subjects of high/low responders in day 3 / day 7 data sets

Year	Total (Day 3 / Day 7)	High Responders (Day 3 / Day 7)	Low Responders (Day 3 / Day 7)
2007	9 (9/9)	7 (7/7)	2 (2/2)
2008	28 (26/26)	21 (19/19)	7 (7/7)
2009	28 (27/28)	16 (16/16)	12 (11/12)
2010	75 (74/66)	64 (63/55)	11 (11/11)
2011	73 (34/71)	33 (17/32)	40 (17/39)
Total	213 (170/200)	141 (122/129)	72 (48/71)

In the initial analyses, 2/3 of the subjects are randomly selected as training set for 10-fold cross-validation and the remaining subjects are used as testing set for blind prediction. With the PSO/DAMIP framework, we identified 13 and 6 sets of genes containing no more than 5 gene signatures which achieve over 80% accuracy in both 10-

fold cross-validation and blind prediction using day 3 and day 7 data sets respectively as shown in Table 2.4.2.2.

Table 2.4.2.2.a. Prediction accuracy using day 3 data (13 sets)

Gene Signature Set	10-Fold Cross-Validation (CV)			Blind Prediction (BT)		
	Overall	High	Low	Overall	High	Low
200628_S_At 202278_S_At 203066_At 212960_At	0.818	0.756	0.839	0.814	0.818	0.812
201089_At 210427_X_At 212960_At 231420_At	0.811	0.756	0.830	0.925	0.818	1.000
1553798_A_At 201089_At 212960_At 214279_S_At	0.811	0.756	0.830	0.814	0.818	0.812
210823_S_At 212760_At 212960_At 226008_At	0.804	0.783	0.811	0.814	0.909	0.750
1569142_At 202443_X_At 212317_At 214723_X_At	0.804	0.756	0.820	0.851	0.909	0.812
206749_At 210756_S_At 212960_At 214723_X_At	0.804	0.756	0.820	0.888	0.909	0.875
1569142_At 209268_At 212960_At 215671_At	0.811	0.756	0.830	0.851	1.000	0.750
1569142_At 212960_At 219237_S_At 221584_S_At	0.804	0.756	0.820	0.851	1.000	0.750
1569142_At 201137_S_At 212960_At 223139_S_At	0.804	0.756	0.820	0.814	0.818	0.812

Table 2.4.2.2.a. (continued)

Gene Signature Set	10-Fold Cross-Validation (CV)			Blind Prediction (BT)		
	Overall	High	Low	Overall	High	Low
209268_At 212960_At 216570_X_At 230013_S_At	0.804	0.756	0.820	0.814	0.909	0.750
208771_S_At 212960_At 214279_S_At 216570_X_At	0.811	0.756	0.830	0.814	0.818	0.812
204311_At 210427_X_At 212960_At 219237_S_At	0.811	0.756	0.830	0.814	0.818	0.812
208771_S_At 212960_At 223189_X_At 226278_At	0.818	0.783	0.830	0.814	0.909	0.750

Table 2.4.2.2b. Prediction accuracy using day 7 data (6 sets)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
1554408_A_At 205692_S_At 211430_S_At 213012_At	0.803	0.800	0.805	0.888	0.818	0.937
202182_At 205692_S_At 211430_S_At 213012_At	0.803	0.800	0.805	0.888	0.818	0.937
203655_At 208639_X_At 211430_S_At 213012_At	0.803	0.800	0.805	0.814	0.818	0.812
203655_At 210046_S_At 211430_S_At 213012_At	0.803	0.800	0.805	0.814	0.818	0.812
205692_S_At 211430_S_At 211868_X_At 213012_At	0.815	0.800	0.823	0.888	0.818	0.937

Table 2.4.2.2b. (continued)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
203655_At 211430_S_At 212902_At 213012_At	0.809	0.800	0.814	0.814	0.818	0.812

Next we used 4 of the trials as the training set and used the remaining one trial as the testing set, and hence 5 runs were executed. The sets of gene signatures that correctly blind predict the testing set in each run using day 3 and day 7 data sets are presented in Table 2.4.2.3 and Table 2.4.2.4, respectively.

Table 2.4.2.3. Prediction accuracy using day 3 data set.

a. Prediction accuracy when the testing set is the trail in 2007 (4 sets)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
1552665_At 213933_At 228389_At 232321_At 242012_At	0.852	0.75	0.89	1	1	1
1565777_At 205538_At 217896_S_At 242012_At 242712_X_At	0.852	0.75	0.89	1	1	1
205538_At 217896_S_At 231088_At 242012_At 242712_X_At	0.852	0.777	0.88	1	1	1
1559679_A_At 224595_At 234632_X_At 242012_At 242712_X_At	0.852	0.777	0.88	1	1	1

b. Prediction accuracy when the testing set is the trail in 2008 (5 sets)

Gene Signature Set	10-fold CV			BT		
	Overall	High	Low	Overall	High	Low
1555007_s_at 200089_s_at 213237_at 223941_at 224595_at	0.801	0.806	0.8	0.909	0.833	0.937
1559679_a_at 211366_x_at 217039_x_at 224595_at 242012_at	0.801	0.806	0.8	0.818	0.833	0.812
1559679_a_at 203725_at 213632_at 224595_at 242712_x_at	0.834	0.774	0.855	0.863	0.833	0.875
200099_s_at 231088_at 242012_at 242712_x_at	0.834	0.774	0.855	0.818	0.833	0.812
1553970_s_at 1559679_a_at 217896_s_at 224595_at 242712_x_at	0.826	0.774	0.844	0.863	0.833	0.875

c. Prediction accuracy when the testing set is the trail in 2009 (9 sets)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
1558431_At 203031_S_At 211976_At 213237_At 243878_At	0.841	0.75	0.869	0.913	0.777	1
217039_X_At 226675_S_At 242012_At 242712_X_At 45714_At	0.825	0.75	0.847	0.913	0.777	1

c. (continued)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
212761_At 219944_At 242012_At 242712_X_At 45714_At	0.808	0.75	0.826	0.913	0.777	1
209140_X_At 219944_At 242012_At 242712_X_At 45714_At	0.808	0.75	0.826	0.869	0.777	0.928
225388_At 227235_At 242012_At 242712_X_At 45714_At	0.808	0.75	0.826	0.826	0.777	0.857
227235_At 229390_At 242012_At 242712_X_At 45714_At	0.808	0.75	0.826	0.869	0.777	0.928
212761_At 226675_S_At 236338_At 242012_At 45714_At	0.808	0.75	0.826	0.826	0.777	0.857
212761_At 213426_S_At 226675_S_At 242012_At 45714_At	0.816	0.75	0.836	0.826	0.777	0.857
213237_At 227235_At 242012_At 242712_X_At 45714_At	0.808	0.75	0.826	0.869	0.777	0.928

d. Prediction accuracy when the testing set is the trail in 2010 (5 sets)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
1557895_At 211976_At 213237_At 224595_At 231272_At	0.831	0.785	0.857	0.803	0.888	0.789
1553970_S_At 1569780_At 200091_S_At 224595_At 239507_At	0.87	0.821	0.897	0.818	0.777	0.824
211976_At 213237_At 217234_S_At 224595_At 231272_At	0.844	0.785	0.877	0.803	0.777	0.807
1559679_A_At 214252_S_At 215671_At 224595_At 242757_At	0.818	0.785	0.836	0.803	0.777	0.807
1559753_At 213237_At 214003_X_At 224595_At 232321_At	0.818	0.75	0.857	0.803	0.777	0.807

e. Prediction accuracy when the testing set is the trail in 2011 (5 sets)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
1569780_At 213237_At 224595_At 231272_At	0.83	0.72	0.86	0.88	0.833	0.923
213237_At 221136_At 224595_At 231272_At	0.805	0.72	0.827	0.84	0.75	0.923
221136_At 222268_X_At 224595_At 231272_At	0.813	0.72	0.838	0.8	0.75	0.846

e. (continued)

Gene Signature Set	10-Fold CV			BT		
	Overall	High	Low	Overall	High	Low
1554241_At 213933_At 224595_At 228438_At	0.813	0.72	0.838	0.84	0.75	0.923
203031_S_At 222527_S_At 224595_At 226941_At 242012_At	0.813	0.72	0.838	0.8	0.75	0.846

Table 2.4.2.4. Prediction accuracy using day 7 data set.

a. Prediction accuracy when the testing set is the trail in 2007 (33 sets)

Gene Signature Set	10-fold CV			BT		
	Overall	High	Low	Overall	High	Low
1554242_a_at 211058_x_at 211430_s_at 228434_at 231309_at	0.843	0.847	0.841	1	1	1
1554242_a_at 211430_s_at 228434_at 231309_at 241248_at	0.843	0.83	0.85	1	1	1
1554242_a_at 205692_s_at 211639_x_at 228814_at 231288_at	0.837	0.813	0.85	1	1	1
1554242_a_at 211430_s_at 228434_at 231309_at 232594_at	0.837	0.813	0.85	1	1	1
1553551_s_at 1554242_a_at 211430_s_at 228434_at 231309_at	0.831	0.83	0.831	1	1	1
…						

b. Prediction accuracy when the testing set is the trail in 2008 (39 sets)

Gene Signature Set	10-fold CV			BT		
	Overall	High	Low	Overall	High	Low
1556489_at 202712_s_at 203655_at 209540_at 211430_s_at	0.821	0.796	0.835	0.863	0.833	0.875
1555892_s_at 1569269_s_at 205692_s_at 228434_at 231309_at	0.827	0.777	0.855	0.818	0.833	0.812
1555892_s_at 1567663_at 202712_s_at 205692_s_at 228434_at	0.821	0.777	0.845	0.863	0.833	0.875
…						

c. Prediction accuracy when the testing set is the trail in 2009 (4 sets)

Gene Signature Set	10-fold CV			BT		
	Overall	High	Low	Overall	High	Low
201398_s_at 211430_s_at 211647_x_at 215565_at 231309_at	0.805	0.76	0.828	0.833	0.8	0.857
202355_s_at 205692_s_at 206641_at 215214_at 217227_x_at	0.812	0.8	0.818	0.833	0.7	0.928
202355_s_at 205692_s_at 211645_x_at 227721_at 231309_at	0.805	0.74	0.838	0.833	0.8	0.857
205692_s_at 217029_at 221184_at 227721_at 231309_at	0.832	0.7	0.898	0.833	0.8	0.857

d. Prediction accuracy when the testing set is the trail in 2010 (4 sets)

Gene Signature Set	10-fold CV			BT		
	Overall	High	Low	Overall	High	Low
201090_x_at 205692_s_at 211637_x_at 231288_at 243478_at	0.817	0.823	0.812	0.81	0.777	0.816
1554242_a_at 215214_at 231288_at	0.808	0.745	0.859	0.844	0.777	0.857
202355_s_at 205692_s_at 228434_at 231309_at	0.8	0.764	0.828	0.827	0.888	0.816
1555092_at 205692_s_at 228434_at 231309_at	0.8	0.725	0.859	0.844	0.888	0.836

e. Prediction accuracy when the testing set is the trail in 2011 (3 sets)

Gene Signature Set	10-fold CV			BT		
	Overall	High	Low	Overall	High	Low
211430_s_at 213046_at 223565_at 231309_at	0.828	0.769	0.847	0.758	0.735	0.785
211430_s_at 217258_x_at 223565_at 231309_at	0.81	0.73	0.835	0.758	0.764	0.75
202557_at 206641_at 231309_at 238875_at	0.783	0.73	0.8	0.758	0.764	0.75

In the next step, we reversed the previous analysis, i.e., used one year trial to predict all of the other trials. Due to lack of subjects in 2007, we combined 2007 trial with 2008 trial. The selected gene signature sets can be different between trials because of two reasons. First, the flu virus and vaccine between each year are different. Second, PSO/DAMIP framework uses a modified PSO algorithm as the feature selection method to

select subsets of gene signatures and cannot identify all good subsets. A gene signature set that predicts the high/low responders for all years can be considered as general predictors for the high/low responders of the flu vaccine for any given year. And we observed that gene signatures that are most frequently shown in the selected sets have significant impacts on the prediction accuracy. Therefore, we designed a new feature selection algorithm based the original PSO algorithm and incorporated with feature ranking method to find the common good sets in all of the 4 trials, i.e., the sets that can correctly predict the subjects in all of the 4 trials. Feature ranking methods such as Pearson's correlation coefficient are frequently used when the number of feature candidates is large, but significant features and subsets can be excluded due to the different measurement between the feature ranking method and DAMIP classifier. In this study, it is critical to find as many good subsets as possible for each trial in order to find the common sets. We designed the algorithm as follows.

In step 1, the PSO/DAMIP framework is executed for 100 replicates, and all of the sets with no more than 10 gene signatures that can correctly predict over 75% of the subjects in blind prediction are selected. Then, the gene signatures in the selected sets of each trial are ranked by their occurrences, and the gene signatures that are commonly shown in top 200 of each of the 4 gene signature rankings are selected as the gene candidate set. Next in step 2, DAMIP is executed for all no-more-than-5 gene signature sets in the gene candidate set to establish their corresponding classification rules for each trial. And we report the sets that can correctly predict on all trials.

We applied the modified PSO/DAMIP framework to discriminate the subjects with high/low responders. 46 gene signatures and 66 gene signatures are found from step 1 for day 3 data and day 7 data, respectively. Table 2.4.2.5 represents the common sets using day 3 data or day 7 data. Specifically, 21 sets that can predict all trials with prediction accuracy over 70% and 944 sets that can predict 3 trials out of 4 with prediction accuracy over 70% are found by the modified feature selection algorithm using day-3 data set. And 175 sets that can predict all trials with prediction accuracy over 70% and 10,686 sets that can predict 3 trials out of 4 with prediction accuracy over 70% are found using day-7 data set.

We highlighted the gene signatures that occur most frequently among the good results as shown in Table 2.4.2.6.

Table 2.4.2.5. Common sets that well predict all trials.

a. Using day 3 data set (21 sets)

Gene Signature	2007-2008		2009		2010		2011	
Set	CV	BT	CV	BT	CV	BT	CV	BT
205987_at 208575_at 217896_s_at 224285_at 231272_at	0.862	0.736	0.826	0.741	0.738	0.792	0.72	0.745
1559679_a_at 211976_at 213237_at 228389_at 231272_at	0.793	0.798	0.913	0.733	0.757	0.868	0.88	0.771
1552665_at 1553970_s_at 200628_s_at 203066_at 224595_at	0.896	0.771	0.826	0.733	0.727	0.766	0.72	0.754
1555082_a_at 213237_at 216380_x_at 223865_at 224595_at	0.827	0.736	0.782	0.733	0.723	0.788	0.72	0.754
211976_at 213237_at 217896_s_at 224285_at 231272_at	0.862	0.745	0.782	0.766	0.787	0.844	0.84	0.72
…								

b. Using day 7 data set (175 sets)

Gene Signature Set	2007-2008		2009		2010		2011	
	CV	BT	CV	BT	CV	BT	CV	BT
1554069_at 1555892_s_at 205692_s_at 228434_at 231309_at	0.793	0.756	0.75	0.758	0.724	0.782	0.79	0.765
209540_at 211430_s_at 213882_at 228434_at 231309_at	0.827	0.75	0.75	0.805	0.844	0.756	0.725	0.819
211430_s_at 213502_x_at 213882_at 228434_at 231309_at	0.862	0.777	0.791	0.751	0.879	0.73	0.741	0.801
211430_s_at 213882_at 215457_at 228434_at 231309_at	0.827	0.777	0.75	0.812	0.862	0.739	0.741	0.837
206063_x_at 215176_x_at 231309_at 236207_at 237867_s_at	0.758	0.743	0.75	0.751	0.81	0.747	0.725	0.792
…								

Table 2.4.2.6. Gene signatures with highest occurrence

Day 3		Day 7	
Gene signature	Occurrence (21 set)	Gene signature	Occurrence (944 set)
213237_at	16	211430_s_at	121
211976_at	10	231309_at	105
1559679_a_at	7	228434_at	43
224595_at	7	221184_at	32
217896_s_at	6	1555892_s_at	31

2.4.3.3 Analysis 3

In Analysis 3, 109 subjects with high responses on day 30 from 2007 to 2011 are given as shown in Table 2.4.2.7. The subjects are split into two classes, "persistent" responders and "temporary" antibody responders. Persistent responders are the subjects are with high responses on day 30 and also day 60+, and temporary antibody responders are the subjects are with high responses on day 30 but low responses on day 60+. For each subject, 6699 gene signatures on day 3 and 5520 gene signatures on day 7 are given. Due to lack of subjects in 2007 and 2011 trials, we combined 2007 trial with 2008 trial and 2011 trial with 2010 trial. And we applied the PSO/DAMIP framework on each of the three trials to find the gene signature sets that achieve the best prediction accuracy. The results are shown in Table 2.4.2.8.

Table 2.4.2.7. Subjects of Analysis 3.

	Total	Persistent (P)	Temporary (T)
2007-2008	23	9	14
2009	16	6	10
2010-2011	70	20	50
Total	109	35	74

Table 2.4.2.8a. Prediction results of 3 trials in Analysis 3 for day 3 data set.

Year	Gene Sign-ature Set	10-fold Cross-Validation (CV)			Blind Prediction (BT)		
		Overall	P	T	Overall	P	T
2007 -2008	211430_s_at 213046_at 223565_at 231309_at	0.95	1	0.909	0.811	0.705	0.846
2009	218972_at 1570001_at 214873_at 225118_at 234405_s_at	0.875	0.833	0.9	0.808	0.75	0.83
2009	218972_at 1570001_at 214873_at 225118_at 234405_s_at	0.875	0.833	0.9	0.808	0.75	0.83

Table 2.4.2.8a. (continued)

Year	Gene Signature Set	10-fold Cross-Validation (CV)			Blind Prediction (BT)		
		Overall	P	T	Overall	P	T
2009	1570001_at 214873_at 225118_at 220556_at 234405_s_at	0.875	0.833	0.9	0.821	0.75	0.849
2009	218972_at 1570001_at 214873_at 225118_at 206326_at	0.875	0.833	0.9	0.808	0.75	0.83
2010 -2011	205997_at 205375_at 202040_s_at 241987_x_at 208664_s_at	0.905	0.909	0.904	0.833	0.8	0.857

Table 2.4.2.8b. Prediction results of 3 trials in Analysis 3 for day 7 data set.

Year	Gene Signature Set	10-fold Cross-Validation (CV)			Blind Prediction (BT)		
		Overall	P	T	Overall	P	T
2007 - 2008	240301_at 1558791_at 223388_s_at 201249_at	0.84	0.705	0.884	0.811	0.705	0.846
2009	206481_s_at 220226_at 214725_at 201852_x_at 1557446_x_at	0.812	0.833	0.8	0.808	0.7	0.849
2009	206481_s_at 220226_at 214725_at 203443_at 1557446_x_at	0.812	0.833	0.8	0.808	0.75	0.83
2009	206481_s_at 220226_at 205003_at 214725_at 1557446_x_at	0.812	0.833	0.8	0.808	0.7	0.849

Table 2.4.2.8b. (continued)

Year	Gene Signature Set	10-fold Cross-Validation (CV)			Blind Prediction (BT)		
		Overall	P	T	Overall	P	T
2009	206481_s_at 220226_at 231131_at 205003_at 214725_at	0.812	1	0.7	0.808	0.85	0.792
2010 -2011	227525_at 218482_at 209588_at 201954_at 227452_at	0.867	0.727	0.904	0.777	0.733	0.809
2010 -2011	237151_s_at 218482_at 203777_s_at 209588_at 227452_at	0.83	0.727	0.857	0.777	0.733	0.809

2.4.3 Predicting Response to Intra-articular Injections of Hyaluronic Acid for Knee Osteoarthritis

2.4.3.1 Introduction

A new CDC study reports that the lifetime risk of knee osteoarthritis (OA) with symptoms is nearly one in two, or 46%. The study authors also found that nearly two in three obese adults will develop painful knee osteoarthritis over their lifetime. The study provides what is likely the first lifetime risk estimate of symptomatic knee osteoarthritis in the United States. Knee osteoarthritis—a common form of arthritis that wears away the cartilage cushioning the knee joint—is a leading cause of arthritis disability. In 2004, $14.3 billion were spent on hospital costs associated with total knee replacements.

At least 18% of out-patient visits to military treatment facilities by active duty personnel are attributed to painful knee disorders. The management of knee pain depends on the diagnosis, inciting activity, underlying medical conditions, body mass, and chronicity. In general, non-operative management is the mainstay of initial treatment and includes rehabilitation, activity modification, weight loss when indicated, shoe orthoses, local modalities, and medication. The oral medication often prescribed is an analgesic, usually with anti-inflammatory properties. Supplements, such as chondroitin sulfate and glucosamine have been shown to have a role. Since 1997, the regimen has expanded to include viscosupplementation. These agents are preparations of hyaluronic acid or their derivatives (HA) which are sterilely injected into the knee. Research studies have clearly demonstrated that HA improves knee function.

The goal of this study is to evaluate two different HA preparations to determine which patient population or patient characteristics would benefit most from their use. The study uses a prospective, double blinded clinical trial. A clinical predictive model is developed to uncover discriminatory patterns that can predict outcome. This predictive model can be implemented as part of a clinical practice guideline for evidence-based intervention. The model enables providers to administer HA products more selectively and

effectively to targeted population to maximize cost effectiveness and the percentage of patients who experience a successful HA trial.

2.4.3.2 Predictive Analysis

We apply the method of discriminant analysis via mixed integer programming (DAMIP) on the HA data to uncover patient and treatment factors that predict optimal response to intra-articular injections of hyaluronic acid for knee osteoarthritis. The model determines which patient variables lead to the best outcomes of HA.

Figure 2.4.3 shows the machine learning framework where features are first selected via particle swarm optimization. The resulting classification rule is subsequently established via the DAMIP classifier. Ten-fold cross validation evaluation is performed. If the results satisfy the pre-set conditions, the classification rule is reported. Blind prediction using this rule is then performed.

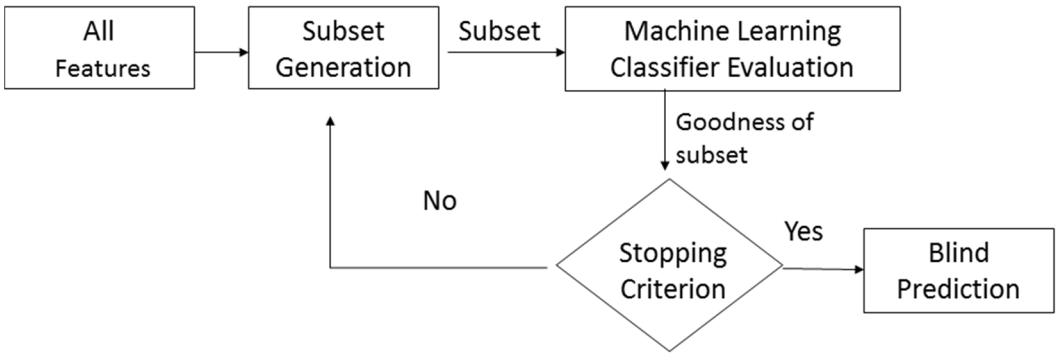

Figure 2.4.3. Machine learning framework for predictive analytics

In 10-fold cross validation, the training set is partitioned into 10 roughly equal parts. In each run, 9-fold are selected to form the rule, and the remaining 1-fold is then tested, counting how many of them are classified into which group. Through 10 folds procedure (where each fold is being validated exactly once), we obtain an unbiased estimate of the classification accuracy.

Blind prediction is performed on patients that are independent of the training set to gauge the predictive power of the established rule. These patients have never been used in the machine learning analysis. We run each patient through the rule, and it returns a status of the patient. The status is then checked against the clinical status to confirm the accuracy.

The above two study aims together will culminate in a clinical decision algorithm for the use of viscosupplementation in the treatment of knee OA. For example, a provider determines that HA is indicated for a particular patient. The provider would then enter specific variables into a clinical computer program and a response set would be generated for the potential outcome after using hyaluronic acid injections. The optimal HA agent(s) would be ranked. The provider would then take this information into account as part of the clinical decision process to select the HA agent for the individual patient.

For this HA study, we perform computational experiments for the following studies:

- predicting reinjection status using data collected up to i) at first injection, ii) after T0, iii) after T5.
- predicting treatment responder status at T5:
- WOMACP20, Treatment Responder Status Using 20% Reduction in WOMAC Pain Scale (0 = Non-Responder, 1 = Responder),
- KOOSP20, Treatment Responder Status Using 20% Reduction in KOOS Pain Scale (0 = Non-Responder, 1 = Responder),
- OARSI,
- predicting recovery status at T5, recovery on any KOOS Scale

We also perform the prediction over each of the two types of injections to gauge their similarities and differences in treatment outcome characteristics.

Table 2.4.3.1 shows the number of patients in the training set and the blind prediction set for prediction reinjection status using data up to i) at first injection, ii) after T0, iii) after T5. In this analysis, for every attribute in which there is any missing data, an

associated binary attribute is created to capture if data is missing or not for this field. The missing data is then be filled with median value of the associated demographic population.

Table 2.4.3.2 shows the training set and blind prediction statistics used for predicting treatment responder status and recovery status. We eliminate those patients where the predictor variables are missing.

Table 2.4.3.1. Training set and blind prediction set characteristics for predicting reinjection status.

Type Of Analysis		Training Set			Blind Prediction Set		
Predicting Re-Injection Status:	Number Of Attributes	Total Subjects	No Reinjection	Reinjection	Total Subjects	No Reinjection	Reinjection
Using Data Up To At First Injection,	27	150	111	39	53	40	13
Up To After T0	483	150	111	39	53	40	13
After T5	1215	150	111	39	53	40	13

Table 2.4.3.2. Training set and blind prediction set characteristics for predicting treatment responder status and recovery status.

Predicting Treatment Responder / Recovery Status At T5		Total Subjects			Training Set			Blind Prediction Set		
	Group:	Total	Non-Responder	Responder	Total	Non-Responder	Responder	Total	Non-Responder	Responder
All	KOOSP20	141	71	70	71	33	38	70	38	32
	KOOSRecovery	141	84	57	71	39	32	70	45	25
	KOOSOARSI	141	80	61	71	37	34	70	43	27
	WOMACP20	141	75	66	71	34	37	70	41	29

Table 2.4.3.2. (continued)

Predicting Treatment Responder / Recovery Status At T5		Total Subjects			Training Set			Blind Prediction Set		
	Group:	Total	Non-Responder	Responder	Total	Non-Responder	Responder	Total	Non-Responder	Responder
Synvisc	KOOSP20	76	36	40	40	17	23	36	19	17
	KOOSRecovery	76	44	32	40	21	19	36	23	13
	KOOSOARSI	76	44	32	40	20	20	36	24	12
	WOMACP20	76	37	39	40	18	22	36	19	17
Euflexxa	KOOSP20	65	35	30	35	20	15	30	15	15
	KOOSRecovery	65	40	25	35	22	13	30	18	12
	KOOSOARSI	65	36	29	35	21	14	30	15	15
	WOMACP20	65	38	27	35	21	14	30	17	13

For every attribute in which there is any missing data, an associated binary attribute is created to capture if data is missing or not for this field. The missing data is then be filled with median value of the associated demographic population. We eliminate those patients where the predictor variables are missing.

2.4.3.3 Results for Predictive Analysis

Analysis Involving Both Injections

We summarize below the best predictive rules and the associated discriminatory attributes for each of the analysis. Tables 2.4.3.3a - 2.4.3.3c show the prediction accuracy for no-reinjection versus re-injection when using attributes collected up to i) at first injection, ii) after T0, iii) after T5 respectively. We can observe the high accuracy in predicting success for patients using screening and T0 attributes alone (86% blind predictive accuracy). This is very promising for identifying patients who should be targeted for HA intervention (with expected success outcome). Including attributes until T5 increases significantly the accuracy for predicting the reinjection group (89%).

Table 2.4.3.3a. Best predictive rules for re-injection status when using only screening attributes.

Using attributes collected up to first injection,	**10-fold cross validation**			**blind prediction**		
Attributes	Overall	No-reinjection	Re-injection	Overall	No-reinjection	Re-injection
	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7
Injector ScreenWeight ScreenCurrentlySmokeCigarettes ScreenNumberPerDay MissIdctr_ScreenEverSmoked ScreenWD	0.71	0.71	0.71	0.72	0.72	0.71
Injector ScreenWeight ScreenCurrentlySmokeCigarettes ScreenNumberPerDay MissIdctr_ScreenEverSmoked ScreenCompleted	0.71	0.71	0.71	0.72	0.72	0.71

Table 2.4.3.3b. Best predictive rule for re-injection status when using screening + T0 attributes.

Using attributes collected up to after T0	**10-fold cross validation**			**blind prediction**		
Attributes	Overall	No-reinjection	Re-injection	Overall	No-reinjection	Re-injection
	>= 0.8	>= 0.7	>= 0.7	>= 0.8	>= 0.7	>= 0.7
ScreenCompleted T0KneeEffusionL T0MarxCutting T0MarxCuttingSymptomFree T0SelfEff8 T0ConfidenceInjector T0Medication2Effective T0Medication3Effective T0ExerciseEffective T0OtherTXEffective	0.85	0.89	0.74	0.81	**0.86**	0.71

Table 2.4.3.3c. Best predictive rule for re-injection status when using screening + T0 to T5 attributes.

Using attributes collected up to after T5	**10-fold cross validation**			**blind prediction**		
Attributes	Overall	No-reinjection	Re-injection	Overall	No-reinjection	Re-injection
	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8
MissIdctr_T0KOOSA16 MissIdctr_T1ROMR MissIdctr_T1MedFrequency MissIdctr_T2MedsPrescribed MissIdctr_T4WomA2 MissIdctr_T5ThighSizeDiff T5KOOSPN8 MissIdctr_T5StaffFriendly	0.84	0.84	0.83	0.83	0.81	**0.89**

Screening + T0 + T1 + T2 + T3 + T4+ T5 with category attributes modified (highlight in yellow) to reflect the demographic success rate ranking	**10-fold cross validation**			**blind prediction**		
Attributes	Overall	No-reinjection	Re-injection	Overall	No-reinjection	Re-injection
	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8
T0EQRateHealth MissIdctr_T0ImproveWalk T0ImproveKneel T0ImproveRun T0Medication3Effective T5KOOSPN8 T5KOOSSP4	0.86	0.88	0.8	0.81	0.81	0.82
T0EQRateHealth T0Medication3Effective T0PhysicalTherapyEffective T5KOOSPN8 T5KOOSSP4 T5SF11d MissIdctr_T0MARXINTERFERE	0.85	0.87	0.8	0.83	0.83	0.82

Tables 2.4.3.4a – 2.4.3.4d present the best predictive rules for predicting various treatment responder status and the recovery status.

Table 2.4.3.4a. Best predictive rule for predicting treatment responder status KOOSP20

	10-fold cross validation			**blind prediction**		
Attributes	Overall	KOOSP 20 = 0	KOOSP2 0 = 1	Overall	KOOS P20 = 0	KOOSP 20 = 1
	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7
T0KOOSPN1	0.80	0.82	0.79	0.74	0.74	0.75
T0MarxDeceleratingSymptomFree						
T0OSTEOARTHRITISDEGENERATIVEARTHRITISproblem						
T2MedsNarcotics						
MissingValue_T3SANE						
MissingValue_T4WomA13						
MissingValue_T0MARXPLUS						
MissingValue_T0Mancuso						

Table 2.4.3.4b. Best predictive rule for predicting recovery status KOOSRecovery

	10-fold cross validation			**blind prediction**		
Attributes	Overall	KOOSRec overy = 0	KOOSRec overy = 1	Overall	KOOSRec overy = 0	KOOSRec overy = 1
	>= 0.75	>= 0.7	>= 0.7	>= 0.75	>= 0.7	>= 0.7
Injector	0.79	0.74	0.84	0.76	0.76	0.76
T0MedsNarcotics						
T0KOOSQ1						
T0RelievePain						
T0ImproveStairs						
T3Alignment						
T3WomA12						
MissingValue_T4ThighSizeDiff						
T4ROMR						
MissingValue_T4WomA5						

Table 2.4.3.4c. Best predictive rule for predicting treatment responder status KOOSOARSI

	10-fold cross validation			blind prediction		
Attributes	Overall	KOOSOA RSI = 0	KOOSOA RSI = 1	Overall	KOOSOA RSI = 0	KOOSOA RSI = 1
	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7
T3MedsNarcotics T4VASPainMoving MissingValue_T4WomA7 T4StaffSkills T0KOOSADL T0BellamyWStif	0.72	0.71	0.73	0.74	0.77	0.703

Table 2.4.3.4d. Best predictive rule for predicting treatment responder status WOMACP20

	10-fold cross validation			blind prediction		
Attributes	Overall	WOMAC P20 = 0	WOMAC P20 = 1	Overall	WOMAC P20 = 0	WOMAC P20 = 1
	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7
ScreenEverSmoked T0EQRateHealth T3MedFrequency T3WomA5 T3WomA16 MissingValue_T4WomP2 T4WomA12 MissingValue_T4WRecoveryADL	0.73	0.71	0.76	0.71	0.71	0.724

Analysis for Synvisc

We next summarize the results when we consider each injection separately. Tables 2.4.3.5a-2.4.3.5d show the predictive rules for predicting treatment responder status and recovery status for patients injected with Synvisc.

Table 2.4.3.5a. Best predictive rule for predicting treatment responder status KOOSP20 with Synvisc.

	10-fold cross validation			**blind prediction**		
Attributes	Overall	KOOSP 20 = 0	KOOSP 20 = 1	Overall	KOOSP 20 = 0	KOOSP 20 = 1
	>= 0.8	>= 0.75	>= 0.75	>= 0.8	>= 0.75	>= 0.75
T0MARXNOSYMP T0MarxDeceleratingSymptomFree T0OSTEOARTHRITISDEGENERATIVEARTHRITISproblem MissingValue_T4OverallTreatment ScreenCurrentlySmokeCigarettes	0.83	0.82	0.83	0.81	**0.84**	0.76

Table 2.4.3.5b. Best predictive rule for predicting treatment responder status KOOSOARSI with Synvisc.

	10-fold cross validation			**blind prediction**		
Attributes	Overall	KOOSOA RSI = 0	KOOSOA RSI = 1	Overall	KOOSOA RSI = 0	KOOSOA RSI = 1
	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7	>= 0.7
Injector T0SF11c T2KneeEffusionR MissingValue_T2SANE T4MedsNone T4WomA10	0.73	0.72	0.73	0.83	0.86	0.80

Table 2.4.3.5c. Best predictive rule for predicting treatment responder status WOMACP20 with Synvisc.

	10-fold cross validation			Blind prediction		
Attributes	Overall	WOMAC P20 = 0	WOMAC P20 = 1	Overall	WOMAC P20 = 0	WOMAC P20 = 1
	>= 0.8	>= 0.75	>= 0.75	>= 0.8	>= 0.75	>= 0.75
T0SF11c T0KOOSPN8 T0MarxStairsSymptomFree	0.83	0.78	0.86	0.83	0.84	0.82

Table 2.4.3.5d. Best predictive rule for predicting recovery status KOOSRecovery with Synvisc.

	10-fold cross validation			blind prediction		
Attributes	Overall	KOOSRecovery = 0	KOOSRecovery = 1	Overall	KOOSRecovery = 0	KOOSRecovery = 1
	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8
T0RelievePain	0.83	0.81	0.84	0.83	0.83	0.85
MissingValue_T0ANEMIAOROTHERB						
LOODDISEASEproblem						
MissingValue_T0MarxPivoting						
T0ImproveWalk						
T310ThighSizeR						
T3WomA9						
T4WomA1						
T0KOOSPain						
T0WStif						
T0EQ5Value						

Analysis for Euflexxa

Tables 2.4.3.6a-2.4.3.6d show the predictive rules for predicting treatment responder status and recovery status for patients injected with Euflexxa.

Table 2.4.3.6a. Best predictive rule for predicting treatment responder status KOOSP20 with Euflexxa.

	10-fold cross validation			blind prediction		
	Overall	KOOSP20 = 0	KOOSP20 = 1	Overall	KOOSP20 = 0	KOOSP20 = 1
Attributes	>= 0.85	>= 0.8	>= 0.8	>= 0.85	>= 0.8	>= 0.8
ScreenAge	0.88	0.88	0.88	0.88	0.82	**0.93**
T0MedsNarcotics						
T0SF3a						
T0SF3b						
T0SF4d						
T0ACLINJURY						
T0SelfEff2						
T4VASPainMoving						
T4WomP2						
MissingValue_T3BellamyWStif						

Table 2.4.3.6b. Best predictive rule for predicting treatment responder status KOOSOARSI with Euflexxa.

	10-fold cross validation			**blind prediction**		
Attributes	Overall	KOOSOA RSI = 0	KOOSOA RSI = 1	Overall	KOOSOA RSI = 0	KOOSOA RSI = 1
	>= 0.85	>= 0.85	>= 0.85	>= 0.85	>= 0.85	>= 0.85
T0SF4d T0SF7 ScreenMaritalStatus T2MedsNone T4BellamyWADL MissingValue_T4WRecoveryStif	0.86	0.86	0.86	0.87	0.88	0.86
T0SF4d T0SF7 ScreenMaritalStatus T2MedsNone T4WomA12 MissingValue_T4WRecoveryStif	0.91	0.95	0.86	0.87	0.88	0.86
ScreenAge T0SF4d T0SF7 T310ThighSizeR T3WomA1 T4VASPainMoving T4WomP2	0.89	0.90	0.86	0.87	0.88	0.86

Table 2.4.3.6c. Best predictive rule for predicting treatment responder status WOMACP20 with Euflexxa.

	10-fold cross validation			blind prediction		
Attributes	Overall	WOMACP20 = 0	WOMACP20 = 1	Overall	WOMACP20 = 0	WOMACP20 = 1
	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8	>= 0.8
ScreenEverSmoked T0PainExpect T0Confidence T1KneePain T2PainOnPalpL T2MedsOTC T3KneeEffusionL T4MedsNone T4WomP4 T4WOMACP20PFS	0.89	0.86	0.93	0.83	0.82	0.85

Table 2.4.3.6d. Best predictive rule for predicting recovery status KOOSRecovery with Euflexxa.

	10-fold cross validation			blind prediction		
Attributes	Overall	KOOSRecovery = 0	KOOSRecovery = 1	Overall	KOOSRecovery = 0	KOOSRecovery = 1
	>= 0.85	>= 0.8	>= 0.8	>= 0.85	>= 0.8	>= 0.8
T0AntalgicGait T0MedFrequency T0KOOSS2 T0ImproveDailyTasks T0TimeSymptoms T3KneePain T4WomA3 T0KOOSSym	0.86	0.86	0.85	**0.90**	**0.89**	**0.92**

2.4.4 Alzheimer's Disease

2.4.4.1 Introduction

Alzheimer's disease (AD), the 7th leading cause of death in the United States, is a progressive and irreversible brain disease which causes memory loss and other cognitive problems severe enough to affect daily life. Dementia is a collection of symptoms of cognitive function problems, such as thinking, remembering, or reasoning problems, and AD is the most common cause of dementia. Mostly AD occurs in people over 65, although familial AD has an earlier onset. Currently, AD is incurable; drugs are used to manage the symptoms or to prevent or slow the progress of the disease.

Mild cognitive impairment (MCI) is a condition that there is clear evidence of cognitive problems, most often involving short term memory, but normal day to day functioning is preserved. In other words, MCI is a situation between normal aging and dementia. People with MCI may or may not develop dementia in the future, but people with MCI are at higher risk of developing dementia than those without MCI.

The evaluation of AD or MCI is based on patient information including complete medical history, Neuropsychological exam, laboratory tests, neuropsychological tests, brain scans (CT or MRI), and information from close family members. Neuropsychological changes in the expression of cognitive declines are important to the diagnosis of AD and MCI. Statistical analyses as predictive analysis tools are applied to neuropsychological data to understand MCI patents [119, 182]. Besides statistical analyses, classification models are applied to neuropsychological data for predicting brain damage [180] and whether nondemented elderly declined to diagnosis of dementia or Alzheimer's disease [91].

In addition to the traditional diagnosis, the clinical diagnosis of MCI and AD is increasingly aided by biomarkers predictive of underlying pathology. A number of recent studies generated additional enthusiasm for a blood-based test to predict non-demented control and AD [156, 157]. Identifying MCI and AD remains challenging. Hu [78]

measured levels of 190 plasma proteins and identified 17 analytes associated with the diagnosis of MCI or AD.

We apply the multi-stage classification model to predict the control, MCI, and AD groups using two data sets: The first one was the neuropsychological test data that conducted by Emory Alzheimer's Disease Research Center in 2011; the second one was plasma biomarkers information collected by 2 independent centers (University of Pennsylvania, Philadelphia; Washington University, St. Louis, MO). We present the prediction results for both data sets in the following sections.

2.4.4.2 Predictive Analysis Using Neuropsychological Data

The neuropsychological tests conducted in this data set includes Mini Mental State Examination (MMSE), Clock drawing test, Word list memory tasks by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), and Geriatric depression scale (GDS). The MMSE is a screening tool for cognitive impairment, which is brief, but covers five areas of cognitive function, including orientation, registration, attention and calculation, recall, and language. The clock drawing test assesses cognitive functions, particularly visuo-spatial abilities and executive control functions. The CERAD word list memory tasks assess learning ability for new verbal information. The tasks include word list memory with repetition, word list recall, and word list recognition. The GDS is a screening tool to assess the depression in older population.

Data of 267 subjects with known groups were collected as shown in Table 2.4.4.1. Among the 267 subjects, 2/3 of the subjects in each group are randomly selected as training set for 10-fold cross validation, while the remaining subjects are selected as testing set for blind prediction. 107 features are included for feature selection and classification. Among the 107 features are 3 features representing age, gender, education year, 15 features from Clock drawing test, 11 features from GDS, 13 features from MMSE, and 65 features from word list memory tasks.

464 feature sets of no more than 10 features that can correctly predict more than 80% of the subjects in both 10-fold cross validation and blind prediction are found by the PSO/DAMIP framework. The feature sets that achieve the best prediction accuracy are presented in Table 2.4.4.2. The overall prediction accuracy of 10-fold cross validation and blind prediction are over 85%, while the prediction accuracy of each group is over 80%. The prediction accuracy no longer improves when more features are used in the classification model. In Table 2.4.4.3, we highlight the features that most frequently occur in the 464 feature sets.

Table 2.4.4.1. Group information of 267 subjects in neuropsychological data set.

	Total	Control	MCI	AD	MCI or AD
Training	178	72	51	55	106
Testing	89	36	26	27	53
Total	267	108	77	82	159

Table 2.4.4.2. Prediction accuracy of the best feature sets.

Feature set	10-fold Cross-Validation				Blind Prediction			
	Overall	Ctrl	MCI	AD	Overall	Ctrl	MCI	AD
cClockNumbers4 cClockCenter GDS6 Score for What is the year? MMSE Total cWL1Arm cWL1Ticket cWL2Ticket cWLrTotal cWRyCabin	86.0%	87.8%	80.0%	88.3%	86.2%	88.2%	80.6%	90.9%
cClockNumbers4 cClockHands4 cClockCenter Score for What is the month? Score for Where are we? MMSE Total cWL1Ticket cWLrTotal cWRyButter cWRnVillage	86.0%	89.2%	80.0%	86.7%	85.1%	85.3%	80.6%	90.9%

Table 2.4.4.2. (continued)

cClockNumbers2 cClockNumbers3 cClockNumbers5 cClockCenter Score for What is the month? Score for Where are we? MMSE Total cWL1Ticket cWLrTotal	86.0%	87.8%	80.0%	88.3%	85.1%	85.3%	80.6%	90.9%
GDS15Total MMSE Total cWL1Arm cWL1Letter cWLcrCabin cWL1Ticket cWLrTotal cWRnMounain ncWRyCabin	85.5%	86.5%	80.0%	88.3%	83.9%	82.4%	80.6%	90.9%
cClockNumbers4 cClockCenter Score for What is the year? Score for What is the month? Score for Where are we? MMSE Total cWL1Ticket cWLrTotal cWRyButter cWRnVillage	86.6%	89.2%	82.2%	86.7%	86.2%	88.2%	80.6%	90.9%

Table 2.4.4.3. Features with the highest occurrences in the 310 feature sets.

Feature	Test	Occurrences
MMSE Total	MMSE	100.0%
cWLrTotal	Word list	94.4%
cWL1Ticket	Word list	94.2%
cClockCenter	Clock	76.1%
Score for What is the year?	MMSE	59.5%
Score for What is the month?	MMSE	53.4%

2.4.4.3 Predictive Analysis Using Plasma Biomarkers

Data of 352 subjects with complete information are collected as shown in Table 2.4.4.4. 2/3 of the subjects in each group are randomly selected as training set for 10-fold cross validation, while the remaining subjects are selected as testing set for blind prediction. We use 31 features for feature selection including gender, age, education years, MMSE, and 10 indicators and 17 analytes that identified by Hu [78].

92 feature sets of no more than 10 features that can correctly predict more than 80% of the subjects in both 10-fold cross validation and blind prediction are found by the PSO/DAMIP framework. The feature sets that achieve the best prediction accuracy are presented in Table 2.4.4.5. The overall prediction accuracy of 10-fold cross validation and blind prediction are over 85%, while the prediction accuracy of each group is over 80%. The prediction accuracy no longer improves when more features are used in the classification model. In Table 2.4.4.6, we highlight the features that most frequently occur in the 464 feature sets.

Table 2.4.4.4. Group information of 352 subjects in plasma biomarkers data set.

	Total	Control	MCI	AD	MCI or AD
Training	250	35	133	82	215
Testing	102	21	62	19	81
Total	352	56	195	101	296

Table 2.4.4.5. Prediction accuracy of the best feature sets.

Feature set	10-fold Cross-Validation				Blind Prediction			
	Overall	Ctrl	MCI	AD	Overall	Ctrl	MCI	AD
MMSE ApoE_1 tTau Ab42 BNP Resistin IGFBP2 tTauG91 LoAbHiTau SAP3	81.6%	91.4%	72.9%	91.5%	80.4%	81.0%	75.8%	94.7%

Table 2.4.4.5. (continued)

MMSE ApoE_1 tTau Ab42 BNP SAP IGFBP2 tTauG91 LoAbHiTau Resistin3	81.6%	91.4%	72.9%	91.5%	80.4%	81.0%	75.8%	94.7%
MMSE ApoE_1 tTau Ab42 IGFBP2 tTauG91 LoAbHiTau BNP3 Resistin3 SAP3	81.6%	91.4%	72.9%	91.5%	80.4%	81.0%	75.8%	94.7%

Table 2.4.4.6. Features with the highest occurrences in the 92 feature sets.

Feature	Occurrences
ApoE_1	100.0%
tTau	100.0%
Ab42	100.0%
IGFBP2	100.0%
tTauG91	100.0%
LoAbHiTau	100.0%
MMSE	100.0%
BNP	59.8%
Resistin	52.2%
SAP3	52.2%

CHAPTER III

CANCER TREATMENT PLANNING

This chapter investigates optimization approaches applied to radiation therapy in cancer treatment. The key to the effectiveness of radiation therapy for the treatment of cancer lies both in the fact that the repair mechanisms for cancerous cells are less efficient than that of normal cells, and the ability to deliver higher doses to the target volume. Thus, a dose of radiation sufficient to kill cancerous cells may not be lethal for nearby healthy tissue. The goal of radiation therapy is to conform the spatial distribution of the prescribed dose to the tumor volume while minimizing the dose to the surrounding normal structures. One key treatment of radiation therapy is high-dose rate (HDR) brachytherapy which allows localized high-dose radiation delivery to the tumor target via short-term implantation of radioactive seeds. HDR brachytherapy treatment plans are assessed by tumor control probability (TCP) which measures the probability of killing all cancerous cells in the affected organ. In the first section, we introduce a TCP-driven HDR treatment planning system for cervical cancer. The treatment planning is facilitated by dose escalation on the tumor that is guided by the advanced imaging techniques. In the second section, we introduce a nonlinear optimization system that maximizes TCP directly in HDR brachytherapy treatment planning. It maximizes the nonlinear function TCP while preserving the healthy tissues by adopting dose volume constraints. To tackle the nonlinear optimization problem, we propose a solution strategy that couples local search with piecewise linear approximation of TCP.

3.1 Biological Planning for High-Dose Rate Brachytherapy: Application to Cervical Cancer Treatment

This section contains the paper appeared in Interfaces - The Daniel H. Wagner Prize for Excellence in Operations Research Practice 2013; 43(5): 462-47. EK Lee, F Yuan, A Templeton, R Yao, K Kiel, JCH Chu. Biological planning for high-dose rate brachytherapy: Application to cervical cancer treatment.

Biological Planning for High-Dose-Rate Brachytherapy: Application to Cervical Cancer Treatment

Eva K. Lee, Fan Yuan
Center for Operations Research in Medicine and HealthCare, Georgia Institute of Technology, Atlanta, Georgia 30332
{eva.lee@gatech.edu, fyuan3@isye.gatech.edu}
Alistair Templeton, Rui Yao, Krystyna Kiel, James CH Chu
Rush University Medical Center, Chicago, Illinois 60612
{Alistair_Templeton@rsh.net, Rui_Yao@rsh.net, Krystyna_Kiel@rsh.net, James_C_H_Chu@rush.edu}

Cervical cancer has a high mortality rate (approximately 35 percent) in the United States and is difficult to treat successfully. One promising treatment is high-dose-rate brachytherapy, which entails delivering high-dose radiation to the tumor via the temporary implantation of radioactive seeds. This treatment promises to be particularly effective in eradicating tumors, while preserving the organs. Yet, major obstacles to successful treatment remain, especially (1) determining the best seed type, spatial configuration of seeds, and seed dwelling time, and (2) improving the probability that the treatment will eliminate all malignant cells. We developed an advanced planning model to simultaneously address both of these issues. To permit taking advantage of the best available information, our model works with inputs from positron emission tomography. We begin with a multiobjective, nonlinear, mixed-integer programming model that is initially intractable. To solve the model, we introduce an original branch-and-cut and local-search approach that couples new polyhedral cuts with matrix reduction and intelligent geometric heuristics. The result has been accurate solutions, which are obtained rapidly. Clinical trials at Rush

University Medical Center have demonstrated superior medical outcomes. These analytical techniques are applicable not only to cervical cancer, but also to other types of cancer, including breast, lung, and prostate cancer.

Key words: cancer therapeutics; cervical cancer; tumor control probability; integer programming; biological optimization; high-dose-rate brachytherapy.

3.1.1 Introduction

Almost a million cancer patients in the United States receive some form of radiation therapy each year [7]. Radiation is delivered using either external beam technology or a procedure known as brachytherapy. Brachytherapy uses a radioactive substance sealed in needles, seeds, wires, or catheters. which are placed directly (permanently or temporarily) into or near the cancer. This allows a full tumoricidal effect to eradicate the tumor from within the cancer site, while ensuring that minimal radiation reaches the healthy surrounding tissues. For high-dose-rate (HDR) brachytherapy, patients receive treatment through catheters during 3 to 10 outpatient sessions over a period of five days to two weeks. Brachytherapy preserves organs, usually with no loss of functionality; thus, it is rapidly becoming the choice of treatment for prostate, breast, cervix, and uterus cancer.

Operations research (OR) has brought breakthrough advances in treatment-planning optimization, as evidenced by the 2007 Franz Edelman award work by Memorial Sloan Kettering Cancer Center (MSKCC), which saves half a billion dollars in yearly operations and delivery costs via intelligent real-time OR-based treatment-planning approaches, while the tumor control probability (TCP) (i.e., the probability of extinction of clonogenic tumor cells by the end of treatment) improves from 65 percent to 95 percent [107].

This work brings two first-of-its-kind advances to HDR brachytherapy treatment design. First, TCP, which depends upon a highly complex function that models the responses of cancer cells and normal cells to radiation, is incorporated in the planning objective. This is distinct from the dose-based planning that is commonly employed in

current treatment design. Second, positron emission tomography (PET) information, which relates cancer cell proliferation and distribution, is incorporated within the constraints, facilitating targeted, escalated dose delivery to improve the overall clinical outcome of HDR treatments.

This work is distinct from the MSKCC in three ways: (1) HDR brachytherapy uses temporary implants that require multiple sessions; in addition to determining seed positions, the dwell time also has to be optimized. (2) This is the first time that TCP, an important measure of desired outcome, has been successfully incorporated in a treatment-planning analytical model; we determine the TCP function from a complex biological model, and place it in the objective. (3) This is the first time that PET tumor cell proliferation and distribution are incorporated within radiation therapy (external beam or brachytherapy) for dose-escalation planning.

The optimization models we develop, which are TCP driven and PET-image guided and permit HDR with dose escalation, initially prove to be intractable. The intractability arises from three sources. First, our models share the denseness properties of previous treatment-planning models [107]. We found that even without the complications that nonlinear TCP functions and PET-based dose escalation introduce, we could not solve the associated treatment-planning instances using competitive optimization software, even after we ran this software for several months of CPU time. Second, the extreme nonlinearity of our TCP functions increases the difficulties. Third, the competing PET-based dose escalation constraints that seek to go between cancer pockets and critical normal tissues offer only a tight solution space.

We focus our discussion on cervical cancer, although the methodologies are applicable to most types of cancer. Cervical cancer ranks as the second most common cancer in women worldwide, with about 500,000 new cases and 250,000 deaths annually. Almost 80 percent of cervical cancer cases are in less-developed countries [194]. The majority of cervical cancer cases (75 percent) are caused by the human papilloma virus [140]. The cancer grows slowly, and in its early stages may not have any symptoms. Thus,

the mortality rate remains high at about 35 percent. In the United States in 2013, an estimated 12,340 women will be diagnosed with cervical cancer and about 4,030 women will die from it [140]. About 0.68 percent of women born today will be diagnosed with cervical cancer at some time during their lifetimes [75].

The five-year survival rate for women diagnosed with cervical cancer is close to 75 percent [142]. The choice of treatment depends on the stage of the cancer, the size of the tumor, the patient's desire to have children, and the patient's age. Standard treatments include surgery, chemotherapy, and radiation therapy. With advances in radiation-therapy modalities and their organ-preserving characteristics, it is rapidly becoming the treatment of choice for cervical cancer [138].

In this paper, we describe our original treatment-planning models as we applied them to cervical cancer, and we describe our computational breakthroughs that permit rapid, accurate solutions. Our planning methods were implemented by the Rush University Medical Center. To the best of our knowledge, Rush University conducted the first and only clinical trial in the United States for HDR brachytherapy with PET-based dose escalation applied to cervical cancer. We report on how, with modeling assistance, the Medical Center was able to increase its treatment success and improve its quality of care, thus reducing both mortality and personal and financial burdens for cervical cancer patients.

3.1.2 Challenges and Objectives

With advances in computed tomography (CT) and magnetic resonance (MR) imaging technology, it is possible to produce contours of gross tumor volume, clinical target volume (CTV), planning target volume (PTV), and organs at risk (OARs), and to view the radiation dose within these contours as radionuclide implant locations and dwell times are adjusted. This in turn enables the use of optimization technology to derive custom treatment plans that best achieve the clinical goals of delivering a full tumoricidal dose to eliminate the cancers, while minimizing the doses to OARs [36, 56, 73, 83, 90, 99, 183]. Inverse

planning and multiobjective optimization with penalty costs have become more commonly used to address the trade-off between treating the tumor and sparing the OARs, while the optimization solution process remains a major challenge [4, 74, 89, 95, 136, 162]. Because treatment planning is intrinsically combinatorial in nature, relaxation and heuristic algorithms (e.g., linear programming or simulated annealing) have been typically employed [4, 16, 74, 84, 89, 136].

PET imaging is an important advance for cervical cancer brachytherapy treatment planning [13, 188]. The ability of PET imaging to accurately define the primary lesion by including positive lymph nodes in the PTV facilitates treatment planning. The use of FDG-PET (i.e., PET with fluorodeoxyglucoseas as a radiopharmaceutical tracer) offers a unique method for visualizing tumors, which permits treatment optimization [121]. Integrated PET and CT for treatment planning for three-dimensional conformal radiation therapy improves the standardization of volume delineation [27]. MR spectroscopy (MRS)-guided dose escalation for prostate cancer indicates that the TCP can be dramatically improved if biological information can be included within a personalized treatment design [107, 203]. This work differs from the dose-escalation work of Zaider et al. [204] in that, in addition to incorporating dose-escalation constraints within the treatment constraints, our model incorporates, within the objective function, data on the radio resistance and sensitivity of both tumor and normal cells to drive the optimization process.

The crux and challenges of HDR brachytherapy dose distributions include the following. (1) The seed type, spatial configuration, and dwell time per treatment must be determined. (2) Tumor control, a very complex biological relationship, depends on the time of the treatment, radioactive decay of the radioisotope, dose received, volume and density of tumor cells, and biological radiosensitivity and radioresistancy of the normal and tumor cells. (3) Current therapies treat the diseased organ as a homogeneous mass; however, advances in PET imaging can now distinguish cell populations based on cell density, and the metabolic activities of tumor cells, clearly differentiating them from the normal healthy cells. Such capability demands advances in treatment-planning optimization where tumor

biological knowledge is incorporated, if true personalized targeted treatment is to be realized and result in improvements in local TCPs.

This work tackles complex biological treatment planning via an OR approach, as we describe below.

- We derive novel OR-based TCP-driven PET-image-guided dose-escalation treatments based on the technology of multiobjective nonlinear mixed-integer programming (NMIP). This marks the first time that TCP is incorporated both within the treatment optimization and as a plan objective; this is also the first time in which PET-image cell-proliferation knowledge is coupled within the treatment-plan solution space.
- We derive generalized conflict hypergraphs and uncover new polyhedral theory and facial structures for these NMIP instances.
- We design a rapid branch-and-cut and local-search solution engine that couples novel cutting planes, matrix reduction, and intelligent geometric heuristics, along with a local hybrid genetic algorithm, to arrive at good solutions to these intrinsically NP-hard and intractable treatment-planning instances.

We test the robustness of the resulting plans. The clinicians evaluate the quality of the plans based on the TCP, dose distribution, and other clinical metrics that are important indicators of treatment outcomes.

3.1.3 Materials and Methods

High-Dose-Rate Brachytherapy

HDR brachytherapy treatment is given in 3 to 10 sessions, depending on the type of cancer being treated. The HDR system uses a single tiny highly radioactive source of Iridium-192, which is laser welded to the end of a thin, flexible stainless steel cable. The source is housed in an afterloader (i.e., a remote control device that mechanically places the radioactive source at predetermined positions within the applicator and stores the source

between treatments). The computer-guided afterloader directs the source into the treatment catheters or applicator, which has been placed in the patient. The source travels through each catheter in predefined steps, called “dwell” positions. The distribution of radiation and dose is determined by the dwell positions at which the source stops and the length of time it dwells there. This ability to vary the dwell times is similar to having an unlimited choice of source strengths. This level of dose control is possible only with HDR brachytherapy.

A major advantage of HDR is that the final doses are known before any radiation treatment is given. Because the patient and implant position is the same as when the treatment plan is devised, the doses are accurate. Further, because of the high radioactivity of the Iridium-192 source, the treatment time takes only minutes, rendering little opportunity for the implant to move and deposit a radiation dose where it is not intended.

The gynecological HDR procedure can be briefly summarized as follows. First, in the operating room, catheters are inserted into a patient who is under local, general, or spinal anesthesia. Interstitial catheters are inserted through the body tissue to encompass the tumor. For cervical treatment, a template is sutured to the skin to hold the treatment catheters in position. A CT scan is taken to determine the exact location of the catheters in relationship to the diseased organ and normal tissues. The CT images are used for treatment-planning optimization. The dosimetrist (i.e., a specialist who has the expertise necessary to generate radiation dose distributions and dose calculations, in collaboration with a medical physicist and a radiation oncologist) designs the plan on a computer and customizes the radiation doses to conform to the target volume, while minimizing the doses to the nearby normal tissues. After the physician has approved the treatment plan, the computer transfers the treatment-plan instructions to the HDR remote afterloader. On the day of the treatment, the patient is moved into the brachytherapy treatment room. The ends of the treatment catheters that protrude outside the body are connected to “transfer” tubes, which are then connected to the afterloader. The programmed instructions guide the afterloader on where to direct the source and how long to leave the source in each dwell position. The patient is alone in the treatment room as the treatment is being given, and the therapists and nurses continually monitor the treatment through an intercom and closed-

circuit TV cameras. The entire treatment process takes approximately 30 to 90 minutes, depending on the size and complexity of the implant and the activity of the source. Upon treatment completion, the sutures holding the catheters in place are clipped and the implant is gently removed. Figure 3.1 depicts the delivery of HDR brachytherapy for cervical cancer.

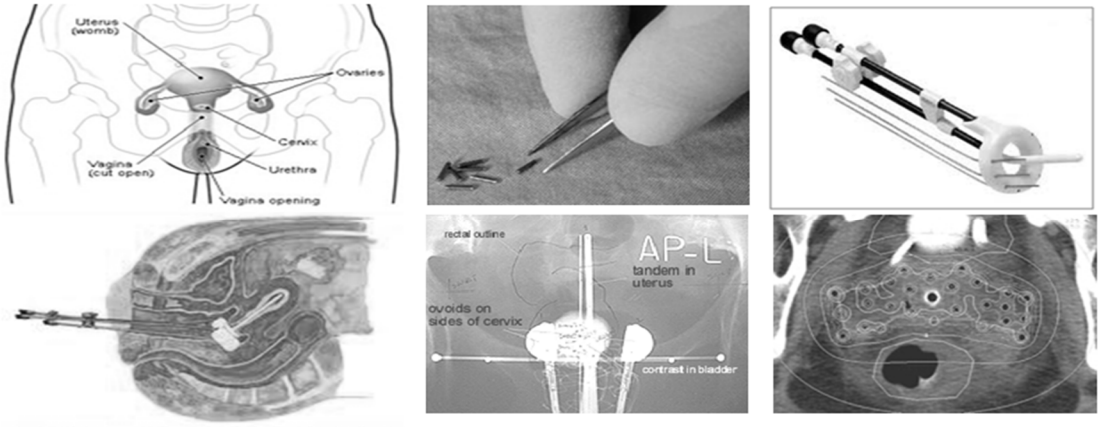

Figure 3.1. On the far left, we show the cervix anatomy (top) and the associated HDR treatment with the applicator (bottom). The remaining images on the top show the Ir^{192} radioactive seeds and the Vienna ring CT-MR applicator. The CT image (bottom middle) shows the catheters and seed positions with respect to the diseased cervix. The image is used for treatment-planning optimization. The bottom right shows a transverse view with isodose curves overlaid.

PET Image

For our study of a group of cervical cancer patients, we obtained both PET images and CT scans. The biological PET image is first fused onto the treatment CT image (see Figure 3.2). PTVs, critical structures, and OARs are delineated from CT images. The enhanced PET signal allows the identification of dense pockets of cancer cells, which define the boost target volume (BTV). HDR plans are optimized to deliver a prescribed dose of 35 Gy Ir^{192} to the PTV and 37 to 40 Gy to the BTV, following 45 Gy of external beam radiotherapy.

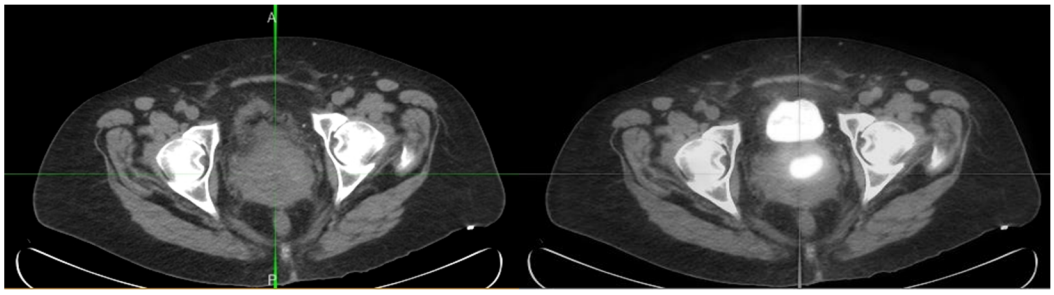

Figure 3.2. This figure shows the CT treatment image for the cervix (left), and the resulting image with an overlay of the PET image for plan design and optimization (right). We can clearly observe the PET tumor pockets (bright spots) inside the cervix.

Novel OR-Based Treatment-Planning Model

The OR challenges we faced are:

1. Effectively modeling the TCP within the treatment-planning objective;
2. Incorporating the PET-image information for biological targeted dose escalation; and
3. Advancing computational strategies to solve the associated intractable nonlinear combinatorial instances.

Dose calculation is based on guidelines from the American Association of Physicists in Medicine (AAPM) task groups for brachytherapy [158], which we describe in detail in the appendix.

To the extent that PET can indicate the presence of faster-proliferating and (or) a higher density of tumor cells, recognizing such regions in the organ could be consequential in terms of tumor control. Therefore, we incorporate the TCP in our treatment-planning process. Specifically, we are interested in: (1) the maximal TCP gain obtainable by incorporating PET information in treatment planning, and (2) the largest fractional tumor-pocket volume for which PET-guided planning remains useful. Clearly, if tumor cells are uniformly spread throughout most of the cervix volume, the gain would be insignificant.

We generalize the TCP based on a reliable biological model [203], which is derived by using birth and death processes. The parameters include the number of tumor cells, their

survival probability, and the birth and death rates. The survival probability depends on the dose delivered, its timing and duration, and the repair of sublethal damage. We describe the full TCP model in the appendix.

In our treatment-planning model, we represent each anatomical structure by a collection of discretized voxels (three-dimensional volumetric pixels), and choose sizes such that they are conducive for modeling. Each dwell location is modeled via two variables: a binary decision variable to indicate whether a radioactive seed will be deposited and a continuous variable to denote the associated dwell time. We impose constraints to ensure 95 percent PTV coverage, while restricting the underdose to the PTV, and upper- and lower-dose bounds for OARs and PTVs. We then strategically choose dose levels and parameters so that the overall PTV dose remains relatively homogeneous, as the clinicians desire, while protecting the OARs by using the maximum dose that the organs can tolerate to avoid inflicting severe and permanent harm. In the escalated-dose case, the PET-identified region receives an escalated dose over the prescribed dose. Not all the dose bounds can be satisfied simultaneously because of the close proximity and conflicting dose targets of cancerous and normal cells; therefore, our first objective is to find a treatment plan that satisfies as many dose bound constraints as possible. This helps to achieve a rapid dose fall-off from the PTV-prescribed dose, ensuring that the prescribed dose conforms to the tumor shape, while minimizing damage to healthy normal tissues. The second objective incorporates the TCP function, which depends on the time of the treatment, radioactive decay of the radioisotope, dose received, volume and density of the tumor cells, and biological radiosensitivity and radioresistancy of the normal and tumor cells. This results in a multiobjective NMIP problem, which is intractable by existing computational techniques. We present the full model in the appendix.

Computational Challenges and Solution Strategies

The treatment model has three objectives: (1) the temporal delivery objective that governs the dwell times; (2) the dose volume-based objective that, along with the temporal objective, guides the optimization engine to a solution that best satisfies the imposed

dosimetric and volumetric constraints for conformal treatment; and (3) the biological and clinical TCP objective.

To apply our multiple-objective mixed-integer programming (MIP) solution strategies, we begin by first solving the MIP instance that requires the minimum PTV coverage, while minimizing the maximum dwell time across all the possible seed locations.

For the dose volume-based objective, these MIP instances inherit the dense dose matrix properties as in the MSKCC brachytherapy instances [107]. Using a competitive commercial solver, the solver does not return a feasible solution, even after running for a month of CPU time on an Intel Xeon E5430 Quad Core Xeon Processors at 2.66 GHz, 1333 MHz FSB, and 12 MB Cache per processor.

We employ hypergraphic polyhedral cuts to accelerate the solution process. In particular, Easton et al. [38] introduce the notion of uniform hypergraph and derived facial structures of uniform hypercliques. In their work, they show that these hyperclique inequalities can help to successfully solve the small, yet 100 percent dense, previously intractable market-share instances. Lee and Zaider [107] show that hypercliques, along with novel matrix-reduction approaches and clever geometric-based heuristics, can help solve these intractable MIP instances to optimality. Furthermore, the solution process can be achieved within seconds; thus, the real-time treatment-planning process, which has since become standard across the United States, for prostate permanent implants, was realized.

In our work, the challenges are more complex; these challenges include the multiobjective nature of our problem, the highly nonlinear TCP objective, and the competing dose escalation and OAR dose distribution within the solution space.

We advance the polyhedral theory work of Easton et al. [38] and introduce the concept of generalized conflict hypergraphs. Within this high-dimension construct, we derive new polyhedral theories, including generalized hyper-clique, hyper-oddholes, hyper-

antioddholes, hyper-webs, hyper-antiwebs, and hyper-star facial structures, and their associated *Chvátal-Gomory (CG) rank*s [100, 101].

Computationally, we tackle the dose volume-based and the biological tumor control objectives simultaneously using a branch-and-cut and local-search approach. We caution that because TCP is highly nonlinear, it is difficult to convexify or linearize it for actual branch-and-cut solution exploration. Specifically, we solve the MIP instance with the dose volume-based objective via a branch-and-cut algorithm that couples new polyhedral cuts, along with matrix reduction and intelligent geometric heuristics algorithms. When we obtain an integer solution, or when a heuristic within the branch-and-cut setting returns a feasible solution, we perform a local search to examine the TCP values across the entire neighborhood. Given a seed configuration with dwell times, we calculate the associated TCP based on the resulting PTV and PET-pocket dose volume histograms. We then keep the best solution (i.e., the solution with the maximum TCP value) as the incumbent solution. The local search involves swapping and a hybrid genetic algorithm, where one can rapidly examine the neighborhood space to identify the best TCP-value solutions.

Such an approach guarantees the return of a feasible solution, while exploiting the best possible TCP values within the neighborhood feasible space. In addition to deriving novel general hypergraphic structures and encapsulating them within a rapid computational engine for solving these instances, we also investigate polyhedral approaches in MIP convexification of posynomial and signomial functions [164, 165]. We expect further advances in directly addressing the TCP objective using these novel cutting planes and polyhedral results, which we have obtained from special classes of nonlinear MIPs.

We examine multiple variations to determine the one with the best performance in terms of dose distribution to various organs and the associated TCP. The variations include minimizing the overdose and underdose to the PTV, and (or) a combination of these. The overdose and underdose can be obtained by transforming the binary variables v_P^L and v_P^U in Constraints (6) and (7) (see the appendix) into continuous variables to capture the dose differences. The weights in the objective function can be nonlinear to the overdose or

underdose amount (e.g., it can be piecewise linear or quadratic penalties). Other variations include maximizing the dose fall-off from the prescribed PTV dose.

3.1.4 Validation and Results

To gauge the feasibility, characteristics, and potential benefit of PET-image-guided dose escalation, our initial validation consists of 15 cases in which the patient has cervical cancer. Each patient had previously received a 45 Gy dose of external radiation. The PTV ranges from 82.8 to 137.47 cm^3 and the BTV ranges from 10 percent to 41 percent. For each case, we contrast three alternative strategies: (1) a standard HDR plan with no dose escalation, (2) a BTV escalation with the same PTV prescription dose, and (3) a BTV escalation with a reduced PTV prescription dose. For both escalation strategies, we consider two variations (a 37 Gy increase and a 40 Gy increase to the BTV), and observe the effects on PTV and OAR dose profiles and TCP quality.

Figure 3.3a illustrates the dose volume histogram and dose profiles for a patient with a BTV that is 19.39% of the cervix. The *y*-axis is the cumulative volume, and the *x*-axis is the radiation dose received. In this figure, the standard plan is labeled by a square, escalated PET $\geq$ 37 Gy with 35 Gy PTV prescribed dose is labeled by circles, escalated PET $\geq$ 37 Gy with 33 Gy reduced PTV is labeled by an x, escalated PET $\geq$ 40 Gy with 35 Gy PTV prescribed dose is labeled by triangles, and escalated PET $\geq$ 40 Gy and 33 Gy reduced PTV dose is labeled by rhombuses. We can observe the escalated dose of the BTV versus the PTV. Further, compared to the standard plan, dose reduction occurs in the rectum and bladder in the escalated plans with a BTV $\geq$ 37 Gy. We observe a larger reduction when we increase the BTV to $\geq$ 40 Gy. This translates to a reduction in normal tissue toxicity and complications. Figure 3.3b illustrates similar trends for a small BTV volume (i.e., 6.3%of the PTV).

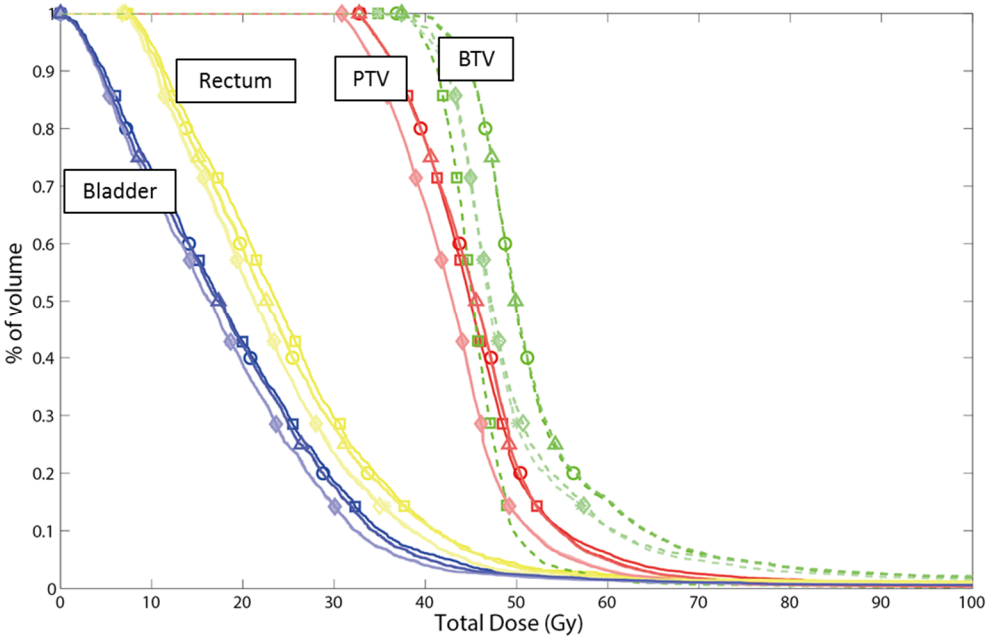

Figure 3.3a. This figure shows the dose volume histogram for a cervical cancer patient with a BTV that is 19.39% of the PTV.

- ❑ : Standard HDR plan (PTV prescribed dose = 35 Gy); there is no separate curve for the PET.
- ○ :Escalated PET (PTV prescribed dose = 35 Gy, PET $\geq$ 37 Gy).
- * :Escalated PET and reduced PTV dose (PTV prescribed dose = 33 Gy, PET $\geq$ 37 Gy).
- Δ :Escalated PET (PTV prescribed dose = 35 Gy, PET $\geq$ 40 Gy).
- ❖ :Escalated PET and reduced PTV dose (PTV prescribed dose = 33 Gy, PET $\geq$ 40 Gy).

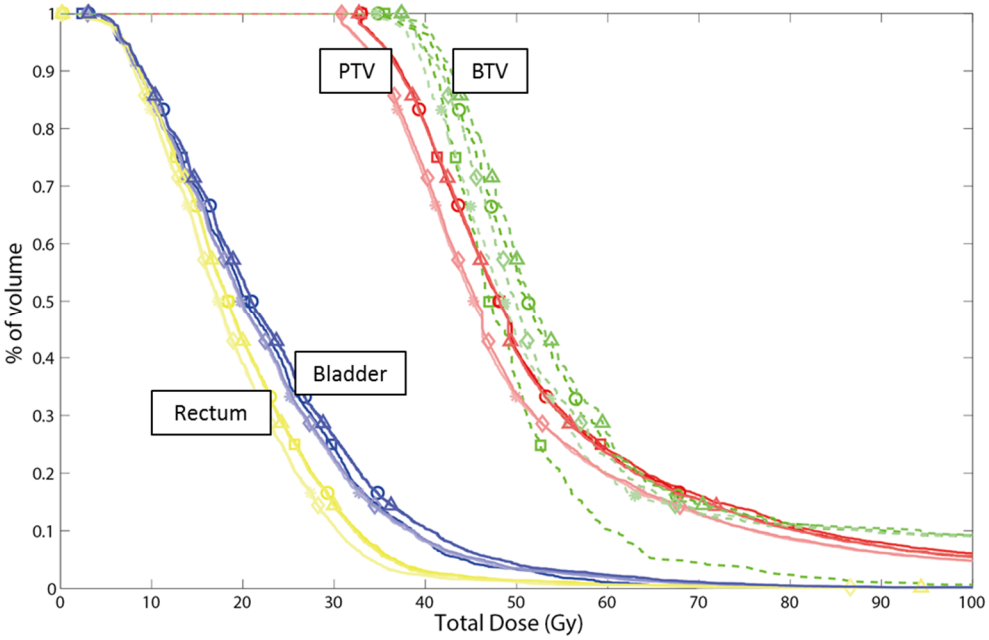

Figure 3.3b. This figure illustrates the dose volume histogram for a cervical cancer patient with a BTV that is 6.3% of the PTV.

To contrast the dose distributions of the standard versus escalated plans, Figure 3.4 illustrates clear hot spots (i.e., high dose, 150 percent isodose curves) around the PET-identified voxels in the escalated plan; however, they are absent in the standard plan.

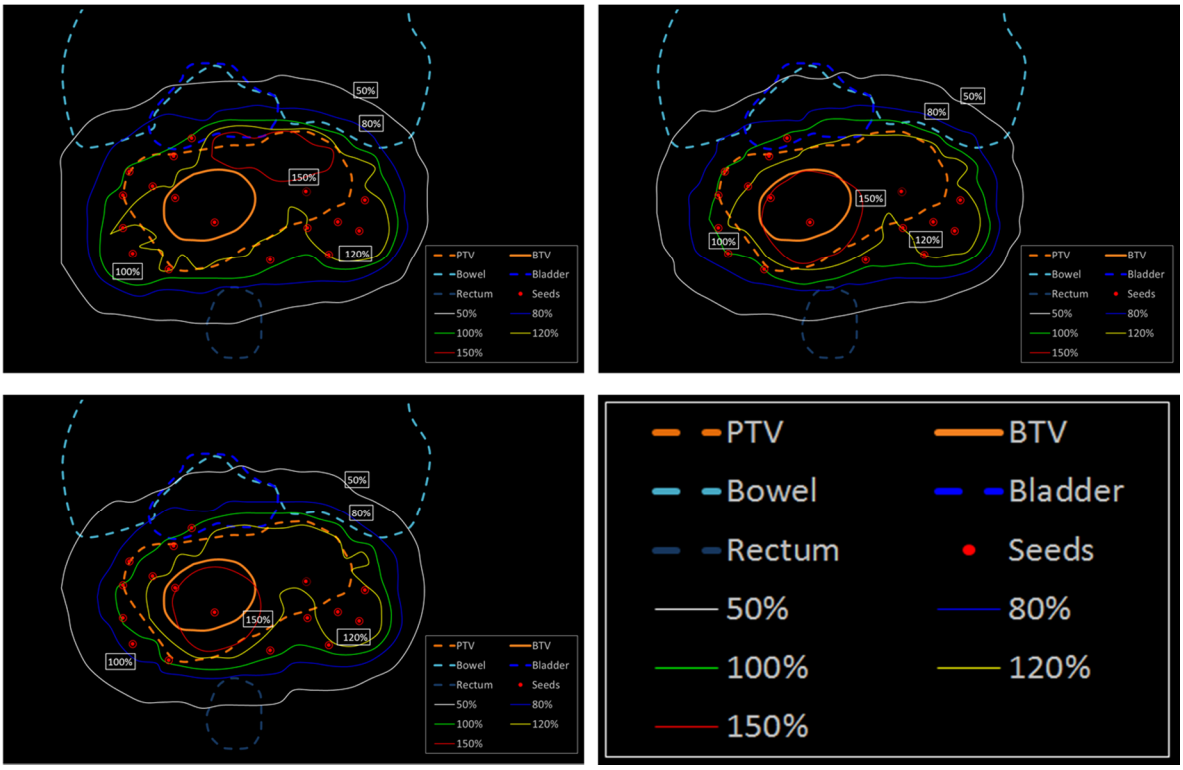

Figure 3.4. This figure contrasts the isodose curves of a standard plan (top left), an escalated plan with 35 Gy of PTV prescribed dose (top right), and an escalated plan with 33 Gy of PTV prescribed dose (bottom left). We can observe the very conformed 100 percent isodose curves to the PTV contour. The 150 percent isodose curves in both escalated plans around the PET-identified pockets are clearly absent from the standard plan. The 120 percent isodose curve is tighter in the bottom escalated plan, reflecting the lower dose to PTV because of the lower prescribed value (33 Gy versus 35 Gy).

For the 15 cases, the TCPs for standard plans range from 52 percent to 64 percent. For dose escalation, when an escalated dose of $\geq$ 37 Gy is placed in the PET-identified tumor pockets, all escalated plans show a slight reduction of 0.5 percent to 12.4 percent in the rectum and bladder dose, while 99 percent of the BTV receives over 40 Gy. The resulting TCP values range from 87 percent to 99 percent. When the BTV is less than 15 percent of the PTV, dose escalation can be delivered with a virtually identical PTV dose, as in the standard plan.

When the BTV is over 20 percent of the PTV, dose escalation to PET-identified voxels intrinsically increases the PTV dose by 1 percent. Boosting the BTV to 40 Gy results in no dose increase to the PTV in all plans. When the PTV is prescribed as a reduced dose of 33 Gy, independent of the size of the BTV, escalation can be achieved, while the dose to the PTV, bladder, and rectum are reduced simultaneously All plans can be generated within a CPU minute. This allows for real-time OR-based treatment planning and on-the-fly dynamic reconfiguration.

Table 3.1 highlights the TCP for three representative patients: BTV $\leq$ 10 percent of PTV (small), 10 to 25 percent (medium), and $\geq$ 25 percent (large). On all patients, the TCP of all escalated plans is over 70 percent. Specifically, when the BTV is boosted to over 40 Gy, the resulting TCP is uniformly high ($\geq$ 87 percent). We list the plans A to E according to improvements to the TCP. The best TCP (Plan E) is achieved when we boost the PET-identified pockets to over 40 Gy, while maintaining the PTV dose at 35 Gy. This can be partially explained by the fact that there remain cancer cells loosely populated outside the PET-identified pockets. Hence, a prescribed dose of 35 Gy is able to eliminate these cancer cells, but a prescribed dose of 33 Gy may not be as effective. Table 3.2 contrasts the dose received by the OARs. For brevity, we focus on the plans where the PTV all receives 35 Gy prescribed dose (i.e., plans A, B, and E).

Table 3.1. This table contrasts the TCPs in various plans. The boldface values represent the TCP associated with escalated plans.

BTV/PTV ratio category	Small	Medium	Large
PTV (in cc)	82.5	131.5	89.7
PET-identified volume (BTV in cc)	5.2	25.5	26.0
Ratio: BTV/PTV	6.3%	19.39%	28.99%
Treatment-planning model	**Tumor control probability (TCP)**		
A: Standard HDR plan (PTV dose = 35 Gy)	0.6052	0.6358	0.5942
B: PET-guided escalated plan (BTV ≥ 37 Gy, PTV = 35 Gy)	**0.8741**	**0.9383**	**0.9663**
C: PET-guided escalated plan (BTV ≥ 37 Gy, PTV = 33 Gy)	**0.8401**	**0.7382**	**0.7849**
D: PET-guided escalated plan (BTV ≥ 40 Gy, PTV = 33 Gy)	**0.9777**	**0.9465**	**0.9638**
E. PET-guided escalated plan (BTV ≥ 40 Gy, PTV = 35 Gy)	**0.9861**	**0.9639**	**0.9730**

Table 3.2. This table displays the dose distribution in standard plans versus escalated plans (using the same prescribed PTV dose). D90 (cGy) represents the dose received by 90 percent of the organ and D 2cc is the minimum dose to the most exposed two cm^3 of OARs.

Patient		D90 (cGy)		D2cc (cGy)		Mean dose (cGy)	
	Plans	PTV	PET pockets	Bladder	Rectum	Bladder	Rectum
Small	STANDARD	3,735.8	4,091.3	2,650.37	2,400.57	2,272.5	2,006.5
	B: BTV ≥ 37 Gy, PTV = 35 Gy	-0.7%	+3.2%	-6.5%	-0.2%	-0.5%	-0.9%
	E: BTV ≥ 40 Gy, PTV = 35 Gy	-0.6%	+4.6%	-6.1%	-0.2%	-0.5%	-0.8%
Medium	STANDARD	3,675.4	4,135	2,602.77	2,654.15	1,782.2	2,006.1
	B: BTV ≥ 37 Gy, PTV = 35 Gy	+0.8%	+3.8%	+3.1%	-3.2%	-1.0%	-0.9%
	E: BTV ≥ 40 Gy, PTV = 35 Gy	+0.4%	+6.5%	+2.3%	-2.0%	-1.8%	-2.6%
Large	STANDARD	3,666.2	4,046	2,791.26	2,881.11	1,189.3	3,042.2
	B: BTV ≥ 37 Gy, PTV = 35 Gy	-0.7%	+3.1%	-8.5%	-7.5%	-6.4%	-6.9%
	E: BTV ≥ 40 Gy, PTV = 35 Gy	+0.1%	+7.1%	-5.0%	-8.5%	-1.7%	-12.4%

3.1.5 Findings, Implementation, and Challenges

With the precision of the HDR brachytherapy delivery, a TCP improvement in treatment plans can be readily realized in the outcomes of actual treatments. TCP offers biological information about tumor and normal cells and their radiosensitivity and radioresistancy to radiation doses; thus, incorporating such information in optimized personalized treatment plans is invaluable. Coupled with TCP knowledge, advances in biological and functional imaging offer new opportunities to incorporate radiobiological parameters within the planning process. The OR-based treatment-planning algorithm we describe herein allows for TCP-driven PET-enhanced personalized treatment, which facilitates the targeted delivery of escalated doses and improvements in overall clinical outcomes.

Our study reveals improvements both in local tumor control and OAR toxicity, two competing and desirable goals that were previously thought to be unachievable simultaneously. The work herein showcases the importance of novel modeling and breakthrough optimization solution strategies in personalized treatment-planning advances. The dose escalation is sensitive to the size of the PET BTV. In addition, we have demonstrated that it is possible to reduce the PTV dose, while escalating doses to the BTV, as plans C and D, in which the PTV receives only 33 Gy.

This work addresses three unique challenges.

- The TCP function is complex and highly parametric, and is sensitive to the density of cancer cells and the radiobiological characteristics of normal and tumor cells.
- The biological-driven MIP treatment models are intractable using existing methodologies and competitive solvers, because they would require breakthroughs in polyhedral theory and computational advances. This is the first time that TCP is being incorporated within a treatment-planning optimization modeling and solution process. This is also the first time that PET-image-guided dose escalation is being performed.
- The actual gain in local tumor control must be validated through clinical trials to quantify the associated treatment outcome improvements that are realized through PET-guided dose escalation.

In practice, the sophisticated modeling and novel and fast optimization algorithm ensures that there is no increase in solution time for escalated dose planning. The radiation oncologists must guide us regarding the proper escalated dose values.

Rush University Medical Center began a clinical trial for PET hot HDR dose resteering in July 2011. All patients enrolled were diagnosed as having International Federation of Gynecology and Obstetrics (FIGO) IIB or IIIB cervical cancer. In a FIGO IIB cancer, the tumor has spread to the parametrial area (i.e., tissue surrounding the uterus); in a FIGO IIIB cancer, the tumor has grown into the pelvic wall and (or) causes hydronephrosis or nonfunctioning kidneys [8]. The patient's treatment included whole pelvic radiation therapy with concurrent chemotherapy. The radiation therapy treatment schemes included external beam treatments for the cervix and parametrium, along with HDR to the cervix with boosted PET-positive doses. The HDR CTVs were delineated based on CT and MR imaging positive volumes. The PET-positive volumes were boosted to 50 percent higher doses than those delivered to CTVs. Some cases were treated with the Syed applicator; others were treated with the tandom-ring applicator, in conjunction with three or four parametrial interstitial needles.

All patients who entered the study received two treatment plans: one using the in-house treatment-planning algorithm for dose escalation and another using a commercial algorithm with options for manual fine-tuning. The attending physician was responsible for selecting the final plan. Acceptable boost plans were achieved for all patients. These boosted plans were then successfully delivered to all patients. For the boosted plans, the PET pockets received elevated doses compared to the standard plans, while doses to the bladder and rectum were reduced. The clinician was pleased with the performance of the PET-image-guided targeted-dose escalation and the successful completion of the clinical study. The hospital is planning to continue the study by enrolling additional patients. The results thus far indicate that such dose escalation is feasible to deliver and is beneficial to the cancer patients. The techniques are applicable to the treatment of other types of cancer, including breast, lung, prostate, and esophagus cancer.

3.1.6 Impact and Significance

As we state above, cervical cancer is the second most common cancer in women worldwide; approximately 500,000 new cases are diagnosed each year [194]. In developing nations, it is often the most common cause of cancer-related death among women and a leading cause of death overall [141]. In this section, we discuss the significance of our work.

Quality of Care and Quality of Life for Patients

- Compared to standard HDR plans, PET-guided dose-escalation plans improve tumor control consistently across all patients. This translates to improvements in cure rates and reductions in mortality.
- Clinical evidence shows a reduction of the radiation dose to the bowel, bladder, and rectum. Thus, our system reduces side effects and complications. This has a profound impact on both health-care costs and the patient's quality of life.
- The planning process requires only seconds to return good treatment plans. This offers quality assurance in treatment delivery (image-guided or not), independent of the training and experience of the operators. It helps to ensure a uniform quality of care among patients and across all hospital sites.

Advancing the Cancer Treatment Frontier

- The work marks the first time that complex TCP is incorporated within treatment planning, and as an objective in driving high-quality treatment plans.
- This work also marks the first time that PET images are incorporated within the treatment-planning environment for targeted dose-escalation planning optimization.
- The fast solution engine and the seamless incorporation of functional imaging information allows treatment-plan optimization and reoptimization in real time based on new images. as the patients receive daily treatments. This opens up the potential for next-generation adaptive real-time image-guided HDR brachytherapy.
- With advances in biological imaging, such as PET and MRS, incorporation of such knowledge within treatment modalities will soon become standard for personalized

treatment planning. To the best of our knowledge, Rush University's radiation therapy clinical trial is the first and only one in the country that delivers PET-based dose- escalation HDR brachytherapy for patients with cervical cancer. Our work has the potential to set the national standard and guidelines for biological image-guided brachytherapy treatment.

- Observing the clinical trend, brachytherapy is rapidly becoming the treatment of choice, because its side effects are generally less severe when compared to external beam radiation therapy and surgery, and because of its effectiveness for early-stage treatment. Further, brachytherapy preserves the organ and its functionality. The latter is of special concern to younger early-stage cancer patients who still look forward to bearing children.
- The methodologies are applicable to brachytherapy (both high-dose rate and low-dose rate) for other types of cancer, including prostate, breast, bileduct, lung, and sarcoma.

Advances in Operations Research Methodologies

- This study marks the first use of sophisticated combinatorial optimization approaches to tackle the complexities inherent in incorporating TCP as a clinical objective within HDR brachytherapy. The resulting treatment plans offer superior TCPs with simultaneous toxicity reduction to OARs. This can be a precursor to subsequent clinical trials.
- This study is the first in which PET images are incorporated within the planning optimization model, giving rise to the competing goals of escalating the dose to the tumor, while simultaneously not increasing the dose to the OARs.
- The highly nonlinear multiobjective MIP environment offers a powerful modeling paradigm. However, the resulting intractable instances demand theoretical and computational breakthroughs to solve these instances for actual clinical delivery.
- We introduce a new concept of generalized conflict hypergraphs and derive polyhedral results, including hypercliques, hyper-oddholes, hyper-antioddholes, hyper-webs, hyper-antiwebs, and hyper-star facial structures, and their associated CG ranks.

- The branch-and-cut and local-search approach described herein couples new polyhedral cuts, along with matrix reduction and intelligent geometric heuristics. This approach can successfully address the highly complex and nonlinear TCP function and the dose-based objective. It can rapidly return good, feasible solutions.
- Independently, this work motivates our polyhedral investigation on MIP convexification of posynomial and signomial functions [141, 142], an area that deserves advances in its own right, because many real-world problems can be modeled as complex NMIPs that demand theoretical and computational advances.

Within the medical community, there is an urgent push to incorporate the radiobiological characteristics of normal and tumor cells, and biological and functional imaging advances, within treatment delivery to realize and improve clinical outcomes and tumor control. Our work provides proof of concept of the feasibility and potential clinical benefits of such personalized, targeted treatment-planning design and delivery. Moreover, the resulting plans offer improvements in tumor control and reduce the radiation to the OARs, two competing and desirable characteristics that were previously thought to be unachievable simultaneously.

Our rapid operator-independent biological treatment-planning system provides the groundwork for advancing the technological frontier of image-guided brachytherapy. It opens up opportunities to conduct complex clinical investigations that may otherwise be impossible, as evidenced in MRS-guided dose-escalation studies [147, 126, 127, 128, 129, 130, 131]. The sophisticated OR modeling paradigm provides great flexibility in realistically modeling the clinical problem, and the novel rapid solution engine objectively returns the best possible plans.

Appendix

Dose Calculation

Let $D(r)$ denote the dose contribution per minute of a seed to a voxel that is r units away. The two-dimensional dose-rate calculation can be represented as $D(r,\theta)$, which is calculated based on the AAPM task groups for brachytherapy TG43U1 and U2 [139].

$$2D: \quad D(r,\theta) = S_K \cdot \Lambda \cdot \frac{G(r,\theta)}{G(r_0,\theta_0)} \cdot g(r) \cdot F(r,\theta), \qquad (1)$$

where S_k is the air kerma strength (U, 1U = 1 cGy cm^2/hour), $\Lambda = D(r_0,\theta_0)/S_k$ represents the dose-rate constant (cGy/hr-U), $g(r)$ is a radial dose function, $G(r,\theta)$ represents a geometry function, with $G(r_0,\theta_0)$ as the geometry function at the reference point (r_0,θ_0), where $r_0 = 1\,\text{cm}$ and $\theta_0 = 90^\circ$, and $F(r,\theta)$ is a two-dimensional anisotropy function. The values for the geometry function $G(r,\theta)$ are obtained from tabulated data.

Incorporating the TCP within the Treatment-Planning Objective

To the extent that PET can be taken to indicate the presence of faster-proliferating and (or) a higher density of tumor cells, recognizing such regions in the organ could be consequential in terms of tumor control. We incorporate the TCP within our treatment-planning process. Specifically, we are interested in (1) the maximal TCP gain obtainable by incorporating PET information in treatment planning, and (2) the largest fractional tumor-pocket volume for which PET-guided planning remains useful. Clearly, if tumor cells are uniformly spread throughout most of the cervix volume, the gain would be insignificant.

We generalize the TCP based on a reliable biological model developed by Zaider and Minerbo [204]. The formulas are derived using the birth and death processes. For brachytherapy, the TCP equation is

$$TCP(t) = \left[1 - \frac{S(t)e^{(b-d)t}}{1 + bS(t)e^{(b-d)t} \int_0^t \frac{dt'}{S(t')e^{(b-d)t'}}} \right]^n, \qquad (2)$$

where n is the initial number (at time $t = 0$) of tumor cells, $S(t)$ is the survival probability of tumor cells at time t, and b and d are the birth and death rates of these cells, respectively.

The birth rate b and the death rate d relate to two parameters: potential doubling time T_{pot} and tumor cell loss factor ϕ, where $b = 0.693/T_{pot}$ and $\phi = d/b$. In the TCP calculation, the time t in Equation (2) is typically taken to be the duration of the treatment period or the expected remaining life span of the patient.

For simplicity and convenience in Equation (2), we use the linear quadratic expression for the survival function: $S(D) = e^{-\alpha D - \beta q(t) D^2}$, where D is the dose delivered over the time interval t, and $q(t)$ makes explicit the repair of sublethal damage. In the case in which the dose rate decreases exponentially,

$$q(t) = \frac{2(\lambda t)^2}{(\mu t)^2 (1 - \lambda^2/\mu^2)(1 - e^{-\lambda t})^2} \left[e^{-(\lambda+\mu)t} + \mu t \left(\frac{1 - e^{-2\lambda t}}{2\lambda t} \right) - \frac{1 + e^{-2\lambda t}}{2} \right],$$

where λ is the radioactive decay constant of the radioisotope ($\lambda = 0.0094$ d^{-1} for ^{192}Ir) and $\mu = 1/t_0$, where t_0 is the average time for the sublethal damage repair, typically in the order of one hour.

To complete the TCP calculation, we give numerical values to the parameters n, T_{pot}, ϕ, α, β, and t_0 to represent the response of the rapidly proliferating and (or) radioresistant segment of tumor cells.

The TCP value is sensitive to the volume and density of cells. We take the volume and density of cells in the cervical tumor ranges from $\rho = 10^8$ to $\rho = 10^{10}$ cells/cm^3. The potential doubling time is taken as $T_{pot} = 15$ or $T_{pot} = 20$ days, and the cell loss factor is taken as $\phi = 0.5$ or $\phi = 0.75$. In addition, $\alpha = 0.487$ Gy^{-1}, $\beta = 0.055$ Gy^{-2} for radiosensitive cells, and $\alpha = 0.155$ Gy^{-1}, $\beta = 0.052$ Gy^{-2} for radioresistant cells, whose values are determined from *in vitro* cell-survival measurements. For the sublethal damage repair constant. we take $t_0 = 1$ hour.

Novel TCP-Driven PET-Image-Guided Treatment-Planning Model

We employ a multiobjective MIP model for HDR brachytherapy treatment planning. The model incorporates the TCP as the objective function, in addition to the rapid dose fall-off function to ensure dose conformity to the tumor region.

Briefly, let x_j be a 0/1 indicator variable for recording placement or nonplacement of a seed in grid position *j* and t_j be the continuous variable for the dwell time of a seed in grid position j. The total radiation dose at voxel P is given by

$$\sum_j D(\| P - X_j \|) t_j, \tag{3}$$

where X_j is a vector corresponding to the coordinates of grid point j, $\| \bullet \|$ denotes the Euclidean norm, and $D(r)$ denotes the dose contribution per minute of a seed to a voxel that is *r* units away.

The target lower and upper bounds, L_P and U_P, for the radiation dose at voxel *P* are represented by the following dose constraints:

$$\sum_j D(\| P - X_j \|) t_j \geq L_P \tag{4a}$$

$$\sum_j D(\| P - X_j \|) t_j \leq U_P. \tag{4b}$$

For each voxel P in each anatomical structure, a binary variable is used to capture whether or not the desired dose level is achieved. For simplicity, we use the BTV to represent the set of tumor voxels identified by the PET images.

The TCP-driven PET-image-guided dose-escalated multiobjective treatment model is given by the following model:

Maximize $\sum_{P}(\eta_P v_P^L + \mu_P v_P^U)$

Maximize TCP

subject to

$$\sum_j D(||P - X_j||)t_j \geq \text{PrDose}\cdot\lambda \qquad P \text{ in BTV} \qquad (5)$$

$$\sum_j D(||P - X_j||)t_j + M_P(1 - v_P^L) \geq L_P \qquad P \text{ in PTV-BTV}, \; P \text{ in OARs} \qquad (6)$$

$$\sum_j D(||P - X_j||)t_j - N_P(1 - v_P^U) \leq U_P, \qquad P \text{ in PTV-BTV}, \; P \text{ in OARs} \qquad (7)$$

$$|\text{BTV}| + \sum_{p\in(\text{PTV}-\text{BTV})} v_P^L \geq \alpha\,|\text{PTV}|, \qquad (8)$$

$$t_j \leq T_j x_j \qquad (9)$$

$$\sum_j x_j \leq MaxSeeds \qquad (10)$$

$$v_P^L,\ v_P^U,\ x_j \in \{0,1\},\ t_j \geq 0$$

PrDose represents the clinical prescribed dose to the tumor, and λ ($\lambda > 1$) represents the dose-escalation factor. This factor is guided by clinicians as well as its effect on normal tissue complication. Constraint (5) ensures that the PET-identified tumor voxels receive escalated doses. In Constraints (6), and (7), v_P^L and v_P^U are 0/1 variables. If a solution is found such that v_P^L = 1, then the lower bound for the dose level at point P is satisfied. Similarly, if v_P^U = 1, the upper bound at point P is satisfied (see Constraint 7).

The constants M_P and N_P are chosen appropriately for the PTV and for various OARS. For PTV, M_P corresponds to the underdose limit, whereas N_P corresponds to

overdose limit, and L_P = *PrDose* corresponds to the prescription dose. M_P and N_P are strategically chosen so that the overall PTV dose remains relatively homogeneous (e.g., $N_P / M_P < 1.2$), as the clinicians desire. For the OARs, N_P represents the maximum dose tolerance that the organs can sustain, without inflicting severe and permanent harm. These values are determined from clinical findings and are part of the planning procedures and guidelines. For cervical cancer treatment, Constraint (6) does not apply to any OAR.

In Constraint (8), α corresponds to the minimum percentage of tumor coverage required (e.g., α = 0.95). Because all the PET-identified tumor voxels satisfy the prescribed dose bound (and beyond), we count those in PTV-BTV and these BTV-voxels to ensure that overall it satisfies α percent of the tumor volume. Here, $|PTV|$ represents the total number of voxels used to represent the PTV of the cervix. Constraint (8) thus corresponds to the coverage level that the clinician desires. In Constraint (9), the duration t_j in grid position j is positive only when this position is selected. Its value is bounded by the maximum time limit T_j. The time usually is bounded by the length of the treatment session, which is usually between 20 and 30 minutes, depending on the tumor stage and prognosis condition. Constraint (10) limits the number of seeds used to *MaxSeeds*. The constant can be omitted; however, in some cases, clinicians know their desired numbers, which they tell the planner.

Note that BTV voxels are excluded in Constraint (7) because there is no reason to place an upper bound on the dose to these tumor voxels. Constraint (5) ensures that no underdose for PET-identified voxels exists; thus, Constraint (6) is unnecessary for these voxels.

The first objective is to find a treatment plan that satisfies as many bound constraints as possible; this is surrogate to rapid dose fall-off, ensuring conformity of the prescribed dose to the tumor. The parameters η_P and μ_P allow us to prioritize the importance of various anatomical structures. Using a weighted sum is important for the cervical cancer cases to balance the volume of the cervix versus the nearby OARs (e.g.,

bladder, rectum, and bowel). The second objective function incorporates the TCP function, which depends on the time of the treatment, radioactive decay of the radioisotope, dose received, volume and density of tumor cells, and the biological radiosensitivity and radioresistancy of the normal and tumor cells.

Acknowledgment

This work is partially supported by a grant from the National Science Foundation.

3.2 TCP Optimization in Cancer Treatment Planning

In this section, it includes a paper that is prepared for submission to the International Journal of Radiation Oncology, Biology and Physics.

3.2.1 Introduction

In radiation therapy, radiation is delivered using either external beam technology or using internal seed implantation which is known as brachytherapy. For high-dose rate (HDR) brachytherapy, the radioactive substance that sealed in the seeds or catheters are placed temporarily into or near the cancer to eradicate the tumor while minimal radiation is delivered to the healthy tissues. Recent advances in functional imaging, such as magnetic resonance spectroscopy (MRS) and positron emission tomography (PET), facilitate identification of the tumor cells based on their molecular characteristics, which allows the treatment planning to deliver escalated dose to the tumor inside the target organ [50, 107]. The improvement on the overall clinical outcome in HDR treatment using dose escalation guided by PET imaging has been shown in our previous studies [93].

Operations research approaches such as linear optimization, mixed integer programming, and multi-objective optimization, have been wildly applied to construct the treatment planning optimization models [30, 32, 34, 43, 48, 93, 94, 162]. The goal of treatment optimization is to deliver the prescriptive dose to the tumor while limiting or minimizing radiation to healthy cells and critical organs in the neighborhood of the tumor. The radiation dose to the healthy cells and critical organs in treatment optimization can be controlled by dose volume constraint and generalized equivalent uniform dose (gEUD). Dose volume constraint on an organ limits the portion of the organ that receives more than a given prescriptive dose [34, 43, 48]. Generalized EUD for an organ is given by $\text{gEUD} = (1/n \sum_i d_i^a)^{1/a}$, where n is the number of voxels in the structure, d_i is the dose to voxel i, and a is a parameter. Craft et al incorporated the gEUD as an objective into the multi-objective optimization model for IMRT treatment planning [32]. The function $f(d) = d^a$ in gEUD is fitted with a three-segment piecewise linear convex function. Thus the multi-objective optimization can be constructed as a two-step linear programming problem:

minimizes the excess of the piecewise linear gEUD function over the maximum desired level in the first step, and uses the solution from the step 1 model as constraints to find the Pareto optimal treatment plan in the second step.

Tumor control probability (TCP) is an estimated function of the probability of killing all malignant cells in the affected organ given a dose of radiation in radiation therapy. In treatment planning of radiation therapy, TCP has been wildly used for assessing the biological response of the treatment plans [43, 50, 80, 147, 191, 205]. Mohan et al investigated the balance between TCP and normal tissue complication probability (NTCP) [134]. Potential improvement on TCP from a treatment plan has been investigated [174]. When an integral total dose volume or energy deposition is given, the maximum TCP value can be obtained by solving the TCP equation with Lagrange multipliers [40, 173, 200], and the obtained dose distribution can be served as the prescription dose in the following inverse planning [201]. Levin-Plotnik et al. discussed about finding the dose distribution that maximizes the TCP function while the mean dose is fixed and given [116]. Zaider et al derived an analytical expression of TCP taking into account the length of the treatment and the stochastic process of eradiating tumor cells [204]. To find the dose distribution with optimal TCP while constraining on the dose requirement on the health tissues remains challenging.

We incorporated the TCP into the objective of the multi-objective optimization model in previous study [43]. The optimal TCP is found within the neighborhood of the solution obtained from various optimization models. In this study, we build an optimization model that maximizes the TCP directly and satisfies the dose requirements on the targeted organ, and healthy tissues in order to preserve their functionalities while delivering an escalated dose to the tumor. We focus on the optimal treatment planning of HDR brachytherapy on cervical cancer using dose escalation guided by PET imaging. We design a solution strategy that first fits the piecewise linear approximation of the TCP function, then solves the piecewise-linear optimization problem that maximizes the piecewise linear approximation of TCP, and finally performs a local search to improve the TCP value within the neighborhood of the solution.

3.2.2 Methodology

TCP is a function computing the probability that no malignant cells are left in the affected organ. It can be estimated in the treatment planning by using spatial dose configuration and the dwell time per treatment. Let x_j be a 0/1 indicator variable for recording placement or non-placement of a seed in grid position j and t_j be a continuous variable for the dwell time of a seed in grid position j. The total radiation dose received by voxel P in a treatment can be calculated as:

$$d_P(T) = \sum_j D(|P - X_j|) \cdot t_j, \quad t_j \le T,$$

where T denotes the duration of the treatment, X_j denotes a vector corresponding to the coordinates of grid point j, $|\cdot|$ denotes the absolute distance, and $D(|\cdot|)$ denotes the dose contribution per minute of a seed to a voxel. Let n_p represent the number of tumor cells in the representing area of voxel P. Therefore, the TCP equation at the end of the treatment according to the model developed by Zaider and Minerbo [204] is given by:

$$TCP\ (T, d_P) = \prod_P \left[1 - \frac{S(T, d_P)\ e^{(b-d)T}}{[1 + bS(T, d_P(T))\ e^{(b-d)T} \int_0^T \frac{dt'}{S(t', d_P(t'))\ e^{(b-d)t'}}]} \right]^{n_P},$$

where the survive probability $S(t, d_P(t))$ can be calculated by

$$S(t, d_P(t)) = e^{-\alpha d_P(t) - \beta q(t) d_P^2(t)}.$$

Our optimization model maximizes the TCP that also satisfies the dose requirement given by the physician. We employ PET-image to facilitate our treatment planning by identifying the dense pockets of cancer cells. The volume delineated by PET-image defines the boost target volume (BTV). Thus escalated dose can be delivered to BTV while the prescriptive dose is delivered to the rest of the volume in the planning target volumes (PTV), which is the cervix in this study. In addition to killing the cancer cells and preserving the functionality of the cervix, a treatment plan should preserve the functionalities of the surrounding critical organs including rectum, bladder, and bowel. For convenience of reading, we denote *PTV-BTV* as the set of voxels in the region of PTV and outside the region of BTV, and *BTV* as the set of voxels in the region of BTV. Let *PrDose*

denote the prescriptive dose for *PTV-BTV*, and *BTV_PrDose* denote the escalated prescriptive dose for *BTV*. Let *OAR* denote the set of voxels in critical organs. In our study, *OAR* can be split into three subsets surrounding the cervix: *RECTUM*, *BLADDER*, and *BOWEL*.

We incorporate partial dose volume constraints on the maximum/minimum dose to *PTV-BTV* and *BTV* and maximum dose constraints to *RECTUM*, *BLADDER*, and *BOWEL*. For example, a constraint of a maximum dose to *RECTUM* can be that 95% of the voxels must receive no more than 90% of the prescriptive dose and 100% of the voxels must receive no more than 120% of the prescriptive dose. Let α_s denote the voxel coverage percentage, β_s denote the percentage of the prescriptive dose, and $L_{s,k}^{P}$ denote a 0/1 indicator to denote whether the dose to voxel $P \in s$ satisfies the dose requirement β_s for set *s* in $\{PTV, PTV\text{-}BTV, RECTUM, BLADDER, BOWEL\}$. The prescriptive dose of tumor that can kill the tumor cells is instructed by physicians. Recall *BTV_PrDose* and *PrDose* denote the prescriptive dose for *BTV* and *PTV-BTV* respectively where *BTV_PrDose* is *PrDose* plus the escalated dose on *BTV*. The multiple sets of dose constraints can be used with different combinations of (α_s, β_s) as shown in Table 3.3. Specifically, a hard constraint is set for the minimum dose to *BTV* that the escalated prescriptive dose of *BTV* must be satisfied. For the minimum dose to *PTV-BTV*, two constraints are set: 100% level at 93% of the prescriptive dose, 95% level at 100% of the prescriptive dose. For the maximum dose to *PTV-BTV* and *BTV*, three levels of coverage are predetermined: 90% level at 150% of the prescriptive dose, 95% level at 160% of the prescriptive dose, and 100% level at 200% of the prescriptive dose. For organs in *OAR*, four levels for the maximum dose are pre-determined: 100% level at 120% of the prescriptive dose, 95% level at 90% of the prescriptive dose, 90% level at 80% of the prescriptive dose, and 80% at 60% of the prescriptive dose.

Table 3.3a. Partial dose volume constraints for minimum dose requirements.

Set s	% of covered voxels (α_s)	% of prescriptive dose (β_s)	Prescriptive dose
BTV	100%	100%	*BTV_PrDose*
PTV-BTV	100%	93%	*PrDose*
	95%	100%	*PrDose*

Table 3.3b. Partial dose volume constraints for maximum dose requirements.

Set s	% of covered voxels (α_s)	% of prescriptive dose (β_s)	Prescriptive dose
BTV	100%	200%	*BTV_PrDose*
	90%	150%	*BTV_PrDose*
	95%	160%	*BTV_PrDose*
PTV-BTV	100%	200%	*PrDose*
	95%	160%	*PrDose*
	90%	150%	*PrDose*
OAR	100%	120%	*PrDose*
	95%	90%	*PrDose*
	90%	80%	*PrDose*
	80%	60%	*PrDose*

Let *TCP*(*D*) denote the TCP value given a dose configuration $D = \{d_P\}$, $P \in BTV$ for simplicity and the fixed treatment duration T = 20 min. The optimization model then can be formulated as a nonlinear mixed integer programming:

Maximize $TCP(D)$
Subject to
Dose received by voxels

$$\sum_j D(|P - X_j|)t_j - d_P = 0 \tag{3.2.1}$$

Maximum dose constraints on *PTV-BTV* and *OAR*

$$d_P - L_s^P\left(\beta_s(\alpha = 1) - \beta_s\right) Pr\ Dose \le \beta_s\ Pr\ Dose. \qquad P \in s,\ s = PTV\text{-}BTV,\ OAR \tag{3.2.2}$$

$$\sum_{P \in s} L_s^P \le \alpha_s\ N_s \qquad P \in s,\ s = PTV\text{-}BTV,\ OAR \tag{3.2.3}$$

Minimum dose constraints on *PTV-BTV*

$$d_P - L_{PTV-BTV}^P\left(\beta_{PTV-BTV}(\alpha = 1) - \beta_{PTV-BTV}\right) Pr\ Dose \ge \beta_{PTV-BTV}\ Pr\ Dose \qquad P \in PTV\text{-}BTV \tag{3.2.4}$$

$$\sum_{P \in PTV-BTV} L_{PTV-BTV}^P \ge \alpha_{PTV-BTV}\ N_{PTV-BTV} \qquad P \in PTV\text{-}BTV \tag{3.2.5}$$

Minimum dose constraints on *BTV*

$$d_P \ge BTV\_PrDose. \qquad P \in BTV \tag{3.2.6}$$

Maximum seeds constraint

$$\sum_j x_j \le max\ Seeds \tag{3.2.7}$$

$$0 \le t_j \le T \cdot x_j,\ x_j \in \{0,1\},\ L_s^P \in \{0,1\} \qquad P \in s \tag{3.2.8}$$

$\beta_s(\alpha)$ denotes β_s value corresponding to a given α_s, and N_s represents the number of voxels in set *s*. Constraints (1) derive the dose at voxels from the duration of seeds and the dose rate function. Constraints (3.2.2) - (3.2.3) ensure that the maximum dose coverage constraints are met for all voxels in *PTV-BTV* and *OAR*. Constraints (3.2.4) - (3.2.5) ensure that the minimum dose coverage constraints are met for all voxels in *PTV-BTV* and *OAR*. Constraints (3.2.6) ensure that the minimum dose requirements for voxels in *BTV* are satisfied. Constraint (3.2.7) limits the number of seeds to *maxSeeds*. This constraint can be omitted, or in some cases, clinicians have their desired numbers in which they can inform the planner. The duration t_j in grid position j is positive only when this position is selected, and its value is bounded by the treatment duration *T*.

It is difficult to solve the optimization problem with the nonlinear objective TCP function, this is true also for the convexified and linearized version of the TCP function due to its sensitivity and highly nonlinear characteristics. We design a solution algorithm described as follows.

The optimization problem is equivalent to a convex optimization problem that maximizes the logarithmic function of the TCP function $log(TCP(T, d_P))$ since the TCP function is a non-decreasing function with:

$$log\,TCP(D) = log\left[\prod_P TCP(d_P)\right] = \sum_P log\ TCP(d_P).$$

Furthermore, the logarithm function $log\ TCP(d_P)$ is a concave function that values in $(-\infty, 0]$ as $TCP(d_P)$ is a non-decreasing function on [0,1]. By fitting $log\ TCP(d_P)$ with a concave piecewise linear function where the numbers and the length of segments are pre-determined by the minimum and maximum dose requirements, the optimization problem can be converted into a piecewise linear optimization problem, and ultimately a linear optimization problem that maximizes the concave piecewise linear approximation of the TCP function.

In the final step, we perform local search within the neighborhood of the obtained optimal solution that maximizes $TCP(D)$ rather than the piecewise approximation of it. A simulated annealing algorithm is executed for 1,000 iterations in which subgradients of the dwell time vector t are found by swapping 0.01 minute of seed duration from one seed location to another to search for an improvement direction.

3.2.3 Results

To gauge the feasibility, characteristics, and potential benefit of PET-image guided dose escalation, initial validation consists of fifteen cervical cancer patient cases. These patients have all received prior 45 Gy of external radiation dose. The PTV ranges from 82.8 to 137.47 cm^3 and the BTV ranges from 10% to 41%. For each patient case, 3 alternative strategies are contrasted: a) standard HDR plan with no escalation, b) BTV escalation with

the same PTV prescription dose, and c) BTV escalation with reduced PTV prescription dose. For both escalated strategies, we consider two variations (37 Gy boost to BTV vs. 40 Gy boost) and observe the effect on PTV and OAR dose profiles and TCP quality.

For the fifteen patients, the tumor control probabilities for standard plans range from 48% to 63%. For dose escalation, when an escalated dose of $\geq$ 37 Gy is placed on the PET-identified tumor pockets, in all escalated plans, there is slight reduction (2-5%) in rectum and bladder dose, while 99% of BTV receives over 40Gy. The resulting TCP values range from 82% to 99%. When the BTV is less than 15% of the PTV, dose escalation can be delivered with virtually identical PTV dose as in the standard plan. When BTV occupies over 20% of the PTV, dose escalation to PET-identified voxels intrinsically increases PTV dose by 1%. Boosting BTV to 40Gy results in no dose increase to PTV in all plans. When PTV is prescribed a reduced dose of 33Gy, independent of the size of the BTV, escalation can be achieved while dose to PTV, bladder and rectum are simultaneously reduced. All plans can be generated within a CPU minute. This allows for real-time OR-based treatment planning and on-the-fly dynamic re-configuration.

Table 3.4 highlights the tumor control probability for three representative patients: BTV 10% of PTV (small), 20-30% (medium), and 41% (large). On all patients, TCP of all escalated plans is over 80%. Specifically, when BTV is boosted to over 40Gy, the resulting TCP is uniformly high ($\geq$ 89%). We list the plans A –E according to improvement in TCP. The best TCP (E) is achieved when we boost the PET-identified pockets to over 40Gy while maintaining the PTV dose at 35 Gy. This can be partially explained by the fact that there remain cancer cells loosely populated outside the PET-identified pockets. Hence a prescribed dose of 35 Gy is able to eliminate these cancer cells, but a prescribed dose of 33 Gy may not be as effective. Figure 3.5 illustrates the dose volume histogram and dose profiles for the three patients listed in Table 3.4.

Table 3.4. Comparison of TCPs in treatment plans. Plan B - E: Minimizing total overdose in cervix incorporated with TCP. Plan B1 and C1: Solving the linear relaxation problem of maximizing TCP.

BTV/PTV Ratio Category	Small	Medium	Large
Planning Target Volume (PTV in cc)	82.5	131.5	89.7
PET-identified volume (BTV in cc)	5.2	25.5	26.0
Ratio: BTV/PTV	6.3%	19.39%	28.99%
Treatment Planning Model	Tumor Control Probability		
A. Standard HDR Plan (PTV dose = 35 Gy)	60.5%	63.6%	59.4%
B. PET-guided Escalated Plan (BTV $\geq$ 37 Gy, PTV = 35Gy)	87.4%	93.8%	76.6%
C. PET-guided Escalated Plan (BTV $\geq$ 37 Gy, PTV = 33Gy)	84.0%	63.8%	68.5%
D. PET-guided Escalated Plan (BTV $\geq$ 40 Gy, PTV = 33Gy)	97.8%	94.7%	96.4%
E. PET-guided Escalated Plan (BTV $\geq$ 40 Gy, PTV = 35Gy)	98.6%	96.4%	97.3%
B1. TCP Optimization Plan (BTV $\geq$ 37 Gy, PTV = 35Gy)	99.0%	97.8%	99.1%
C1. TCP Optimization Plan (BTV $\geq$ 37 Gy, PTV = 33Gy)	96.3%	91.9%	96.5%

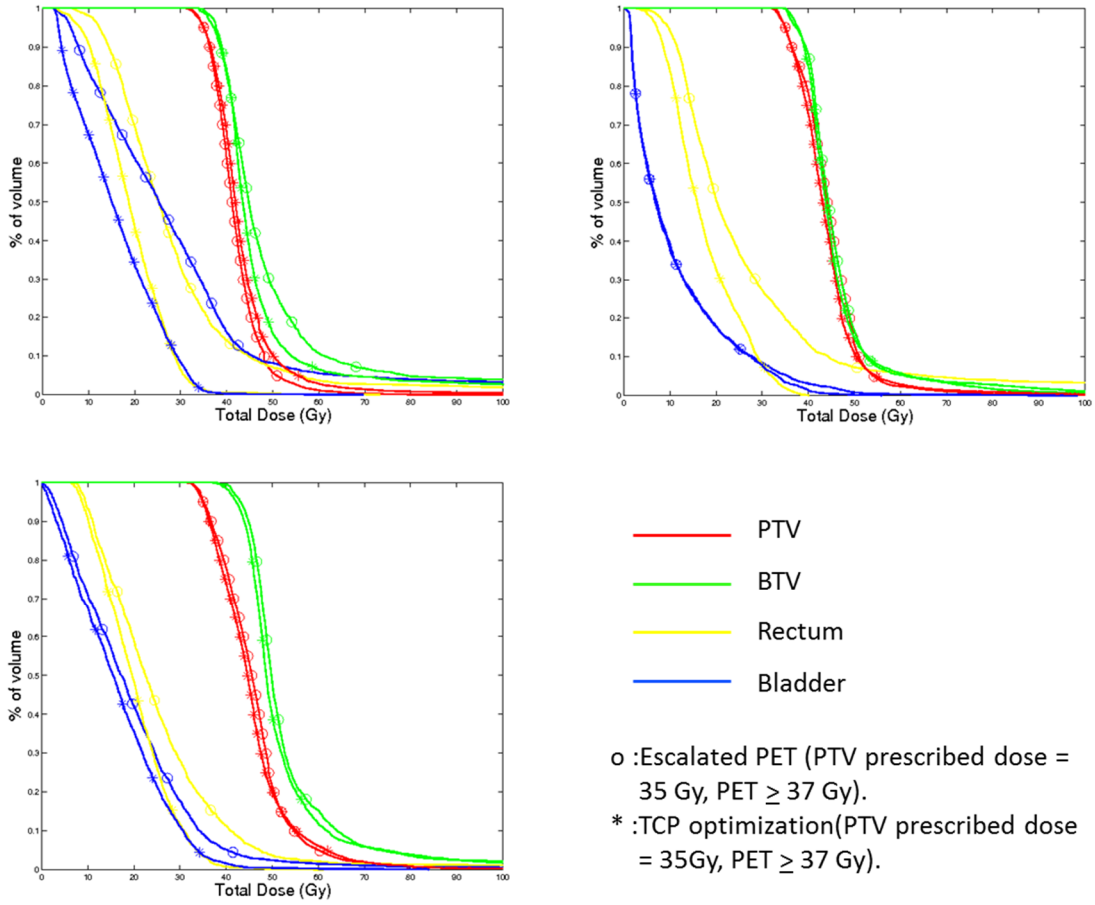

Figure 3.5. Comparison of Dose volume histogram between dose escalated plan (B) and TCP optimization plan (B1).

CHAPTER IV

PUBLIC HEALTH DECISION MAKING

Public health emergencies such as pandemic flu and bioterrorism attack require public health administrators to assess the upcoming event and determined the needed resources to mitigate the impact of the emergencies. When the resources are limited, it is critical to allocate the resources optimally and strategically such that the loss from the emergency is minimized. For example, the number of infections in a flu pandemic can be minimized by vaccination that prioritizing the high-risk groups. On the other hand, when the resources are adequate to prevent the emergency, it is beneficial for the public health department to adopt the most efficient strategy such that the effort and budget can be saved for other incidents.

In this chapter we discuss two topics: In the first section, we investigate vaccination strategies against a pandemic flu to find the optimal strategy when limited vaccines are available by constructing a mathematical model for the course of the flu development and the process of the vaccination; in the second section we analyze the cost-effectiveness of emergency response strategies against a large-scale anthrax attack where the entire population in a city must be protected within 48 hours.

4.1 Strategies for Vaccine Prioritization

This section contains the paper to be appeared in Interfaces - The Daniel H. Wagner Prize for Excellence in Operations Research Practice 2015; EK Lee, F Yuan, F Pietz, BA Benecke. Vaccine Prioritization for Effective Pandemic Response.

Vaccine Prioritization for Effective Pandemic Response

Eva K Lee [*], Fan Yuan [*], Ferdinand Pietz [†], Bernard Benecke [‡]

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, † Strategic National Stockpile,‡ Global Disease Detection and Emergency Response, Centers for Disease Control and Prevention, GA.

Abstract

When vaccine availability is limited, prioritized vaccination is considered the best strategy to contain a (flu) pandemic. We derive a mathematical decision framework to track the effectiveness of prioritized vaccination through the course of a pandemic. Our approach couples a disease propagation model with both a vaccine queuing model and optimization engine to determine the optimal prioritized coverage in a mixed vaccination strategy. This approach demonstrably minimizes infection and mortality.

The study reveals there is an optimal coverage for the high-risk group that results in the lowest overall attack and mortality rates, given known outbreak characteristics, vaccine inventory levels, and individual risk factors, Such information is critical to public health policy makers as they determine the best strategies for population protection. This is particularly important in determining when to switch from a prioritized strategy focusing on high risk groups to a non-prioritized strategy where the vaccine becomes publicly available. This analysis highlights the importance of non-interrupted vaccine supply. Although the 2009-H1N1 supply eventually covered over 30% of the population, the resulting attack and mortality rates are inferior to a scenario in which only 20% of the population is covered by a non-interrupted supply. Early vaccination is also important: a 3-week delay diminishes a 9.9% infection reduction to a mere 0.9%.

The optimal trigger for switching from prioritized to non-prioritized vaccination is sensitive to infectivity and vulnerability of the high-risk groups. Our study underscores

the importance of throughput efficiency in dispensing and its effects on the overall attack and mortality rates. The more transmissible the virus is, the lower the threshold for switching to non-prioritized vaccination. Our model is generalizable, and allows incorporation of seasonality and virus mutation of the biological agents. The system empowers policy makers to *make the right decisions at the appropriate time to save more lives, better utilize limited resources, and reduce the health service burden during a pandemic event.*

*The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

4.1.1 Introduction

When limited vaccines are available, prioritized vaccination is considered the best strategy to mitigate the impact of a (flu) pandemic [132, 187]. Traditionally, healthcare workers and volunteers involved in vaccinations [150], children, elderly, and others with underlying medical conditions receive priority for receiving vaccine. The effectiveness of prioritizing elderly, however, has been under debate [170]. Previous prioritized vaccination strategies focused largely on spatial distribution of the population and age, with the goal of minimizing the severity of the pandemic measured by infections, deaths, years of life lost, and cost [130, 133, 163, 196]. Prioritizing children from ages 5 to 17 years old and an early start of the vaccination campaign are stressed for reducing morbidity and mortality [28]. Tuite et al investigated the optimal ranking for age stratified groups and risk stratified groups under unlimited vaccine availability [186]. Willinga et al reported an importance leveling strategy to decide the vaccination coverage for risk groups under limited vaccine availability [189]. Mylius et al compared complications ranking and infection ranking strategies [137]. Yang et al. report that vaccinating children before adults may mitigate a severe epidemic if vaccine coverage of the population is high and vaccination starts early enough [199].

Previous work has proposed strategies for optimal vaccine distribution by an iterative guessing routine [118], by linear approximation [130,131], and via a stochastic simulation framework [148]. Most United States public health departments use a mixed vaccination strategy that starts as prioritized then switches to a non-prioritized strategy after a certain time interval [26]. Determining *when* that switch should occur is critical and vital.

Past studies assumed that vaccination takes no time and all vaccinees receive vaccine simultaneously once vaccines arrive [130, 131, 186]. Others assumed that vaccination occurs before the pandemic [117]. A proper disease spread model must consider the distribution of vaccinations over time, including such important factors as supply interruptions and varying patient arrival rates.

This project was initiated ahead of the H1N1 flu when the lead author was invited to a White House public health event. The White House Director for Medical Preparedness Policy sought technical advice and recommendations for better optimizing population protection initiatives. The computational system and the paper emerged from that effort.

The goal is to design a system that empowers policy makers with strategies for mass infection and casualty mitigation during pandemic under strained time and limited vaccine supplies. Our work takes all the critical factors into account by coupling a vaccine dispensing queuing model with a vaccination optimization module. These are driven by a disease propagation model that aims to determine the optimal trigger for switching from prioritized to non-prioritized vaccination so as to minimize infection and mortality severity throughout the pandemic's course when vaccine supplies are limited. We contrast two propagation models: systems of ordinary differential equations (ODE) versus an individualized agent-based model. In both cases, a 6-stage disease transmission model is used, and the vaccination process is characterized by operations and service distribution data collected via time-motion studies in seasonal and H1N1vaccination events at various sites across the nation [98].

We analyze the sensitivity of the optimal switch trigger (i.e., trigger for switching from a prioritized to a non-prioritized dispensing strategy) over these parameters: vaccine availability, start of vaccination, dispensing throughput efficiency, triage accuracy, outbreak characteristics, infectivity rate, and risk factors.

We believe this work is the first mathematical-computational model to combine disease propagation, dispensing operations, and optimization capability; and the first that allows rapid determination of optimal switch triggers. Moreover, it includes innovative computational strategies to derive good near-optimal solutions. To the best of our knowledge, previous studies on vaccine prioritization did not include all the critical details considered herein in their models, nor were they used by the CDC, or used otherwise, to influence public-health policy making.

CDC confirms that this is the first time an actionable and operational switch trigger has been defined, an advance that is critical and vital to better mitigation of infections and mass casualties. The study confirmed the importance in the establishment of the National Institute of Translational Medicine for rapid medical countermeasures and drug design. The system, RealOpt-VacOpt, has been in use since 2011 for advising on vaccine distribution, prioritization, and triage strategies. In particular, the CDC Strategic National Stockpile escalated their dispensing training with a team of dedicated CDC public health experts who are providing hands-on training with RealOpt-POD to improve operational performance.

4.1.2 Materials and Methods

Models and Parameters

To study the infectious status of the population, we develop mathematical computational models that capture a 6-stage SEPAIR disease propagation process (susceptible (S), exposed (E, infected but not infectious), infectious (P, infectious but not yet symptomatic), asymptomatic (A, infectious and asymptomatic), symptomatic (I,

infectious and symptomatic), and recovered and immune (R)) along with the stochastic vaccination process inside point-of-dispensing sites (PODs) (Figure 4.1). For comparison, we take two approaches. The first integrates a system of ordinary differential equations for disease progression with a queueing model for vaccine dispensing (ODE-queueing), and an optimization module for finding the optimal switch trigger while tracking the vaccination and disease propagation process. The second replaces the ODE-queueing model by agent-based simulation [97]. The disease spread here is tracked in great detail as individuals go through the vaccination process. Optimization is performed over the ODE system or within the simulation environment to determine the optimal mixed strategy.

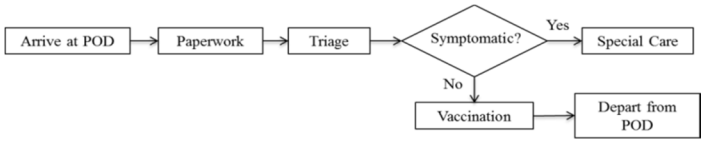

Figure 4.1. This figure shows the patient flow inside a typical point-of-dispensing (POD) site for flu vaccination. Based on collected H1N1 mass vaccination data, triage service time is fitted with a triangular distribution of (0.25, 2.642, 3.75) (minute), whereas dispensing is estimated with a triangular distribution of (0.2, 0.684, 1.633) (minute)

We stratify the population into five risk groups: normal adults, children 10 years-old and younger, healthcare workers, pregnant women, and patients with underlying disease conditions. The latter four are considered high-risk groups for influenza related complications; i.e., they have higher probability of being infected, and have higher mortality rate once they are infected [37].

The novel 6-stage SEPAIR transmission model is used to track the disease progression of each individual [97,195]. The measure of transmission is the basic reproduction number R_0 [35, 51, 70]. We create three more stages: special care (SC), vaccinated and immune (V), and deceased (D). SC tracks the individuals with influenza-like symptoms who choose to self-quarantine at home, or are treated by the doctors for the illness. Consequently, these symptomatic individuals do not infect others. V tracks the

individuals who have received vaccines, developed immunity, and are protected by the vaccines.

PODs are set up for dispensing the flu vaccine; and each individual receives only one shot of the vaccine [26]. Due to heterogeneous system behaviors inside and outside the POD, the vaccination model is divided into three components: outer-POD, intra-POD, and post-POD. All individuals stay in outer-POD at the beginning. Once a prioritized vaccination strategy begins, high-risk individuals will arrive at the POD and receive vaccine.

The computational model characterizes the services inside the POD, including paperwork, triage, vaccination, and special care as shown in Figure 1. Other stations, such as orientation and medical and mental evaluation can also be incorporated. Individuals in any of the SEPAIR stages will arrive at the POD to be vaccinated. Everyone except symptomatic individuals receive vaccines. Inside the POD, health care workers triage individuals with influenza-like symptoms and assign them to consultation, treatment, or hospitalization as necessary. Vaccines only work on individuals in susceptible stage. However efficacy is not inherently100% [26]. Therefore, some vaccinated individuals (about 10%) return to the community with their immune status unchanged. With only one vaccine per individual allowed, they do not come back to the POD. (The model itself can accommodate the use of 2 or more vaccine doses per individual.) Appendix A presents the ordinary differential equations and queueing model for disease propagation and POD operations

Past studies assumed that vaccination takes no time and all vaccinees receive vaccine simultaneously once the vaccines arrive [130,186]. Others assumed that vaccination occurs before the pandemic [118]. Neither assumption reflects realistic or typical on-the-ground situations. For more practical views, we integrate vaccination operations from actual events within our transmission models to study the overall disease propagation process.

During CDC time-motion studies for flu and H1N1 mass dispensing, we collected arrival rates and service times. These observed data are incorporated into our model. Specifically, we consider the case where the arrival rate decreases as the number of non-vaccinated individuals in the population decreases. During the prioritized vaccination stage, only high-risk, not-yet-vaccinated people will come.

We assume that the initial arrival rates for prioritized strategy and non-prioritized strategy are similar. Since the high-risk population is significantly smaller than the non-high-risk one with this assumption, we infer that high-risk individuals are more willing to receive vaccine under the prioritized strategy than the overall population. This appropriately aligns with the assumption that the high-risk individuals are more willing to receive vaccine compared to the general population [39, 145]. Motivated public health campaigning helps facilitate this process.

Let J be the total number of risk groups, j=1, 2, ..., J, where "1" represents the low-risk group (normal adults). The arrival rate $\lambda(t)$ at time t can be expressed as a linear function of the current non-vaccinated live population. Let N denote the total population, N' denote the high-risk population:

$$N' = \sum_{j \neq 1,\, j \in J} \pi_j N$$

where π_j is the percentage of population in group j. Specifically, $\lambda(t)$ can be expressed as follows:

$$\lambda(t) = \begin{cases} \lambda_0 N_0(t) / N & \text{if non-prioritized strategy is being used,} \\ \lambda_0 [\sum_{j \neq 1,\, j \in J} N_{0,j}(t)] / N' & \text{if prioritized strategy is being used.} \end{cases}$$

Here, λ_0 is the initial arrival rate, $N_0(t)$ is the non-vaccinated non-special care live population outside the POD at time t, and $N_{0,j}(t)$ denotes the number of individuals of group j in $N_0(t)$. Figure 4.2 illustrates an example of the variation of arrival rates under mixed strategy. $N_0 = N$ initially.

We use the symbol g to denote the switch trigger. The prioritized strategy is halted when the percentage of vaccine covering the high-risk individuals reaches an optimal value for g.

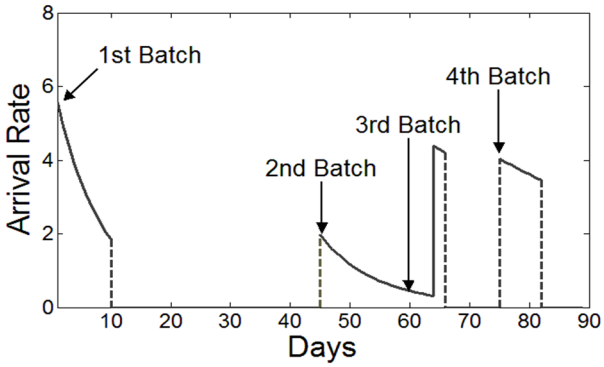

Figure 4.2. Arrival rate for 2009 H1N1 supply when mixed strategy was used: 4 batches of vaccine supply were available at day 1, day 45, day 60, and day 75, respectively. The mixed strategy switched from prioritized to non-prioritized on Day 63. This is just an illustration; this is not an

Let AR represent the overall attack rate, and AR_j represent the overall attack rate in group j. The attack rate is a form of incidence that measures the proportion of persons in a population who experience an acute health event during a limited period (e.g., during an outbreak), calculated as the number of new cases of a health problem during an outbreak divided by the size of the population at the beginning of the period. The optimal switch trigger g^* that minimizes the total number of infections and mortality can be obtained by solving the following nonlinear optimization problem:

$$Min\ AR(g) = \sum_{j=1}^{J} AR_j(g)$$

$$s.t. \sum_{j=2}^{J} VE_j(t) \le g \cdot VS \quad \text{during prioritized strategy,} \qquad (1)$$

$$\sum_{j=1}^{J} VE_j(t) \le VS(t) \quad \forall t \ge 0 \qquad (2)$$

$$0 \le g \le u = \min\ \{N'/VS \cdot 100\%,\ 100\%\} \qquad (3),$$

where $VE_j(t)$ represents the number of vaccinees in group j, N' represents the high-risk population, $VS(t)$ represents the total vaccine supply by time t, VS represents the overall vaccine supply level, and u is an upper bound of the prioritized coverage. (1) ensures that the number of high-risk individuals who receive vaccine is no greater than g fraction of total available vaccine supply under the prioritized strategy. (2) ensures that the total number of vaccinees will not exceed the total vaccine supply at any time. Both VE_j and AR_j are output from the ODE system. The system collects the value of VE_j continuously to maintain feasibility of the constraints. Thus, once the total number of vaccinees

reaches the current vaccine supply, prioritized vaccination stops. At that point, the PODs are shut down until the next batch of vaccines arrives. This reflects the actual practice where POD planners keep track of their total vaccine inventory, and maintain headcounts of arrival to ensure proper closing of the entrance door.

In our model, the optimal switch trigger is obtained by a line search algorithm for a given scenario where the objective function value is calculated through the ODE-queueing or the simulation system. We perform multiple iterations with random initial switch trigger values. It is computationally intensive to solve the ODE-queueing-optimization system, with each instance requiring about 4000 CPU minutes to solve.

Optimization within a simulation environment is also difficult and remains a challenge. We employ our large-scale computational system developed at CDC, RealOpt©, to achieve the results [98].

Design of Experiments

The model allows us to stratify the risk groups and their vulnerabilities. We use η_j to denote the *risk-factor* of the infection rate for group *j*, $\eta_j \geqslant 1$, where 1 represents the standard risk factor for normal adults. Then, $\eta_2 = 2$ implies that the infection rate for children 10 and younger is two times that of normal adults. In our study, we also contrast the optimal switch trigger with high-risk group including children (0-10 years-old) versus the scenario that excludes school children (6-10 years-old).

We perform our analysis based on the 2009 H1N1 vaccine supply, which arrived in four batches: the first batch covered 10.0% of the population; the subsequent three batches covered 6.67%, 4.0%, and 10.0% of the population, and were available 45 days, 60 days, and 75 days respectively after the first batch [26].

Early vaccination has been shown to be crucial to contain the pandemic [67]. We consider early vaccination from two factors: i) vaccine supply without interruption and ii) vaccination start time. We study their joint impact on optimal switch trigger in a mixed

strategy.

We contrast the 2009 H1N1 vaccine supply schedule to two other levels: 20% and 40% continuous supply. In the mathematical model, the time horizon is set to 90 to 120 days, and the time unit is in minutes. The reproduction number R_0 ranges from 1.0 to 2.0. We analyze vaccination effectiveness under different start times (no delay to 3-week wait until arrival of vaccines) as well as different initial numbers of infection, 0.5% to 2.0% of the population.

To illustrate the importance of integrating vaccination operations into the disease propagation study, we contrast the attack rates and mortality rates when vaccination throughput varies. Our default triage accuracy is set to 95%. We also experiment with triage false positive and false negative rates ranging from 50% to 100% to gauge their importance. To validate the performance of our models, we contrast the results from the agent-based stochastic simulation-optimization model to the ODE-queueing-optimization model. We also compare our results against the actual infection estimates of the H1N1 pandemic as reported by the CDC [26, 144].

4.1.3 Results

We report optimal switch triggers for the mixed vaccination strategy for three scenarios of vaccine supply: 1) The 2009 *H1N1 vaccine supply* schedule; 2) 20% vaccine supply; and 3) 40%. The analysis is performed for the state of Georgia with a population of 9,687,663 [25]. Forty-three point-of-dispensing sites are setup strategically for mass vaccination to protect the region [98].

The high-risk groups consist of 14.8% children 10 years-old and younger, 3.0% healthcare workers and volunteers, 1.2% pregnant women, and 0.3% patients with underlying disease conditions. Each individual is assumed to receive only one shot of vaccine [26].

The *switch trigger* is expressed in terms of *the percentage of vaccine used for the high-risk population before vaccination is opened up to the general population*. The term *full non-prioritized strategy* refers to the case when the switch trigger is 0%, meaning a non-prioritized strategy is employed from the outset, where no distinction is made of the risk factors among the population. The term *full prioritized strategy* means that vaccine will be given first to all high-risk individuals before it will be distributed to the general population. Thus, the switch trigger for a full prioritized strategy is 100% if the number of available vaccine doses is less than the number of individuals in the high risk population, and is equal to the ratio of the high risk population to available doses otherwise. Note that when the vaccine supply is less than that of the high-risk population, a full prioritized strategy means all vaccine will be used for the high-risk group and nothing is left for the general population. The term *mixed strategy* refers to a situation where a specified percentage of vaccine is used first to vaccinate the high risk group, after which any remaining vaccine is used to vaccinate the general public.

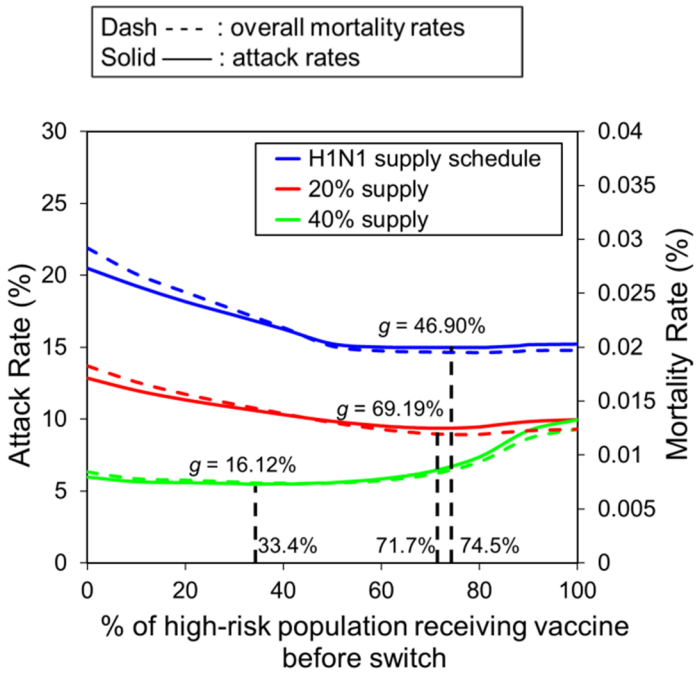

Figure 4.3. Plot of the overall (day 0 to day 90) attack rate and mortality rate as a function of percentage of high-risk coverage when R_0 = 1.2 (blue = 2009 H1N1 supply level, red = 20% level, green = 40% level). Solid

Figure 4.3 shows plots of the overall attack rate and mortality rate as a function of the high-risk group coverage when the reproduction value R_0 = 1.2, and the initial infection of the population is 1%. We indicate the global minimum of each curve and

mark on it the associated optimal switch trigger, g, with respect to the three vaccine supply levels. We note that an increase of attack rate of 1% in our study amounts to 96,876 sick individuals.

Table 4.1 shows the optimal mixed strategies under the three vaccine supply scenarios as R_0 varies, and compares the associated attack rate for the optimal mixed strategy to those resulting from the full prioritized strategy, and the full non-prioritized strategy. When R_0 = 1.2, under the 2009 H1N1 vaccine supply schedule, the optimal mixed strategy occurs with switch trigger of 46.9%. The attack rate increases by 1.54% and 36.78%, respectively, for the full prioritized coverage and the full non-prioritized strategies**.** Across all R_0, these strategies produce an average attack rate increase of 0.84% and 16/96% respectively, when compared to the optimal mixed strategy for H1N1 vaccine supply. Note that for the 40% vaccine supply level, the full prioritized strategy is strikingly inferior to the optimal mixed strategy. (Attack rates increase by 77% or more for all R_0.)

Independent of the vaccine supply level, the optimal switch trigger decreases as R_0 increases. This indicates that the start time for non-prioritized vaccination should begin earlier when the pandemic is more severe. Likewise, when the level of vaccine supply is relatively high, the start time for non-prioritized vaccination should also begin earlier. In general, the percentage of high-risk population receiving vaccine decreases as the vaccine supply level increases and as the pandemic is more severe. There is a slight abnormality for 2009 H1N1 vaccine supply (74.5% high-risk coverage versus 71.7% for 20% vaccine level). This is partly due to the supply interruption. For the same reason, for a given R_0, the 2009 H1N1 vaccine supply schedule contributes to the highest attack rate when compared to the other two supply levels.

Table 4.1. Optimal mixed strategy and associated attack rate, and comparison with attack rates for the full prioritized strategy, and the full non-prioritized strategy, as vaccine supply and R_0 vary.

R_0	Vaccine Supply	Optimal mixed strategy			% increase in attack rate w.r.t. optimal attack rate, AR^*	
		Optimal switch trigger g = % of vaccine used for high-risk	% of high-risk receiving vaccine	Attack rate AR^*	Full prioritized	Full non-prioritized
1.2	20%	69.19%	71.7%	9.37%	+6.30%*	+37.14%
	2009 H1N1	46.90%	74.5%	14.98%	+1.54%	+36.78%
	40%	26.12%	54.1%	5.48%	+81.75%	+9.12%
1.6	20%	69.00%	71.5%	33.72%	+8.16%	+15.90%
	2009 H1N1	32.73%	52.0%	46.48%	+0.88%	+10.20%
	40%	20.27%	42.0%	16.20%	+125.12%	+12.59%
2.0	20%	59.25%	61.4%	52.76%	+8.23%	+5.14%
	2009 H1N1	32.67%	51.9%	64.75%	+0.11%	+3.89%
	40%	16.45%	34.1%	32.25%	+77.05%	+5.89%

* The % increase in attack rate is computed as $(AR - AR^*)/AR^*$, where AR is the attack rate of the indicated strategy, and AR^* is the rate for the optimal mixed strategy.

Figure 4.3 and Table 4.1 highlight one very important result: although the 2009 H1N1 vaccine supply eventually covered 30.67% of the population, the resulting attack and mortality rates are inferior to the scenario when vaccine supply level is only 20% but available for continuous use with no supply disruptions.

Figure 4.4a contrasts results obtained from the ODE-queueing model with those obtained from the individual-based stochastic simulation model for the optimal mixed strategy (46.9%), the full prioritized strategy, and the full non-prioritized strategy under the 2009 H1N1 vaccine schedule when R_0 = 1.2. From simulation, the attack rates are 14.32%, 14.51%, and 19.53% respectively for the optimal mixed strategy, the full prioritized strategy, and the full non-prioritized strategy. Note that 0.19% infection amounts to 18,406 infections. We observe similar trends: by day 94, optimal mixed strategy results in 1.54% and 36.78% infection reduction with respect to the other two strategies in the ODE-queueing model, and 1.36% and 36.38% reduction in the individual-based stochastic simulation model. Figure 4.4b zooms in to show the difference between the optimal mixed strategy and the full prioritized strategy.

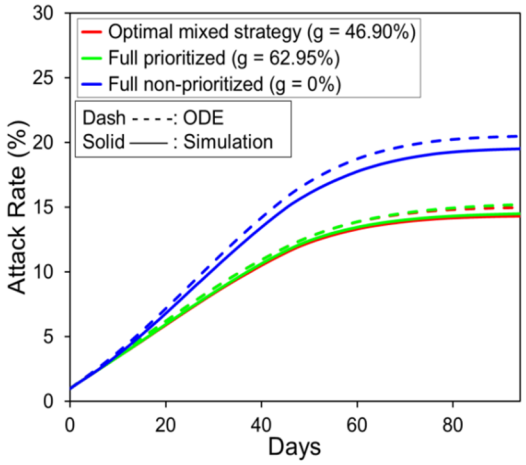

Figure 4.4a. Attack rate curves for the ODE-queueing model (dotted) versus the individual-based stochastic simulation model (solid) when $R_0 = 1.2$ for 3 vaccination strategies under the 2009 H1N1

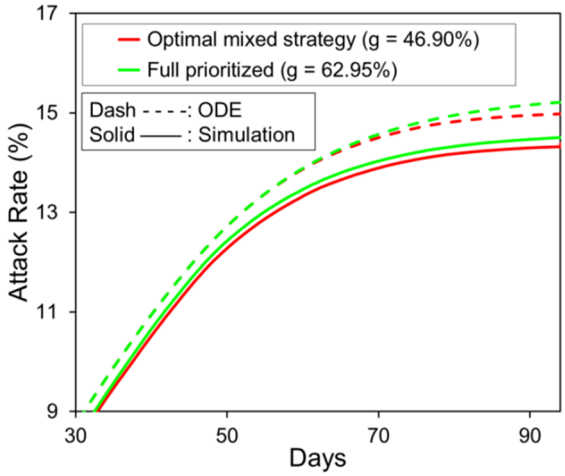

Figure 4.4b. A closer look at the difference between the optimal mixed strategy versus full prioritized strategy

In Table 4.2**,** we contrast optimal mixed strategies for various risk-factor combinations for pre-school children and school children in two scenarios: i) all children 10 years and younger are included in the high-risk group (14.8%); ii) exclude school children (age 6-10, 6.6%) from the high-risk group. In the first scenario the optimal switch trigger increases when the risk factor of preschool children and/or of school children increases.

If school children (age 6-10) are excluded from the high-risk group, the optimal switch trigger increases when the risk factor of pre-school children increases or the risk factor of school children decreases.

Table 4.2. The table contrasts the optimal switch trigger (%) against risk factor combinations for pre-school and school children. Risk factor 1 represents no added risk, 2 means twice as likely to get infected if exposed, and 4 means four times as likely to get infected.

Vaccine supply level:		20%			2009 H1N1 vaccine schedule			40%		
High risk group	School children risk factor	Preschool children risk factors								
		1	2	4	1	2	4	1	2	4
Include all ≤ 10 years old	1	32.04	62.63	75.46	32.98	38.97	50.42	15.97	16.07	20.99
	2	56.26	69.19	77.30	34.06	46.90	50.48	16.07	16.12	23.40
	4	72.09	76.04	77.88	50.42	50.48	50.75	20.27	23.40	25.81
Exclude 6-10 years-old	1	28.85	60.99	72.28	47.40	47.59	47.65	13.94	14.14	24.70
	2	28.18	60.78	72.18	47.34	47.53	47.65	13.90	14.09	24.70
	4	10.42	60.69	70.25	3.15	47.40	47.53	5.45	14.09	24.66

To quantify the importance of early vaccine availability, Figure 5a compares the daily prevalence of infection for optimal mixed strategies given a single-batch 10% supply versus the 2009 H1N1 vaccine supply schedule for different scenarios related to when initial vaccination begins. For the case of no delay, with vaccination beginning on day 0, the H1N1 supply schedule begins to show reduction in daily prevalence starting at day 45 and eventually achieves over two-fold reduction (at day 90, 0.08% vs. 0.27%). There is no difference prior to day 45 since the H1N1 schedule starts out at only 10%, but eventually covers 30% of the population. However, in the cases for which there is a delay in initial vaccination, the reduction becomes nearly negligible.

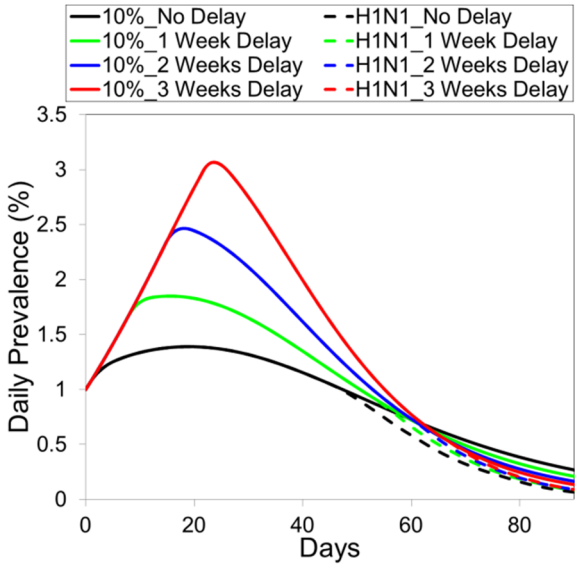

Figure 4.5a. This figure shows the percentage of population that are infectious at the end of each day resulting from the 2009 H1N1 supply (dashed curves) versus 10% vaccine supply (solid curves) when vaccination begins on time,1 week delay, 2 weeks delay, and 3 weeks delay and is dispensed according to the associated optimal mixed strategy. Here, $R_0 = 1.2$, initial infected population is 1.0%. The x-axis represents the number of days since the pandemic began.

Figure 5b compares the overall attack rates (left figure) for optimal mixed strategies given four supply levels and different scenarios as to when initial vaccination begins. It shows that when vaccination starts on time, the 2009 H1N1 supply results in 9.9% reduction in overall infections compared to a single-batch 10% supply (14.98% versus 16.63%)**.** However, when vaccination is delayed for 3 weeks, the associated reduction is only 0.9%. A similar trend is observed for the mortality rates (right figure). These figures highlight the importance of starting vaccination as early as possible.

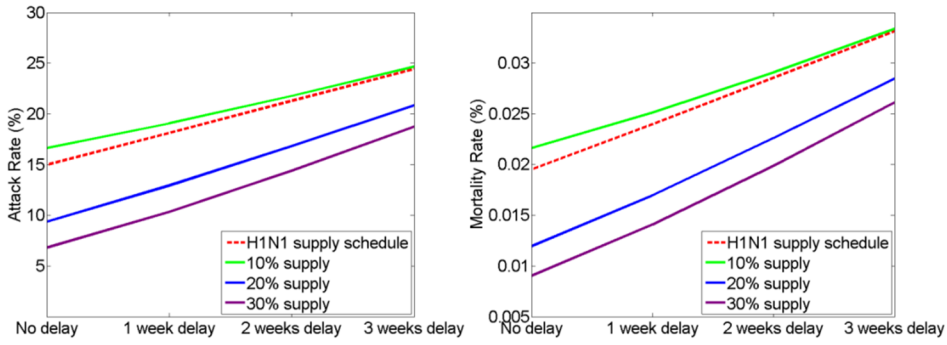

Figure 4.5b. This figure contrasts the overall (day 0 to day 90) attack rates (left) and mortality rate (right) associated with four different supply scenarios in the face of an initial delay in availability ranging from zero to three weeks. The optimal switch triggers are 100%, 69.19% and 31.20%, respectively, for 10%, 20%, and 30% vaccine supply levels. The differences in both attack and mortality rates are less profound as the delay increases.

Figure 4.6 (left) shows the optimal switch triggers for several combinations of reproduction number and initial number of infected individuals with respect to the initial

vaccine supply level. The optimal switch triggers appear to converge as the vaccine supply level reaches around 50%. Recall that the percentage of high risk population in our study is 19.3%. Figure 4.6 (right) shows that when the vaccine supply is less than 13.8% of the population, independent of the pandemic severity and the initial population infection, all vaccines will be given to the high-risk individuals. Table 3 shows the maximum vaccine supply level under which all will be dispensed to the high-risk population. This table reinforces that to achieve optimal infection and mortality reduction it is not always necessary to vaccinate all high-risk individuals first.

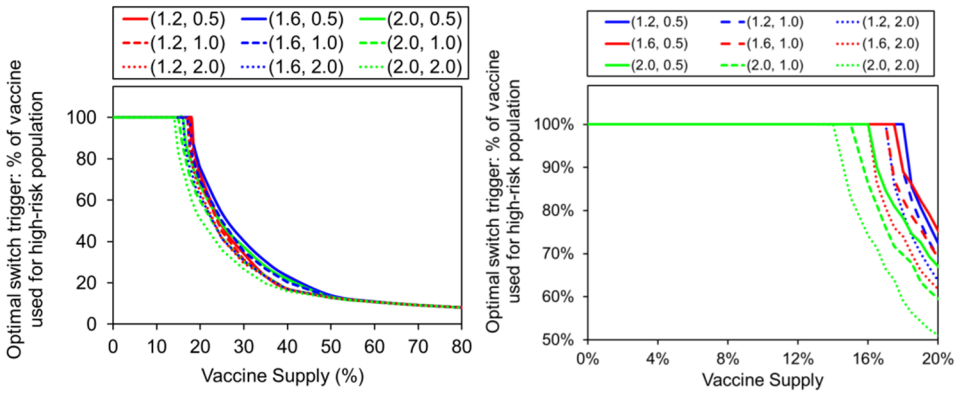

Figure 4.6. This figure (left) shows optimal switch triggers (percentage of vaccine dispensed to the high-risk group) against the vaccine supply levels in the optimal mixed strategy for nine different scenarios (R_0, α) of initial reproduction rates, R_0, and percentage of initial infection, α. The high-risk population is at 19.3%. The graph shows that, even when there is quite limited supply of vaccine, it is advantageous to begin vaccination of the general population before vaccinating all those in the high risk group. A magnified view (right) regarding the vaccine level versus percentage used for high-risk individuals before switch.

Table 4.3. From Figure 4.6 (right), the maximum vaccine supply level under which all will be dispensed to high-risk individuals.

(R_0, α)	$R_0 = 1.2$	$R_0 = 1.6$	$R_0 = 2.0$
$\alpha = 0.5$	17.9%	17.5%	16.0%
$\alpha = 1.0$	17.7%	17.2%	15.1%
$\alpha = 2.0$	17.2%	16.0%	13.8%

Figure 4.7 shows the resulting attack rates and mortality rates for the three vaccine supply levels and varying dispensing throughput efficiency rates. Observe that

the attack rate increases as the dispensing throughput decreases, with the increase more significant when the dispensing throughput is low. A similar trend is observed across all three vaccine supply levels. A similar trend is also observed for the mortality rates. This underscores the importance of dispensing efficiency, even within an optimal mixed strategy with no vaccination delay. Accepting that server variability in dispensing is unavoidable, dispensing throughput efficiency is highly influenced by optimal resource allocation at the dispensing sites, as shown in recent CDC large-scale dispensing studies [189, 190].

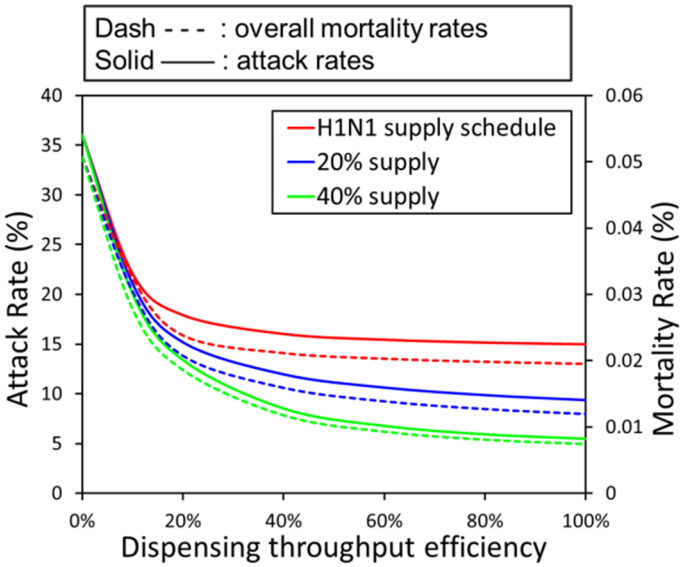

Figure 4.7. Contrast of the overall (day 0 to day 90) attack rates and mortality rates associated with three supply scenarios with respect to dispensing throughput efficiency. Solid lines denote attack rates, and dotted lines

When triage fails to identify individuals with flu perhaps due to lack of symptoms (false-negative), the resulting attack and mortality rates increase slightly. However, the optimal switch trigger remains rather constant. This trend is observed on all vaccine supply levels and we briefly illustrate the results for vaccine level 20%. Figure 4.8a shows the attack rate decreases from 9.46% for 50% triage accuracy to 9.27% at 100% triage accuracy. The optimal switch trigger remains at 69.19% throughout. However, when healthy people are being triaged as infected (false positive), the optimal switch trigger decreases, signaling earlier vaccination to the general public (Figure 4.8b).

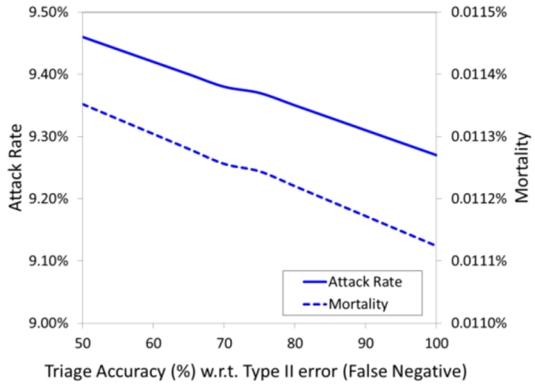

Figure 4.8a. Effect of triage accuracy (results from false negative) on overall attack rate and mortality rate. Attack rate increases slightly from 9.27% to 9.46%. A triage error of 10%

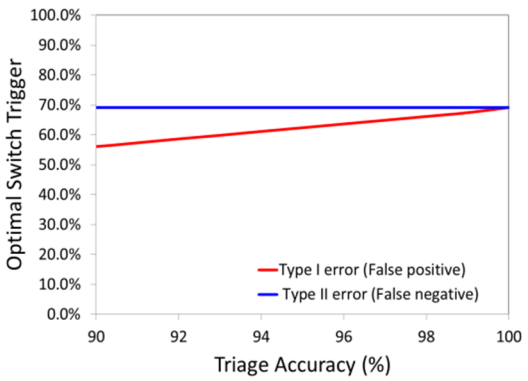

Figure 4.8b. The optimal switch trigger drops from 69.19% to 56.16% when triage accuracy for healthy individuals is reduced by 10% (false positive, red line). It remains at 69.19% independent of any Type II errors committed.

Figure 4.9 contrasts the CDC reported attack rates for the 2009 H1N1 pandemic versus our simulation results. The curves are obtained by varying the values for three parameters: the initial infection rate, the reproduction number, and the switch trigger.

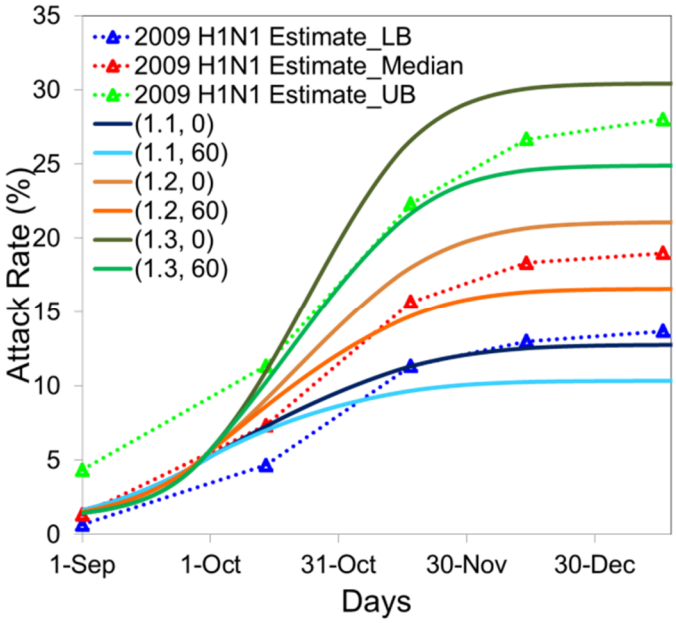

Figure 4.9. Contrast of attack rates for the 2009 H1N1 pandemic from our study versus actual CDC reported data. The solid curves are our results. The dashed curves represent the CDC reported attack rates from the 2009 pandemic. Each line of our results represents a combination of two parameters: the reproduction number, and the switching time. Example: (1.1, 60) refers to R0 = 1.1, with optimal switch trigger (g=52.1%) happens at Day 60 after vaccination begins.

4.1.4 Impact and Significance

In this study, we propose a 'mixed model' approach to quickly determine the optimal 'switch trigger' in a mixed vaccination strategy to minimize both infection and mortality over the course of a pandemic. We do this by interoperably coordinating a disease

propagation model with a vaccine queueing model and an optimization engine. For realism and pragmatic purposes, we integrate vaccination operations from actual events within our models to study overall disease propagation process. We perform our study using census data and hospital information from Georgia and the 2009 H1N1 vaccine supply. This work advances both the scientific and public health policy frontiers.

Mathematical OR Advances

This study offers unique features that have not been previously investigated / incorporated:

- This is the first mathematical-computational model marrying disease propagation with optimization capability that allows rapid determination of the optimal switch triggers – that is, the timing for switching from the prioritized vaccination strategy to the non-prioritized strategy during the course of the pandemic. It is also the first model that mathematically defines "optimal switch trigger" that minimizes the overall attack and mortality rates. Using the computational decision framework, optimal switch triggers can be obtained based on the parameters of the high-risk individuals, the characterization of the pandemic, the vaccine supply availability, and the vaccination rate. The implicit relationship between the optimal switch trigger and the parameters is made explicit and appropriately modeled.
- This is also the first model which incorporates actual vaccination operations and dispensing processes. Mathematical models have been developed to evaluate various vaccination strategies, including early vaccination [67], sequencing the order of the vaccination for the risk groups, and aiming on various objectives to mitigate the influenza [186]. Our study determines the optimal switch trigger, accounting for supply levels and actual dispensing operations. The findings highlight the importance of these factors within the policy making process.
- The ODE+queueing+optimization model are computationally intensive. We derive rapid solution techniques to obtain good near-optimal solutions. Similarly, when replacing the ODE disease propagation and POD queueing model with an individualized agent-based simulation model (ABM), the resulting

ABM+optimization model remains computationally challenging. We derive efficient computational strategies in solving these instances, as well. [98].

- The model is generalizable for other types of infectious diseases. It can also handle seasonality and virus mutation of the biological agents. This information is critical to public health policy makers and decision makers as they determine the best strategies for population protection during pandemics when vaccine supplies are limited and time is critical.

To the best of our knowledge, previous studies on vaccine prioritization did not include all the critical details considered herein in their models, nor were they used by the CDC, or used otherwise, to influence public-health policy making.

Implementation, Public Policy, and Health Impacts

The system, RealOpt-VacOpt, was designed and implemented in Java for portability. It has been in use since 2011 for advising on vaccine distribution, prioritization, and triage strategies. The public health implications are that policy makers can evaluate better trade-offs faster *to save more lives and better utilize limited resources during a pandemic event.* We outline briefly some of these decisions (obtained from our system) adopted by policy makers and their implications below.

- *Optimal switch time to minimize overall attack and mortality rates* CDC confirms that this is the first model to define and determine an actionable and operational switch trigger, an advance that is critical and vital to better mitigation of infections and mass casualties. Given outbreak characteristics, vaccine inventory level, and individual risk factors, there is an optimal switch trigger for prioritizing vaccines that results in lowest overall attack and mortality rates. This optimal strategy is a significant improvement over full prioitized and full non-prioritized strategies. Such information is critical to public health policy makers as they determine the best strategies for population protection, and in particular when to switch from a prioritized strategy focusing on high risk groups to a non-prioritized strategy in which vaccine is made available to the general public. Our model is the first that incorporates virus characteristics, disease spread, and dispensing

operations to optimize vaccine strategies.

The optimal switch trigger is sensitive to the infectivity and vulnerability of the high-risk groups. Specifically, we illustrate the effect to the optimal switch trigger when the vulnerability of the pre-school and school children vary. A full non-prioritized strategy may be preferred if some high-risk individuals are excluded from the prioritized groups (i.e., school children or elderly). Furthermore, we present the optimal switch trigger under different levels of pandemic severity as a function of the early vaccine availability, where the importance of early vaccine availability is emphasized. (Figures 4.5a, 4.5b)

The results reflect that a full prioritized strategy to cover all of the high-risk individuals is not an optimal strategy (Figure 4.3, Table 4.1). When early vaccine supply is limited at a low level, the doses should all be given to the high-risk individuals (Figure 4.6, Table 4.3). But the optimal switch trigger decreases as more vaccine becomes available. And this value converges as the vaccine supply is above 50% (Figure 4.6). Uniformly under various reproductive values, population infection levels, and vaccine supply levels, vaccinating a fraction of the high-risk individuals in the mixed strategy offers the best disease mitigation results compared to a full prioritized or full non-prioritized strategy.

The analysis indicates that timely vaccination and a mixed strategy are critical. Further, delayed vaccination and batched vaccine supply reduce effectiveness and raise overall attack and mortality rates. The more transmissible the virus is, the lower the optimal switch trigger. For very transmissible viruses, the non-prioritized strategy should begin promptly.

- *Improved understanding of vaccine supply and importance of timeliness of vaccination:* The analysis highlights the importance of non-interrupted vaccine supply: although the 2009 H1N1 vaccine supply eventually covered over 30% of the population, the resulting attack and mortality rates are inferior to the scenario when a supply covering only 20% of the population is available without supply interruption. This finding challenges the traditional rolling-out policy where supplies are acquired and shipped in batches. It also highlights the importance of early vaccination: the 2009 H1N1 supply results in 9.9% reduction in overall

infections when compared to a single-batch 10% supply, but the reduction is only marginal (0.9 %) when vaccination is delayed for 3 weeks. These findings are critical to the important policy decisions that public health leaders must make: the acquisition and the timing for vaccination. The system and the results confirmed the decision in the establishment of the National Institute of Translational Medicine for rapid medical countermeasures and drug design.

- *Importance of efficient dispensing operations:* Our study underscores the importance of dispensing throughput efficiency and its effects on the overall attack rates and mortality rates (Figure 4.7). As a result, the CDC Strategic National Stockpile escalated their dispensing training with a team of dedicated CDC public health experts. The CDC team provides regular hands-on RealOpt-POD mass dispensing training to state/local/tribal public health emergency response coordinators to improve operational performance.
- *Understand the effect of triage:* Triage operations have proven to be expensive for mass dispensing. CDC has been undecided regarding triage since their own analysis indicated that addition of triage within a POD can slow down dispensing throughput by over 20%. Our analysis herein informed CDC decision makers that errors made by false negatives pose negligible effect on the overall attack and mortality rates (Figure 4.8a). Removing triage saves resources and improves dispensing throughput, which is critical to improving the overall attack and mortality rates.
- *An objective tool for policy analysis and decision making:* The computational decision framework is generalizable and flexible and can be used for analysis of any type of infectious disease. Such an objective decision framework empowers policy makers to better understand the health implications of their decisions. The resulting policy directly impacts the overall ability to protect population health of this nation and beyond.

Acknowledgement

The authors would like to thank Dr. James Lawler, Dr. Carter Mecher, and Dr. Duane Caneva, former Directors for Medical Preparedness Policy at the White House National Security Staff, for their insightful discussion regarding vaccine dispensing strategies in the United States. The work is partially supported by a grant from the Centers for Disease Control and Prevention and the National Science Foundation.

Appendix: The ODE-queuing disease propagation model

We define the disease progression stage space as Φ = {S(susceptible), E(exposed), P(infectious), A(asymptomatic), I(symptomatic), RC(recovered and immune)}, along with SC = special care, V = vaccinated and immune individuals, and D = deceased. The individual flow between these stages is illustrated in Figure 4.S1. The risk group space Γ = {1, 2, 3, 4, 5} represents the five risk groups: normal adults, children 10 years-old and younger, healthcare workers, pregnant women, and patients with underlying disease conditions. We assume POD has K service stations. To track individuals from each risk group and where these individuals are, we use double indices, i, j, where i runs from Ω = {0, 1, …, K, K+1} and $j \in \Gamma$ to indicate live individuals a) outside the POD (outer-POD, $i = 0$), b) inside the POD (inside-POD, in one of the K service stations, $i = 1, ..., K$), and c) exiting the POD when the vaccine fails (post-POD, $i = K+1$). Individuals who are vaccinated and are immune will move to stage V. Let $\vartheta_{i,j}$ be the number of individuals of group j at station i with disease stage ϑ, where $i \in \{0,1,\ldots,K, K+1\}$, $j \in \Gamma$, and $\vartheta \in \Phi$. Finally, let N_i, $i \in \{0,1,\ldots,K, K+1\}$ denote the number of live individuals (under different disease stages) at each station: outside the POD, inside the POD, and post-POD.

The transmission model is shown in Figure 10. We apply our model to a population of 9,687,653 with specific percentages for each risk group. Table S1 summarizes the input model parameters.

For discussion brevity, we let K = 2, so that station 0 is outer-POD, station 1 is triage station, station 2 is vaccination station, and station 3 is post-POD. The allocation of

nurses and their service times at triage and vaccination listed in Table 4.S1 are the real numbers obtained from time motion study in H1N1 flu campaigns. The outer-POD disease progression is expressed via equations (1-1) to (1-7):

$$\frac{d}{dt}S_{0,j} = -\lambda_j(t)\frac{S_{0,j}}{N_{0,j}} - \beta_j \cdot \frac{(P_0 + A_0 + I_0 + P_3 + A_3 + I_3)}{N_0 + N_3 + V} \cdot S_{0,j} \quad (1\text{-}1)$$

$$\frac{d}{dt}E_{0,j} = \beta_j \cdot \frac{(P_0 + A_0 + I_0 + P_3 + A_3 + I_3)}{N_0 + N_3 + V} \cdot S_{0,j} - \lambda_j(t)\frac{E_{0,j}}{N_{0,j}} - \mu_E E_{0,j} \quad (1\text{-}2)$$

$$\frac{d}{dt}P_{0,j} = \mu_E E_{0,j} - \lambda_j(t)\frac{P_{0,j}}{N_{0,j}} - \mu_P P_{0,j} \quad (1\text{-}3)$$

$$\frac{d}{dt}A_{0,j} = (1-p_S)\mu_P P_{0,j} - \lambda_j(t)\frac{A_{0,j}}{N_{0,j}} - \mu_A A_{0,j} \quad (1\text{-}4)$$

$$\frac{d}{dt}I_{0,j} = p_S\mu_P P_{0,j} - \lambda_j(t)\frac{I_{0,j}}{N_{0,j}} - \mu_I I_{0,j} - p_H\mu_H I_{0,j} \quad (1\text{-}5)$$

$$\frac{d}{dt}RC_{0,j} = (1-m_j)(\mu_A A_{0,j} + \mu_I I_{0,j}) + (1-m\_sc_j)\mu_{SC} SC_j - \lambda_j(t)\frac{RC_{0,j}}{N_{0,j}} \quad (1\text{-}6)$$

$$\frac{d}{dt}SC_{0,j} = p_H\mu_H I_{0,j} - \mu_{SC}\, SC_{0,j} \quad (1\text{-}7)$$

Equations (2-1) to (3-7) represent the rates of change in the population inside the POD:

$$\frac{d}{dt}S_{1,j} = \lambda_j(t)\frac{S_{0,j}}{N_{0,j}} - h_{1,j}(S) - \beta_j \cdot \frac{(P_1 + A_1 + I_1)}{N_1} \cdot S_{1,j} \quad (2\text{-}1)$$

$$\frac{d}{dt}E_{1,j} = \lambda_j(t)\frac{E_{0,j}}{N_{0,j}} + \beta_j \cdot \frac{(P_1 + A_1 + I_1)}{N_1} \cdot S_{1,j} - h_{1,j}(E) - \mu_E E_{1,j} \quad (2\text{-}2)$$

$$\frac{d}{dt}P_{1,j} = \lambda_j(t)\frac{P_{0,j}}{N_{0,j}} + \mu_E E_{1,j} - h_{1,j}(P) - \mu_P P_{1,j} \quad (2\text{-}3)$$

$$\frac{d}{dt}A_{1,j} = \lambda_j(t)\frac{A_{0,j}}{N_{0,j}} + (1-p_S)\mu_P P_{1,j} - h_{1,j}(A) - \mu_A A_{1,j} \quad (2\text{-}4)$$

$$\frac{d}{dt}I_{1,j} = \lambda_j(t)\frac{I_{0,j}}{N_{0,j}} + p_S\mu_P P_{1,j} - h_{1,j}(I) - \mu_I I_{1,j} \quad (2\text{-}5)$$

$$\frac{d}{dt}RC_{1,j} = \lambda_j(t)\frac{RC_{0,j}}{N_{0,j}} + (1-m_j)(\mu_A A_{1,j} + \mu_I I_{1,j}) - h_{1,j}(RC) \quad (2\text{-}6)$$

$$\frac{d}{dt}SC_{1,j} = h_{1,j}(I) - \mu_{SC}\, SC_{1,j} \quad (2\text{-}7)$$

$$\frac{d}{dt}S_{2,j} = h_{1,j}(S) - h_{2,j}(S) - \beta_j \cdot \frac{(P_2 + A_2 + I_2)}{N_2} \cdot S_{2,j} \quad (3\text{-}1)$$

$$\frac{d}{dt}E_{2,j} = h_{1,j}(E) + \beta_j \cdot \frac{(P_2 + A_2 + I_2)}{N_2} \cdot S_{2,j} - h_{2,j}(E) - \mu_E E_{2,j} \quad (3\text{-}2)$$

$$\frac{d}{dt}P_{2,j} = h_{1,j}(P) + \mu_E E_{2,j} - h_{2,j}(P) - \mu_P P_{2,j} \quad (3\text{-}3)$$

$$\frac{d}{dt}A_{2,j} = h_{1,j}(A) + (1-p_S)\mu_P P_{2,j} - h_{2,j}(A) - \mu_A A_{2,j} \quad (3\text{-}4)$$

$$\frac{d}{dt}I_{2,j} = p_S\mu_P P_{2,j} - h_{2,j}(I) - \mu_I I_{2,j} \quad (3\text{-}5)$$

$$\frac{d}{dt}RC_{2,j} = h_{1,j}(RC) + (1-m_j)(\mu_A A_{2,j} + \mu_I I_{2,j}) - h_{2,j}(RC) \quad (3\text{-}6)$$

$$\frac{d}{dt}SC_{2,j} = 0 \quad (3\text{-}7)$$

where service rate $h_{i,j}$ is given by

$$h_{i,j}(\vartheta) = ST_i^{-1} \cdot min(n_i, N_i)\frac{\vartheta_{i,j}}{N_i} \quad (4)$$

and represent the real service rate for the individuals staying at transmission stage ϑ from group j at POD station i. Equations (5-1) to (5-7) represent the rates of change in the population who go back to the community (outside of the POD) without immunity (that is, they received vaccine but did not become immune):

$$\frac{d}{dt}S_{3,j} = (1-p_V)h_{2,j}(S) - \beta_j\frac{(P_0+A_0+I_0+P_3+A_3+I_3)}{N_0+N_3+V}S_{3,j} \quad (5\text{-}1)$$

$$\frac{d}{dt}E_{3,j} = h_{2,j}(E) + \beta_j\frac{(P_0+A_0+I_0+P_3+A_3+I_3)}{N_0+N_3+V}S_{3,j} - \mu_E E_{3,j} \quad (5\text{-}2)$$

$$\frac{d}{dt}P_{3,j} = h_{2,j}(P) + \mu_E E_{3,j} - \mu_P P_{3,j} \quad (5\text{-}3)$$

$$\frac{d}{dt}A_{3,j} = h_{2,j}(A) + (1-p_S)\mu_P P_{3,j} - \mu_A A_{3,j} \quad (5\text{-}4)$$

$$\frac{d}{dt}I_{3,j} = h_{2,j}(I) + p_S\mu_P P_{3,j} - \mu_I I_{3,j} - p_H\mu_H I_{3,j} \quad (5\text{-}5)$$

$$\frac{d}{dt}RC_{3,j} = h_{2,j}(RC) + (1-m_j)(\mu_A A_{3,j} + \mu_I I_{3,j}) + (1-m\_sc_j)\mu_{SC}\sum_{i=1}^{3}SC_{i,j} \quad (5\text{-}6)$$

$$\frac{d}{dt}SC_{3,j} = p_H\mu_H I_{3,j} - \mu_{SC}\,SC_{3,j} \quad (5\text{-}7)$$

Equations (6) and (7) represent respectively two outcomes for the transmission flow: deceased and vaccinated.

$$\frac{d}{dt}D_j = m_j\sum_{i=0}^{3}(\mu_A A_{i,j} + \mu_I I_{i,j}) + m\_sc_j\sum_{i=0}^{3}(\mu_{SC}SC_{i,j}) \quad (6)$$

$$\frac{d}{dt}V_j = p_V h_{2,j}(S) \quad (7)$$

Equation (8) represents the infection flow, which relates to the objective function of the optimization problem. Equation (9) cumulates the vaccine consumption, which is used to keep the model feasible.

$$\frac{d}{dt}AR_j = \beta_j \cdot \frac{P_0 + A_0 + I_0 + P_3 + A_3 + I_3}{N_0 + N_3 + V}(S_{0,j} + S_{3,j}) + \beta_j \frac{P_1 + A_1 + I_1}{N_1} S_{1,j} + \beta_j \frac{P_2 + A_2 + I_2}{N_2} S_{2,j} \quad (8)$$

$$\frac{d}{dt}VE_j = \sum_{\vartheta \in \Phi} h_{2,j}(\vartheta) \quad (9)$$

Thus, everyone who goes to the POD gets vaccinated (if they are in the proper risk group) except those being triaged as infected, who will be sent for treatment in the hospitals or for special care.

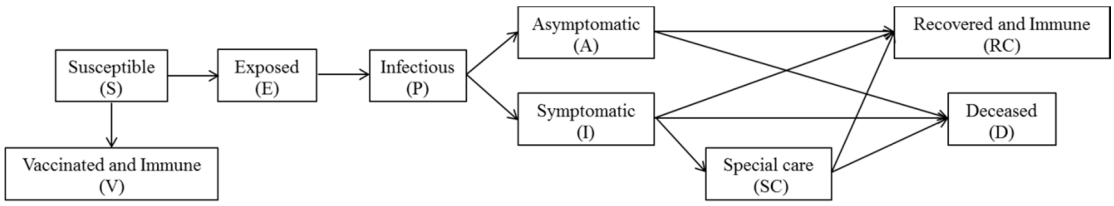

Figure 4.S1. This figure shows the diseases progression stages. Each stage is further stratified by the risk groups.

Table 4.S1. Model parameter description.

Parameter	Description	Base case value (at time t=0)
Ω	Station space	Ω = {0, 1, …, K, K+1}, where 0 is outer-POD, 1, …, K are service stations inside the POD. Station K is the vaccination station, station K+1 represents post-POD.
Φ	Disease progression stage space	Φ = {S(susceptible), E(exposed), P(infectious), A(asymptomatic), I(symptomatic), RC(recovered and immune)}.
Γ	Risk group space	{1, 2, 3, 4, 5} ({normal, children 10 or younger, healthcare workers, pregnant women, patients with underlying disease conditions})
Π	Risk group population distribution	{0.807, 0.148, 0.03, 0.013, 0.002}. [26]
N	Total population of the region	
N'	High-risk population	$N' = (1 - \pi_1) N$, where $\pi_1 = 0.807$.
V_j	Number of vaccinated and immune individuals in group j.	At time $t = 0$, $V_j = 0$.
V	Total number of vaccinated and immune individuals	$= \sum_{j \in \Gamma} V_j$.
SC	Special care	At time $t = 0$, $SC_j = 0$.
D_j	Number of deaths in group j.	At time $t = 0$, $D_j = 0$.

Table 4.S1. (continued)

N_i	Number of non-special care live individuals at station i	$N_i = S_i + E_i + P_i + A_i + I_i + RC_i, \ \forall i \in \{0,1,...,K, K+1\}$ At time $t = 0$, $N_0 = N$, $N_1 = ... = N_K = N_{K+1} = 0$.
$N_{i,j}$	Number of non-special care live individuals of group j at station i	$N_{i,j} = S_{i,j} + E_{i,j} + P_{i,j} + A_{i,j} + I_{i,j} + RC_{i,j}, \ \forall i \in \{0,1,...,K, K+1\}, j$ $\sum_{j \in \Gamma} N_{i,j} = N_i$.
α	Initial infection percentage	1.0%, range 0.5% - 2.0%
AR	Attack rate	At time $t = 0$, $AR = \alpha N$.
AR_j	Attack rate in group j.	$\sum_{j \in \Gamma} AR_j = AR$. At time $t = 0$, $AR_j = \pi_j \cdot AR$.
η	Relative risk factor for each risk group	{1, 2, 1, 3, 3}.
$1/\mu_E$	Mean exposed duration	1 day
$1/\mu_P$	Mean pre-symptomatic duration	1 day
$1/\mu_A$	Mean asymptomatic duration	3 days
$1/\mu_I$	Mean symptomatic duration	3 days
$1/\mu_H$	Mean duration before special care from showing symptom	2 days
$1/\mu_{SC}$	Mean special care duration	5 days
p_S	Symptomatic probability	2/3
β_j	Baseline infectivity rate for group j	$= \eta_j \cdot R_0 \frac{1}{1/\mu_P + (1-p_S)/\mu_A + p_S \cdot 1/\mu_I}$
p_V	Vaccine effectiveness	90%
p_H	Percentage of symptomatic patients being hospitalized	20%
m_j	Mortality rate for group j	0.12%
m_sc_j	Mortality rate for group j after special care (or hospitalization)	$m_j / 2$
ϑ_i	Number of individuals at station i with disease stage ϑ	At time $t = 0$, $S_0 = (1-\alpha)N$, $E_0 = \alpha N \cdot \Delta \cdot 1/\mu_E$, $P_0 = \alpha N \cdot \Delta \cdot 1/\mu_P$, $A_0 = \alpha N \cdot \Delta \cdot (1-p_S) \cdot 1/\mu_A$, $I_0 = \alpha N \cdot \Delta \cdot p_S \cdot 1/\mu_I$, $V_0 = 0$, $SC_0 = 0$, $D_0 = 0$, $RC_0 = 0$, where $\Delta = (1/\mu_E + 1/\mu_P + (1-p_S) \cdot 1/\mu_A + p_S \cdot 1/\mu_I)^{-1}$, $\vartheta_i = 0$, for $i \neq 0$.

Table 4.S1. (continued)

$\vartheta_{i,j}$	Number of individuals of group j at station i with disease stage ϑ	$\vartheta_{i,j} \geq 0$, $i \in \{0, 1, \ldots, K, K+1\}$, $j \in \Gamma$, and $\vartheta \in \Phi$. $\sum_{j=1}^{J} \vartheta_{i,j} = \vartheta_i$. At time $t = 0$, $\vartheta_{i,j} = \pi_j \cdot \vartheta_i$, for $i \in \{0, 1, \ldots, K, K+1\}$, $j \in \Gamma$.
VE_j	Number of vaccinees in group j.	At time $t = 0$, $VE_j = 0$.
$VS(t)$	Cumulative vaccine supply	For 2009 H1N1 vaccine supply $VS(t) = \begin{cases} 0.1 \cdot N & 0 \leq t < 45\,days \\ 0.1667 \cdot N & 45\,days \leq t < 60\,days \\ 0.2067 \cdot N & 60\,days \leq t < 75\,days \\ 0.3067 \cdot N & t \geq 75\,days. \end{cases}$
$\lambda(t)$	Arrival rate at the POD	$\lambda(t) = \begin{cases} \lambda_0 N_0(t)/N & \text{if non-prioritized strategy is used} \\ \lambda_0[\sum_{j \neq 1, j \in J} N_{0,j}(t)]/N' & \text{if prioritized strategy is used,} \end{cases}$ where N_0 is the number of non-vaccinated live individuals outside the POD, N is the total population, and $N_{0,j}$ denotes the number of non-vaccinated live individuals of group j outside the POD. λ_0 is the initial arrival rate, $\lambda_0 = 12.018$ / min at time 0.
$\rho_j(t)$	Arrival allocation for high-risk groups	$= \begin{cases} \rho_j(t) = N_{0,j}(t) / \sum_{j \in \Gamma} N_{0,j}(t) & \text{if non-prioritized strategy is being used,} \\ \rho_1(t) = 0, & \\ \rho_j(t) = N_{0,j}(t) / \sum_{j \neq 1, j \in \Gamma} N_{0,j}(t) \ \forall j \neq 1 & \end{cases}$ $\sum_{j \in \Gamma} \rho_j(t) = 1.$ if prioritized strategy is being used.
$\lambda_j(t)$	Arrival rate for group j	$= \lambda(t)\, \rho_j(t)$
n_i	Number of nurses at station i	
ST_i	Service time at station i	

4.2 Dispensing Medical Countermeasures in Response to an Anthrax Attack

4.2.1 Introduction

Once a large-scale anthrax attack happens, people may die by the end of 48 hours if they do not receive medical countermeasures against anthrax. It is crucial that all individuals living in the metropolitan area where the attack occurs receive medications within the 48 hour timeframe. An individual needs to take 2 pills of ciprofloxacin or doxycycline per day for 60 days until the epidemic has terminated to be kept from infection. In addition to the medications, individuals must be registered with the government and be instructed by physicians in a face-to-face meeting in order to ensure the he/she is fully protected.

Strategies for dispensing medical countermeasures against a large-scale anthrax attack, including postal delivery [6] and mass dispensing by Point-of-Dispensing (POD) [5, 12, 22, 31, 112, 192, 193] have been examined by numerous researchers. Postal delivery is to deliver the medical countermeasures by USPS postal carriers to each household in the city. Mass dispensing at POD is to let the population pick up the medical countermeasures at PODs, which are set up at public places in the city. However, these two strategies are not flawless: Postal carriers can only deliver a fixed amount of pills to each household rather than providing the enture prescriptive dose to each individual; to dispense the medical countermeasures in PODs within 48 hours may increase the risk of infection due to the congestion in the PODs. We herein consider a postal delivery and POD refill strategy, named Postal+Refill, to protect the entire population by delivering initial medications by postal carriers within the first 24 hours and refilling medications at PODs in the following days. People who finish their initial medications need to pick up sufficient medications at PODs such that they are kept from infection till the end of the epidemic. This strategy takes the advantage of both postal delivery and mass dispensing at POD: postal delivery provides fast temporary prevention to the entire population; and mass dispensing ensures people are instructed by the physicians and receive sufficient medications.

4.2.2 Model Description

In a Postal+Refill strategy, the initial medical countermeasures must be dispensed by postal delivery to the entire affected population in the first 24 hours after the attack. The medications are prepared by the public health departments within the first 12 hours, and then postal carriers deliver medication to each household in the following 12 hours. A bottle of 20 pills is delivered to each household, which provides temporary protection from anthrax. The length of the protection depends on the size of a household, i.e., 20 pills provide 6-day prevention to a family of 3, or 5-day prevention to a family of 4.

Households are required to refill their medications before they finish their temporary medications. Every individual must be registered with the government and be instructed by physicians face-to-face in order to make sure he/she receives enough medications and understand its full usage. During the postal delivery of the initial therapy, mass dispensing events in PODs are set up and organized in the city. POD locations are selected from a candidate list that consists of schools, gyms, clinics, and other public common areas. The employees in these candidate locations are pre-trained by CDC such that the locations are well prepared. Starting from the second day which is 24 hours after the attack, the PODs are open to public and dispense the sufficient medications ($\geq$ 60 pills per person per instruction of physicians) to the population.

To plan a Postal+Refill strategy in the city, public health emergency planners must determine within the first 12 hours:

1. Number of postal carriers and police force required,
2. The POD locations and the covered population of each location, and
3. The staff resource required at each POD and their responsibilities.

In our analysis, four metropolitan areas are considered in evaluating the cost and the staff resource requirement of the Postal+Refill strategy: Washington DC, New York City, City of Chicago, and City of Los Angeles as shown in Figure 4.10.

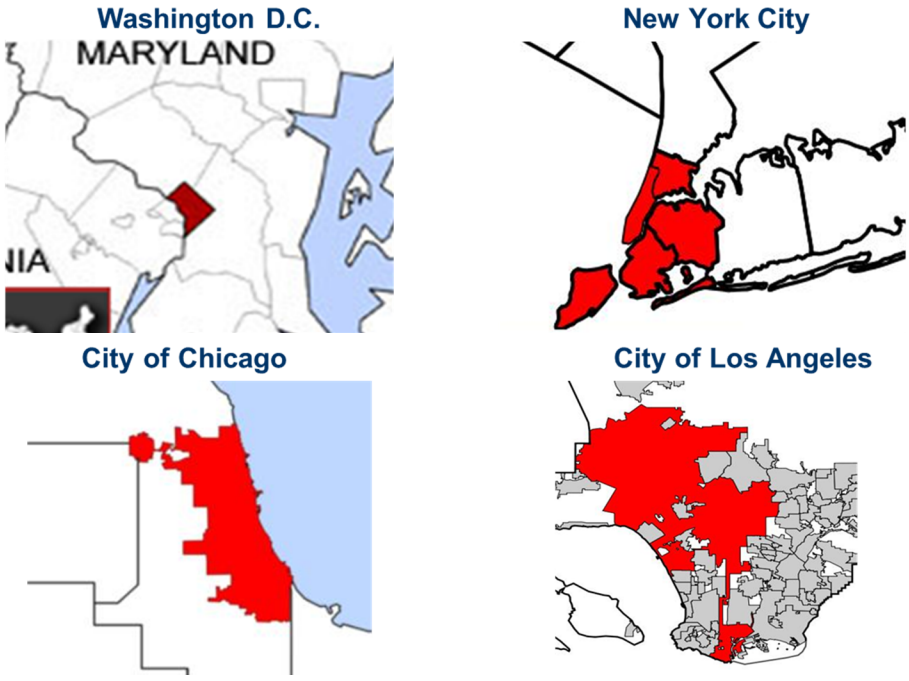

Figure 4.10. Metropolitan area of the four metropolitan areas

In postal delivery, USPS postal carriers follow the postal carrier routes to deliver one bottle of pills (20 pills) to every household in the city regardless of the household size. Postal delivery routes are the regular mailing routes that go through the city and serve each of the households along the route. To estimate the required staff and budget for postal delivery, we make the following assumptions. First, each of the carriers covers two regular postal carrier routes in a zip-code area within 12 hours. Extra carriers are required at a higher cost (150% of the regular cost) if the current workforce cannot cover all of the postal carrier routes in the city. Second, additional routes are used to deliver the pills to the population who do not reside at a permanent address in the city. Third, one police officer is assigned to each carrier for security, and 20% additional security officers for consolidating the delivery units in the city. The estimation of the required work forces and the operation cost is listed in Table 4.4.

Table 4.4. Required work forces in Day 1 postal delivery

Metropolitan Area	Washington DC	New York City	City of Chicago	City of Los Angeles
Population	599,657	8,363,710	2,853,114	3,833,995
Number of households	233,330	3,254,362	1,110,161	1,491,827
Number of USPS carriers in the city	531	3067	1755	1108
Ratio of postal carriers to households (10^{-3})	1.51	0.85	1.14	1.02
Number of postal carriers required	352	2,762	1,268	1,936
Number of postal carriers on regular routes	320	2200	1018	1428
Number of postal carriers on additional routes	30	562	250	94
Number of additional postal carriers required	0	0	0	414
Number of security officers required	422	3,314	1,522	1,826
Unit cost per postal carrier (per 12-hour shift, Estimated by hourly salary + rough estimate of fuel cost	320.2	330	326.1	335.15
Unit cost per security officer	372	408	418.8	494.4
Total labor cost	243,654	2,038,356	944,533	1,606,009

Starting from Day 2 (the second day after the anthrax attack), PODs are open to the public so that households can register and pick up the remaining pills that will provide protection from anthrax. To determine the locations of PODs and the covered population for each POD, we build a mixed integer programming model in RealOpt-Regional© [98] that minimizes the number of PODs in the city and subjects to the travel distance to the POD less than or equal to 15 miles. Households are required to pick up the refilled pills in the POD on any day no later than the refill due day. Given the initial medications of 20 pills, the refill due day of a household depends on the household size (Table 4.5). PODs are designed to open 12 hours a day until Day 10 (20 pills prevent a family of 1 person for 10 days). For simplicity, we assume that a household will arrive on any day between Day 2 and the refill due day with equal probability. Thus, the

designed throughput at PODs on any open day must satisfy the expected arrivals (percentage of population) on this day as shown in Table 4.5.

In a general dispensing process, households go through the following steps: arrive, register and file forms, get triaged by healthcare workers, pick up pills from physicians, and leave. The process at each city can be slightly different (Figure 4.11) according to the feedbacks from the previous POD operational drills. The distribution of the processing time at each step is estimated using CDC time motion studies in full-scale drills or actual dispensing operations. Six types of staff are recruited to operate a POD: volunteer, nurse/MD, health technician, security, traffic controller, and interpreter. Their responsibilities are different as shown in Table 4.6. For some major ethic groups, e.g., Hispanic population or Chinese population, interpreters are required at each step when over 5% of the arrivals speak Spanish/Chinese rather than English.

To avoid the increased risk of infection, a household is designed to stay in the POD with no more than 30 minutes including all waiting times and processing times. The minimum number of worker-shift of each type of staff at a POD that can serve the expected arrivals and the maximum duration of each household is determined by a simulation optimization model in RealOpt-POD© [44,42,98] using the POD layout used by the city. The required numbers of worker shifts of all cities are shown in Table 4.7.

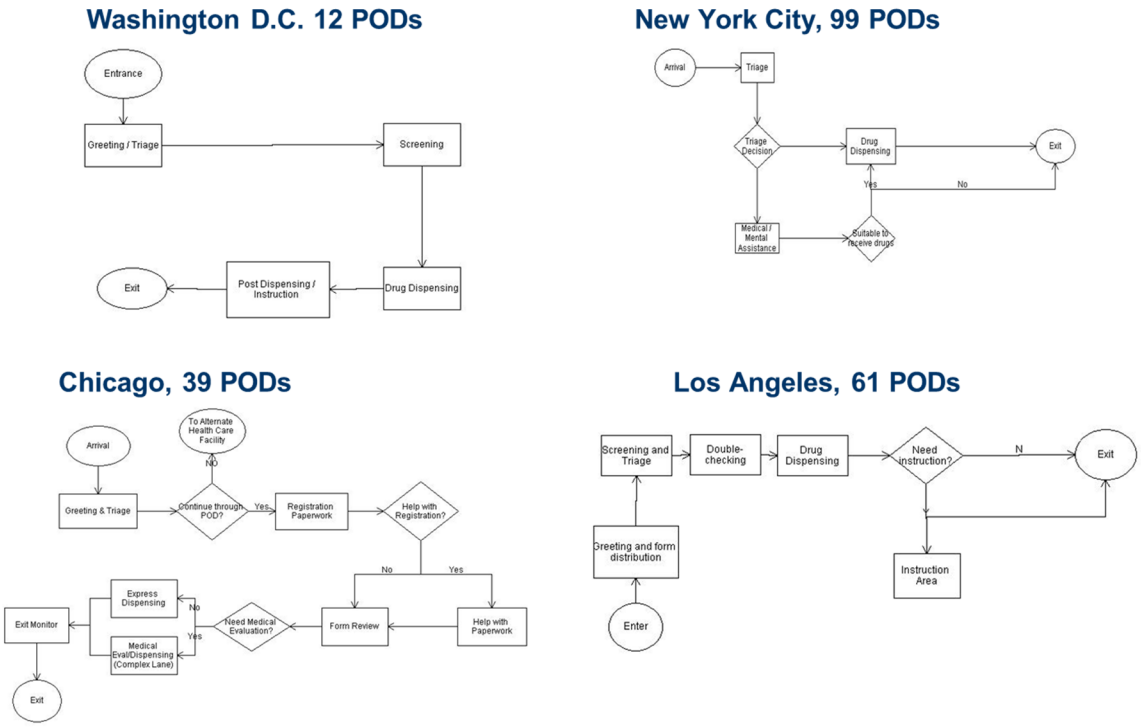

Figure 4.11. Layout of the PODs in the four cities.

Table 4.5. The required throughput at PODs (% of population).

Household size	Refill due	Day									
		2	3	4	5	6	7	8	9	10	Total
4+ (10%)	Day 2	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%
3 (55%)	Day 3	50%	50%	0%	0%	0%	0%	0%	0%	0%	100%
2 (30%)	Day 5	25%	25%	25%	25%	0%	0%	0%	0%	0%	100%
1 (5%)	Day 10	11%	11%	11%	11%	11%	11%	11%	11%	11%	100%
Required throughput:		45.6%	35.6%	8.1%	8.1%	0.5%	0.5%	0.5%	0.5%	0.5%	100%

Table 4.6. Staff responsibilities

	Greeting and form filing	Triage	Drug dispensing	Security	Traffic Control
Volunteer*	Y				
Nurse/MD	Y	Y	Y		
Health technician	Y				
Security	Y			Y	
Traffic controller	Y				Y
Interpreters	Y	Y	Y	Y	Y

*: no volunteer recruited in New York City

Table 4.7. Worker shifts required for Postal+Refill strategy

		Washington DC	New York City	City of Chicago	City of Los Angeles
Number of PODs		9	99	39	61
Work force required	Volunteer	432	0	2,188	3,838
	Nurse/MD	621	5,188	2,456	3,390
	Health Technician	472	3,148	2,432	2,664
	Security	532	5,898	2,000	2,612
	Traffic Control	540	5,346	3,861	3,843
	Interpreters	344	4,295	750	995
	Total	2,941	23,875	13,687	17,342

4.2.3 Cost-effective Analysis

We compare the Postal+Refill strategy to another strategy -- the 36-hour mass dispensing strategy which has been adopted by many cities. 36-hour mass dispensing strategy prepares PODs and medical countermeasures in the first 12 hours and dispenses the medications in the PODs in the following 36 hours so that the population is protected within 48 hours after the anthrax attack. The two strategies are evaluated by the estimated required work force and the total cost including labor cost and operation cost.

The Postal+Refill strategy operates 9 days (from Day 2 to Day 10), while the 36-hour mass dispensing strategy operates three consecutive 12-hour shifts and the healthcare workers cannot cover two consecutive shifts. Postal+Refill requires more worker shifts compared to the 36-hour POD mass dispensing in the scenarios of three cities with the exception of the City of Chicago as shown in Table 4.8a. However, in 36-hour mass dispensing strategy, the healthcare workers cannot cover two consecutive shifts, the number of workers required to participate in training is more than those in Postal+Refill strategy in the scenarios of three cities except Washington DC as shown in Table 4.8b.

The cost of each strategy consists of two parts: training cost and operation cost. Training cost is the cost for setting up drills and training healthcare workers for learning and practicing their responsibilities during the real event, while operation cost is the cost that occurs during the real event. We assume each POD must complete 3 days training. The labor cost during an event can be estimated by the average hourly salary in the city for a particular position (including vehicle cost for postal carriers) as shown in Table 4.9. The site cost spent in an event, such as power, water, and rent, is difficult to measure, and can range from $0 to $10,000 depending on the location and the size of the event. When a default daily operation cost per POD is assumed to be $5,000, the total cost of Postal+Refill strategy are lower than that of the 36-hour mass dispensing strategy in the three cities with the exception of Washington DC as shown in Figure 4.12, Figure 4.13, and Table 4.10. The 36-hour mass dispensing strategy costs more in training because it uses more PODs and requires more workers. When the daily POD operation cost ranges

from $0 to $10,000, the Postal+Refill strategy always costs more than the 36-hour mass dispensing strategy in Washington DC, while it always costs more in City of Chicago and City of Los Angeles (Figure 4.14). In New York City, the Postal+Refill strategy is the cost-effective strategy when the daily POD operation cost is less than $8,751.

Table 4.8a. Number of worker shifts

	Washington DC	New York City	City of Chicago	City of Los Angeles
Postal+Refill	**3,293**	**26,636**	**14,954**	**19,277**
postal carriers	352	2,762	1,268	1,936
healthcare workers	2,941	23,874	13,686	17,341
36-hour mass dispensing	**1,929**	**23,763**	**13,389**	**19,530**

Table 4.8b. Number of workers.

	Washington DC	New York City	City of Chicago	City of Los Angeles
Postal+Refill	**1,692**	**13,638**	**7,503**	**9,836**
postal carriers	352	2,762	1,268	1,936
healthcare workers	1,340	10,876	6,235	7,900
36-hour mass dispensing	**1,286**	**15,842**	**8,926**	**13,020**

Table 4.9. Labor cost ($ per shift)

	Washington DC	New York City	City of Chicago	City of Los Angeles
Postal carrier	320.2	330	326.1	335.15
Police officer	372	408	418.8	494.4
Volunteer	0	0	0	0
Nurse/MD	363.6	427.2	256.8	405.6
Health Technician	232.92	279.6	249.6	247.2
Security	372	408	418.8	494.4
Traffic Control	218.4	168	145.2	157.2
Interpreters	540	351.6	230.4	321.6

Table 4.10. Cost of the two strategies

	Washington DC	New York City	City of Chicago	City of Los Angeles
Postal+Refill	2,945,205	26,700,041	9,931,429	15,321,289
Postal delivery	243,654	2,038,356	944,533	1,606,009
POD refill cost	2,701,551	24,661,685	8,986,896	13,715,280
training cost	1,324,217	12,295,557	4,423,150	6,721,271
operation cost	1,377,334	12,366,128	4,563,745	6,994,009
number of PODs	12	99	39	61
36-hour mass dispensing	2,253,130	28,928,318	11,182,939	19,827,174
training cost	1,442,087	18,790,546	7,260,293	12,913,116
operation cost	811,043	10,137,773	3,922,646	6,914,058
number of PODs	12	115	47	73

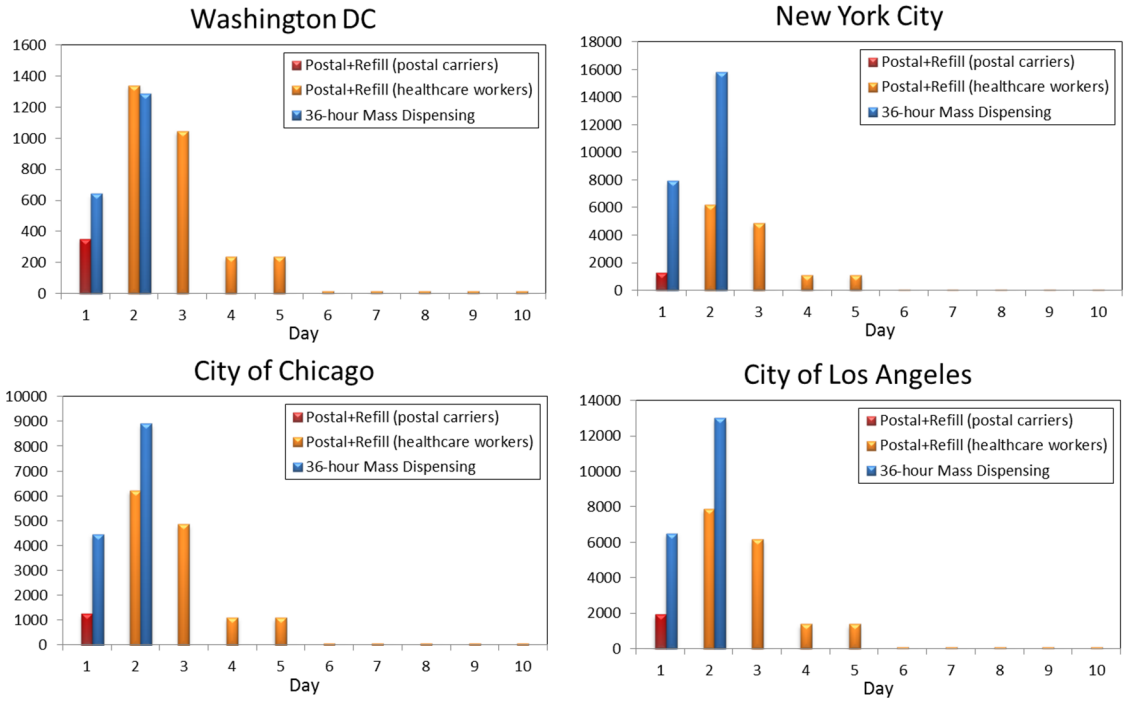

Figure 4.12. Worker allocation in 4 metropolitan areas under Postal+Refill strategy and 36-hour mass dispensing strategy.

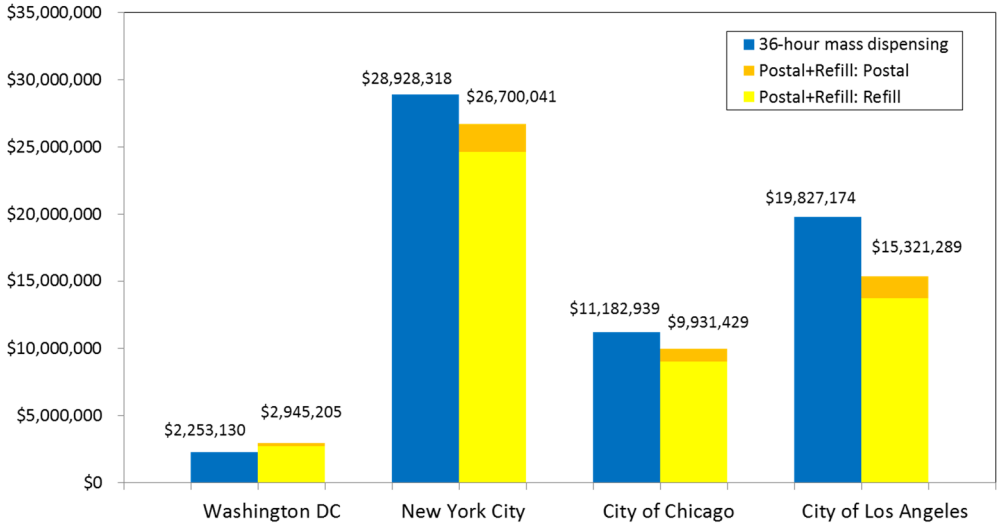

Figure 4.13. Cost comparison of the Postal+Refill and the 36-hour mass dispensing.

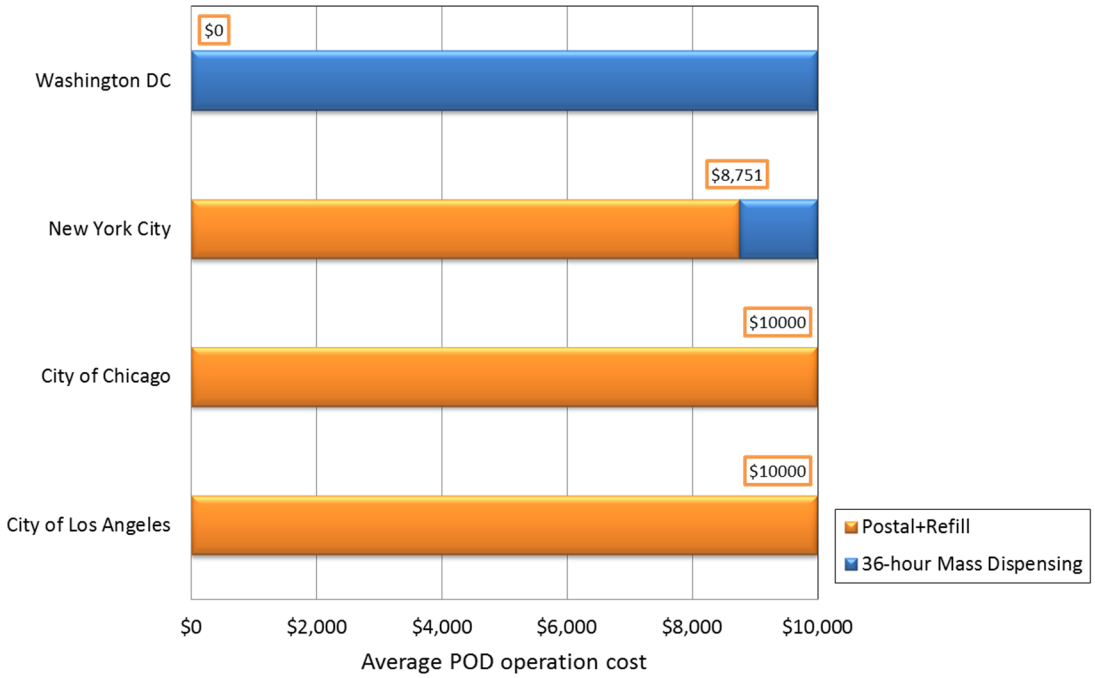

Figure 4.14. The cost-effective strategy when the POD operation cost varies. Blue: 36-hour mass dispensing costs less. Orange: Postal+Refill costs less.

4.2.4 Conclusion

In summary, the Postal+Refill strategy delivers the initial medications to the population by postal carriers within 24 hours and dispenses the remaining pills at PODs in the following days. It allows extra time to set up the PODs for pill refill by providing temporary protection to the entire population by postal delivery, and reduces the infection risk of congestions in the PODs by extending the dispensing process from less than 2 days to 9 days. As shown in our analysis, this strategy is more effective in the cities with high population density in terms of the required work forces and the cost. Compared to the mass dispensing strategy, it benefits from the reduced work load for training the PODs and workers for dealing with the potential congestions in the first two days. However, security concerns may be an issue since postal workers could become target of hoax or opportunistic criminals. Citizens may not take the pills if there is rumored or reported illness due to individuals taken tempered pills.

REFERENCES

[1]. Abad, P. and Banks, W., "New LP based heuristics for the classification problem," *European Journal of Operational Research*, vol. 67, pp. 88-100, 1993.

[2]. Agrafiotis, D. and Cedeno, W., "Feature selection for structure-activity correlation using binary particle swarms," *Journal of Medicinal Chemistry*, vol. 45, no. 5, pp. 10981107, 2002.

[3]. Alessandrini, E.A., Lavelle, J.M., Grenfell, S. M., Jacobstein, C. R., and Shaw, K. N., "Return Visits to a Pediatric Emergency Department," *Pediatric Emergency* Care, vol. 20, no. 3, pp. 166-171, 2004.

[4]. Alterovitz, R., Lessard, E., Pouliot, J., Hsu, I.C., O'Brien, J.F., and Goldberg, K., "Optimization of HDR brachytherapy dose distributions using linear programming with penalty costs," *Medical Phys*, vol. 33, no. 11, pp4012–4019, 2006.

[5]. Altevogt, B. M., Kammersall, M. S., and Davis, M., "Dispensing Medical Countermeasures for Public Health Emergencies: Workshop Summary," *National Academies Press*, 2008.

[6]. Altevogt, B.M., Nadigm, L., and Stroud, C., "Medical Countermeasures Dispensing: Emergency Use Authorization and the Postal Model: Workshop Summary," *National Academies Press*, 2010.

[7]. American Cancer Society (2012) Cancer Facts & Figures 2012 Accessed July 1, 2013, http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-031941.pdf.

[8]. American Joint Committee on Cancer (2009) Cervix uteri cancer staging. Accessed July 1, 2013, http://www.cancerstaging.org/staging/posters/cervix8.5x11.pdf.

[9]. Anderson, J., "Constrained discrimination between k populations," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 31, no. 1, pp. 123-139, 1969.

[10]. Asparoukhov, O. and Stam, A., "Mathematical programming formulations for two-group classification with binary variables," *Annals of Operations Research*, vol. 74, pp. 89-112, 1997.

[11]. Aujesky, D., Maria, K., et al., "Predictors of early hospital readmission after acute pulmonary embolism," *Arch Intern Med.*, vol. 169, no. 3, pp. 287-293, 2009.

[12]. Baccam, P., and Boechler, M., "Public health response to an anthrax attack: an evaluation of vaccination policy options," *Biosecurity and Bioterrorism*, vol. 5, no. 1, 26-34 2007.

[13]. Bailey, D.L., Townsend, D.W., Valk, P.E., Maisey, M.N., "Positron Emission Tomography: Basic Sciences," Springer-Verlag, Secaucus, NJ, 2003.

[14]. Bajgier, S. and Hill, A., "An experimental comparison of statistical and linear programming approaches to the discriminant problem," *Decision Sciences*, vol. 13, pp. 604-618, 1982.

[15]. Banks, W. and Abad, P., "An efficient optimal solution algorithm for the classification problem," *Decision Sciences*, vol. 22, pp. 1008-1023, 1991.

[16]. Beliën, J., Colpaert, J., De Boeck, L., and Demeulemeester, E., "A hybrid simulated annealing linear programming approach for treatment planning in HDR

brachytherapy with dose volume constraints," *In Proceedings of the 35th International Conference on Operational Research Applied to Health Services*, 2009.

[17]. Benbassat, J., Taragin, M., "Hospital readmissions as a measure of quality of health care," *Arch Intern Med*, vol. 160, pp. 1074-1081, 2000.

[18]. Bennett, K. and Mangasarian, O., "Multicategory discrimination via linear programming," *Optimization Methods and Software*, vol. 3, pp. 27-39, 1994.

[19]. Bennett, K. and Mangasarian, O., "Robust linear programming discrimination of two linearly inseparable sets," *Optimization Methods and Software*, vol. 1, pp. 23-34, 1992.

[20]. Bertolazzi, P., Felici, G., Festa, P., and Lancia, G., "Logic classification and feature selection for biomedical data," *Computers and Mathematics with Applications*, vol. 55, no. 5, pp. 889-899, 2008.

[21]. Bradley, P. and Mangasarian, O., "Massive data discrimination via linear support vector machines," *Optimization Methods and Software*, vol. 13, no. 1, pp. 1-10, 2000.

[22]. Bravata, D. M., Zaric, G. S., Holty, J. E. C., Brandeau, M. L., Wilhelm, E. R., McDonald, K. M., and Owens, D. K., "Reducing mortality from anthrax bioterrorism: strategies for stockpiling and dispensing medical and pharmaceutical supplies," *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, vol. 4, no. 3, pp. 244-262, 2006.

[23]. Brooks, J.P., Lee, E.K., "Analysis of the consistency of a mixed integer programming-based multi-category constrained discriminant model," *Annals of Operations Research on Data Mining*, vol. 174, no. 1, pp. 147-168, 2010.

[24]. Brooks, J., "Solving a mixed-integer programming formulation of a classification model with misclassification limits," PhD thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 2005.

[25]. Bureau of the Census UDoC. Census 2010. Available: http://www.census.gov/ (Accessed Oct 15, 2012).

[26]. Centers for Disease Control and Prevention (2009) H1N1 Flu. Available: http://www.cdc.gov/h1n1flu/.

[27]. Ciernik, I.F., Dizendorf, E., Baumert, B.G., Reiner, B., Burger, C., Davis, J.B., Lütolf, U.M., Steinert, H.C., Von Schulthess, G.K., "Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): A feasibility study," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 57, no. 3, pp. 853–863, 2003.

[28]. Conway, J.M., Tuite, A.R., "Vaccination against 2009 pandemic H1N1 in a population dynamical model of Vancouver, Canada: timing is everything," *BMC Public Health*, vol. 11, no. 1, pp. 932, 2011.

[29]. Correa, E., Freitas, A., and Johnson, C., "A new discrete particle swarm algorithm applied to attribute selection in a bioinformatics data set," in *Genetic and Evolutionary Computation Conference*, (New York, NY), pp. 3542, ACM, 2006.

[30]. Craft, D., Halabi, T., and Bortfeld, T. "Exploration of tradeoffs in intensity-modulated radiotherapy," *Phys. Med. Biol.*, vol. 50, pp. 5857-68, 2005.

[31]. Craft, D. L., Wein, L. M., and Wilkins, A. H., "Analyzing bioterror response logistics: The case of anthrax," *Management Science*, vol. 51, no. 5, pp. 679-694, 2005.

[32]. Craft, D., et al. "An approach for practical multiobjective IMRT treatment planning," *International journal of radiation oncology* Biology* Physics*, vol. 69, no. 5, pp. 1600-1607, 2007.

[33]. Dash, M., and Liu, H., "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, pp. 131-156, 1997.

[34]. Deasy, J. O., "Multiple local minima in radiotherapy optimization problems with dose–volume constraints," *Medical physics*, vol. 24, no. 7, pp. 1157-1161, 1997.

[35]. Diekmann, O., Heesterbeek, J.A.P., "Mathematical Epidemiology of Infectious Disease Model Building, Analysis and Interpretation," Wiley, Chichester, 2000.

[36]. Duan, J., Kim, R.Y., Elassal, S., Lin, H.Y., Shen, S., "Conventional high-dose-rate brachytherapy with concomitant complementary IMRT boost: A novel approach for improving cervical tumor dose coverage," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 71, no. 3, pp. 765-771, 2008.

[37]. Dushoff, J., et al., "Vaccinating to protect a vulnerable subpopulation," *PLoS Med*, vol.4, pp. e174, 2007.

[38]. Easton, T., Hooker, K., Lee, E.K., "Facets of the independent set polytope," *Math. Programming Series B*, vol. 98, pp. 177–199, 2003.

[39]. Eastwood, K., Durrheim, D.N., Jones, A., and Butler, M., "Acceptance of pandemic (H1N1) 2009 influenza vaccination by the Australian public," *MJA*, Volume 192, 2010.

[40]. Ebert, M.A., and Hoban, P.W., "Some characteristics of tumour control probability for heterogeneous tumours," *Physics in medicine and biology*, vol. 41, no. 10, pp. 2125, 1996.

[41]. Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407-499, 2004.

[42]. Lee, E.K., Chen, C.H., Pietz, F., Benecke, B, "Modeling and optimizing the public health infrastructure for emergency response, " *Interfaces - The Daniel H. Wagner Prize for Excellence in Operations Research Practice*, vol. 39, no. 5, pp. 476-490, 2009.

[43]. Lee, E.K., Yuan, F., Templeton, A., Yao, R., Kiel, K., Chu, J., "Biological planning for high-dose rate brachytherapy: Application to cervical cancer treatment," *Interfaces - The Daniel H. Wagner Prize for Excellence in Operations Research Practice*, vol. 43, no. 5, pp. 462-47, 2013.

[44]. Lee, E.K., Maheshwary, S., Mason, J., Glisson, W., "Decision support system for mass dispensing of medications for infectious disease outbreaks and bioterrorist attacks," *Annals of Operations Research- Computing and Optimization in Medicine and Life Sciences*, vol. 148, pp. 25-53, 2006.

[45]. Erenguc, S. and Koehler, G., "Survey of mathematical programming models and experimental results for linear discriminant analysis," *Managerial and Decision Economics*, vol. 11, pp. 215-225, 1990.

[46]. Feltus, F., Lee, E., Costello, J., Plass, C., and Vertino, P., "DNA signatures associated with CpG island methylation states," *Genomics*, vol. 87, pp. 572-579, 2006.

[47]. Feltus, F., Lee, E., Costello, J., Plass, C., and Vertino, P., "Predicting aberrant CpG island methylation," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 12253-12258, 2003.

[48]. Fiorino, C., et al., "Rectal dose–volume constraints in high-dose radiotherapy of localized prostate cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 57, no. 4, pp. 953-962, 2006.

[49]. Fisher, R., "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179-188, 1936.

[50]. Fraass, B.A. and Moran, J.M., "Quality, technology and outcomes: evolution and evaluation of new treatments and/or new technology," *Seminars in radiation oncology*, Vol. 22., No. 1., WB Saunders, 2012.

[51]. Fraser, C., Riley, S., Anderson, R., Ferguson, N., "Factors that make an infectious disease outbreak controllable," *Proc Natl Acad Sci USA*, vol. 101, no. 16, pp. 6146-6151, 2004.

[52]. Freed, N. and Glover, F., "A linear programming approach to the discriminant problem," *Decision Sciences*, vol. 12, pp. 68-74, 1981.

[53]. Freed, N. and Glover, F., "Evaluating alternative linear programming models to solve the two-group discriminant problem," *Decision Sciences*, vol. 17, pp. 151-162, 1986.

[54]. Freed, N. and Glover, F., "Simple but powerful goal programming models for discriminant problems," *European Journal of Operational Research*, vol. 7, pp. 44-60, 1981.

[55]. Fung, G. and Mangasarian, O., "Proximal support vector machine classifiers," in *Proceedings KDD-2001*, (San Francisco), August 26-29 2001.

[56]. Gallagher, R. J. and Lee, E. K., "Mixed integer programming optimization models for brachytherapy treatment planning," In *Proceedings of the AMIA Annual Fall Symposium* (p. 278), American Medical Informatics Association, 1997.

[57]. Gallagher, R., Lee, E., and Patterson, D., "An optimization model for constrained discriminant analysis and numerical experiments with iris, thyroid, and heart disease datasets," in *Proceedings of the 1996 American Medical Informatics Association*, October 1996.

[58]. Gallagher, R.J, Lee, E. K., and Patterson, D. A., "Constrained discriminant analysis via 0/1 mixed integer programming," *Annals of Operations Research*, vol. 74, pp. 65-88, 1997.

[59]. Gehrlein, W., "General mathematical programming formulations for the statistical classification problem," *Operations Research Letters*, vol. 5, no. 6,pp. 299-304, 1986.

[60]. Glen, J., "An iterative mixed integer programming method for classification accuracy maximizing discriminant analysis," *Computers and Operations Research*, vol. 30, pp. 181-198, 2003.

[61]. Glen, J., "Classification accuracy in discriminant analysis: A mixed integer programming approach," *Journal of the Operational Research Society*, vol. 52, pp. 328-339, 2001.

[62]. Glen, J., "Integer programming methods for normalisation and variable selection in mathematical programming discriminant analysis models," *Journal of the Operational Research Society*, vol. 50, pp. 1043-1053, 1999.

[63]. Glen, J., "Mathematical programming models for piecewise-linear discriminant analysis," *Journal of the Operational Research Society*, vol. 56, pp. 331-341, 2005.

[64]. Glover, F., "Improved linear programming models for discriminant analysis," *Decision Sciences*, vol. 21, pp. 771-785, 1990.

[65]. Glover, F., Keene, S., and Duea, B., "A new class of models for the discriminant problem," *Decision Sciences*, vol. 19, pp. 269-280, 1988.

[66]. Gochet, W., Stam, A., Srinivasan, V., and Chen, S., "Multigroup discriminant analysis using linear programming," *Operations Research*, vol. 45, no. 2, pp. 213-225, 1997.

[67]. Gojovic, Z., Sander, B., Fisman, D., Krahn, M.D., Bauch, C.T., "Modeling mitigation strategies for pandemic (H1N1) 2009," *CMAJ*, vol. 181, pp. 673-680, 2009.

[68]. Goksu, E., Oktay, C., "Factors affecting revisit of COPD exacerbated patients presenting to emergency department," *European Journal of Emergency Medicine* vol. 17, no. 5, pp. 283-285, 2010.

[69]. Gordon, J.A., An, L.C., Hayward, R.A., et al., "Initial emergency department diagnosis and return visits: risk versus perception," *Ann Emerg Med*, vol. 32, pp. 569-573, 1998.

[70]. Grassly, N.C., Fraser, C., "Mathematical models of infectious disease transmission," *Nat Rev Microbiol*, vol. 6, pp. 477-487, 2008.

[71]. Hand, D., Discrimination and classification. New York: John Wiley, 1981.

[72]. Hocking, R. R., and R. N. Leslie, "Selection of the best subset in regression analysis," *Technometrics*, vol. 9, no. 4, pp. 531-540, 1967.

[73]. Holloway, C.L., Racine, M., Cormack, R.A., O'Farrell, D.A., Viswanathan, A.N., "Sigmoid dose using 3D imaging in cervical-cancer brachytherapy," *Radiotherapy Oncology*, vol. 93, no. 2, pp. 307–310, 2009.

[74]. Holm, A., Larsson, T., Carlsson, T.A., "Impact of using linear optimization models in dose planning for HDR brachytherapy," *Medical Phys.*, vol. 39, no. 2, pp. 1021, 2012.

[75]. Howlader, N., Noone, A.M., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S.F., Kosary, C.L., et al., eds., "SEER cancer statistics review, 1975-2010," 2012.

[76]. Hsu, C.W. and Lin, C.J., "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[77]. Hu, S.C., "Analysis of patient revisits to the emergency department," *Am J Emerg Med*, vol. 10, pp. 366-370, 1992.

[78]. Hu, W.T., et al., "Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease," *Neurology*, vol. 79, no. 9, pp. 897-9, 2012.

[79]. Iannarilli, F. and Rubin, P., "Feature selection for multiclass discrimination via mixed-integer linear programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 779-783, 2003.

[80]. Jensen, I., et al. "Radiobiological impact of reduced margins and treatment technique for prostate cancer in terms of tumor control probability (TCP) and normal tissue complication probability (NTCP)," *Medical Dosimetry*, vol. 36, no. 2, pp. 130-137, 2010.

[81]. Joachimsthaler, E. and Stam, A., "Mathematical programming approaches for the classification problem in two-group discriminant analysis," *Multivariate Behavioral Research*, vol. 25, no. 4, pp. 427-454, 1990.

[82]. Joynt, K.E., et al., "Thirty-day readmission rates for Medicare beneficiaries by race and site of care," *JAMA*, vol. 305, no. 7, pp. 675-681, 2011.

[83]. Kang, H.C., Shin, K.H., Park, S.Y., Kim, J.Y., "3D CT-based high-dose-rate brachytherapy for cervical cancer: Clinical impact on late rectal bleeding and local control," *Radiotherapy Oncology*, vol. 97, no.3, pp. 507–513, 2010.

[84]. Karabis, A., Belotti, P., Baltas, D., "Optimization of catheter position and dwell time in prostate HDR brachytherapy using HIPO and linear programming," In *World Congress on Medical Physics and Biomedical Engineering*, September 7-12, 2009.

[85]. Keith, K.D., Bocka, J.J., Kobernick, M.S., et al., "Emergency department revisits," *Ann Emerg Med*, vol. 18, pp. 964-968, 1989.

[86]. Kelly, A.M., Chirnside, A.M., Curry, C.H., "An analysis of unscheduled return visits to an urban emergency department," *NZ Med J*, vol. 106, pp. 334-336, 1993.

[87]. Kennedy, J. and Eberhart, R., "A discrete binary version of the particle swarm algorithm," in *1997 IEEE International Conference on Systems, Man, and Cybernetics*, (Piscataway, NJ), pp. 4104-4108, IEEE, 1997.

[88]. Kennedy, J. and Eberhart, R., "Particle swarm optimisation," in *Proceedings of the IEEE International Conference on Neural Networks*, (Piscataway, NJ), pp. 1942-1948, IEEE, 1995.

[89]. Kim, D.H., Wang-Chesebro, A., Weinberg, V., Pouliot, J., Chen, L.M., Speight, J., Littell, R., Hsu, I.C., "High-dose rate brachytherapy using inverse planning simulated annealing for locoregionally advanced cervical cancer: A clinical report with 2-year follow-up," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 75, no. 5, pp. 1329–1334, 2009.

[90]. Kirisits, C., Pötter, R., Lang, S., Dimopoulos, J., Wachter-Gerstner, N., Georg, D., "Dose and volume parameters for MRI-based treatment planning in intracavitary brachytherapy for cervical cancer," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 62, no. 3, pp. 901–911, 2005.

[91]. Kluger, A., Ferris, S., Golomb, J., Mittelman, M., and Reisberg, B., "Neuropsychological prediction of decline to dementia in nondemented elderly," *Journal of Geriatric Psychiatry and Neurology*, vol. 12, pp. 168-179, 1999.

[92]. Krzysztof, J.C., Moore, G.W., "Uniqueness of medical data mining," *Artificial intelligence in medicine*, vol. 26, no. 1, pp. 1-24, 2002.

[93]. Kufer, K.H., Monz, M., Scherrer, A., Suss, P., Alonso, F., Sultan, A.S.A., Bortfeld, T.H., Craft, D., and Thieke, C.H.R., "Multicriteria optimization in intensity modulated radiotherapy planning," *Berichte des Fraunhofer ITWM 7*, 2005.

[94]. Lahanas, M., Baltas, D. and Zamboglou, N., "Anatomy-based three-dimensional dose optimization in brachytherapy using multiobjective genetic algorithms," *Med. Phys.*, vol. 26, pp. 1904–18, 1999.

[95]. Lahanas, M., Baltas, D., Giannouli, S., "Global convergence analysis of fast multiobjective gradient based dose optimization algorithms for high dose rate brachytherapy," *Phys. Medicine Biol.*, vol. 48, no. 5, pp. 599–617, 2003.

[96]. Lavrač, N., "Selected techniques for data mining in medicine," *Artificial Intelligence in Medicine*, vol. 16, no. 1, pp. 3-23, 1999.

[97]. Lee, E.K., Chen, C.H., Pietz, F., Benecke, B., "Disease propagation analysis and mitigation strategies for effective mass dispensing," *American Medical Informatics Association 2010 Symposium Proceedings*, pp. 427-431, 2010.

[98]. Lee, E.K., Pietz, F., Benecke, B., Mason, J., Burel, G., "Advancing public health and medical preparedness with operations research," *Interfaces -- The Franz Edelman Award Achievement in Operations Research*, vol. 43, no. 1, pp. 79-98, 2013.

[99]. Lee, E.K., Gallagher, R.J., Silvern, D., Wuu, C.S., Zaider, M., "Treatment planning for brachytherapy: An integer programming model, two computational approaches and experiments with permanent prostate implant planning," *Phys. Medicine Biol.*, vol. 44, no. 1, pp. 145–165, 1999.

[100]. Lee, E.K., Maheshwary, S., "Facets of conflict hypergraph," *SIAM J. Optim.*, 2012.

[101]. Lee, E.K., Shapoval, A., Surana, V., "Conflict hypergraphs," *SIAM J. Optim.*, 2012.

[102]. Lee, E.K., Wu, T.L., "Classification and disease prediction via mathematical programming. In Data Mining, Systems Analysis, and Optimization in Biomedicine," *AIP Conference Proceedings*, vol. 953, pp. 1-42, 2007.

[103]. Lee, E.K., Zaider, M., "On the determination of an effective planning volume for permanent prostate implants," *International J. Radiation Oncology Biol. Phys.*, vol. 49, no. 4, pp. 1197–1206, 2001.

[104]. Lee, E.K., Zaider, M., "Intraoperative dynamic dose optimization in permanent prostate implants," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 56, no. 3, pp. 854–861, 2003.

[105]. Lee, E.K., Zaider, M., "Mixed integer programming approaches to treatment planning for brachytherapy: Application to permanent prostate implants," *Ann. Oper. Res. Optim. Medicine*, vol. 119, pp. 147–163, 2003.

[106]. Lee, E.K., Zaider, M., "Incorporating biological metabolite information within treatment of prostate carcinoma and analysis of dose escalation effect," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 66, no. 3, pp. 572–573, 2006.

[107]. Lee, E.K., Zaider, M., "Operations research advances cancer therapeutics," *Interfaces*, vol. 38, no. 1, pp. 5–25, 2008.

[108]. Lee, E.K., "Large-scale optimization-based classification models in medicine and biology," *Annals of Biomedical Engineering - Systems Biology, Bioinformatics, and Computational Biology*, vol. 35, no. 6, pp. 1095-1109, 2007.

[109]. Lee, E.K., Fung, A., Brooks, J., and Zaider, M., "Automated planning volume definition in soft-tissue sarcoma adjuvant brachytherapy," *Physics in Medicine and Biology*, vol. 47, pp. 1891-1910, 2002.

[110]. Lee, E.K., Gallagher, R., and Patterson, D., "A linear programming approach to discriminant analysis with a reserved-judgment region," *INFORMS Journal on Computing*, vol. 15, no. 1, pp. 23-41, 2003.

[111]. Lee, E.K., Gallagher, R., Campbell, A., and Prausnitz, M., "Prediction of ultrasound-mediated disruption of cell membranes using machine learning techniques and statistical analysis of acoustic spectra," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 1, pp. 82-89, 2004.

[112]. Lee, Y. M., "Analyzing dispensing plan for emergency medical supplies in the event of bioterrorism," In *Proceedings of the 40th Conference on Winter Simulation (pp. 2600-2608). Winter Simulation Conference*, 2008.

[113]. Lee, Y.J. and Mangasarian, O., "RSVM: Reduced support vector machines," in *Proceedings of the SIAM International Conference on Data Mining*, (Chicago), April 5-7 2001.

[114]. Lee, Y.J., Mangasarian, O., and Wolberg, W., "Breast cancer survival and chemotherapy: A support vector machine analysis," in *DIMACS Series in Discrete Mathematical and Theoretical Computer Science*, vol. 55, pp. 1-10, American Mathematical Society, 2000.

[115]. Lerman, B., Kobernick, M.S., "Return visits to the emergency department," *J Emerg Med*, vol. 5, pp. 359-362, 1987.

[116]. Levin-Plotnik, D. and Hamilton, R.J., "Optimization of tumour control probability for heterogeneous tumours in fractionated radiotherapy treatment protocols," *Physics in medicine and biology*, vol. 49, no. 3, pp. 407, 2004.

[117]. Longini Jr, I.M., Halloran, M.E., Nizam, A., Yang, Y., "Containing pandemic influenza with antiviral agents," *Am J Epidemiol*, vol. 159, no.7, pp. 623–33, 2004.

[118]. Longini Jr, I.M., Ackerman, E., Elveback, L.R., "An optimization model for influenza A epidemics," *Math Biosci*, vol. 38, pp. 141–157, 1978.

[119]. Lopez, O., Becker, J., Jagust, W., Fitzpatrick, A., Carlson, M., DeKosky, S., Breitner, J., Lyketsos, C., Jones, B., Kawas, C., and Kuller, L., "Neuropsychological characteristics of mild cognitive impairment subgroups," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 77, pp. 159-165, 2006.

[120]. Luedi, P., Hartemink, A., and Jirtle, R., "Genome-wide prediction of imprinted murine genes," *Genome Research*, vol. 15, pp. 875-884, 2005.

[121]. Malyapa, R.S., Mutic, S., Low, D.A., Zoberi, I., Bosch, W.R., Laforest, R., Miller, T.R., Grigsby, P.W., "Physiologic FDG-PET three-dimensional brachytherapy treatment planning for cervical cancer," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 54, no. 4, pp. 1140–1146, 2002.

[122]. Mangasarian, O., "Generalized support vector machines," *Advances in Neural Information Processing Systems*, 135-146, 1999.

[123]. Mangasarian, O., "Linear and nonlinear separation of patterns by linear programming," *Operations Research*, vol. 13, pp. 444-452, 1965.

[124]. Mangasarian, O., Setiono, R., and Wolberg, W., "Pattern recognition via linear programming: Theory and application to medical diagnosis," in *Large-Scale Numerical Optimization* (Coleman, T. and Li, Y., eds.), (Philadelphia, Pennsylvania), pp. 22-31, SIAM, 1990.

[125]. Mangasarian, O., Street, W., and Wolberg, W., "Breast cancer diagnosis and prognosis via linear programming," *Operations Research*, vol. 43, no. 4, pp. 570-577, 1995.

[126]. Markowski, E. and Markowski, C., "Some difficulties and improvements in applying linear programming formulations to the discriminant problem," *Decision Sciences*, vol. 16, pp. 237-247, 1985.

[127]. Martin-Gill, C., Reiser, R.C., "Risk factors for 72-hour admission to the ED," *Am J Emerg Med*, vol. 22, pp. 448-53, 2004;

[128]. McCusker, J., Cardin, S., Bellavance, F., et al., "Return to the emergency department among elders: patterns and predictors," *Acad Emerg Med*, vol. 7, pp. 249-259, 2000.

[129]. McCusker, J., Healey, E., Bellavance, F., et al., "Predictors of repeat emergency department visits by elders," *Acad Emerg Med*, vol. 4, pp. 581-588, 1997.

[130]. Medlock, J., Galvani, A.P., "Optimizing influenza vaccine distribution," *Science*, vol. 325, pp. 1705–1708, 2009.

[131]. Medlock, J., Meyers, L.A., "Optimizing allocation for a delayed influenza vaccination campaign," *PLoS Curr Influenza*, RRN1134, 2009.

[132]. Meltzer, M.I., Cox, N.J., and Fukuda K., "The economic impact of pandemic influenza in the United States: priorities for intervention," *Emerg Infect Dis.*, vol. 5, no. 5, pp. 659–671, 1999.

[133]. Miller, M.A., Viboud, C., Olson, D.R., Grais, R.F., Rabaa, M.A., and Simonsen, L., "Prioritization of Influenza Pandemic Vaccination to Minimize Years of Life Lost," *The Journal of Infectious Diseases*, vol. 198, pp. 305–11, 2008.

[134]. Mohan, R., et al., "Clinically relevant optimization of 3-D conformal treatments," *Medical physics*, vol. 19, no. 4, pp. 933-944, 1992.

[135]. Monteiro, S. and Kosugi, Y., "Particle swarms for feature extraction of hyperspectral data," *IEICE Transactions on Information and Systems*, vol. E90-D, no. 7, pp. 10381046, 2007.

[136]. Morton, G.C., Sankreacha, R., Halina, P.A., Loblaw, A., "A comparison of anatomy-based inverse planning with simulated annealing and graphical optimization for high-dose-rate prostate brachytherapy," *Brachytherapy*, vol. 7, no. 1, pp. 12–16, 2008.

[137]. Mylius, S.D., Hagenaars, T.J., Lugner, A.K., Wallinga, J., "Optimal allocation of pandemic influenza vaccine depends on age, risk and timing," *Vaccine*, vol. 26, pp. 3742–3749, 2008.

[138]. Nag, S., Orton, C., Young, D., "The American Brachytherapy Society survey of brachytherapy practice for carcinoma of the cervix in the United States," *Gynecologic Oncology*, vol. 73, no. 1, pp. 111–118, 1999.

[139]. Narendra, P. and Fukunaga, K., "A branch and bound algorithm for feature subset selection," *IEEE Transactions on Computers*, vol. C-26, no. 9, pp. 917-922, 1977.

[140]. National Cancer Institute (2013a) National Cancer Institute fact sheet: HPV and cancer. Accessed July 1, 2013, http://www.cancer.gov/cancertopics/factsheet/Risk/HPV.

[141]. National Cancer Institute (2013b) Cancer advances in focus. Accessed July 1, 2013, http://www.cancer.gov/cancertopics/factsheet/cancer-advances-in-focus/cervical.

[142]. National Institutes of Health (2013) NIH research portfolio online reporting tools. Accessed July 1, 2013, http://report.nih.gov/nihfactsheets/viewfactsheet.aspx?csid=76.

[143]. Newton, A.S., Ali, S., Johnson, D.W., et al., "Who Comes Back? Characteristics and Predictors of Return to Emergency Department Services for Pediatric Mental Health Care," *Academic Emergency Medicine*, vol. 17, pp. 177–186, 2010.

[144]. Nichol, K.L., Wuorenma, J., von Sternberg, T., "Benefits of Influenza Vaccination for Low-, Intermediate-, and High-Risk Senior Citizens," *Arch Intern Med*, vol. 158, pp. 1769 - 1776, 1998.

[145]. Nordqvist, J., "2009 Swine Flu Pandemic Infected 1 In Every 5 People," Medical News Today. 2013, January 25. http://www.medicalnewstoday.com/articles/255456.php.

[146]. Obenshain, M.K., "Application of Data Mining Techniques to Healthcare Data," *Infection Control and Hospital Epidemiology*, vol. 25, no. 8, pp. 690-695, 2004.

[147]. Kallman, P., Agren, A., and Brahme, A., "Tumor and normal tissue responses to fractionated non-uniform dose delivery," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 62 , pp. 249–262, 1992.

[148]. Patel, R., Longini Jr, I.M., and Halloran, M.E., "Finding optimal vaccination strategies for pandemic influenza using genetic algorithms," *Journal of Theoretical Biology*, vol. 234, pp. 201-212, 2005.

[149]. Pavur, R., Wanarat, P., and Loucopoulos, C., "Examination of the classificatory performance of MIP models with secondary goals for the two group discriminant problem," *Annals of Operations Research*, vol. 74, pp. 173-189, 1997.

[150]. Pearson, M.L., Bridges, C.B., Harper, S.A., "Influenza Vaccination of Health-Care Personnel," *MMWR*, vol. 55, (RR02) 1-16, 2006.

[151]. Pierce, J.M., Kellerman, A.L., Oster, C., "'Bounces': an analysis of short-term revisits to a public hospital emergency department," *Ann Emerg Med*, vol. 19, pp. 752-757, 1990.

[152]. Poli, R., Kennedy, J., Blackwell, T., "Particle swarm optimization: an overview," *Swarm Intelligence*, vol. 1, pp. 33-57, 2007.

[153]. Pudil, P., Novovicova, J., and Kittler, J., "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, 1994.

[154]. Querec, T.D., Akondy, R., Lee, E.K, et al., "Systems biology approaches predict immunogenicity of the yellow fever vaccine in humans," *Nature Immunology*, vol. 10, pp. 116-125, 2009.

[155]. Raphael, D., Bryant, T., "The state's role in prompting population health: public health concerns in Canada, USA, UK, and Sweden," *Health Policy*, vol. 78, pp. 39-55, 2006.

[156]. Ray, S., Britschgi, M., Herbert, C., et al., "Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins," *Nat Med*, vol. 13, pp. 1359 –1362, 2007.

[157]. Reddy, M.M., Wilson, R., Wilson, J., et al., "Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening," *Cell*, vol. 144, pp. 132–142, 2011.

[158]. Rivard, M.J., Coursey, B.M., DeWerd, L.A., Hanson, W.F., Huq, M.S., Ibbott, G.S., Mitch, M.G., Nath, R., Williamson, J.F., "Update of AAPM task group No.

43 Report: A revised AAPM protocol for brachytherapy dose calculations," *Medical Phys.*, vol. 31 no. 12, pp. 633–674, 2004.

[159]. Rubin, P., "A comparison of linear programming and parametric approaches to the two-group discriminant problem," *Decision Sciences*, vol. 21, pp. 373-386, 1990.

[160]. Rubin, P., "Separation failure in linear programming discriminant models," *Decision Sciences*, vol. 22, pp. 519-535, 1991.

[161]. Rubin, P., "Solving mixed integer classification problems by decomposition," *Annals of Operations Research*, vol. 74, pp. 51-64, 1997.

[162]. Ruotsalainen, H., Miettinen, K., Palmgren, J., Lahtinen, T., "Interactive multiobjective optimization for anatomy-based three-dimensional HDR brachytherapy," *Phys. Medicine Biol.*, vol. 55, no. 16, pp. 4703–4719, 2010.

[163]. Scherer, A., Mclean, A., "Mathematical models of vaccination," British Medical Bulletin, vol. 62, no. 1, pp. 187-99, 2002.

[164]. Shapoval, R., Lee, E.K., "Mixed integer programming convexification of posynomial functions," (2012a), Submitted.

[165]. Shapoval, R., Lee, E.K., "Mixed integer programming convexification ii. Extension from posynomial to signomial functions," (2012b), Submitted.

[166]. Siedlecki, W. and Sklansky, J., "A note on genetic algorithms for largescale feature selection," *Pattern Recognition Letters*, vol. 10, pp. 335-347, 1989.

[167]. Silva, A. D. and Stam, A., "A mixed integer programming algorithm for minimizing the training sample misclassification cost in two-group classification," *Annals of Operations Research*, vol. 74, pp. 129-157, 1997.

[168]. Silva, A. D. and Stam, A., "Second order mathematical programming formulations for discriminant analysis," *European Journal of Operational Research*, vol. 72, pp. 4-22, 1994.

[169]. Silva, P.D. and Stam A., "Second order mathematical programming formulations for discriminant analysis," *European Journal of Operational Research*, vol. 72, no. 1, pp. 4–22, 1994.

[170]. Simonsen, L., Taylor, R.J., Viboud, C., Miller, M.A., Lisa, A.J., "Mortality benefits of influenza vaccination in elderly people: an ongoing controversy," *The Lancet Infectious Diseases*, Volume 7, Issue 10, Pages 658-666, 2007.

[171]. Smith, C., "Some examples of discrimination," *Annals of Eugenics*, vol. 13, pp. 272-282, 1947.

[172]. Somol, P., Pudil, P., Novovicova, J., and Paclik, P., "Adaptive floating search methods in feature selection," *Pattern Recognition Letters*, vol. 20, pp. 1157-1163, 1999.

[173]. South, C. P., Partridge, M., and Evans, P.M., "A theoretical framework for prescribing radiotherapy dose distributions using patient-specific biological information," *Medical physics*, vol. 35, no. 10, pp. 4599-4611, 2008.

[174]. Søvik, Å, et al., "Optimization of tumour control probability in hypoxic tumours by radiation dose redistribution: a modelling study," *Physics in medicine and biology*, vol. 52, no. 2, pp. 499, 2007.

[175]. Stam, A. and Joachimsthaler, E., "A comparison of a robust mixed-integer approach to existing methods for establishing classification rules for the

discriminant problem," *European Journal of Operational Research*, vol. 46, pp. 113-122, 1990.

[176]. Stam, A. and Joachimsthaler, E., "Solving the classification problem in discriminant analysis via linear and nonlinear programming methods," *Decision Sciences*, vol. 20, pp. 285-293, 1989.

[177]. Stam, A. and Ungar, D., "RAGNU: A microcomputer package for two group mathematical programming-based nonparametric classification," *European Journal of Operational Research*, vol. 86, pp. 374-388, 1995.

[178]. Stam, A., "Nontraditional approaches to statistical classification: Some perspectives on Lp-norm methods," *Annals of Operations Research*, vol. 74, pp. 136, 1997.

[179]. Stearns, S., "On selecting features for pattern classifiers," in *Proceedings of the Third International Conference on Pattern Recognition*, (Coronado, California), pp. 71-75, 1976.

[180]. Stuss, D. and Trites, R., "Classification of neurological status using multiple discriminant function analysis of neuropsychological test scores," *Journal of Consulting and Clinical Psychology*, vol. 45, no. 1, pp. 145-145, 1977.

[181]. Sueyoshi, T., "DEA-discriminant analysis in the view of goal programming," *European Journal of Operational Research*, vol. 115, pp. 564-582, 1999.

[182]. Tabert, M., Manly, J., Liu, X., Pelton, G., Rosenblum, S., Jacobs, M., Zamora, D., Goodkind, M., Bell, K., Stern, Y., and Devanand, D., "Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment," *Archives of General Psychiatry*, vol. 63, pp. 916-924, 2006.

[183]. Tanderup, K., Nielsen, S.K., Nyvang, G.B., Pedersen, E.M., Røhl, L., Aagaard, T., Fokdal, L., Lindegaard, J.C., "From point A to the sculpted pear: MR image guidance significantly improves tumour dose and sparing of organs at risk in brachytherapy of cervical cancer," *Radiotherapy Oncology*, vol. 94, no. 2, pp. 173–180, 2010.

[184]. Tang, E., Suganthan, P., and Yao, X., "Feature selection for microarray data using least squares SVM and particle swarm optimization," in *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, (Piscataway, NJ), IEEE, 2005.

[185]. Tibshirani, R., "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267-288, 1996.

[186]. Tuite, A.R., Fisman, D.N., Kwong, J.C., and Greer, A.L., "Optimal Pandemic Influenza Vaccine Allocation Strategies for the Canadian Population," *PLoS One*, vol. 5, no. 5, pp. e10520, 2010.

[187]. Uscher-Pines, L., Omer, S.B., Barnett, D.J., Burke, T.A., Balicer, R.D., "Priority Setting for Pandemic Influenza: An Analysis of National Preparedness Plans," *PLoS Med*, vol. 3, no. 10, pp. e436, 2006.

[188]. Wachter-Gerstner, N., Wachter, S., Reinstadler, E., Fellner, C., Knocke, T.H., Pötter, R., "The impact of sectional imaging on dose escalation in endocavitary HDR-brachytherapy of cervical cancer: results of a prospective comparative trial," *Radiotherapy Oncology*, vol. 68, no. 1, pp. 51–59, 2003.

[189]. Wallinga, J., van Bovena, M., and Lipsitch, M., "Optimizing infectious disease interventions during an emerging epidemic," *PNAS*, vol. 107, no. 2, pp. 923-928, 2010.

[190]. Wanarat, P. and Pavur, R., "Examining the effect of second-order terms in mathematical programming approaches to the classification problem," *European Journal of Operational Research*, vol. 93, pp. 582-601, 1996.

[191]. Webb, S., and Nahum, A.E., "A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density," *Physics in Medicine and Biology*, vol. 38, no. 6, pp. 653, 1993.

[192]. Wein, L.M., Craft, D.L., and Kaplan, E.H., "Emergency response to an anthrax attack," *Proceedings of the National Academy of Sciences*, vol. 100, no. 7, pp. 4346-4351, 2003.

[193]. Bicknell, W.J., "How to best fight against bioterrorism," Cato Institute, August 2003.

[194]. World Health Organization (2013) Comprehensive cervical cancer prevention and control: A healthier future for girls and women. Accessed July 1, 2013. http://www.who.int/reproductivehealth/topics/cancers/en/.

[195]. Wu, J.T., Riley, S., Fraser, C., Leung, G.M., "Reducing the Impact of the Next Influenza Pandemic Using Household-Ba---sed Public Health Interventions," *PLoS Med*, vol. 3, no. 9, pp. e361, 2006.

[196]. Wu, J.T., Riley, S., Leung, G.M., "Spatial considerations for the allocation of pre-pandemic influenza vaccination in the United States," *Proc. R. Soc. B*, vol. 274, pp. 2811-2817, 2007.

[197]. Wu, T.L., "Solving a mixed-integer programming formulation of a classification model with misclassification limits," PhD thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 2008.

[198]. Xiao, B. and Feng, Y., "Alternative discriminant vectors in LP models and a regularization method," *Annals of Operations Research*, vol. 74, pp. 113-127, 1997.

[199]. Yang, Y., Sugimoto, J.D., Halloran, M.E., Basta, N.E., Chao, D.L., Matrajt, L., Potter, G., Kenah, E., and Longini Jr, I.M., "The transmissibility and control of pandemic influenza A (H1N1) virus," *Science*, vol. 326, pp. 729, 2009.

[200]. Yang, Y., and Xing, L., "Optimization of radiotherapy dose-time fractionation with consideration of tumor specific biology," *Medical physics*, vol. 32, no. 12, pp. 3666-3677, 2005.

[201]. Yang, Y., and Xing, L., "Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution," *Medical physics*, vol. 32, no.6, pp. 1473-1484, 2005.

[202]. Zaider, M., Lee, E.K., "Treatment planning for low dose rate and high dose rate brachytherapy," In *Basic and Advanced Techniques in Prostate Brachytherapy, Chapter 12*. Taylor & Francis, The United Kingdom, pp. 142–156, 2005.

[203]. Zaider, M., Minerbo, G.N., "Tumour control probability: A formulation applicable to any temporal protocol of dose delivery," *Phys. Medicine Biol.*, vol. 45, no. 2, pp. 279–293, 2000.

[204]. Zaider, M., Zelefsky, M.J., Lee, E.K., Zakian, K.L., Amols, H.I., Dyke, J., Cohen, G., Hu, Y., Endi, A.K., Chui, C., Koutcher, J.A., "Treatment planning for prostate implants using MR spectroscopy imaging," *Internat. J. Radiation Oncology Biol. Phys.*, vol. 47, no. 4, pp. 1085–1096, 2000.

[205]. Zhang, P.P, et al., "Volumetric modulated arc therapy: planning and evaluation for prostate cancer cases," *International Journal of Radiation Oncology* Biology* Physics*, vol. 76, no.5, pp. 1456-1462, 2010.