VALUE IN HEALTH 20 (2017) 310–319

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

# Constrained Optimization Methods in Health Services Research—An Introduction: Report 1 of the ISPOR Optimization Methods Emerging Good Practices Task Force

*William Crown, PhD*[1,*], *Nasuh Buyukkaramikli, PhD*[2], *Praveen Thokala, PhD*[3], *Alec Morton, PhD*[4], *Mustafa Y. Sir, PhD*[5], *Deborah A. Marshall, PhD*[6,7], *Jon Tosh, PhD*[8], *William V. Padula, PhD, MS*[9], *Maarten J. Ijzerman, PhD*[10], *Peter K. Wong, PhD, MS, MBA, RPh*[11], *Kalyan S. Pasupathy, PhD*[12,*]

[1]OptumLabs, Boston, MA, USA; [2]Scientific Researcher, Institute of Medical Technology Assessment, Erasmus University Rotterdam, Rotterdam, The Netherlands; [3]Research Fellow, University of Sheffield, Sheffield, UK; [4]Professor of Management Science, Department of Management Science, Strathclyde Business School, University of Strathclyde, Glasgow, Scotland, UK; [5]Assistant Professor, Health Care Policy & Research, Information and Decision Engineering, Mayo Clinic Kern Center for the Science of Health Care Delivery, Rochester, MN, USA; [6]Canada Research Chair, Health Services & Systems Research; Arthur J.E. Child Chair in Rheumatology Research; Director, HTA, Alberta Bone & Joint Health Institute; [7]Associate Professor, Department Community Health Sciences, Faculty of Sciences, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada; [8]Senior Health Economist, DRG Abacus, Manchester, UK; [9]Assistant Professor, Department of Health Policy & Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [10]Professor of Clinical Epidemiology & Health Technology Assessment (HTA); Head, Department of Health Technology & Services Research, University of Twente, Enschede, The Netherlands; [11]Vice President and Chief Performance Improvement Officer, Illinois Divisions and HSHS Medical Group, Hospital Sisters Health System (HSHS), Belleville, IL. USA; [12]Associate Professor - Healthcare Policy & Research, Lead, Information and Decision Engineering, Mayo Clinic Kern Center for the Science of Health Care Delivery, Rochester, MN, USA

ABSTRACT

Providing health services with the greatest possible value to patients and society given the constraints imposed by patient characteristics, health care system characteristics, budgets, and so forth relies heavily on the design of structures and processes. Such problems are complex and require a rigorous and systematic approach to identify the best solution. Constrained optimization is a set of methods designed to identify efficiently and systematically the best solution (the optimal solution) to a problem characterized by a number of potential solutions in the presence of identified constraints. This report identifies 1) key concepts and the main steps in building an optimization model; 2) the types of problems for which optimal solutions can be determined in real-world health applications; and 3) the appropriate optimization methods for these problems. We first present a simple graphical model based on the treatment of "regular" and "severe" patients, which maximizes the overall health benefit subject to time and budget constraints. We then relate it back to how optimization is relevant in health services research for addressing present day challenges. We also explain how these mathematical optimization methods relate to simulation methods, to standard health economic analysis techniques, and to the emergent fields of analytics and machine learning.

*Keywords:* decision making, care delivery, modeling, policy.



## Introduction

In common vernacular, the term "optimal" is often used loosely in health care applications to refer to any demonstrated superiority among a set of alternatives in specific settings. Seldom is this term based on evidence that demonstrates that such solutions are, indeed, *optimal*—in a mathematical sense. By "optimal" solution we mean the *best possible solution* for a given problem given the complexity of the system inputs, outputs/outcomes, and constraints (budget limits, staffing capacity, etc.). Failing to identify an "optimal" solution represents a missed opportunity to improve clinical outcomes for patients and economic efficiency in the delivery of care.

Identifying optimal health system and patient care interventions is within the purview of mathematical optimization models. There is a growing recognition of the applicability of constrained optimization methods from operations research to health care problems. In a review of the literature [1], note more than 200 constrained optimization and simulation studies in health care. For example, constrained optimization methods have been applied in problems of capacity management and location selection for both health care services and medical supplies [2–5].

* *Address correspondence to:* Kalyan S. Pasupathy, Mayo Clinic, Health Sciences Research, 200 First Street SW, Harwick 2-43, Rochester, MN 55905.

E-mail: William.Crown@optum.com; Pasupathy.Kalyan@mayo.edu.



http://dx.doi.org/10.1016/j.jval.2017.01.013

Constrained optimization is an interdisciplinary subject, cutting across the boundaries of mathematics, computer science, economics, and engineering. Analytical foundations for the techniques to solve the constrained optimization problems involving continuous, differentiable functions and equality constraints were already laid in the 18th century [6]. However, with advances in computing technology, constrained optimization methods designed to handle a broader range of problems trace their origin to the development of the simplex algorithm—the most commonly used algorithm to solve linear constrained optimization problems—in 1947 [7–11]. Since that time, various constrained optimization methods have been developed in the field of operations research and applied across a wide range of industries. This creates significant opportunities for the optimization of health care delivery systems and for providing value by transferring knowledge from fields outside the health care sector.

In addition to capacity management, facility location, and efficient delivery of supplies, patient scheduling, provider resource scheduling, and logistics are other substantial areas of research in the application of constrained optimization methods to health care [12–16]. Constrained optimization methods may also be very useful in guiding clinical decision making in actual clinical practice where physicians and patients face constraints such as proximity to treatment centers, health insurance benefit designs, and the limited availability of health resources.

Constrained optimization methods can also be used by health care systems to identify the optimal allocation of resources across interventions subject to various types of constraints [17–23]. These methods have also been applied to disease diagnosis [24,25], the development of optimal treatment algorithms [26,27], and the optimal design of clinical trials [28]. Health technology assessment using tools from constrained optimization methods is also gaining popularity in health economics and outcomes research [29].

Recently, the ISPOR Emerging Good Practices Task Force on Dynamic Simulation Modeling Applications in Health Care Delivery Research published two reports in *Value in Health* [30,31] and one in *Pharmacoeconomics* [32] on the application of dynamic simulation modeling (DSM) to evaluate problems in health care systems. Although simulation can provide a mechanism to evaluate various scenarios, by design, they do not provide optimal solutions. The overall objective of the ISPOR Emerging Good Practices Task Force on Constrained Optimization Methods is to develop guidance for health services researchers, knowledge users, and decision makers in the use of operations research methods to optimize health care delivery and value in the presence of constraints. Specifically, this task force will 1) introduce constrained optimization methods for conducting research on health care systems and individual-level outcomes (both clinical and economic); 2) describe problems for which constrained optimization methods are appropriate; and 3) identify good practices for designing, populating, analyzing, testing, and reporting results from constrained optimization models.

The ISPOR Emerging Good Practices Task Force on Constrained Optimization Methods will produce two reports. In this first report, we introduce readers to constrained optimization methods. We present definitions of important concepts and terminology, and provide examples of health care decisions in which constrained optimization methods are already being applied. We also describe the relationship of constrained optimization methods to health economic modeling and simulation methods. The second report will present a series of case studies illustrating the application of these methods including model building, validation, and use.

## Definition of Constrained Optimization

*Constrained optimization* is a set of methods designed to efficiently and systematically find the best solution to a problem characterized by a number of potential solutions in the presence of identified constraints. It entails maximizing or minimizing an objective function that represents a quantifiable measure of interest to the decision maker, subject to constraints that restrict the decision maker's freedom of action. Maximizing/minimizing the objective function is carried out by systematically selecting input values for the decision from an allowed set and computing the objective function, in an iterative manner, until the decision yields the best value for the objective function, a.k.a optimum. The decision that gives the optimum is called the "optimal solution." In some optimization problems, two or more different decisions may yield the same optimum. Note that, *programming* and *optimization* are often used as interchangeable terms in the literature, for example, linear programming and linear optimization. Historically, programming referred to the mathematical description of a plan/schedule, and optimization referred to the process used to achieve the optimal solution described in the program.

The components of a constrained optimization problem are its objective function(s), its decision variable(s), and its constraint(s). The *objective function* is a function of the decision variables that represents the quantitative measure that the decision maker aims to minimize/maximize. *Decision variables* are mathematical representation of the constituents of the system for which decisions are being taken to improve the value of the objective function. The *constraints* are the restrictions on decision variables, often pertaining to resources. These restrictions are defined by equalities/inequalities involving functions of decision variables. They determine the allowable/feasible values for the decision variables. In addition, *parameters* are constant values used in objective function and constraints, like the multipliers for the decision variables or bounds in constraints. Each parameter represents an aspect of the decision-making context: for example, a multiplier may refer to the cost of a treatment.

## A Simple Illustration of a Constrained Optimization Problem

Imagine you are the manager of a health care center, and your aim is to benefit as many patients as possible. Let us say, for the sake of simplicity, you have two types of patients—regular and severe patients, and the demand for the health service is unlimited for both these types. Regular patients can achieve two units of health benefits and severe ones can achieve three units. Each patient, irrespective of severity, takes 15 minutes for consultation; only one patient can be seen at any given point in time. You have 1 hour of total time at your disposal. Regular patients require $25 of medications, and severe patients require $50 of medications. You have a total budget of $150. What is the greatest health benefit this center can achieve given these inputs and constraints?

At the outset, this problem seems straightforward. One might decide on four regular patients to use up all the time that is available. This will achieve eight units of health benefit while leaving $50 as excess budget. An alternate approach might be to see as many severe patients as possible because treating each severe patient generates more per capita health benefits. Three patients (totaling $150) would generate nine health units leaving 15 minutes extra time unused. There are other combinations of regular and severe patients that would generate different levels of health benefits and use resources differently.

This is graphically represented in Figure 1, with regular patients on the x-axis and severe patients on the y-axis. Line CF is the time constraint limiting total time to 1 hour. Line BG is the budget constraint limiting to $150. Any point to the southwest of these constraints (lines), respectively, will ensure that

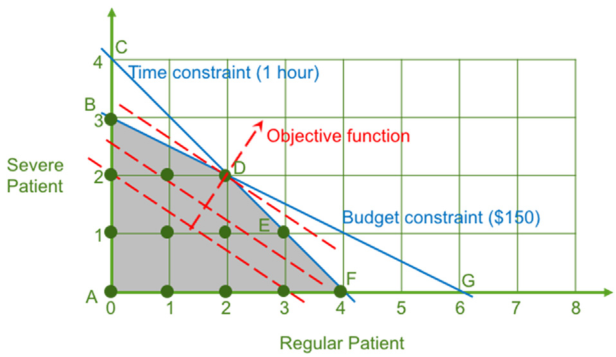


**Fig. 1 – Graphical representation of solving a simple integer programming problem.**

time and budget do not exceed the respective limits. The combination of these, together with non-negativity of the decision variables, gives the feasible region.

The lines AB-BD-DF-FA form the boundary of the feasibility space, shown shaded in the figure. In problems that are three- or more dimensional, these lines would be hyperplanes. To obtain the optimal solution, the dashed line is established, the slope depends on the relative health units of the two decision variables (i.e., the number of regular and severe patients seen). This dashed line moves from the origin in the northeast direction as shown by the arrow. The optimal solution is two regular patients and two severe patients. This approach uses the entire 1-hour time as well as the $150 budget. Because regular and severe patients achieve two- and three-unit health benefits, respectively, we are able to achieve 10 units of health benefit and still meet the time and budget constraints.

No other combination of patients is capable of achieving more benefits while still meeting the time and budget constraints. Note that not all resource constraints have to be completely used to attain the optimal solution. This hypothetical example is a small-scale problem with only two decision variables; the number of regular and severe patients seen. Hence, they can be represented graphically with one variable on each axis.

With the difficulty in representing larger problems graphically, we turn to mathematical approaches, such as the simplex algorithm, to find the solutions. The simplex algorithm is a structured approach of navigating the boundary (represented as lines in two dimensions and hyperplanes in three or more dimensions) of the feasibility space to arrive at the optimal solution. Table 1 summarizes the main components of the example and notes several other dimensions of complexity (linear vs nonlinear, deterministic vs stochastic, static vs dynamic, discrete/integer vs continuous) that can be incorporated into constrained optimization models.

The mathematical formulation of the model is as follows:

Max $f_R\, x_R + f_L\, x_L$ (objective function)
subject to $c_R\, x_R + c_L\, x_L \le B$ (budget constraint)
$t_R\, x_R + t_L\, x_L \le T$ (time constraint)
$x_R, x_L \ge 0$ and integer (decision variables)

where $c_R, c_L$ is the cost of regular and severe patients, respectively; $B$ is the total budget available; $t_R, t_L$ is time to see regular and severe patients, respectively; $T$ is the total time available; $f_R, f_L$ is health benefits of regular and severe patients, respectively; and $x_R, x_L$ is the number of regular and severe patients, respectively.

In the current version of the problem, the parameters are as follows:

$f_R$ = two health benefit units, $f_L$ = three health benefit units
$c_R = \$25$, $c_L = \$50$, $B = \$150$
$t_R = 0.25$ hours, $t_L = 0.25$ hours, $T = 1$ hour
So the model is as follows:

Max $2\, x_R + 3x_L$ (objective function)
subject to $25x_R + 50x_L \le 150$ (budget constraint)
$0.25x_R + 0.25x_L \le 1$ (time constraint)
$x_R, x_L \ge 0$ and integer

As described above, Figure 1 illustrates the graphical solution to this model. However, problems with higher dimensionality must use mathematical algorithms to identify the optimal solution. The problem described above falls into the category of *linear optimization*, because although the constraints and the objective function are linear from an algebraic standpoint, the decision variables must be in the form of integers. As will be discussed further in the section “Steps in a Constrained Optimization Process,” there are other optimization modeling frameworks, such as combinatorial, nonlinear, stochastic, and dynamic optimization.

Because the algorithms for *integer optimization* problems can take much longer to solve computationally than those for linear optimization problems, one alternative is to set the integer optimization problem up and solve it as a linear one. If fractional values are obtained, the nearest feasible integers can be used as the final solution. This should be done with caution, however. First, rounding the solution to the nearest integers can result in an infeasible solution or, and second, even if the rounded solution is feasible, it may not be the optimal solution to the original integer optimization problem. *Nonlinear optimization* is suitable when the constraints or the objective function is nonlinear. In problems in which there is uncertainty, such as the estimated health benefit that each patient might receive in the above example, *stochastic optimization techniques* can be used.

*Dynamic optimization (known commonly as dynamic programming)* formulation might be useful when the optimization problem is not static, the problem context and parameters change in time, and there is an interdependency among the decisions at different time periods (for instance, when decisions made at a given time interval, say number of patients to be seen now, affects the decisions for other time periods, such as the number of patients to be seen tomorrow). Table 1 summarizes the model components in the hypothetical problem, relates it to health services with examples, and identifies the specific terminology.

## Problems That Can Be Tackled with Constrained Optimization Approaches

In this section, we list several areas within health care in which constrained optimization methods have been used in health services. The selected examples do not represent a comprehensive picture of this field, but provide the reader a sense of what is possible. In Table 2, we compare problems using the terminology of the previous section, with respect to decision makers, decisions, objectives, and constraints.

## Steps in a Constrained Optimization Process

An overview of the main steps involved in a constrained optimization process [33] is described here and presented in Table 3. Some of the steps are common to other types of modeling methods. It is important to emphasize that the process of optimization is iterative, rather than comprising a strictly sequential set of steps.

### *Problem Structuring*

This involves specifying the objective, that is, goal, and identifying the decision variables, parameters, and the constraints involved.

**Table 1 – Model summary and extensions.**

| | Hypothetical problem | Real-life health services | Terminology |
|---|---|---|---|
| Aim | Maximize health/health care benefits | Maximize health/health care benefits | Objective function |
| Options available | Regular or severe patients | Service lines, case mix, service mix, etc. | Decision variables |
| Constraints | Total cost ≤$15<br>Total time ≤1 h | Budget constraint<br>Time constraint<br>Resource constraint (e.g., staff and beds) | Constraints |
| Evidence base | Cost of each patient, health benefits of each patient, and the time taken for consultation | Costs, health benefits, and other relevant data associated with each intervention to be selected | Model (to determine the objective function and constraints) |
| Complexity | *Static*<br>The problem does not have a time component; decision made in one time period does not affect decisions made in another | *Dynamic*<br>The optimization problem and parameters may change in different time points, and the decision made at any point in time can affect decisions at later time points (e.g., there can be a capacity constraint defined on 2 mo, whereas the planning cycle is 1 mo) | Optimization method |
| | *Deterministic*<br>All the information is assumed to be certain (e.g., cost of each patient, health benefits of each patient, and the time taken for consultation) | *Stochastic*<br>Know that the information is uncertain (i.e., uncertainty in the costs and benefits of the interventions) | |
| | *Linear* (i.e., each additional patient costs the same and achieves same health benefits) | *Nonlinear* (objective function or constraints may have a nonlinear relationship with the model parameters, e.g., total costs and QALYs typically have a nonlinear relationship with the model parameters) | |
| | *Integer/discrete*<br>The decision variables (number of patients) can take only discrete and integer values | *Continuous*<br>The decision variables can take fractional values (e.g., number of hours) | |

QALYs, quality-adjusted life-years.

These can be specified using words, ideally in nontechnical language, so that the optimization problem is easily understood. This step needs to be performed in collaboration with all the relevant stakeholders, including decision makers, to ensure all aspects of the optimization problem are captured. As with any modeling technique, it is also crucial to surface key modeling assumptions and appraise them for plausibility and materiality.

### *Mathematical Formulation*

After the optimization problem is specified in words, it needs to be converted into mathematical notation. The standard mathematical notation for any optimization problem involves specifying the objective function and constraint(s) using decision variables and parameters. This also involves specifying whether the goal is to maximize or minimize the objective function. The standard notation for any optimization problem, assuming the goal is to maximize the objective, is as shown below:

Maximize $z = f(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_k)$

subject to

$c_j(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_k) \leq C_j$

for $j = 1, 2, \ldots, m$

where, $x_1, x_2, \ldots, x_n$ are the decision variables, $f(x_1, x_2, \ldots, x_n)$ is the objective function, and $c_j(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_k) \leq C_j$ represent the constraints. Note that the constraints can include both inequality and equality constraints and that the objective function and the constraints also include parameters $p_1, p_2, \ldots, p_k$, which are not varied in the optimization problem. Specification of the optimization problem in this mathematical notation allows clear identification of the type (and number) of decision variables, parameters, and the constraints. Describing the model in mathematical form will be useful to support model development.

### *Model Development*

The next step after mathematical formulation is model development. Model development involves solving the mathematical problem described in the previous step, and often performed iteratively. The model should estimate the objective function and the left-hand side values of the constraints, using the decision variables and parameters as inputs. The complexity of the model can vary widely. Similar to other types of modeling, the complexity of the model will depend on the outputs required, and the level of detail included in the model, whether it is linear or nonlinear, stochastic or deterministic, static or dynamic.

### *Perform Model Validation*

As with any modeling, it is important to ensure that the model developed represents reality with an acceptable degree of fidelity [33]. The requirements of model validation for optimization are more stringent than for, for example, simulation models, due to the need for the model to be valid for all possible combinations of the decision variables. Thus, appropriate caution needs to be taken to ensure that the model assumptions are valid and that the model produces sensible results for the different scenarios. At the very least, the validation should involve checking of the face validity (i.e., experts evaluate model structure, data sources,

**Table 2 – Examples of health care decisions for which constrained optimization is applicable.**

| Type of health care problem | Typical decision makers | Typical decisions | Typical objectives | Typical constraints |
|---|---|---|---|---|
| Resource allocation within and across disease programs | Health authorities, insurance funds | List of interventions to be funded | Maximize population health | Overall health budget, other legal constraints for equity |
| Resource allocation for infectious disease management | Public health agencies, health protection agencies | Optimal vaccination coverage level | Minimize disease outbreaks and total costs | Availability of medicines, disease dynamics of the epidemic |
| Allocation of donated organs | Organ banks, transplant service centers | Matching of organs and recipients | Maximize matching of organ donors with potential recipients | Every organ can be received by at most one person |
| Radiation treatment planning | Radiation therapy providers | Positioning and intensity of radiation beams | Minimizing the radiation on healthy anatomy | Tumor coverage and restriction on total average dosage |
| Disease management models | Leads for a given disease management plan | Best interventions to be funded, best timing for the initiation of a medication, best screening policies | Identify the best plan using a whole disease model, maximizing QALYs | Budget for a given disease or capacity constraints for health care providers |
| Workforce planning/ staffing/shift template optimization | Hospital managers, all medical departments (e.g., ED and nursing) | Number of staff at different hours of the day, shift times | Increase efficiency and maximize utilization of health care staff | Availability of staff, human factors, state laws (e.g., nurse-to-patient ratios), budget |
| Inpatient scheduling | Operation room/ICU planners | Detailed schedules | Minimize waiting time | Availability of beds, staff |
| Outpatient scheduling | Clinical department managers | Detailed schedules | Minimize overutilization and underutilization of health care staff | Availability of appointment slots |
| Hospital facility location | Strategic health planners | Set of physical sites for hospitals | Ensure equitable access to hospitals | Maximum acceptable travel time to reach a hospital |

ED, emergency department; ICU, intensive care unit; QALYs, quality-adjusted life-years.

assumptions, and results), and verification or internal validity (i.e., checking accuracy of coding).

### Select Optimization Method

This step involves choosing the appropriate optimization method, which is dependent on the type of optimization problem that is addressed. Optimization problems can be broadly classified, depending on the nature of the objective functions and the constraints, for example, into linear versus nonlinear, deterministic versus stochastic, continuous versus discrete, or single versus multiobjective optimization. For instance, if the objective function and constraints consist of linear functions only, the corresponding problem is a linear optimization problem. Similarly, in deterministic optimization, the parameters used in the optimization problem are fixed whereas in stochastic optimization, uncertainty is incorporated. Optimization problems can be continuous (i.e., decision variables are allowed to have fractional values) or discrete (e.g., a hospital ward may be either open or closed; the number of computed tomography scanners that a hospital buys must be a whole number).

Most optimization problems have a single objective function; however, when optimization problems have multiple conflicting objective functions, they are referred to as multiobjective optimization problems. The optimization method chosen needs to be in line with the type of optimization problem under consideration. Once the optimization problem type is clear (e.g., discrete or nonlinear), a number of texts may be consulted for details on solution methods appropriate for that problem type [33–36].

Broadly speaking, optimization methods can be categorized into *exact approaches* and *heuristic approaches*. Exact approaches iteratively converge to an optimal solution. Examples of these include simplex methods for linear programming and the Newton method or interior point method for nonlinear programming [34,37]. Heuristic approaches provide approximate solutions to optimization problems when an exact approach is unavailable or is computationally expensive. Examples of these techniques include relaxation approaches, evolutionary algorithms (such as genetic algorithms), simulated annealing, swarm optimization, ant colony optimization, and tabu-search. Besides these two approaches (i.e., exact or heuristic), other methods are available to tackle large-scale problems as well (e.g., decomposition of the large problems to smaller subproblems).

There are software programs that help with optimization; interested readers are referred to the Web site of the Institute for Operations Research and the Management Sciences

**Table 3 – Steps in an optimization process.**

| Stage | Step | Description |
|---|---|---|
| Modeling | Problem structuring | Specify the objective and constraints, identify decision variables and parameters, and list and appraise model assumptions |
| | Mathematical formulation | Present the objective function and constraints in mathematical notation using decision variables and parameters |
| | Model development | Develop the model; representing the objective function and constraints in mathematical notation using decision variables and parameters |
| | Model validation | Ensure the model is appropriate for evaluating all possible scenarios (i.e., different combinations of decision variables and parameters) |
| Optimization | Select optimization method | Choose an appropriate optimization method and algorithm on the basis of characteristics of the problem |
| | Perform optimization/ sensitivity analysis | Use the optimization algorithm to search for the optimal solution and examine the performance of the optimal solution for reasonable values of parameters |
| | Report results | Report the results of optimal solution and sensitivity analyses |
| | Decision making | Interpret the optimal solution and use it for decision making |

(www.informs.org) for a list of optimization software. The users need to specify, and more importantly understand, the parameters used as an input for these optimization algorithms (e.g., the termination criteria such as the level of convergence required or the number of iterations).

### *Perform Optimization/Sensitivity Analysis*

Optimization involves systematically searching the feasible region for values of decision variables and evaluating the objective function, consecutively, to find a combination of decision variables that achieve the maximum or minimum value of the objective function, using specific algorithms. Once the optimization algorithm has finished running, in some cases, the identified solution can be checked to verify that it satisfies the "optimality conditions" (i.e., Karush-Kuhn-Tucker conditions) [38], which are the mathematical conditions that define the optimality. Once the optimality is confirmed, the results need to be interpreted.

First, the results should be checked to see whether there is actually a feasible solution to the optimization problem, that is, whether there is a solution that satisfies all the constraints. If not, then the optimization problem needs to be adjusted (e.g., relaxing some constraints or adding other decision variables) to broaden the feasible solution space. If a feasible optimal solution has been found, the results need to be understood—this involves interpretation of the results to check whether the optimal solution, that is, values of decision variables, constraints, and objective function, makes sense.

It is also good practice to repeat the optimization with different sets of starting decision variables to ensure the optimal solution is the global optimum rather than local optimum. Sometimes, there may be multiple optimal solutions for the same problem (i.e., multiple combinations of decision variables that provide the same optimal value of objective function). For multiobjective optimization problems (i.e., problems with two or more conflicting objectives), Pareto optimal solutions are constructed from which optimal solution can be identified on the basis of subjective preferences of the decision maker [39,40].

It is good practice to run the optimization problem using different values of parameters, to verify the robustness of the optimization results. Sensitivity analysis is an important part of building confidence in an optimization model, addressing the structural and parametric uncertainties in the model by analyzing how the decision variables and optimum value react to changes in the parameters in the constraints and objective function, which ensures that the optimization model and its solution are good representations of the problem at hand.

Sometimes a solution may be the mathematically optimal solution to the specified mathematical problem, but may not be practically implementable. For example, the "optimal" set of nurse rosters may be unacceptable to staff as it involves breaking up existing teams, deploying staff with family responsibilities on night shifts, or reducing overtime pay to level where the employment is no longer attractive. Analysts should resist the temptation to spring their optimal solution on unsuspecting stakeholders, expecting grateful acceptance: rather, those affected by the model should be kept in the loop through the modeling process. The optimal solution may come as a surprise: it is important to allow space in the modeling process to explore fully and openly concerns about whether the "optimal" solution is indeed the one the organization should implement.

### *Report Results*

The final optimal solution, and if applicable, the results of the sensitivity analyses, should be reported. This will include the results of the optimum "objective function" achieved and the set of "decision variables" at which the optimal solution is found. Both the numerical values (i.e., the mathematical solution) and the physical interpretation, that is, the nontechnical text describing the meaning of numerical values, should be presented. The optimal solution identified can be contextualized in terms of how much "better" it is compared with the current state. For example, the results can be presented as improvement in benefits such as quality-adjusted life-years or reduction in costs.

It is often necessary to report the optimization method used and the results of the "performance" of the optimization algorithm, for example, number of iterations to the solution, computational time, and convergence level. This is important because it helps users understand whether a particular algorithm can be used "online" in a responsive fashion, or only when there is significant time available, for example, in a planning context.

Dashboards can be useful to visualize these benefits and communicate the insights gained from the optimal solution and sensitivity analyses.

### *Decision Making*

The final optimal solution and its implications for policy/service reconfiguration should be presented to all the relevant stakeholders. This typically involves a plan for amending the "decision variables" (e.g., shift patterns, screening frequency—see Table 2 for examples of decision variables—to those identified in the optimal solution). Before an optimal solution can be implemented, it will require getting the "buy-in" from the decision makers and all the stakeholders, for example, frontline staff such as nurses and hospital managers, to ensure that the numerical "optimal" solution found can be operationalized in a "real" clinical setting. It is important to have the involvement of decision makers throughout the whole optimization process to ensure that it does not become a purely numerical exercise, but rather something that is implemented in real life. After the decision is made, data should still be collected to assess the efficiency and demonstrate the benefits of the implementation of the optimal solution.

If decision makers are not directly involved in model development, they may choose not to implement the "optimal" solution as it comes from the model. This is because the model may fail to capture key aspects of the problem (e.g., the model may maximize aggregate health benefits but the decision maker may have a specific concern for health benefits for some disadvantaged subgroup). This does not (necessarily) mean that the optimization modeling has not been useful—enabling a decision maker to see how much health benefit must be sacrificed to satisfy her equity objective may prove to be beneficial toward the overall objective. After the decision is made, the story does not come to an end: data should continue to be collected to demonstrate the benefits of whatever solution is implemented, as well as guiding future decision making.

Table 3 presents the two different stages in optimization, that is, the modeling stage and the optimization stage, highlighting that model development is necessary before optimization can be performed. The goal of constrained optimization is to identify an optimal solution that maximizes or minimizes a particular objective subject to existing constraints.

## Relationship of Constrained Optimization to Related Fields

The use of constrained optimization can be classified into two categories. The first category is the use of constrained optimization as a decision-making tool. The simple illustration in the section "A Simple Illustration of a Constrained Optimization Problem" and all the examples in the section "Problems That Can Be Tackled with Constrained Optimization Approaches" are considered to fall under this category. The second category is the use of constrained optimization as an auxiliary analysis tool. In this category, optimization is an embedded tool and the results of which are often not the end results of a decision problem, but rather they are used as inputs for other analysis/modeling methods (e.g., optimization used in the multiple criteria decision making; in calibrating the inputs for health economic or dynamic simulation models; in machine learning and other statistical analysis methods such as solving regression models or propensity score matching).

As a decision-making tool, optimization is complementary to other modeling methods such as health economic modeling, simulation modeling, and descriptive, predictive (e.g., machine learning), and prescriptive analytics. Most modeling methods typically evaluate only a few different scenarios and determine a good scenario *within* the available options. In contrast, the aim of optimization methods is to efficiently identify the *best* solution overall, given the constraints. In the absence of using optimization methods, a brute force approach, in which all possible options are sequentially evaluated and the best solution is identified among them, might be possible for some problems. However, for most problems, it is too complex and too time consuming to identify and evaluate all possible options. Optimization methods and heuristic approaches might use efficient algorithms to identify the optimal solution quickly, which would otherwise be very difficult and time consuming.

Also, model development using these other methods might be necessary before optimization, especially in situations in which the objective function or constraints cannot be represented in a simple functional form. Thus, all models currently used in health care such as health economic models, dynamic simulation models, and predictive analytics (including machine learning) can be used in conjunction with optimization methods.

### *Constrained Optimization Methods Compared with Traditional Health Economic Modeling in Health Technology Assessments*

Constrained optimization methods differ substantially from health economic modeling methods traditionally used in health technology assessment processes [41]. The main difference between the two approaches is that traditional health economic modeling approaches, such as Markov models, are built to estimate the costs and effects of different diagnostic and treatment options. If decision makers are basing their judgments on modeling results, they may not formally consider the constraints and resource implications in the system. Constrained optimization methods provide a structured approach to optimize the decision problem and to present the best alternatives given an optimization criterion, such as constrained budget or availability of resources.

These differences have major implications. There is an opportunity to learn from optimization methods to improve health technology assessment processes [42–46]. Optimization is a valuable means of capturing the dynamics and complexity of the health system to inform decision making for several reasons. Constrained optimization methods can do the following:

1. *Explicitly take budget constraints into account*: Informed decision making about resource allocation requires an external estimate of the decision maker's willingness to pay for a unit of health outcome, the threshold. Decision making based on traditional health economic models then relies on the principle that by repeatedly applying the threshold to individual health technology assessment decisions, optimization of the allocation of health resources will be achieved.
   However, the focus of health economics is usually about relative efficiency without explicit consideration of budget because many jurisdictions do not explicitly implement a constrained budget nor do they use mechanisms to evaluate retrospectively cost-effectiveness of medical technologies currently in use.
2. *Address multiple resource constraints in the health system, such as resource capacity*: Constrained optimization methods also allow consideration of the effect of other constraints in the health system, such as capacity or short-term inefficiencies. Capacity constraints are usually neglected in health economic models. In health economics models, the outcomes are central to decision makers while the process to arrive at these outcomes is most of the time ignored.
   For health policy makers and health care planners, such capacity considerations are critical and cannot be neglected. Likewise, some technologies are known for short-term inefficiencies; for example, large equipment such as positron

emission tomography/magnetic resonance imaging devices are usually not taken into consideration. It takes a certain amount of time before a new device operates efficiently, and such short-term inefficiencies do influence implementation [47].

3. *Account for system behavior and decisions over time*: Traditional health economic models are often limited to informing a decision of a single technology at a single point in time. Health economic models with a clinical perspective, such as a whole disease model [48,49], or a treatment sequencing model, may allow the full clinical pathway to be framed as a constrained optimization problem that accounts for both intended and unintended consequences of health system interventions over time with feedback mechanisms in the system.
   Each combination of decisions within the pathway can be a potential solution, constrained by the feasibility of each decision, for example, the licensed indication for various treatments within a clinical pathway. These whole disease and treatment sequencing models can evaluate alternative guidance configurations and report the performance in terms of an objective function (cost per quality-adjusted life-year, net monetary benefit) [50,51].
4. *Inform decision makers about implementability of solutions that are recommended*: Health economic models are not typically constrained—it is assumed that resources are available as required and are thus affordable; similarly, the evidence used in the models come from controlled clinical settings, which are idealized settings compared with real clinical setting. An advantage of constrained optimization is the ability to obtain optimal solutions to decision problems and have sensitivity analyses performed. Such analyses inform decision makers about alternate realistic solutions that are feasible and close to the optimal solution.

Thus, in some sense, classic health economics models are "hypothetical" to illustrate the potential value as measured by a specific outcome with respect to cost, whereas optimization is focused on what can be achieved in an operational context. This suggests that constrained optimization methods have great value for informing decisions about the ability to implement a clinical intervention, program, or policy as they actually consider these constraints in the modeling approach.

### Constrained Optimization Methods Compared with Dynamic Simulation Models

DSM methods, such as system dynamics, discrete event simulation, and agent-based modeling, are used to design and develop mathematical representations, that is, formal models, of the operation of processes and systems. They are used to experiment with and test interventions and scenarios and their consequences over time to advance the understanding of the system or process, communicate findings, and inform management and policy design [30–32,52–54]. These methods have been broadly used in health applications [55–57].

Unlike constrained optimization methods, DSMs do not produce a specific solution. Rather they allow for the evaluation of a range of possible or feasible scenarios or intervention options that may or may not improve the system's performance. Constrained optimization methods, in general, seek to provide the answer to which of those options is the "best." Hence, the types of problems and questions that can be addressed with DSMs [30–32] are different from those that are addressed with optimization methods. However, both types of methods can be complementary to each other in helping us to better understand systems.

Traditionally, constrained optimization methods have served two distinct purposes in DSM development. 1) model calibration —fitting suitable model variables to past time series is discussed elsewhere [30–32]; 2) evaluating a policy's performance/effect relative to a criterion or set of criteria. However, the complexity of DSMs compared with simple analytic models may render exact constrained optimization approaches cumbersome, inappropriate, and potentially infeasible because of the large search space, for example, using methods of optimal control.

Because of this complexity, alternatives to exact approaches such as heuristic search strategies are available. Historically, these types of methods have been used in system dynamics and other DSMs. Because of their heuristic nature, there is no certainty of finding the "best" or optimal parameter set rather "good enough" solutions. Hence, the ranges assigned need careful consideration to get "good" solutions, that is, previous knowledge of sensible ranges both from knowledge about the system and from knowledge gained from model building.

Optimization is used as part of system dynamics to gain insight about policy design and strategy design, particularly when the traditional analysis of feedback mechanisms becomes risky due to the large number of loops in a model [58]. Similar procedures to evaluate policies and strategies can be used in discrete event simulation and agent-based modeling, for example, simulated annealing algorithms and genetic algorithms.

### Constrained Optimization Methods as Part of Analytics

Constrained optimization methods fall within the area of analytics as defined by the Institute for Operations Research and the

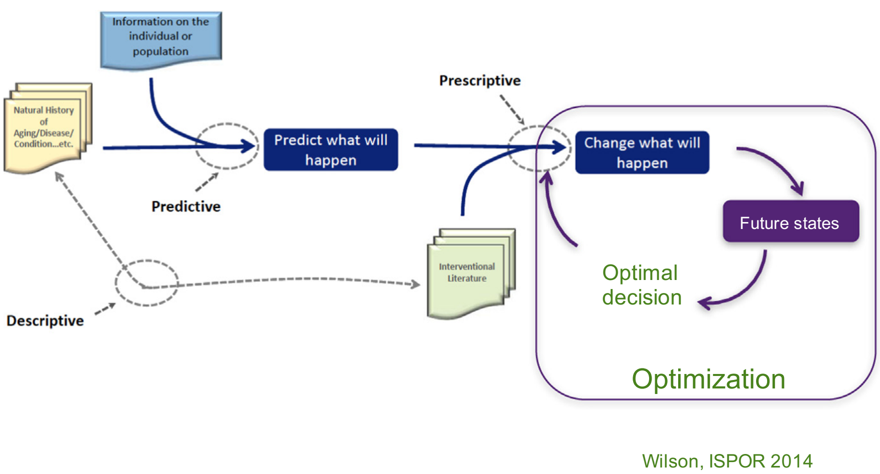


**Fig. 2 – Descriptive, predictive, and prescriptive analytics.**

Management Sciences (https://www.informs.org/Sites/Getting-Started-With-Analytics). Analytics can be classified into descriptive, predictive, and prescriptive analytics (Fig. 2), and discussed below. Constrained optimization methods are a special form of prescriptive analytics.

1. *Descriptive analytics* concern the use of historical data to describe a phenomenon of interest, with a particular focus on visual displays of patterns in the data. Descriptive analytics is differentiated from descriptive analysis, which uses statistical methods to test hypotheses about relationships among variables in the data. Health services research typically uses theory and concepts to identify hypotheses, and historical data are used to test these hypotheses using statistical methods. Examples may include natural history of aging, disease progression, evaluation of clinical interventions, policy interventions, and many others. Traditional health services for the most part falls within the area of descriptive analytics.
2. *Predictive analytics* and machine learning focus on forecasting the future states of disease or states of systems. With the increased volume and dimensions of health care data, especially medical claims and electronic medical record data, and the ability to link to other information such as feeds from personal devices and sociodemographic data, big data methods such as machine learning are garnering increased attention [59].

   Machine learning methods, such as predictive modeling and clustering, have an important intersection with constrained optimization methods. Machine learning methods are valuable for addressing problems involving classification, as well as data dimension reduction issues. And maybe most importantly, optimization often needs forecasts and estimates as inputs, which can be obtained from the results of machine learning algorithms. A discussion of machine learning methods is beyond the scope of this article.

   However, the interested reader will find a detailed introduction elsewhere [60,61]. Machine learning has the ability to "mine" data sets and discover trends or patterns. These are often valuable to establish thresholds or parameter values in optimization models, where it is otherwise difficult to determine the values. Constrained optimization can also leverage the ability of machine learning to reduce high dimensionality of data, say with thousands or millions of variables to key variables.
3. *Prescriptive analytics* uses the understanding of systems, both the historical and future based on historical (descriptive) and predictive analytics, respectively, to determine future course of action/decisions. Traditional (without optimization) clinical trials and interventions fall under the category of prescriptive analytics ("Change what will happen" in figure). Constrained optimization is a specialized form of prescriptive analytics because it helps with determining the *optimal* decision or course of action in the presence of constraints (https://www.informs.org/Sites/Getting-Started-With-Analytics/Analytics-Success-Stories).

## Summary and Conclusions

This is the first report of the ISPOR Constrained Optimization Methods Emerging Good Practices Task Force. It introduces readers to the application of constrained optimization methods to health care systems and patient outcomes research problems. Such methods provide a means of identifying the best policy choice or clinical intervention given a specific goal and given a specified set of constraints. Constrained optimization methods are already widely used in health care in areas such as choosing the optimal location for new facilities and making the most efficient use of operating room capacity.

However, they have been less widely used for decision making about clinical interventions for patients. Constrained optimization methods are highly complementary to traditional health economic modeling methods and DSM, providing a systematic and efficient method for selecting the best policy or clinical alternative in the face of large numbers of decision variables, constraints, and potential solutions. As health care data continues to rapidly evolve in terms of volume, velocity, and complexity, we expect that machine learning techniques will also be increasingly used for the development of models that can subsequently be optimized.

In this report, we introduce readers to the vocabulary of constrained optimization models and outline a broad set of models available to analysts for a range of health care problems. We illustrate the formulation of a linear program to maximize the health benefit generated in treating a mix of "regular" and "severe" patients subject to time and budget constraints and solve the problem graphically. Although simple, this example illustrates many of the key features of constrained optimization problems that would commonly be encountered in health care.

In the second task force report, we describe several case studies that illustrate the formulation, estimation, evaluation, and use of constrained optimization models. The purpose is to illustrate actual applications of constrained optimization problems in health care that are more complex than the simple example described in the current article and make recommendations on emerging good practices for the use of optimization methods in health care research.

## REFERENCES


[1] Rais A, Viana A. Operations research in healthcare: a survey. Int Transact Oper Res 2011;18:1–31.
[2] Araz C, Selim H, Ozkarahan I. A fuzzy multi-objective covering-based vehicle location model for emergency services. Comput Oper Res 2007;34:705–26.
[3] Bruni ME, Conforti D, Sicilia N, et al. A new organ transplantation location–allocation policy: a case study of Italy. Health Care Manag Sci 2006;9:125–42.
[4] Ndiaye M, Alfares H. Modeling health care facility location for moving population groups. Comput Oper Res 2008;35:2154–61.
[5] Verter V, Lapierre SD. Location of preventive health care facilities. Ann Oper Res 2002;110:123–32.
[6] Lagrange JL. Théorie des fonctions analytiques: contenant les principes du calcul différentiel, dégagés de toute considération d'infiniment petits, d'évanouissans, de limites et de fluxions, et réduits à l'analyse algébrique des quantités finies. Ve. Courcier, 1813.
[7] Dantzig GB. Programming of interdependent activities, II: mathematical model. Econometrica 1949;17:200–11.
[8] Dantzig GB. Linear Programming and Extensions. Princeton, NJ: Princeton University Press, 1963.
[9] Gass SI, Assad AA. An Annotated Timeline of Operations Research: An Informal History. Boston, MA: Springer Science & Business Media, 2005.
[10] Kirby MW. Operational Research in War and Peace: The British Experience from the 1930s to 1970. London, UK: Imperial College Press, 2003.
[11] Wood MK, Dantzig GB. Programming of interdependent activities, I: general discussion. Econometrica 1949;17:193–9.
[12] Burke EK, Causmaecker PD, Petrovic S, et al. Metaheuristics for handling time interval coverage constraints in nurse scheduling. Appl Artificial Intelligence 2006;20:743–66.
[13] Cheang B, Li H, Lim A, et al. Nurse rostering problems—a bibliographic survey. Eur J Oper Res 2003;151:447–60.
[14] Lin R-C, Sir MY, Pasupathy KS. Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: specific application to determining optimal resource levels in surgical services. Omega 2013;41:881–92.
[15] Lin R-C, Sir MY, Sisikoglu E, et al. Optimal nurse scheduling based on quantitative models of work-related fatigue. IIE Transact Healthcare Systems Eng 2013;3:23–38.

[16] Sir MY, Dundar B, Steege LMB, et al. Nurse–patient assignment models considering patient acuity metrics and nurses' perceived workload. J Biomedical Inform 2015;55:237–48.
[17] Alistar SS, Long EF, Brandeau ML, et al. HIV epidemic control—a model for optimal allocation of prevention and treatment resources. Health Care Manag Sci 2014;17:162–81.
[18] Earnshaw SR, Dennett SL. Integer/linear mathematical programming models. Pharmacoeconomics 2003;21:839–51.
[19] Earnshaw SR, Richter A, Sorensen SW, et al. Optimal allocation of resources across four interventions for type 2 diabetes. Med Decis Making 2002;22:s80–91.
[20] Lasry A, Sansom SL, Hicks KA, et al. A model for allocating CDC's HIV prevention resources in the United States. Health Care Manag Sci 2011;14:115–24.
[21] Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. J Health Econ 1996;15:641–53.
[22] Thomas BG, Bollapragada S, Akbay K, et al. Automated bed assignments in a complex and dynamic hospital environment. Interfaces 2013;43:435–48.
[23] Zaric GS, Brandeau ML. A little planning goes a long way: multilevel allocation of HIV prevention resources. Med Decis Making 2007;27:71–81.
[24] Lee EK, Wu TL. Disease diagnosis: Optimization-based methods. In: Floudas CA, Pardalos PM, eds. Encyclopedia of Optimization. (2nd) The Netherlands: Springer, 2009. p. 753–84.
[25] Liberatore MJ, Nydick RL. The analytic hierarchy process in medical and health care decision making: a literature review. Eur J Oper Res 2008;189:194–207.
[26] Ehrgott M, Güler Ç, Hamacher HW, et al. Mathematical optimization in intensity modulated radiation therapy. 4OR 2008;6:199–262.
[27] Lee H, Granata KP, Madigan ML. Effects of trunk exertion force and direction on postural control of the trunk during unstable sitting. Clin Biomech 2008;23:505–9.
[28] Bertsimas D, Farias VF, Trichakis N. Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. Oper Res 2013;61:73–87.
[29] Thokala P, Dixon S, Jahn B. Resource modelling: the missing piece of the HTA jigsaw? PharmacoEcon 2015;33:193–203.
[30] Marshall DA, Burgos-Liz L, IJzerman MJ, et al. Selecting a dynamic simulation modeling method for health care delivery research—part 2: report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force. Value Health 2015;18:147–60.
[31] Marshall DA, Burgos-Liz L, IJzerman MJ, et al. Applying dynamic simulation modeling methods in health care delivery research—the simulate checklist: report of the ISPOR Simulation Modeling Emerging Good Practices Task Force. Value Health 2015;18:5–16.
[32] Marshall DA, Burgos-Liz L, Pasupathy KS, et al. Transforming healthcare delivery: integrating dynamic simulation modelling and big data in health economics and outcomes research. PharmacoEconomics 2016;34:115–26.
[33] Hillier FS, Lieberman GJ. Introduction to Operations Research. US: McGraw-Hill Education, 2012.
[34] Minoux M. Mathematical Programming: Theory and Algorithms. Paris: John Wiley & Sons, 1986.
[35] Puterman ML. Markov Decision Processes: Discrete Stochastic Dynamic Programming. Hoboken, NJ: John Wiley & Sons, 2014.
[36] Winston WL, Goldberg JB. Operations Research: Applications and Algorithms. Belmont, CA: Duxbury Press, 2004.
[37] Kelley CT. Iterative Methods for Optimization. Philadelphia, PA: Siam, 1999.
[38] Kuhn H, Tucker A. Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press, 1951.
[39] Branke J, Deb K, Miettinen K, et al. Multiobjective Optimization: Interactive and Evolutionary Approaches. Germany: Springer-VErlag BErlin Heidelberg, 2008.
[40] Miettinen K. Nonlinear Multiobjective Optimization. New York, NY: Springer Science & Business Media, 2012.
[41] Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the Economic Evaluation of Health Care Programmes. Oxford, UK: Oxford University Press, 2015.
[42] Chalabi Z, Epstein D, McKenna C, et al. Uncertainty and value of information when allocating resources within and between healthcare programmes. Eur J Oper Res 2008;191:530–9.
[43] Epstein DM, Chalabi Z, Claxton K, et al. Efficiency, equity, and budgetary policies informing decisions using mathematical programming. Med Decis Making 2007;27:128–37.
[44] McKenna C, Chalabi Z, Epstein D, et al. Budgetary policies and available actions: a generalisation of decision rules for allocation and research decisions. J Health Econ 2010;29:170–81.
[45] Morton A. Aversion to health inequalities in healthcare prioritisation: a multicriteria optimisation perspective. J Health Econ 2014;36: 164–73.
[46] Morton A, Thomas R, Smith PC. Decision rules for allocation of finances to health systems strengthening. J Health Econ 2016;49:97–108.
[47] Van de Wetering G, Woertman WH, Adang EM. A model to correct for short-run inefficiencies in economic evaluations in healthcare. Health Econ 2012;21:270–81.
[48] Tappenden P, Chilcott J, Brennan A, et al. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. Value Health 2012;15:1127–36.
[49] Vanderby SA, Carter MW, Noseworthy T, et al. Modelling the complete continuum of care using system dynamics: the case of osteoarthritis in Alberta. J Simulation 2015;9:156–69.
[50] Kim E. Sequential drug decision problems in long-term medical conditions: a case study of primary hypertension. University of Sheffield, 2015.
[51] Tosh J. Simulation optimisation to inform economic evaluations of sequential therapies for chronic conditions: a case study in rheumatoid arthritis. University of Sheffield, 2015.
[52] Banks J. Handbook of Simulation. Wiley Online Library, 1998.
[53] Harrison JR, Lin Z, Carroll GR, et al. Simulation modeling in organizational and management research. Acad Manag Rev 2007;32:1229–45.
[54] Sokolowski JA, Banks CM. Principles of Modeling and Simulation: A Multidisciplinary Approach. Hoboken, NJ: John Wiley & Sons, 2011.
[55] Macal CM, North MJ, Collier N, et al. Modeling the transmission of community-associated methicillin-resistant *Staphylococcus aureus*: a dynamic agent-based simulation. J Translational Med 2014;12:1.
[56] Milstein B, Homer J, Briss P, et al. Why behavioral and environmental interventions are needed to improve health at lower cost. Health Aff 2011;30:823–32.
[57] Troy PM, Rosenberg L. Using simulation to determine the need for ICU beds for surgery patients. Surgery 2009;146:608–20.
[58] Sterman JD. Business dynamics: systems thinking and modeling for a complex world. New York, NY: McGraw Hill, 2000.
[59] Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. Value Health 2015;18:137–40.
[60] Kubat M. An Introduction to Machine Learning. Switzerland: Springer, 2015.
[61] Witten, IH, Frank E, Hall MA, Pal CJ. Data MIning: Practical Machine Learning Tools and Techniques. Cambridge, MA. 2017.