METHODS REPORT

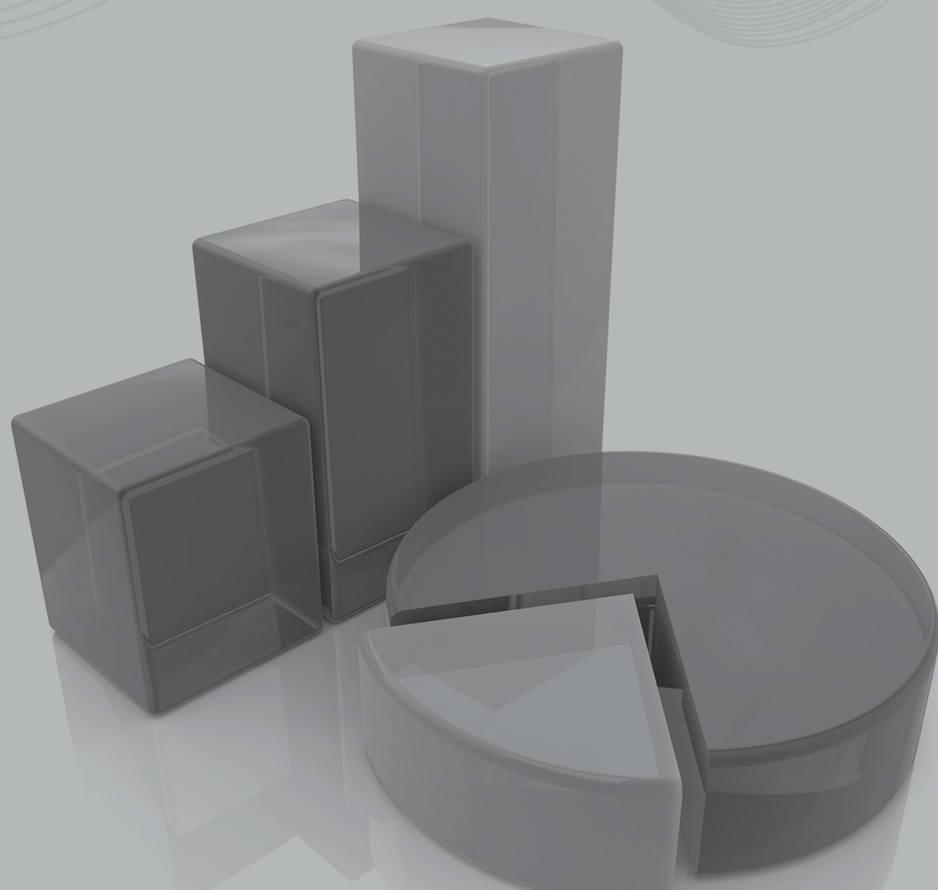
August 2019

Converge3

Integrating health, economic and equity evidence to inform policy

Economic Evaluation Methods for Allocating Resources within a Portfolio of Programs with Fixed Budgets and Additional Considerations

A Converge3 Methods Report



About this Report

This report was prepared by Converge3. Converge3 receives funding from the Province of Ontario. The views expressed in this report are those of Converge3 and do not necessarily reflect those of the Province of Ontario.

Suggested Citation

Converge3. Economic Evaluation Methods for Allocating Resources within a Portfolio of Programs with Fixed Budgets and Additional Considerations: A Converge3 Methods Report. Converge3: Toronto, Canada. 30 August 2019. Available from: https://converge3.ca/publication/methods-economic-evaluation-methods.

About Converge3

Converge3 is a policy research centre based in the Institute of Health Policy, Management and Evaluation at the University of Toronto that focuses on integrating health, economic and equity evidence to inform policy. The Centre is funded by the Province of Ontario and includes multiple partner organizations, including Li Ka Shing Knowledge Institute at St. Michael's Hospital, McMaster University, Ottawa Hospital Research Institute, ICES, Health Quality Ontario, Public Health Ontario, and the Ontario Ministry of Health.

Contact Information

Converge3
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College Street – 4th Floor
Toronto, Ontario M5T 3M6 Canada

https://converge3.ca

info@converge3.ca

@converge3_ca

Contents

Abstract

Cost-effectiveness analysis is an important input into resource allocation decisions in many health contexts and systems but seems to be used less frequently for decision making at meso- or micro-levels. One reason for this limited uptake may be the conventional decision rules of cost-effectiveness make some assumptions that are unrealistic for typical decisions made at these levels. Specifically, cost-effectiveness analysis assumes that programs are perfectly divisible, that there are constant returns to scale, that treatment options are independent, and that other considerations are irrelevant. Constrained optimization is an alternative method for performing economic evaluation that can guide selection of an optimal portfolio of programs with the constraint of a fixed budget and can address programs that are indivisible, introduce interactions between programs, and address additional constraints and objectives, including some equity concerns. Constrained optimization may be most relevant for a decision maker with a fixed budget who must decide which programs to fund or exclude within that constraint. We provide a hypothetical example and illustrate how the optimal selection of programs that a decision maker would fund depends on assumptions, objectives, and constraints. For some decision makers, such as directors of regional authorities or program managers, who often encounter such scenarios, the constrained optimization approach may offer more informed guidance for resource allocation decision problems compared with conventional cost-effectiveness analysis, provided that decision makers can articulate well-defined objectives and constraints.

Economic evaluation within health care

Resource allocation frequently requires simultaneous consideration of multiple outcomes. Within health care, these outcomes commonly include costs, health, and sometimes additional considerations such as equity. Cost is usually viewed as a constraint whereas health is typically viewed as an objective. Equity and other considerations are sometimes viewed as constraints and sometimes as objectives, which we will address further below. Health is often measured by integrating survival with quality of life in the form of a quality-adjusted life year (QALY).[1]

Economic evaluation is the formal process of comparing costs to outcomes and is commonly performed by expressing the trade-off as a ratio (in cost-effectiveness analysis) or as a difference (in cost-benefit analysis).[2] In cost-effectiveness analysis, the health effects may be measured in natural units (such as life years or as number of infections averted) or in QALYs, in which case the economic analysis is often termed a cost-utility analysis. In cost-benefit analysis, health effects are monetized (assigned a dollar value). Although both cost-effectiveness and cost-benefit analyses have strong arguments in their favour, the large majority of economic evaluations in health care use the methods of cost-effectiveness analysis. In this report, we similarly direct our main focus for conventional analyses on the methods of cost-effectiveness analysis.

The limitations of cost-effectiveness analysis

While cost-effectiveness analysis has proven to be a powerful method for assessing the efficiency of competing interventions within health care, the method has significant limitations that restrict its usefulness for guiding decisions when resources are being allocated within the context of fixed budgets for indivisible programs or when decision makers are considering multiple objectives alongside health. An indivisible program is one that can only be funded in its entirety or not at all; perfectly divisible programs can be funded in whole or in part, such as to specific portions of the population.[2] Stated more formally, cost-effectiveness analysis assumes that programs are perfectly divisible, that there are constant returns to scale (the ratio of costs to effects is constant, even if the program is only partially implemented), that treatment options are independent (the costs and effects gained by one program do not influence the costs and effects of another program), and that externalities are irrelevant (the costs and effects are the only considerations).[3,4] In practice, none of these assumptions are likely to be true. While several authors have illustrated how constrained optimization using mathematical programming can address such limitations, there has been limited uptake of such methods.[4-6] In this paper, we illustrate, using simple examples, how constrained optimization using mathematical programming can address concerns regarding indivisibility, independence, and considerations outside of health.

The decision rules of cost-effectiveness analysis

The decision rules of cost-effectiveness have generally been stated in two ways. The first formulation of the decision rule of cost-effectiveness analysis is to “maximize the health effects gained for each additional dollar spent, without exceeding a given threshold.” This rule can be written using mathematical notation, in which intervention Y is preferred over intervention X if:

$$\frac{C_Y - C_X}{E_Y - E_X} < \lambda$$

where C_Y and C_X represent the costs of program Y and X, respectively, E_Y and E_X represent the health effects of program Y and X, and λ represents the willingness-to-pay threshold. In this formulation, the decision about whether a cost-effectiveness analysis ratio is favourable or not only makes sense in reference to a standard, λ, which is the maximum amount that a decision maker would be willing to spend to buy an extra unit of effectiveness. When the perspective is that of a decision maker who is using a societal perspective (considering all cost and outcomes without regard to who is incurring them), λ is often called the “societal willingness-to-pay threshold." Several different methods for determining the value of λ have been discussed and are beyond the scope of this paper.[7] A frequently cited (although contested) threshold in Canada is $50,000/QALY.[8]

The second formulation of the decision rule of cost-effectiveness analysis is to "maximize the health effects gained without exceeding a given budget." This rule can be written using mathematical notation, following Stinnett and Paltiel,[4] as:

$$
\begin{aligned}
maximize&: \sum x_i e_i \\
subject\ to&: 0 \le x_i \le 1 \text{ (for all } i) \\
&\sum x_i c_i \le C \\
&\sum x_i \le 1 \text{ (for i } \in M)
\end{aligned}
$$

where x_i represents the proportion of program i that is implemented, e_i and c_i represent the health effects and costs of each program, C represents the total budget, and M represents the set of mutually exclusive programs under consideration. In plain language, this approach assumes that the decision maker is attempting to maximize the health effects by allocating resources across programs (line 1), while specifying that that programs can be implemented either fully or partially (line 2), that the total cost across programs cannot exceed the global budget (line 3), and that the sum of the portions across programs cannot be greater than 100% (line 4).

Note that this approach does not assume an external cost-effectiveness threshold (λ). However, the value of λ can be inferred from examining the effects gained relative to the cost of the last program funded under the budget constraint, termed the "shadow price." The reciprocal of the shadow price is the cost-effectiveness threshold.[3]

Opportunity costs

Central to the decision rule of cost-effectiveness analysis is the concept of opportunity cost, which states that decision makers should focus not only on the health effects gained from a given investment but also the health effects foregone by not investing in an alternative. This cost is implicit in the incorporation of λ into the cost-effectiveness decision rule. For example, consider a new drug for a condition that has an incremental cost-effectiveness ratio, compared with an existing drug, of $100,000/QALY. Applying the first formulation of the rule of cost-effectiveness analysis would indicate that a decision maker using a threshold value of $50,000/QALY (i.e. λ=$50,000/QALY) should not invest in the new drug since it exceeds the threshold or, stated another way, the extra dollars that are being proposed to "buy" better health for this drug could be better spent on an alternative (although unspecified) investment which is lower than the threshold and therefore represents better value for money. Stated yet another way, the opportunity cost of investing in the new drug is too high. In this paper, we will discuss opportunity "costs" primarily in terms of QALYs forgone; that is, decision makers may adopt certain rules that maximize specified objectives but that nevertheless lead to a lower aggregate number of QALYs gained. Making this trade-off explicit may be instructive for decision makers as a check on whether a specified objective is actually desirable.

Mathematical programming, constrained optimization, and portfolio optimization

The second formulation of the cost-effectiveness decision rule has been considered in the literature as a problem that can be addressed by mathematical programming, constrained optimization, or portfolio optimization.[3,4,9] Mathematical programming refers to construction of a problem in terms of mathematical objectives and constraints and the definition of a set of decision variables which are varied to meet the objective.[10] Mathematical programs frequently do not have simple algebraic solutions. Instead, solutions are found iteratively using search algorithms. These programs are commonly linear or quadratic, reflecting the terms in the objective function. Inputs may be integers, real numbers, or a mixture. In this paper, we restrict ourselves to linear programs and explore both real numbers and integers (restricted to 0 and 1) for certain input parameters. While programming and optimization are frequently used as synonyms, in this paper, we use constrained optimization to refer to the process of finding the best possible solution for a given mathematical program. More specifically, it is a systematic approach to finding the optimal value (i.e. minimum or maximum) of all possible solutions for a set of decision choices that are subject to well-defined preconditions. Constrained optimization has three components: 1) setting the objective (the problem that is to be solved); 2) defining the decision variables (the choices within the problem); and 3) establishing constraints (i.e. decision rules, such as budgetary or equity requirements, within which the solution of the problem should be found).[9] Portfolio optimization refers to a specific form of constrained optimization, in which a decision maker is selecting among several discrete options to maximize an objective.

In this paper, we adopt the perspective of a manager who is seeking to maximize the number of QALYs gained within a fixed budget and is choosing from a portfolio of possible programs. She is constrained by her budget cap rather than by the constraint of meeting a pre-defined willingness-to-pay threshold. Such scenarios may be most relevant to a director or manager in a Ministry of Health, for example, who is assigned a fixed budget and must decide which programs to fund or exclude within that constraint. To the extent that such circumstances mirror the real-life choices that decision makers face, understanding the limitations of cost-effectiveness analysis and the benefits and limitations of a constrained optimization approach could guide decision makers more effectively when selecting the methods that best address their resource allocation decision problems. We will take a primarily non-mathematical approach to explaining examples. Interested readers are invited to view details of the code used to generate all examples and the corresponding output in the Appendix. Examples were completed using Gurobi version 8.1.1 and Python version 3.7.2.

A hypothetical motivating example

Throughout this paper, we will use the following hypothetical example. Consider a manager who has a budget of $900,000 to allocate between 10 possible programs. The total cost of all potential programs is $2,850,000. Accordingly, the decision maker must select a portfolio of programs to fund and other programs to forego. Assume further that the programs differ in terms of their effectiveness (which is measured in QALYs) and their costs, that the programs are distributed among 5 sub-regions, that some programs target youth while other target adults, and that some programs vary in terms of their priority for the decision maker (this may represent an equity or fairness consideration, for example). Priority is

Table 1. Features of hypothetical programs for the motivating example

Program	Sub-region	Target audience	Effectiveness (QALYs gained)	Cost	Priority
A	1	Youth	3.0	$450,000	High
B	1	Adult	2.0	$85,000	Medium
C	2	Adult	6.3	$515,000	Low
D	2	Youth	4.2	$145,000	High
E	2	Adult	1.3	$30,000	Medium
F	3	Adult	6.0	$310,000	Low
G	4	Adult	2.1	$240,000	High
H	4	Youth	2.8	$114,000	Medium
I	5	Youth	4.5	$640,000	Low
J	5	Youth	1.7	$321,000	High
Total			**33.9**	**$2,850,000**	

categorized into 3 levels (High, Medium, and Low). The programs, labelled A to J, are summarized in Table 1. We also assume that there are no existing programs; that is, the status quo consists of zero costs and no QALYs. To begin, we assume that the programs are independent of each other; that is, the implementation of any program has no effect on the costs or QALYs of another program (we explore a scenario where program outcomes are dependent on each other below).

Decision rule 1: conventional cost-effectiveness analysis

Objective	Maximize QALYs
Constraint	Total cost cannot exceed the global budget

We first analyze our example using conventional rules of cost-effectiveness analysis. That is, we are aiming to maximize QALYs under a budget constraint. Note that our programs are mutually exclusive at the individual level (someone enrolled in one program cannot be enrolled in another) but not at the aggregate level; that is, our objective is not to determine the single most attractive program to fund (if that was our objective, we would calculate the incremental cost-effectiveness of each program in sequence, define the cost-effectiveness frontier, and select the intervention that is on the frontier which both buys the greatest number of QALYs and has an incremental cost-effectiveness ratio less than λ. Because our objective is, instead, to determine the optimal portfolio of programs to fund, we are not concerned with incremental cost-effectiveness ratios).

A decision maker who is interested in maximizing QALYs would first invest in the program that yields the greatest number of QALYs per dollar spent – that is, the most efficient program. She would then continue to invest in programs or portions of programs until she has exhausted her global budget. In our example, the order of programs is determined by calculating the cost per QALY gained (Column 5 of Table 2) and then ranking the programs from lowest to highest according to this calculation.

Under a total budget of $900,000, programs E, D, H, B, and F are fully implemented. However, there are insufficient funds to fully implement Program C. Because we are using the decision rules of cost-effectiveness, we will make two assumptions. First, we assume that program C is divisible; that is, the program manager can direct that program C be only partially implemented (in this case, 42% of the total funds required for Program C should be allocated). We are also assuming that 42% of the benefits of Program C are realized. Accordingly the decision rule of cost-effectiveness will involve an expenditure of the entire budget of $900,000 and a gain of 18.94 QALYs. Note that this is the maximum number

Table 2. Programs sorted by cost-effectiveness

Rank	Program	QALY	Cost	Cost/QALY	Cumulative Budget	Cumulative QALYS	Proportion of Program Funded
1	E	1.3	$30,000	$23,077	$30,000	1.30	100%
2	D	4.2	$145,000	$34,524	$175,000	5.50	100%
3	H	2.8	$114,000	$40,714	$289,000	8.30	100%
4	B	2.0	$85,000	$42,500	$374,000	10.30	100%
5	F	6.0	$310,000	$51,667	$684,000	16.30	100%
6	C	6.3	$515,000	$81,746	$900,000	18.94	42%
7	G	2.1	$240,000	$114,286			0%
8	I	4.5	$640,000	$142,222			0%
9	A	3	$450,000	$150,000			0%
10	J	1.7	$321,000	$188,824			0%

of QALYs that can be gained when allocating resources across these programs.

We return to the two important assumptions. First, that Program C is divisible and second, that funding Program C at 42% will also yield 42% of the benefits; that is, we are assuming that Program C yields constant returns to scale. In practice, both of these assumptions may be problematic. Assumptions of divisibility and constant returns to scale may be particularly inappropriate, for example, for programs that have large fixed, relative to variable, costs. Furthermore, programs that hire staff on salaries (rather than on a contractual basis) may similarly be indivisible. The question of constant return to scale becomes moot if we assume indivisibility (since there is only one scale).

The last program funded in this approach has a cost-effectiveness ratio of $81,746, which represents the threshold value for this portfolio of programs. Note, however, that this value may be higher than the societal willingness-to-pay threshold. If, for example, our societal value of λ is $50,000/QALY, we might impose a constraint that no program with a cost-effectiveness ratio greater than this should be funded. The rationale for a decision rule such as this is that money spent on these programs would be optimally reinvested on other portfolios outside of this funding platform; that is, the opportunity costs of investing with this program are too high given competing demands in other portfolios, sectors, or ministries. Of course, it is uncertain whether such reinvestments would be made in practice. Imposing a further restriction to only fund programs that have a cost-effectiveness ratio below $50,000/QALY, programs E, D, H, and B would be funded and the total budget invested would be $374,000 (the program manager would need to return $526,000 to be invested by other managers). Although consideration of efficiency at a societal level is an important consideration, we leave it aside for the remainder of this paper and focus solely on maximizing objectives within a portfolio of programs.

Decision rule 2: constrained optimization with indivisible programs

Objective	Maximize QALYs
Constraint	Total cost cannot exceed the global budget Funding variables are binary

Next, we examine how the resource allocation decision changes under the assumption that the programs are indivisible. Our objectives are similar to those under Decision Rule 1: We wish to maximize the number of QALYs subject to a budget constraint. However, we also now have an additional constraint, namely that each program is either funded or it is not funded; no program can be only partially funded.

Under this decision rule, programs B, D, F, G and H are funded. The total cost of funding all programs is $894,000 – note that the remaining $6000 cannot be spent since programs are indivisible and no program costs less than $6000. In practice, this $6000 might be spent on other programs; however, the only relevant consideration for this example is that it is not invested within this portfolio. Note that program E, which was the most cost-effective program under Decision Rule 1, is no longer funded. Although program E represents an efficient use of resources, it also has a small overall cost. Once we are constrained to considering how to maximize QALY gains across 10 indivisible programs, we must consider both costs and budgets. In this example, the optimal solution, perhaps surprisingly, excludes the single most cost-effective program (under a different budget constraint, program E would be included again).

The total QALY gain under Decision Rule 2 is 17.10 QALYs, which is lower than the theoretical maximum number of QALYs (18.94) that could be gained under Decision Rule 1. However, because there is no change in the objective (our goal is still to maximize QALYs), we do not consider this a QALY loss. Rather, we consider this an assessment of how many QALYs can be gained under more realistic circumstances. Accordingly, we will use this gain (17.10) as the benchmark QALY measure to calculate the number of QALYs foregone under different objective functions.

Decision rule 3: constrained optimization with regional equity

Objective	Maximize QALYs
Constraint	Total cost cannot exceed the global budget Funding variables are binary Each region has 1 or more programs

Our third decision rule assumes that our decision maker wishes to ensure that there is at least one program funded in each region. This may be seen as a form of horizontal equity, if people in different regions have similar health needs for the programs. Recall that Region 3 has only one program, program F; hence, this rule guarantees that program F will be funded. Each of the other regions has 2 or 3 programs. We continue with a model that analyzes programs as indivisible with a global budget constraint and with an objective of QALY maximization.

With these constraints, the programs that maximize the number of QALYs gained are B, E, F, H, and J. The total expenditure is $860,000 and the number of QALYs gained is 13.80. Recall that Decision Rule 2 yielded 17.10 QALYs; thus, there is a “cost” of 3.30 QALYs in order to meet the regional equity objective. In other words, there is a trade-off here between efficiency (maximizing QALYs) and horizontal equity (making sure each region gets a program). Our decision maker would need to decide whether the opportunity cost (a 19% loss in the number of QALYs that could be gained) is acceptable or not.

Decision rule 4: constrained optimization with at least 3 youth services

Objective	Maximize QALYs
Constraint	Total cost cannot exceed the global budget Funding variables are binary Fund at least 3 youth programs

Our fourth decision rule assumes that our decision maker wishes to ensure that at least three youth programs are funded. This may be an equity concern (for example if youth are perceived to be at higher need for services than adults) but it could also be a constraint that is motived by considerations other than equity, such as political demands. As before, we continue with a model that analyzes programs as indivisible with a global budget constraint and with an objective of QALY maximization.

With these constraints, the programs that maximize the number of QALYs gained are D, F, H, and J. In contrast to each of our previous decision rules, only 4 programs are funded rather than 5; of these 4, only one is a program for adults. The total expenditure is $890,000 and the number of QALYs gained is 14.70. Compared to Decision Rule 2, 2.40 QALYs are foregone in order to meet this objective.

What happens if we attempt to combine Decision Rules 3 and 4 – that is, if we impose a constraint that each region must have at least one program funded and that at least 3 youth programs must be funded? No solution is possible for this scenario, illustrating that it is possible to impose too many constraints and create “impossible” scenarios.

Decision rule 5: constrained optimization with weighted QALYs

Objective	Maximize weighted QALYs
Constraints	Total cost cannot exceed the global budget Funding variables are binary
Comment	Weights for low, moderate and high programs are 1.0, 1.2, and 1.6

So far, our decision rules have conceptualized decision maker concerns other than QALY maximization as constraints. However, it is also possible to consider such concerns within the objective function. To illustrate, our fifth decision rule assumes that our decision maker has greater priority for some programs rather than others. If this prioritization reflected levels of need, this would be an example of vertical equity (prioritizing QALY gains according to level of need). We assume that moderate priority programs are valued more highly than low priority programs, with QALYs gained by such programs given a 20% increased weight (a relative weight of 1.20). We similarly assume that high priority programs are assigned a weight of 1.60. Note that weighting the gain in QALYs by a factor, w, is conceptually and algebraically equivalent to saying that the willingness to pay for a QALY is higher by the same factor (to see this, recall that an intervention is cost-effective if $\Delta C/\Delta E < \lambda$, or equivalently $\Delta C - \lambda \Delta E < 0$. Substituting the weighted QALY gain, $w\Delta E$, changes this equation to $\Delta C - \lambda w \Delta E < 0$, which is the same equation as if we had substituted the weighted willingness to pay threshold, $w\lambda$, for λ). Our model analyzes programs as indivisible with a global budget constraint and with an objective of weighted QALY maximization.

With these constraints, the programs that maximize the number of QALYs gained are B, C, D, E, and H. The total expenditure is $889,000 and the number of unweighted QALYs gained is 16.60. Compared to Decision Rule 2, 0.50 unweighted QALYs are foregone in order to meet this objective.

Decision rule 6: constrained optimization with programs that are not independent

Objective	Maximize QALYs
Constraints	Total cost cannot exceed the global budget
	Funding variables are binary
Comment	Interaction between programs A and B: increase of 1.0 QALYs and decrease of $200,000

Recall that an additional assumption of the decision rules of cost-effectiveness is that programs are independent. Our constrained optimization approach can also be used to explore how to relax such assumptions. Consider, for example, programs A and B, both of which are in the same region but one of which targets youth and the other targets adults. Perhaps the programs are offered by the same agency, such that funding both enables families to be treated together and that treating families results in additional QALY gains beyond treating family members individually. We might therefore assume that the QALY gains from funding programs A and B together are greater than the sum of funding either program independently. In our example, we assume that there is an additional gain of 1.0 QALYs if both programs are funded. We might similarly assume that funding both programs might be less expensive than the sum of the cost of each program because of economies of scale. We assume that there are cost savings of $200,000 from such an approach. We include these effects as “interaction” terms in the objective and constraint functions. In this scenario, programs A, B, D, E, and F are funded. The expenditure is $820,000 and the total number of QALYs gained is 17.50. We do not calculate a QALY loss in this scenario since it is not directly comparable to other decision rules.

Summary

Table 3 summarizes the scenarios according to the decision rules, the programs that are funded under each rule, the total expenditure, the total number of QALYs across funded programs, and, when relevant, the QALY loss relative to the maximum possible number of QALYs gained if indivisible programs were funded solely with the objective of maximizing QALYs.

Several points are noteworthy. First, the optimal portfolio varies considerably according to the assumptions, constraints, and objectives that are selected. In our hypothetical example, no program is always selected although one (program I) is never selected. Second, the number of optimal programs can also vary. Third, the total expenditure under the assumption of indivisibility never reaches the maximum budget. Fourth, apart from decision rule 1 (the traditional rule of cost-effectiveness), there is no simple algorithmic approach to deriving the optimal allocation for the other decision rules. In general, optimization software that incorporates integer linear programming is needed to replicate these results.

Table 3. Summary of decision rules and optimal solutions

Rule	Comment	Programs (X denotes funded)										Expenditures	QALYs	QALY loss
		A	B	C	D	E	F	G	H	I	J			
1	QALY maximization, divisible programs, constant return to scale		X	42%	X	X	X		X			$900,000	18.94	N/A
2	QALY maximization, indivisible		X		X		X	X	X			$894,000	17.10	0.00
3	QALY maximization, indivisible, at least 1 per region		X			X	X		X		X	$860,000	13.80	3.30
4	QALY maximization, indivisible, at least 3 youth programs				X		X		X		X	$890,000	14.70	2.40
5	Weighted QALY maximization, indivisible		X	X	X	X			X			$889,000	16.60	0.50
6	QALY maximization, indivisible, interaction between A and B (gain 1 QALY and save $200,000)	X	X		X	X	X					$820,000	17.50	N/A

Discussion

Cost-effectiveness analysis is an important input into resource allocation decisions in many health contexts and systems, including, for example, public funding of drugs in Canada.[11] However, cost-effectiveness analyses seem to be used less frequently for decision making at meso- or micro-levels, such as by regional authorities or by program managers.[12] This limited uptake may be due to several reasons, including the lack of available data and accompanying analyses. However, it may also be because the decision rules of cost-effectiveness analyses, even when framed as health maximization within a constrained budget, are not optimal for such decision making. Two features of the conventional decision rules of cost-effectiveness are particularly important in this regard – the assumption of perfect divisibility of health care programs and the neglect of additional policy objectives beyond maximizing health.

Integer and mixed mathematical programming and constrained optimization methods have been proposed as solutions to the problem of divisibility of health care programs for over 20 years.[4,5,13,14] This approach has several strengths, most notably that it provides an optimal solution under conditions of indivisibility. Importantly, this optimal solution may be significantly different than the solution suggested by cost-effectiveness analysis. In addition, explicit consideration of additional objectives and constraints more accurately reflects the reported reality of many health care decisions. In addition to the constraints that we have illustrated, other applications have addressed issues such as workforce capacity and return to scale.[15]

While there is sufficient interest in constrained optimization within health to motivate a major academic society to convene a group to describe best practices in conducting such analyses, there are relatively few published applications.[9,10] We believe that three reasons may explain this low uptake. First, both decision makers and analysts may be unfamiliar with the methods or software used for optimization, which has historically been used in operations research and industrial engineering much more frequently than in health economics and health services research. Second, many funding decisions are made without consideration of how the investment fits within a constrained budget. For example, jurisdictions in Canada typically do not have fixed budgets for pharmaceuticals and do not evaluate the entire portfolio (the whole drug formulary) when deciding which drugs to list. Thus, an important, but likely under-appreciated, requirement for economic evaluation to be useful is that the method should be "fit to purpose" – that is, the optimal method depends on the decision problem under consideration. We believe that the methods of constrained optimization are a better fit for many decisions made by sub-national or regional decision makers. Third, the use of constrained optimization techniques may introduce new requirements for quantitative data or explicit decision making. In our examples, decision makers had to specify the number of youth services that were to be funded (Decision Rule 4) and the relative importance weights for QALYs for low, medium, and high priority programs (Decision Rule 5). Many advocate for measuring such weights to reflect societal preferences but measurement may be time-consuming or imprecise and may also be unfamiliar to decision makers, further limiting uptake.[16]

Real-world applications may be more complex than our straightforward and relatively simple examples. For example, additional analyses may be required to address the stochastic nature of input parameters (that is, the uncertainty in the constrained optimization model) as well as addressing sensitivity analyses regarding model structures.[17,18] Additional methodological advances could address the dynamic nature of an optimization model in which inputs and parameters change over time, individuals move between programs, or where programs have more complex interdependencies than the relatively simple example we presented in Decision Rule 6.

We have illustrated two ways that constrained optimization can incorporate equity concerns into economic evaluations. The first method (Decision Rule 3) illustrates how this method could be used to address horizontal equity by addressing the needs of individuals with similar levels of need but differential access.[19] In our example, we assumed that the decision maker would want at least one program in each region. The second method illustrates how this method could be used to address vertical equity by prioritizing individuals with greater levels of need, illustrated as choice between youth and adults (Decision Rule 4) or by a prioritization exercise (Decision Rule 5).[20,21] Decision Rules 3 and 4 conceptualize equity as an additional constraint on the primary objective, which is efficiency (maximizing health). Decision Rule 5 conceptualizes equity as a component of the objective, alongside health and quality of life.[22] Similar considerations could consider issues such as the number of people affected by a program. Different methods to combine multiple objectives are also possible. For example, objectives could be linearly weighted and combined, they could be assigned priorities and optimized in order of the priorities, or some combination of methods.

All resouce allocation decisions are ultimately political, rather than technical, problems. Hence, any technical solution will only be useful as far as it is able to quantify decision makers' objectives and constraints and provide outputs that have face validity and are interpretable. In some situations, decision makers may not be able to express their concerns in quantitative terms; thus, constrained optimization solutions will incompletely characterize all of the important considerations. In other situations, constrained optimization may not find exact solutions to problems. In contrast to our examples, where an exact solution was always found, solving some problems may be impractical or impossible. In such situations, heuristic approaches can be used to provide a set of best-fitting approximate solutions. We also demonstrate the trade-off between efficacy and equity, in that equity-optimal solutions yield fewer QALYs than the most efficient solution – decision makers will need to decide what level of trade-off is acceptable. In this and indeed in all situations, including when constrained optimization yields exact solutions, decision makers will need to decide how to incorporate these results into their deliberations.

References

1 Drummond M, Brixner D, Gold M, et al. Toward a consensus on the QALY. *Value Health*. 2009;12 Suppl 1:S31-35.
2 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Fourth ed. New York: Oxford University Press; 2015.
3 Epstein DM, Chalabi Z, Claxton K, Sculpher M. Efficiency, equity, and budgetary policies: informing decisions using mathematical programming. *Med Decis Making*. 2007;27(2):128-137.
4 Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *JHealth Econ*. 1996;15(5):641-653.
5 Sendi P, Al MJ. Revisiting the decision rule of cost–effectiveness analysis under certainty and uncertainty. *Soc Sci Med*. 2003;57(6):969-974.
6 Al MJ, Feenstra TL, Hout BA. Optimal allocation of resources over health care programmes: dealing with decreasing marginal utility and uncertainty. *Health Econ*. 2005;14(7):655-667.
7 Woods B, Revill P, Sculpher M, Claxton K. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. *Value Health*. 2016;19(8):929-935.
8 Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. *Expert Rev Pharmacoecon Outcomes Res*. 2008;8(2):165-178.
9 Crown W, Buyukkaramikli N, Thokala P, et al. Constrained Optimization Methods in Health Services Research-An Introduction: Report 1 of the ISPOR Optimization Methods Emerging Good Practices Task Force. *Value Health*. 2017;20(3):310-319.
10 Crown W, Buyukkaramikli N, Sir MY, et al. Application of Constrained Optimization Methods in Health Services Research: Report 2 of the ISPOR Optimization Methods Emerging Good Practices Task Force. *Value Health*. 2018;21(9):1019-1028.
11 Rocchi A, Miller E, Hopkins RB, Goeree R. Common Drug Review recommendations: an evidence base for expectations? *Pharmacoeconomics*. 2012;30(3):229-246.
12 Waithaka D, Tsofa B, Barasa E. Evaluating healthcare priority setting at the meso level: A thematic review of empirical literature. *Wellcome Open Res*. 2018;3:2-2.
13 Earnshaw SR, Dennett SL. Integer/linear mathematical programming models: a tool for allocating healthcare resources. *Pharmacoeconomics*. 2003;21(12):839-851.
14 Birch S, Gafni A. Cost effectiveness/utility analyses. Do current decision rules lead us to where we want to be? *J Health Econ*. 1992;11(3):279-296.
15 van Baal P, Morton A, Severens JL. Health care input constraints and cost effectiveness analysis decision rules. *Soc Sci Med*. 2018;200:59-64.
16 Bleichrodt H, Diecidue E, Quiggin J. Equity weights in the allocation of health care: the rank-dependent QALY model. *J Health Econ*. 2004;23(1):157-171.
17 McKenna C, Chalabi Z, Epstein D, Claxton K. Budgetary policies and available actions: A generalisation of decision rules for allocation and research decisions. *J Health Econ.* 2010;29(1):170-181.
18 Chalabi Z, Epstein D, McKenna C, Claxton K. Uncertainty and value of information when allocating resources within and between healthcare programmes. *Eur J Oper Res.* 2008;191(2):530-539.
19 Culyer AJ. Equity - some theory and its policy implications. *JMedEthics.* 2001;27(4):275-283.
20 Soares MO. Is the QALY blind, deaf and dumb to equity? NICE's considerations over equity. *Br Med Bull.* 2012;101:17-31.
21 Mooney G. Vertical equity in health care resource allocation. *Health Care Anal.* 2000;8(3):203-215.
22 Morton A. Aversion to health inequalities in healthcare prioritisation: a multicriteria optimisation perspective. *J Health Econ*. 2014;36:164-173.

Appendix

Optimization code

```
#maximize QALYs
from gurobipy import *

try:

    #Create a model object
    m = Model()

    # Create variables indicating whether a program is funded or not
    a = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="A")
    b = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="B")
    c = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="C")
    d = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="D")
    e = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="E")
    f = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="F")
    g = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="G")
    h = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="H")
    i = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="I")
    j = m.addVar(lb=0,ub=1,vtype=GRB.CONTINUOUS, name="J")

    # QALYs, costs, priority
    ina=[3,   450000, 0]
    inb=[2,    85000, 2]
    inc=[6.3, 515000, 1]
    ind=[4.2, 145000, 0]
    ine=[1.3,  30000, 2]
    inf=[6,   310000, 1]
    ing=[2.1, 240000, 0]
    inh=[2.8, 114000, 2]
    ini=[4.5, 640000, 1]
    inj=[1.7, 321000, 0]

    Budget = 900000

    # DECISION RULE 1: maximize QALYs, divisible
    m.setObjective(a*ina[0]+b*inb[0]+c*inc[0]+d*ind[0]+e*ine[0]+f*in-
f[0]+g*ing[0]+h*inh[0]+i*ini[0]+j*inj[0] , GRB.MAXIMIZE)

    # Add budget constraint:
    c1=m.addConstr(a*ina[1]+b*inb[1]+c*inc[1]+d*ind[1]+e*ine[1]+f*in-
f[1]+g*ing[1]+h*inh[1]+i*ini[1]+j*inj[1]<=Budget,name="budget")

      # Optimize
    m.optimize()
      # Calculate actual budget
    budget=quicksum([a.X*ina[1], b.X*inb[1],  c.X*inc[1],
               d.X*ind[1],  e.X*ine[1], f.X*inf[1],  g.X*ing[1],
h.X*inh[1],  i.X*ini[1],  j.X*inj[1] ])
    for v in m.getVars():
        if v.x!=0:
            print(v.varName, v.x)

    print('Maximum QALYs:', m.objVal)
    print("Budget",budget.getValue())
      # DECISION RULE 2: recast variables to be binary
```

```
    for v in m.getVars():
        v.setAttr("Vtype",GRB.BINARY)

      # reset and rerun
    m.reset(0)
    m.optimize()
    budget=quicksum([a.X*ina[1], b.X*inb[1],  c.X*inc[1],
               d.X*ind[1],  e.X*ine[1], f.X*inf[1],  g.X*ing[1],
h.X*inh[1],  i.X*ini[1],  j.X*inj[1] ])
    for v in m.getVars():
        if v.x!=0:
            print(v.varName, v.x)

    print('Maximum QALYs:', m.objVal)
    print("Budget",budget.getValue())


      # DECISION RULE 2: regions (horizontal equity)
    #regions
    reg1=a+b
    reg2=c+d+e
    reg3=f
    reg4=g+h
    reg5=i+j

    r1= m.addConstr(reg1>=1)
    r2=m.addConstr(reg2>=1)
    r3=m.addConstr(reg3>=1)
    r4=m.addConstr(reg4>=1)
    r5=m.addConstr(reg5>=1)

    m.reset(0)
    m.optimize()
    budget=quicksum([a.X*ina[1], b.X*inb[1],  c.X*inc[1],
               d.X*ind[1],  e.X*ine[1], f.X*inf[1],  g.X*ing[1],
h.X*inh[1],  i.X*ini[1],  j.X*inj[1] ])
    for v in m.getVars():
        if v.x!=0:
            print(v.varName, v.x)
    print('Maximum QALYs:', m.objVal)
    print("Budget",budget.getValue())

      # DECISION RULE 4: Youth
    m.remove(r1)
    m.remove(r2)
    m.remove(r3)
    m.remove(r4)
    m.remove(r5)

    #youth
    youth=a+d+h+i+j
    y1=m.addConstr(youth>=3)
    m.reset(0)
    m.optimize()
    budget=quicksum([a.X*ina[1], b.X*inb[1],  c.X*inc[1],
               d.X*ind[1],  e.X*ine[1], f.X*inf[1],  g.X*ing[1],
h.X*inh[1],  i.X*ini[1],  j.X*inj[1] ])
    for v in m.getVars():
```

```
        if v.x!=0:
            print(v.varName, v.x)
    print('Maximum QALYs:', m.objVal)
    print("Budget",budget.getValue())
    m.remove(y1)

      # DECISION RULE 5: WEIGHTED QALYs
    w=[1,1.2,1.6]

    m.setObjective(a*ina[0]*w[ina[2]]
                   +b*inb[0]*w[inb[2]]
                   +c*inc[0]*w[inc[2]]
                   +d*ind[0]*w[ind[2]]
                   +e*ine[0]*w[ine[2]]
                   +f*inf[0]*w[inf[2]]
                   +g*ing[0]*w[ing[2]]
                   +h*inh[0]*w[inh[2]]
                   +i*ini[0]*w[ini[2]]
                   +j*inj[0]*w[inj[2]]
                   , GRB.MAXIMIZE)
    m.reset(0)
    m.optimize()
    budget=quicksum([a.X*ina[1], b.X*inb[1],  c.X*inc[1],
d.X*ind[1],  e.X*ine[1], f.X*inf[1],  g.X*ing[1],  h.X*inh[1],  i.X*ini[1],
j.X*inj[1] ])

    u=quicksum([a.x*ina[0],b.x*inb[0],c.x*inc[0],
d.X*ind[0],  e.X*ine[0], f.X*inf[0],  g.X*ing[0],  h.X*inh[0],  i.X*ini[0],
j.X*inj[0] ])

    for v in m.getVars():
        if v.x!=0:
            print(v.varName, v.x)

    print('Maximum QALYs:', m.objVal)
    print("Budget",budget.getValue())
    print("Unweighted QALYs",u.getValue())

      # DECISION RULE 6: RELAX DEPENDENCE ASSMPTION
    interactq=1
    interactc=-200000
    m.setObjective(a*ina[0]+b*inb[0]+c*inc[0]+d*ind[0]+e*ine[0]+f*in-
f[0]+g*ing[0]+h*inh[0]+i*ini[0]+j*inj[0] + a*b*interactq , GRB.MAXIMIZE)

m.remove(c1)
    c1=m.addConstr(a*ina[1]+b*inb[1]+c*inc[1]+d*ind[1]+e*ine[1]+f*in-
f[1]+g*ing[1]+h*inh[1]+i*ini[1]+j*inj[1]+a*b*interactc<=Budget,name="bud-
get")

m.reset(0)
m.optimize()
budget=quicksum([a.X*ina[1], b.X*inb[1],  c.X*inc[1],
d.X*ind[1],  e.X*ine[1], f.X*inf[1],  g.X*ing[1],  h.X*inh[1],  i.X*ini[1],
j.X*inj[1] , a*b*interactc])

    for v in m.getVars():
        if v.x!=0:
            print(v.varName, v.x)
```

```
    print('Maximum QALYs:', m.objVal)
    print("Budget",budget.getValue())

except GurobiError:
    print('Error reported')
```

Program output

```
Python 3.7.2 (v3.7.2:9a3ffc0492, Dec 24 2018, 02:44:43)
[Clang 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license()" for more information.
>>>
============= RESTART: /Users/ahmedbayoumi/Dropbox/converge3.py
=============
Academic license - for non-commercial use only
Optimize a model with 1 rows, 10 columns and 10 nonzeros
Coefficient statistics:
  Matrix range     [3e+04, 6e+05]
  Objective range  [1e+00, 6e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [9e+05, 9e+05]
Presolve time: 0.10s
Presolved: 1 rows, 10 columns, 10 nonzeros

Iteration    Objective       Primal Inf.    Dual Inf.      Time
       0    3.9000000e+01   4.640000e+02   0.000000e+00      0s
       1    1.8942330e+01   0.000000e+00   0.000000e+00      0s

Solved in 1 iterations and 0.20 seconds
Optimal objective  1.894233010e+01
B 1.0
C 0.41941747572815535
D 1.0
E 1.0
F 1.0
H 1.0
Maximum QALYs: 18.942330097087382
Budget 900000.0
Optimize a model with 1 rows, 10 columns and 10 nonzeros
Variable types: 0 continuous, 10 integer (10 binary)
Coefficient statistics:
  Matrix range     [3e+04, 6e+05]
  Objective range  [1e+00, 6e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [9e+05, 9e+05]
Found heuristic solution: objective 13.3000000
Presolve removed 0 rows and 1 columns
Presolve time: 0.00s
Presolved: 1 rows, 9 columns, 9 nonzeros
Variable types: 0 continuous, 9 integer (9 binary)

Root relaxation: objective 1.886893e+01, 1 iterations, 0.00 seconds

    Nodes    |    Current Node    |     Objective Bounds      |     Work
 Expl Unexpl |  Obj  Depth IntInf | Incumbent    BestBd   Gap | It/Node
Time

     0     0   18.86893    0    1   13.30000   18.86893  41.9%     -    0s
```

```
H    0     0                        16.6000000   18.86893  13.7%     -    0s
H    0     0                        17.1000000   18.86893  10.3%     -    0s
     0     0 infeasible    0         17.10000   17.10000  0.00%     -    0s

Explored 1 nodes (1 simplex iterations) in 0.59 seconds
Thread count was 4 (of 4 available processors)

Solution count 3: 17.1 16.6 13.3

Optimal solution found (tolerance 1.00e-04)
Best objective 1.710000000000e+01, best bound 1.710000000000e+01, gap
0.0000%
B 1.0
D 1.0
F 1.0
G 1.0
H 1.0
Maximum QALYs: 17.099999999999998
Budget 894000.0
Optimize a model with 6 rows, 10 columns and 20 nonzeros
Variable types: 0 continuous, 10 integer (10 binary)
Coefficient statistics:
  Matrix range     [1e+00, 6e+05]
  Objective range  [1e+00, 6e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [1e+00, 9e+05]
Presolve removed 6 rows and 10 columns
Presolve time: 0.00s
Presolve: All rows and columns removed

Explored 0 nodes (0 simplex iterations) in 0.33 seconds
Thread count was 1 (of 4 available processors)

Solution count 1: 13.8

Optimal solution found (tolerance 1.00e-04)
Best objective 1.380000000000e+01, best bound 1.380000000000e+01, gap
0.0000%
B 1.0
E 1.0
F 1.0
H 1.0
J 1.0
Maximum QALYs: 13.8
Budget 860000.0
Optimize a model with 2 rows, 10 columns and 15 nonzeros
Variable types: 0 continuous, 10 integer (10 binary)
Coefficient statistics:
  Matrix range     [1e+00, 6e+05]
  Objective range  [1e+00, 6e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [3e+00, 9e+05]
Presolve removed 0 rows and 1 columns
Presolve time: 0.00s
Presolved: 2 rows, 9 columns, 14 nonzeros
Variable types: 0 continuous, 9 integer (9 binary)
Found heuristic solution: objective 14.7000000
```

```
Root relaxation: objective 1.583544e+01, 2 iterations, 0.00 seconds

    Nodes    |    Current Node    |     Objective Bounds      |     Work
 Expl Unexpl |  Obj  Depth IntInf | Incumbent    BestBd   Gap | It/Node
Time

     0     0 infeasible    0        14.70000   14.70000  0.00%     -    0s

Explored 0 nodes (2 simplex iterations) in 0.64 seconds
Thread count was 4 (of 4 available processors)

Solution count 1: 14.7

Optimal solution found (tolerance 1.00e-04)
Best objective 1.470000000000e+01, best bound 1.470000000000e+01, gap
0.0000%
D 1.0
F 1.0
H 1.0
J 1.0
Maximum QALYs: 14.7
Budget 890000.0
Optimize a model with 1 rows, 10 columns and 10 nonzeros
Variable types: 0 continuous, 10 integer (10 binary)
Coefficient statistics:
  Matrix range     [3e+04, 6e+05]
  Objective range  [2e+00, 8e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [9e+05, 9e+05]
Found heuristic solution: objective 16.9600000
Presolve removed 0 rows and 1 columns
Presolve time: 0.00s
Presolved: 1 rows, 9 columns, 9 nonzeros
Variable types: 0 continuous, 9 integer (9 binary)

Root relaxation: objective 2.424272e+01, 1 iterations, 0.00 seconds

    Nodes    |    Current Node    |     Objective Bounds      |     Work
 Expl Unexpl |  Obj  Depth IntInf | Incumbent    BestBd   Gap | It/Node
Time

     0     0   24.24272    0    1   16.96000   24.24272  42.9%     -    0s
H    0     0                      21.5200000   24.24272  12.7%     -    0s
     0     0     cutoff    0        21.52000   21.52000  0.00%     -    0s
Cutting planes:
  Cover: 1

Explored 1 nodes (2 simplex iterations) in 0.80 seconds
Thread count was 4 (of 4 available processors)

Solution count 2: 21.52 16.96

Optimal solution found (tolerance 1.00e-04)
Best objective 2.152000000000e+01, best bound 2.152000000000e+01, gap
0.0000%
B 1.0
C 1.0
D 1.0
```

```
E 1.0
H 1.0
Maximum QALYs: 21.52
Budget 889000.0
Unweighted QALYs 16.6
Optimize a model with 0 rows, 10 columns and 0 nonzeros
Model has 1 quadratic objective term
Model has 1 quadratic constraint
Variable types: 0 continuous, 10 integer (10 binary)
Coefficient statistics:
  Matrix range     [0e+00, 0e+00]
  QMatrix range    [2e+05, 2e+05]
  QLMatrix range   [3e+04, 6e+05]
  Objective range  [1e+00, 6e+00]
  QObjective range [2e+00, 2e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [0e+00, 0e+00]
  QRHS range       [9e+05, 9e+05]
Found heuristic solution: objective -0.0000000
Presolve time: 0.00s
Presolved: 0 rows, 10 columns, 0 nonzeros
Presolved model has 3 quadratic objective terms
Variable types: 0 continuous, 10 integer (10 binary)

Root relaxation: objective 3.490000e+01, 11 iterations, 0.00 seconds

    Nodes    |    Current Node    |     Objective Bounds      |     Work
 Expl Unexpl |  Obj  Depth IntInf | Incumbent    BestBd   Gap | It/Node 
Time

     0     0   20.11609    0    2   -0.00000   20.11609      -     -    0s
H    0     0                      17.5000000   20.11609  14.9%     -    0s
     0     0   19.06811    0    3   17.50000   19.06811  8.96%     -    0s
     0     0   18.92544    0    4   17.50000   18.92544  8.15%     -    0s
     0     0   18.29520    0    3   17.50000   18.29520  4.54%     -    0s
     0     0   18.07500    0    3   17.50000   18.07500  3.29%     -    1s
     0     0     cutoff    0         17.50000   17.50000  0.00%     -    1s

Cutting planes:
  MIR: 2
  StrongCG: 1

Explored 1 nodes (46 simplex iterations) in 1.21 seconds
Thread count was 4 (of 4 available processors)
Solution count 2: 17.5 -0

Optimal solution found (tolerance 1.00e-04)
Best objective 1.750000000000e+01, best bound 1.750000000000e+01, gap 
0.0000%
A 1.0
B 1.0
D 1.0
E 1.0
F 1.0
Maximum QALYs: 17.5
Budget 820000.0
>>>
```

Converge3

Integrating health, economic and equity evidence to inform policy

Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College Street – 4th Floor
Toronto, Ontario M5T 3M6 Canada

https://converge3.ca

info@converge3.ca

@converge3_ca