IMPROVING PUBLIC HEALTH IN DEVELOPING COUNTRIES THROUGH OPERATIONS RESEARCH

PRASHANT YADAV
MIT-Zaragoza International Logistics Program, Zaragoza Logistics Center, Zaragoza, Spain

Public health involves improving the health of a community (as opposed to an individual) by preventive measures, control and treatment of communicable diseases, health education, regular monitoring, and surveillance of diseases. Activities in ensuring public health can be generally classified as "public goods." The benefits from public health activities are of a community nature and its mode of delivery does not always allow targeting and exclusion. It typically costs very little for an additional individual to enjoy the benefits accruing from a public health intervention and it is generally difficult or impossible to exclude individuals from consuming the benefits. A fundamental problem with public health (and more generally all public goods) is the inability of all segments of the population to pay for health and in some cases, the difficulty of motivating individuals to pay because a portion of the resulting benefits from public health expenditures are benefit externalities for them. Poverty and more competing needs for expenditure severely limit the ability of large segments of the population in the developing world to pay for better health. Public health is therefore mostly financed by national governments with the objective of controlling the spread of disease and ensuring better health of their citizens or by supranational agencies to ensure better global health. The resources committed to public health are generated from tax and other governmental revenues and thus lend themselves as ideal candidates for using operations research (OR) techniques in resource allocation and deployment. In the last century, we have seen remarkable progress in the field of public health in large parts of the world, a part of which can be attributed to the use of scientific methods to allocate and deploy public health interventions.

Unfortunately, the spread of infectious diseases such as HIV/AIDS and the resurgence of malaria combined with severe resource constraints have resulted in little progress on this front in low income countries, especially in sub-Saharan Africa. As an example, the average life expectancy is merely 39.6 years in Swaziland and 42.1 years in Zambia as compared to 82.6 years in Japan or 78.2 years in the United States (the world average is 67 years) [1]. The public health situation in low income countries today corresponds in many respects to the situation observed in the rich OECD countries a few decades ago. However, the technical and scientific knowledge for overcoming some of the diseases such as diphtheria, diarrhea, typhoid fever, and malaria already exists now. In addition, reasonable (although not sufficient) financial resources have been committed by institutions such as the Global Fund to Fight AIDS, TB, and Malaria and the US President's Emergency Program for AIDS Relief to counter these diseases. Unlike in the OECD countries where health care is financed either through government or individual private expenditure, financial resources for improvement of health in developing countries come from a variety of sources including bilateral government aid, multilateral aid institutions, private philanthropic organizations, other non-governmental organizations (NGOs), and government loan readjustments with the World Bank (Fig. 1).

This variety in the sources of financing further exacerbates the complexities in resource allocation because not-for-profit, government, and for-profit stakeholders very often have different objectives. Given the overall scarcity of resources and a complex multistakeholder environment, the foremost need in developing countries is now to apply OR techniques to ensure the optimal

Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran

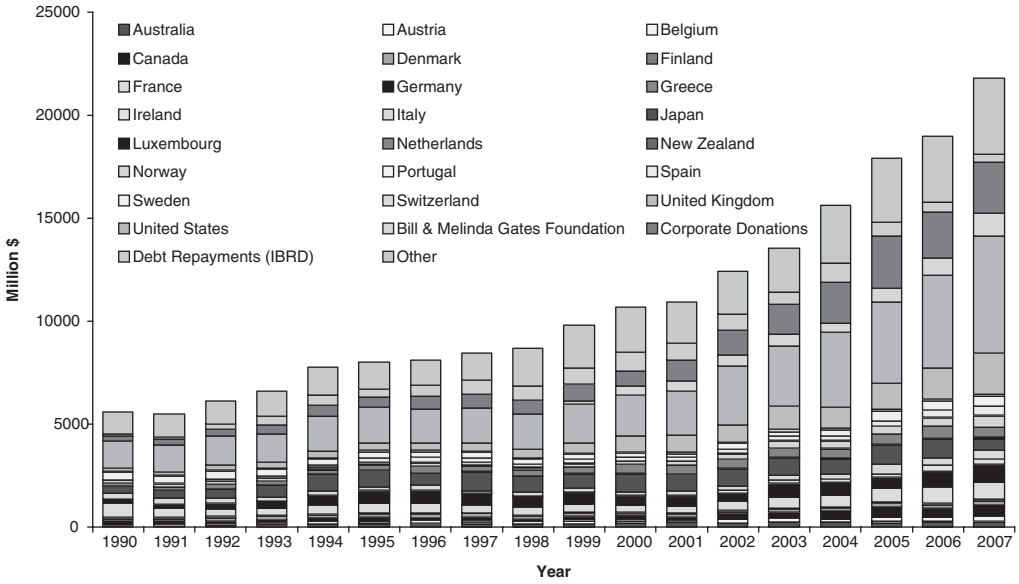

Figure 1. Sources of financing for global health. [*Source*: Source data obtained from the Institute for Health Metrics and Evaluation. http://www.healthmetricsandevaluation.org/resources/datasets/2009/dah.html]

allocation, utilization, and deployment of the available economic and material resources. OR can be applied to optimize efficiency of different public health programs in countries with high disease burdens and scarce resources. As an illustration, the budget allocated to OR in grants allocated by the Global Fund increased from a total of US $42 million for all grants between 2001 and 2005 to US $27.3 million for grants in 2006 alone [2].

OR techniques have traditionally been used in the field of health care for a range of problems such as hospital capacity planning, health-care facility location, outcomes and pharmaco-economic analysis [3]. This article will provide an introductory overview for an OR practitioner/researcher considering applications of OR which have specific relevance to public health in the developing world. The purpose is not to provide a comprehensive survey nor a taxonomy of OR problems in public health in developing countries but instead to provide the reader an overview of the nature of problems in global public health that are suitable for the use of OR techniques.

OPERATIONS RESEARCH IN PUBLIC HEALTH: DEFINITIONAL ISSUES

A review of literature and technical reports in public health reveals that the definition of public health and even more so the definition of OR varies significantly amongst public health practitioners and researchers. A quick review of the existing definitions is presented before we delve further into the topic.

Winslow [4] first defined public health as "the science and art of preventing disease, prolonging life, and promoting health through the organized efforts and informed choices of society, organizations, public and private, communities and individuals." In a more recent definition, the Institute of Medicine [5] defines public health as "fulfilling society's interest in assuring conditions in which people can be healthy by applying scientific and technical knowledge to prevent disease and promote health." For a more detailed discussion on activities that are considered as public health, see Ref. 6.

Similarly, a review of the OR activity description of the Global Fund, the WHO

and various other agencies involved in public health revealed that a commonly accepted definition is "the use of systematic research techniques to provide policy makers and program managers with evidence that they can use to improve program performance." The definition in public health practice is in tune with the early definition by Morse and Kimball [7] and Waddington [8], which includes activities such as impact assessment, outcome measurement, and metric design as a key part of OR. In this article, we work with a more restrictive definition, that is, "the use of mathematical models, statistics and algorithms to aid in decision making with the goal of optimizing performance on specific metrics."

KEY METRICS IN PUBLIC HEALTH

Although the core techniques of OR do not change much upon their application in public health, the nature of the objective function and output metrics are considerably different in many cases. As public health deals with the overall state of health of the population, almost all analysis hinges on being able to measure the state of health. Public health operational research also focuses on the efficiency of an intervention through the use of cost-effective measures. The costs and public health effects of an intervention are assessed to determine whether the public health intervention is worthwhile from an economic perspective. When resources are inadequate to meet all possible health needs—as they almost always are in the case of developing countries—being able to quantify the outcomes using a common health metric aids the efficiency of resource allocation. Several different measures have thus been proposed and we present the most commonly used ones. Some of these metrics are also widely used in health care OR.

Mortality

Mortality rate is a key measure of the health of a population and captures the number of individuals who die each year by age, sex, and the cause of death classified according to standard medical criterion. Most countries collect such data but the quality, coverage, and completeness varies significantly. A 2003 WHO study [9] found that there are 28 countries where less than 70% of the mortality data are complete or an incorrect cause of death was assigned to more than 20% of deaths.

Morbidity

The number of cases of a particular disease occurring in a year is usually reported as cases per 1000. When looking at morbidity it is important to differentiate between *prevalence* and *incidence*.

Incidence describes the occurrence of new disease in the population in a given time period, whereas *prevalence* is a static measure of the proportion of a population that has the given diseased, whether the disease cases occurred recently or at some previous point in time. For infectious diseases of short duration such as malaria or diarrhea, the differences are not significant between the two, but for more chronic communicable diseases such as HIV/AIDS and sexually transmitted infections, there are significant differences between prevalence and incidence.

Life Expectancy

Life expectancy measures the average remaining life span of individuals in a given population group. A commonly used measure is life expectancy at birth.

The above metrics only measure the impact of disease or ill-health as either death or prevalence/incidence. The longer term impacts of disease such as disability and loss of quality of life have a significant cost to the society and an overall public health impact but they are not considered in the three metrics presented earlier. Quality-adjusted life year (QALY) and Disability-adjusted life year (DALY) are two economic measures of health that combine the duration and quality of life.

Quality-Adjusted Life Year (QALY)

QALY is a metric which measures the duration and quality of life [10]. The number of QALYs lived by an individual in one year

is equal to the health-related quality of life weight attached to that year of the individual's life. An individual's quality-adjusted life expectancy at age a can be written as $\sum_{t=a}^{a+L} Q_t$ where L is the remaining life expectancy at a. However, the value of life in a given year is not the same as life in future years and hence a discount factor δ is used. Typically a 3% discount factor is used [11].

Thus, QALYs gained as a result of a specific intervention can be written as $\sum_{t=a}^{a+L^i} \frac{Q_t^i}{(1+\delta)^t} - \sum_{t=a}^{a+L} \frac{Q_t}{(1+\delta)^t}$, where Q_t^i and Q_t represent the quality of life weight with and without the intervention respectively for each time period t and L^i is the remaining life expectancy after the intervention. It is important to note that QALYs do not have an age-weighting function, that is, apart from discounting one QALY has the same value regardless of the age at which it is lived.

Disability-Adjusted Life Year (DALY)

A DALY is a health outcome measure which like the QALY measures both the quality of life reduced due to a disability after the disease and the lifetime lost due to premature mortality. One DALY can be viewed as one lost year of "healthy" life. The concept of DALY was first presented in World Bank [12]. DALY $= YLL + YLD$, where YLL are the years of life lost due to premature mortality (average life expectancy at age of death due to disease) and YLD are the years lost due to disability. YLD are calculated by multiplying the number of cases of the disease by the average duration of a case and a specifically chosen disability weight for each health condition. The disability weight is determined by expert valuations on a scale from 0 (perfect health) to 1 (death).

A key difference between the DALY and the QALY is that in DALY calculations the value attached to life is weighted so that years of life in childhood and old age are counted less [13]. Also, DALY involves continuous discounting, whereas QALY is based on discrete discounting. Ignoring continuous discounting, age weighting and the way the disability/quality scores are computed, one DALY saved compares to one QALY gained, that is, QALY $= 1 -$ DALY.

There has been considerable critique of the DALY (and QALY) as a metric for resource allocation decisions in public health [13,14]. One of the main critiques is its use of different age weights in order to capture lesser economic value creation as a child or as an older person. However, such weighting may work against resource allocations that alleviate diseases prevalent only in children in favor of those diseases that are more prevalent in working adults. Additionally, public health interventions often result in nonhealth related economic benefits and using DALY minimization as the objective in resource allocation may ignore some of these interrelationships [14]. Also, when used for resource allocation, DALY discriminates against those who are already disabled as their DALYs are lower.

Public health OR modelers require metrics so that health outcomes can be expressed in common units to analyze the trade-offs between health and economic benefits. Measuring health outcomes in common units is such a complex process that some of these shortcomings are likely to be present in any economic metric created for this purpose. Despite the shortcomings, QALY and DALY continue to be used as a metric for measuring the impact of health interventions. While QALYs are more commonly used in developed countries, DALYs are used for measuring burden of disease and cost-effectiveness of health interventions in low income countries. Threshold values on cost per DALY saved with a public health intervention in developing countries have emerged as thumb-rules for practitioners in the field of public health to accept or reject an intervention. A common threshold for interventions in low income countries is $100 per DALY saved [15].

Table 1 shows some common interventions in developing countries and their cost-effectiveness ratios measured in US$/DALY averted. The ranges of the cost-effectiveness ratios for many of these interventions are wide and vary according to the local setting due to varying ability to target specific populations. It is also evident that although the $100 per DALY averted thumb rule is mostly used, in instances such as HIV/AIDS where the disease can have a much wider

Table 1. Cost-effectiveness (US$/DALY) for Some Common Public Health Interventions in Developing Countries

Disease	Intervention	Cost Effectiveness (US$/DALY)	Cost Effectiveness Range (US$/DALY)
HIV/AIDS	Antiretroviral therapy	922	350–1494
HIV/AIDS	Condom promotion and distribution	82	52–112
HIV/AIDS	Mother-to-child transmission prevention	192	7–377
HIV/AIDS	Peer and education programs for high-risk groups	37	6–68
HIV/AIDS	Treatment of opportunistic infections	156	3–310
HIV/AIDS	Voluntary counseling and testing	47	10–85
Malaria	Insecticide treated bed-nets	11	5–17
Malaria	Intermittent preventive treatment in pregnancy with sulfadoxinepyrimethamine	19	13–24
Malaria	Residual household spraying	17	9–24
Tuberculosis	Directly observed short-course chemotherapy	301	84–551
Traffic accidents	Increased speeding penalties, enforcement, media campaigns, and speed bumps	21	3–38

[*Source*: Data obtained from disease control priorities project. http://www.dcp2.org.]

social impact than captured in DALYs, resources are allocated to recommended interventions such antiretro viral treatment despite their high cost per DALY averted.

DISTRIBUTIONAL EQUITY CONSIDERATIONS IN PUBLIC HEALTH

OR models for public health are distinct from other OR models because apart from commonly used objectives such as profit, output, and costs, they often also consider *equity of outcomes*. Effective public health requires that there are no large disparities between the availability of medicines, preventive health services, and other inputs across different socioeconomic segments of the population. OR models commonly incorporate equity constructs by setting constraints on minimum and/or maximum levels of an output provided to a given socioeconomic segment and in some cases by minimizing the sum of absolute deviations from the mean level of product or service provision.

A commonly used measure is the concentration curve, which is a plot of the cumulative proportion of the specific health-related variable being measured (income, health, medicine availability, distance to nearest facility etc.) against the cumulative proportion of the population ranked by income, beginning with the poorest, and ending with the richest. Equality on the curve is represented by a diagonal line of slope 1, and the greater the deviation of the curve from this line, the greater the inequality. So for instance, if everyone, irrespective of his or her income, has exactly the same value of the variable which measures poor health status, the concentration curve will be a 45-degree line. On the other hand, if the health variable takes higher values (reflecting poorer health status) among poorer people, the concentration curve will lie above the line of equality. The farther the curve is above the line of equality, the more concentrated is poor health among the poor.

Figure 2 shows the infant mortality measured as deaths of children under five years of age against the cumulative births ranked by income. The curve for India lies above the line of equality, indicating that under-five child deaths in India are concentrated among the poor. In comparison, the Mali curve lies everywhere below that of India implying there is less inequality in under-five child deaths in Mali than in India.

Such concentration curves provide useful visual representation of inequality but provide limited ability to rigorously compare the

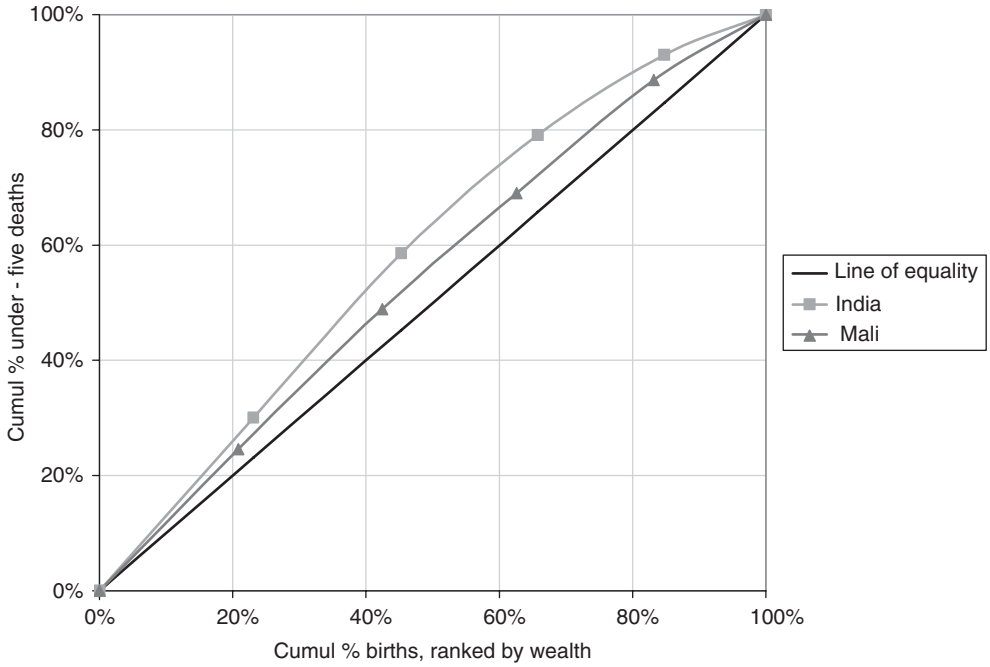

Figure 2. Lorenz curves for inder-five deaths in India and Mali. *Source*: Doorslae. *et al*. 2008 [16].

Table 2. Gini Index of Health Facilities in the Lagos State of Nigeria

Facility Type	Gini Index
Primary health center	0.12107
Public secondary health facility	0.32567
Private hospital	0.23163
Public and private secondary health facilities	0.14015

[*Source*: Data obtained from Oguntade and N.A. Yusuph [18].]

extent of inequality across units of geographical analysis and over time. Different curves can be superimposed on each other as in Fig. 2 to see the difference but the number of such comparisons is limited.

An alternative method is thus to compute the Gini coefficient [17] of the public health input or output metric. The Gini coefficient G is defined as the mean of absolute differences between all pairs of individuals for a chosen measure.

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\overline{x}},$$

where $x_i =$ observed value of the measure, $\overline{x} =$ mean of the observations, and $n =$ number of observations. The minimum value of G is 0 implying perfect distributional equality and the maximum value is 1 which is perfect inequality (implying one individual possesses all of that measure).

A very common use of the Gini coefficient is to measure the distribution of access to health facilities by population in facility location problems or the distribution of availability of drugs by socioeconomic quintiles. Table 2 shows the Gini coefficient for the state of Lagos in Nigeria for different types of health facilities. If the entire population has perfectly equal access to a health facility, the Gini coefficients would be 0.

The success of vaccination programs in low income countries typically is measured not just on the total number or proportion of children vaccinated but the Gini coefficient is calculated to understand any distributional inequities. Design of targeted programs for vaccination, the location of new vaccination clinics and mass distribution campaigns are all driven by such analysis.

BASICS OF DISEASE MODELING

Understanding and quantifying the impact of public health interventions requires some preliminary understanding of disease transmission. Although a detailed discussion of disease transmission model is beyond the scope of this article, a basic model of disease transmission is discussed to give the OR practitioner a general idea. Ross [19] first developed a mathematical model of infectious disease transmission, which was used extensively by Macdonald [20] for studying malaria transmission. Commonly known as the Ross and McDonald model, it is now the basic building block in our understanding of disease transmission.

Most disease models are compartmental models and in the simplest sense they split the human population into a proportion of those who are susceptible to an infection, those who are infected, and those who have recovered and acquired temporary or permanent immunity (Fig. 3). Recovered and immune individuals can return back to the susceptible state after a certain period and birth and natural or disease induced death occur in the different compartments. The changes in the proportions of these three categories are described by differential equations where the parameters are estimated from field studies and vary depending upon the nature of the disease and infection dynamics. Models with more number of states (or compartments) are closer at depicting reality but the resulting system of differential equations becomes complex to solve. For a detailed discussion of dynamic models of disease progression, see Anderson and May [21].

A key parameter of the disease model is the basic reproduction rate R_0, which is the number of secondary infections that can arise when a single infected individual is introduced into a population where everyone is susceptible. Very simply stated, if $R_0 = 2$, there will be two new individuals who will be infected in the first round of transmission, 4 in the second, 8 in the third, and so on. If $R_0 = 100$, there will be 100 secondary infections in the first round, 10,000 in the second, and so on. In reality, the dynamics are more complicated as the susceptible population itself changes. Theoretically, if $R_0 > 1$, the number of infected people will grow exponentially until the entire population is infected. In reality, however, the development of immunity and other factors may arrest the development of disease. If $R_0 < 1$, the number of cases declines exponentially until the infectious disease can be eliminated or eradicated. Using malaria as an example we present some more details of the Ross and McDonald model

$$R_0 = \frac{ma^2p^n l}{-\log_e p},$$

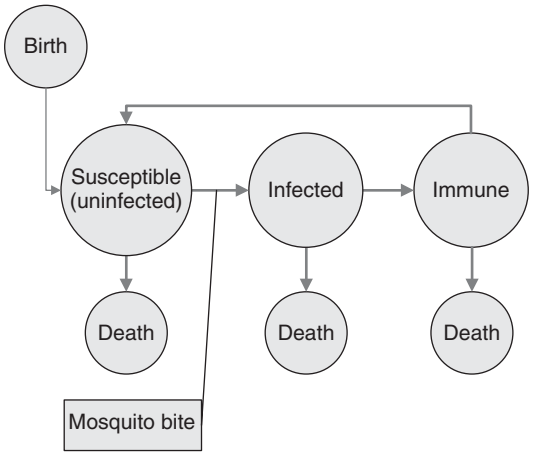

Figure 3. Basic compartmental model of malaria transmission.

where m is the number of female mosquitoes per person in the region, a is the rate at which each mosquito bites, p is the proportion of mosquitoes that survive from one day to the next, n is the maturation period of the parasite in the mosquito, and l is the duration of a single infection.

The rate a at which mosquitoes bite appears as a squared term because a cycle of transmission requires one bite to infect the mosquito and one to pass the infection back to a human. More involved versions of the model also include a probability that a bite from an infected mosquito results in human infection.

The interventions to control malaria can be broadly categorized into three main types: indoor residual spraying, distributing insecticide treated bed-nets, and treatment or prophylaxis with drugs. Transmission of malaria is influenced by many factors and an operations researcher needs to understand which interventions will have the strongest impact on reducing transmission in order to be able to develop allocation models. The above model makes it clear that measures which can reduce p, the proportion of (infected) mosquitoes that survive, is far more beneficial in reducing disease transmission than, for example, insecticidal spraying on mosquito breeding sites, which only impacts m. We can also see that parasite elimination through prompt treatment of infected individuals is also very beneficial as it decreases the duration of an infection l but may also indirectly decrease a. This demonstrates why a good understanding of the underlying disease transmission model is essential for building usable OR models for public health.

Admittedly, the above model was a simplified model only to illustrate how understanding and modeling of disease models remain key for optimal resource allocation. In reality, however, most infection transmission models have nonlinearities that can either lead to stable or unstable equilibrium points. Analysis of resource allocation decisions where disease progression follows a nonlinear transmission presents a complex problem which requires sophisticated OR techniques to solve.

OPERATIONS RESEARCH AND RESOURCE ALLOCATION FOR PUBLIC HEALTH INTERVENTIONS

Public health policy makers at different levels (global, national regional) face a range of resource allocation problems. First, they have to decide the allocation of resources between different diseases (e.g., malaria, HIV/AIDS, diarrhea, tuberculosis, maternal health); secondly, the allocation of resource between multiple interventions within each disease area (e.g., indoor insecticide spraying, distribution of bed-nets, distribution of medicines for treatment, and chemoprophylaxis); and thirdly, the allocation of resources to specific operational and tactical models for delivery of each intervention. As discussed in the section titled "Key Metrics in Public Health," a commonly used method for resource allocation is cost-effectiveness analysis based on cost per DALYs saved or cost per QALY gained. Notwithstanding the shortcomings of QALYs and DALYs, other issues further exacerbate the complexity of public health resource allocation decisions. Some of these include nonconstant returns to scale; differences in measures of cost-effectiveness; high dependence of cost-effectiveness on the spatial distribution of incidence; trade-off between increased efficiency and effectiveness through intervention targeting versus ensuring distributional equity; and interconnectedness of returns from each intervention. We start with a simple example to illustrate resource allocation in public health.

Imagine there are two diseases: Disease A kills 5000 people per year, and Disease B kills only 50 people per year. It costs \$100 per person to prevent people from dying of Disease A and \$50 per person to prevent death from Disease B. The public health resource allocator has \$10,000 for disease control. How should he spend the money? Note that a naïve public resource allocator would set priorities based on the high mortality from Disease A and allocate all the resources available (\$10,000) to Disease A and save 100 lives. If the objective is to maximize the total number of lives saved, a simple linear programming formulation would give us the optimal resource

allocation.

x_A = Budget allocated to Disease A,

x_B = Budget allocated to Disease B

$$\text{Maximize number of lives saved} = \frac{x_A}{100} + \frac{x_B}{50}$$

Subject to

$$x_A + x_B = 10{,}000$$

$$\frac{x_A}{100} \leq 5{,}000\,;\ \frac{x_B}{50} \leq 50.$$

Clearly, the optimal allocation is $x_A =$ \$7500 and $x_B =$ \$2500 with a total of 125 lives saved.

Typically, the cost of interventions are not linear. For example, it is easier and cheaper to conduct insecticide spraying (or drug distribution) in urban areas as populations are easier to reach and population density is higher. The cost of spraying (or drug distribution) is much higher for rural populations, which are more difficult to reach [22]. Thus, the cost per-unit of output for an intervention depends upon the level of that intervention. A piecewise linear model with decreasing returns is commonly used to capture this in public health resource allocation models. Often, it is also difficult to separate the benefits from each intervention.

The cost per DALY saved in Table 3 is an average measure for each intervention and they vary considerably across and within countries depending on the degree of malaria incidence [24]. The wide ranges in the cost per DALY saved is attributed to challenges in measurement due to varying scales of the intervention and nonlinearity in the costs. This shows that the design of optimal resource allocation models has to be region- and context-specific as they will depend on the level of disease incidence, current scale of existing interventions, geo-spatial distribution of populations at risk, and other health infrastructure related issues. Resource allocation managers and public health OR analysts must acknowledge that it is not going to be easy to obtain specific recent estimates of the cost-effectiveness of every intervention for each specific context. Therefore, OR models need to be built based on results from other contexts using carefully selected extrapolation and statistical matching methods.

Table 3. Cost-Effectiveness of Selected Malaria Interventions

Intervention	Cost per DALY Saved (Range)
Insecticide treated bed-nets (net + insecticide treatment)	\$11–\$17
Indoor residual spraying (1 round)	\$9–\$12
Indoor residual spraying (2 rounds)	\$17–\$24
Intermittent preventive therapy during pregnancy	\$13–\$24

[*Source*: Data obtained from Jamieson *et al*. [23], *Disease Control Priorities in Developing Countries,* 2nd edition, 2006.]

One approach to enhance the allocative efficiency between different interventions is to stratify regions in terms of their epidemiology, sociology, and other characteristics and select targeted region-specific interventions. However, this approach may also lead to lower distributional equity of the interventions or their outcomes. Carefully managing the dual objectives of efficiency and equity is a challenge that resource allocation managers in public health constantly face. In addition to the distributional equity considerations, epidemiological or sociodemographic data is hard to obtain in low income countries, which often makes the costs of identifying and/or stratifying the target populations extremely high.

In the long term, resource allocation decisions are a vector of allocations to each intervention for each time period and the optimal allocation will vary considerably based on whether the objective is to eradicate a disease or control a disease. The dynamics of disease progressions imply that there is high degree of complementarity between the interventions and the presence of one intervention positively impacts the benefits from the other. Such interrelationships become complex to model in traditional OR models and sometimes require systems dynamic models [25]. Similarly, there are positive

complementarities across disease-specific interventions implying that a disease-specific public health intervention has benefits for other diseases.

EXAMPLES OF OR PROBLEMS IN PUBLIC HEALTH

With a background on the metrics, objectives, and other notable considerations in OR for public health, we now turn to a few application areas. In Table 4, we present the most commonly observed challenges in developing country public health programs and the nature of OR models that could potentially be used to address that problem.

We describe a few important classes of problems which have particular relevance to public health in the developing world.

Facility Location Problems

Facility location problems have been a cornerstone of OR since the work of Weber [26], who considered the problem of locating a warehouse to minimize the total travel distance between the warehouse and a set of customers. Much later, Hakimi [27] considered the more general problem of locating one or more facilities to minimize either the sum of distances or the maximum distance between facilities and customers. Determining the optimal number of facilities was also considered in many location models. For a detailed review on facility location problems, see Brandeau and Chiu [28].

In traditional facility location models, the objective is to minimize either the weighted sum of distance or the maximum distance traveled. However, a public health planner may want to minimize the maximum distance (or a nonlinear function of maximum distance) traveled to capture the equity considerations presented earlier. The model of Maimon [29], which looks at facility location for public sector applications by incorporating equity and variance of distance traveled in the formulation is particularly applicable to public health settings. Similarly, the model of Schilling [30] considers multiple objectives in locating facilities with public sector objectives.

Traditional location models work under the assumption that individuals choose the facility that is closest to them. Given the social stigma associated with diseases such as HIV/AIDS and sexually transmitted infections, individuals in a public health setting may not always choose facilities closest to them. Literature in public health also shows that attendance at public health clinics falls with increasing distance to the facility [31,32]. For diseases such as TB and HIV/AIDS that require long-term treatment nonadherence to treatment can lead to higher costs as nonadhering patients migrate from first-line to more costly second-line treatment. For TB, Shargie and Lindtjørn [32] show that treatment adherence decreases as distance traveled by patients to the clinic increases. Using that premise [33] investigated the impact of patients' travel

Table 4. Public Health Challenges in Developing Countries and Potential OR Models

Public Health Challenge in Developing Countries	Potential OR Models
Poor physical accessibility to health facilities	• Facility location
High waiting time at public health facilities	• Capacity management
Acute shortage of health care workers	• Capacity management • Staff deployment
Nonavailability/shortages of drugs, vaccines, and other health products at public health facilities	• Network design • Optimal multiechelon inventory control • Inventory rationing

distance to HIV/AIDS clinics on treatment adherence and developed a model for locating new HIV/AIDS clinics in Zambia to maximize adhering population. Their simulation model shows that taking into account patient attendance and adherence in optimal ARV facility location placement could lead to a potential 3% increase in patients starting and adhering to treatment.

Another area of concern to public health is the average or maximum response time of a facility location model to contain disease outbreaks. In such circumstances, opening new temporary facilities to quickly distribute (vaccines, drugs, and therapeutics) to the affected population [34,35] is an effective outbreak containment strategy. The economic trade-off is between the number of facilities to open and the facility capacity versus the distance traveled and its impact on outbreak containment. OR models [35] determine *how many* new facilities should be opened and where and also rules on the assignment of individuals to the facilities. Heier *et al*. [36] consider the more general problem of decentralized facility assignment in such a scenario when the users pick the facility themselves and establish the cost of "anarchy" in such a system. They show that in some cases when patients pick their facilities based on full information about facility status and travel distances, it results in greater total travel time than a centralized solution where patients are allocated to facilities or when patients do not have full information about facility status.

Another key concern for public health in developing countries is the performance of a facility model when hospital infections such as MRSA (methicillin-reistant *Staphylococcus aureus*) or machine and equipment failures that are common in developingcountries could temporarily shut down a health clinic. Berman *et al*. [37] look at the classical *p*-median facility location problem and introduce the possibility that a facility might suffer a disruption. Using the example of Toronto hospitals, they show that considering the probability of failure in the facility location model results in a higher degree of centralization. Their results corroborate the colocation and centralization observed in Toronto and many other hospital systems.

The choice of the location of public health facilities is a key strategic decision variable in the design of public health systems. OR techniques with carefully selected objectives functions can help find facility locations which can maximize public health impact through better access to those requiring treatment or preventive interventions. Future research could focus on incorporating patient's health facility choice models and other benefit externalities into the optimal facility location problem.

Capacity Management Models

The emergence of new infectious diseases or the resurgence of old ones like malaria suddenly presented additional demand on the public health system, but the system could not develop as quickly due to financial and human resource constraints. The lack of human resource capacity is illustrated by the fact that sub-Saharan Africa which has over 25% of the global burden of disease has only 2–3% of the world's health workers. Unfortunatey, the money to hire, train, and sustain new health care workers is not available and is not likely to be available in the medium term [38]. This acute scarcity of health care workers in developing countries, especially in sub-Saharan Africa, offers limited ability in the short- to medium-term to apply models that help determine optimal staffing or capacity levels for clinics. The proper deployment of health staff and careful management of the available capacity in clinics can be a key approach to maximizing public health outcome in such an environment. This is currently lacking and long queues continue to form at many points within the public health system.

Many different queuing models for capacity management in health-care setting have been developed [39,40]. An important point of difference in the case of public health clinics in developing countries is that in many cases average capacity to serve patients is less than the arrival capacity. Thus, in such systems, patients reneging or service rationing is the only way that a system can attain the state of equilibrium;

otherwise, queue lengths keep increasing with time (the analysis of transient queue behavior is beyond the scope of this article). When server utilization is very high, average waiting time can be minimized by giving priority to patients who require shorter service times. Thus, the shortest processing time rule instead of the first-in first-out queue discipline could minimize total waiting times. However, such a rule is difficult to implement in practice due to its perceived unfairness amongst the patients especially in single server systems. If there were two servers, a system like the one observed commonly in supermarkets where there is dedicated express service counter for shorter processing time customers could be envisaged. Using a lesser trained health worker to treat shorter processing time patients can be an effective means of capacity management in public clinics. Thus, a triage step to know patient needs when they enter the queue is important whether to enforce the shortest processing time discipline or direct them to the lesser trained health worker. In reality, health systems are organized as tiered networks with referrals from one stage to the other. Thus, an overall analysis of public health systems requires analyzing queuing networks. Discrete event simulation could thus be a very important OR technique to utilize in the public health system. An overview of applications of discrete event simulation in health care can be found in Ref. 41. Also Hopp and Spearman [42] provide easy to use queuing approximations which can be used for modeling multistation public health systems and to analyze the impact of different parameters and flow arrangements in a public health clinic on waiting time, throughput, and variability propagation.

Since public health is usually free and does not involve any price mechanisms to regulate demand and supply, revenue management based methods of service differentiation or capacity management are not applicable in this setting. Pure rationing to restrict demand is not ethical and creates many sociocultural problems. The reneging process works in a manner that individuals who can afford to obtain care at other locations quickly drop out of the system. The long waiting time in public health clinics in some developing countries acts as a rationing system which favors those who have a low opportunity cost of time and penalizes those who have short or long term formal employment [43]. Thus, on the surface it appears that long queuing automatically leads to socially efficient rationing on the basis of socioeconomic factors. In reality, however, informal arrangements lead to queue jumping which benefits those with more resources making the queue reneging phenomenon into an inequitable and unjust rationing approach. Queuing based analysis, nevertheless, can be an important tool to devise capacity management strategies for public health clinics in developing countries. Future research could focus on modeling the impact of different task shifting interventions (point-of-care testing, nurse triage, telemedicine, decoupling drug dispensing from care provision) and varying queue disciplines in developing country health programs with severe capacity constraints.

Stock Rationing Models

Availability of drugs, vaccines, and other health supplies is very low at the clinic levels in many developing countries. Average availability at public health facilities in certain regions within a group of 36 surveyed countries was as low as 29.4% [44]. The typical structure for distribution of medicines in the public sector in most developing countries consists of a central/primary warehousing and distribution point which supplies to one or two downstream stock holding points depending upon the distribution of population and administrative structure of a country. The level of availability is low at each of the stocking points in the supply chain and thus each stage in the system has to engage in stock rationing. In systems where stock rationing is very common, if the rationing is done based on a fixed proportion of the order size, each downstream unit (clinics or sub-warehouses) inflates its orders to get a bigger share of the available supply [45,46]. If past consumption is used to allocate stock [46], changes in seasonality

and consumption patterns as a result of epidemiological patterns or differential coverage of preventive interventions are not adequately captured. In addition, the degree of lost sales at each clinic continues to get repeated in the future.

The traditional models of inventory rationing in OR analyze the problem of how to allocate inventory to different customer classes [47,48] by setting a threshold of inventory for each class of customers. Most of these models assume a known revenue or cost for each customer class and define the optimal threshold levels for rationing based on overall revenue maximization or cost minimization. Pibernik and Yadav [49] consider a model of capacity reservation, based on service-level objectives when long-term cost or revenue measures of each customer class are not available. Zenios *et al*. [50], in their model for allocating cadaveric kidneys among various patient segments, consider an objective function incorporating QALYs. Also, in most of these models, the customer segments are independent with no overlap. Optimal rationing of drugs at the clinics to achieve better public health outcomes will be based on customer segments that are defined by demographic and health characteristics (medical threshold e.g., CD4 count or severity of disease or income level, sex, age). This implies that their membership in a customer segment is itself dependent upon the rationing decision. Deo [51] considers such a model where a public health planner who wants to maximize the total expected QALYs over a long time horizon, creates the optimal rationing strategy for HIV/AIDS patients. The determination of future QALYs from the rationing decision is captured through a simple disease transmission model. Through numerical analysis he shows that a rationing policy that follows open enrollment with enforced prioritization of current patients out-performs rationing heuristics that are commonly used in practice.

Stock rationing and allocation problems in public health require a model which considers the long-term objective of maximizing public health outcomes (DALYs, QALYs) with disease models to understand the impact of the rationing strategy on long-term disease progression. Better understanding of behavioral issues involved in rationing health commodities is also needed for the rationing models to be put into practice.

CONCLUSION

This article presents an overview of the use of OR in public health in developing countries by first introducing the objective functions and output metrics commonly used by public health decision makers. A discussion of DALY and QALY metrics reveals that although they are commonly used in OR models for public health, their design is not flawless. The simple Ross and Macdonald model of disease transmission applied to malaria illustrates that resource allocation problems in OR need basic understanding of disease transmission. The concept of distribution equity is key to many public health OR problems and besides simple max–min formulations, commonly used measures of inequity such as the Gini coefficient and concentration curve are also required in many public health OR models. The number and scope of OR models for public health in developing countries continues to grow as work in this field continues apace. Understanding the main needs for OR in developing country public health programs and working closely with implementation partners are crucial for the OR/MS practitioner. Application areas of OR/MS in developing country health programs are highlighted along with opportunities for future research in this area.

Acknowledgments

The author gratefully acknowledges the comments of two referees and the associate editor in improving the content and presentation of this article.

REFERENCES

1. World Population Prospects. 2006. United Nations. Available at http://www.un.org/esa/population/publications/wpp2006/WPP2006_Highlights_rev.pdf. Accessed 2009 June 25.

2. Korenromp E, Komatsu R, Katz I, *et al*. Operational Research on HIV/AIDS, tuberculosis and malaria control in Global Fund-supported programs: rounds 1–6 grants. In: Proceedings of the 5th European Conference on Tropical Medicine and International Health; Amsterdam: 2007 Aug.
3. Brandeau ML, Pierskalla WP, Sainfort F, editors. Operations research and health care: a handbook or methods and applications, International Series in Operations Research and Management Science, 70. Boston (MA): Kluwer Academic Publishers; 2004;
4. Winslow CEA. The untilled field of public health. Mod Med 1920;2:183–191.
5. The Future of Public Health. Committee for the study of the future of public health; division of health care services 1998. Washington (DC): National Academy of Science Press.
6. Turnock BJ. Public health: what it is and how it works. Sudbury (MA): Jones & Bartlett Publishers; 2004.
7. Morse PM, Kimball GE. Methods of operations research. New York: Wiley; 1951.
8. Waddington CH. Operational Research. Nature 1948;161:404.
9. Mathers CD, Fat DM, Inoue M, *et al*. Counting the dead and what they died from: an assessment of the global status of cause of death data. Bull World Health Organ 2005;83: 171–177.
10. Zeckhauser R, Shepard DS. Where now for saving lives? Law Contemp Probl 1976;40:5–45.
11. World Health Organization (WHO). The global burden of disease: 2004 update. Geneva: World Health Organization.
12. World Bank. World Bank Development Report 1993: investing in health. New York: Oxford University Press; 1993.
13. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy Plann 2006;21(5):402–408. DOI: 10.1093/heapol/czl018 2006.
14. Anand S, Hanson K. Disability-adjusted life years: a critical review. J Health Econ 1997;16:685–702.
15. Banerjee AV, Benabou R, Mookherjee D, editors. Understanding poverty. New York: Oxford University Press; 2006.
16. van Doorslaer E, Wagstaff A, Lindelow M. The World Bank. Washington (DC); 2008. Available at www.worldbank.org/analyzing healthequity. Accessed 2009 July 01.
17. Gini C. Variabilitá e mutabilita. In: Pizetti E, Salvemini T, editors. 1912 reprinted in Memorie di metodologica statistica. Rome: Libreria Eredi Virgilio Veschi; 1955.
18. Oguntade AE, Yusuph NA. Health infrastructure inequality-a case study of lagos state, Nigeria. Soc Sci 2007;2(1):51–55.
19. Ross R. The prevention of malaria. New York: E.P. Dutton & Co.; 1910.
20. Macdonald G. The analysis of equilibrium in malaria. Trop Dis Bull 1952;49:813–829.
21. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. USA: Oxford University Press; 1992.
22. Barat LM, Palmer N, Basu S, *et al*. Do malaria control interventions reach the poor? A view through the equity lens. Am J Trop Med Hyg 2004;71(2 Suppl):174–178.
23. Jamison DT, *et al*. editors. Disease control priorities in developing countries. USA: Oxford University Press; 2006.
24. Goodman CA, Coleman PG, Mills AJ. Cost-effectiveness of malaria control in sub-Saharan Africa. Lancet 1999;354:378–385.
25. Homer J, Hirsch G. Opportunities and demands in public health systems. Am J Public Health 2006;96(3):452.
26. Weber A. Uber den Standort der Industrien, 1909; translated as Alfred Weber's Theory of the Location of Industries. Chicago (IL): University of Chicago; 1929.
27. Hakimi SL. Optimum locations of switching centres and the absolute centres and medians of a graph. Oper Res 1964;12:450–459.
28. Brandeau ML, Chiu SS. An overview of representative problems in location research. Manage Sci 1989;35:645–674.
29. Maimon O. The variance equity measure in locational decision theory. Ann Oper Res 1986;6(5):147–160.
30. Schilling DA. Dynamic location modeling for public sector facilities: a multicriteria approach. Decis Sci 1980;11:714–724.
31. Rahaman MM, Aziz KM, Munsh MH, *et al*. A diarrhea clinic in rural Bangladesh: influence of distance, age, and sex on attendance and diarrheal mortality. Am J Public Health 1982;72(10):1124–1128.
32. Shargie EB, Lindtjørn B. Determinants of treatment adherence among smear- positive pulmonary tuberculosis patients in southern ethiopia. PLoS Med 2007;4(2):37.
33. Carr CC, Jallah JD. Improving spatial accessibility to antiretroviral treatments for

HIV/AIDS [MS thesis]. Zaragoza, Spain; MIT-Zaragoza International Logistics Program; 2008.

34. Kaplan EH, Craft DL, Wein LM. Emergency response to a smallpox attack: the case for mass vaccination. Proc Natl Acad Sci U S A 2002;99(16):10935–10940.
35. Lee EK, Smalley HK, Zhang Y, *et al*. Facility location and multi-modality mass dispensing strategies and emergency response for biodefense and infectious disease outbreaks. Int J Risk Assess Manage 2008;12:311–351.
36. Heier J, Ergun O, Keskinocak P, *et al*. Decentralized facility location models. In: INFORMS Annual Conference 2008. Washington (DC); 2008.
37. Berman O, Krass D, Menezes M. Facility reliability issues in network p-median problems: strategic centralization and co-location effects. Oper Res 2007;55(2):2332–350. DOI: 10.1287/opre.1060.0348.
38. Conway MD, Gupta S, Khajavi K. Addressing Africa's health workforce crisis. McKinsey Q 2007.
39. Bailey NTJ. Queuing for medical care. Appl Stat 1954;3:137–145.
40. Green L. Queueing analysis in healthcare. In: Hall RW, editor. Patient flow: reducing delay in healthcare delivery. New York: Springer; 2006. pp. 281–308.
41. Jacobson S, Hall S, Swisher J. Discrete-event simulation of health care systems. In: Hall RW, editor. Patient flow: reducing delay in healthcare delivery. New York: Springer Science; 2006. pp. 211–252,281–308.
42. Hopp W, Spearman L. Factory physics. New York: McGraw-Hill; 2000.
43. Rosen S, Sanne I, Collier A, *et al*. Rationing antiretroviral therapy for HIV/AIDS in Africa: choices and consequences. PLoS Med 2005;2:11.
44. Cameron A, Ewen M, Ross-Degnan D, *et al*. Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis. Lancet 2008;373 (9659):240–249.
45. Gonçalves P. Why do shortages inflate to huge bubbles? Working Paper. Cambridge (MA): Sloan School of Management, Massachusetts Institute of Technology; 2008.
46. Lariviere M. Capacity allocation using past sales: when to turn-and-earn. Manage Sci 1999;45(5):685–703.
47. Topkis DM. Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes. Manage Sci 1968;15:160–176.
48. Nahmias S, Demmy WS. Operating characteristics of an inventory system with rationing. Manage Sci 1981;27:1236–1245.
49. Pibernik R, Yadav P. Dynamic capacity reservation and due date quoting in a make-to-order system. Nav Res Log 2009;55(7): 593–611.
50. Zenios SA, Wein LM, Chertow GM. Dynamic allocation of kidneys to candidates on the transplant waiting list. Oper Res 2000;49: 549–569.
51. Deo S. Rationing of HIV treatment in resource-constrained settings under supply uncertainty. In: Proceedings of the Manufacturing & Service Operations Management Society Meeting; 2008; Baltimore (MD).