Chapter 3

Value of Treatment Heterogeneity for Infectious Diseases

Ramanan Laxminarayan and Martin L. Weitzman

Treatment homogeneity is valued in the medical profession. Uniform treatment guidelines are often used to ensure that all physicians prescribe a safe, efficacious, and cost-effective drug in treating a medical condition. However, such a policy may be undesirable when drug resistance is endogenous. In the case of infectious diseases, selection pressure imposed by the use of any single drug (antibiotic, antiviral, or antimalarial) sooner or later leads to the evolution of resistance (by bacteria, viruses, or parasites) to that drug. In this chapter, we show that a "mixed strategy" of multiple drug use is generally desirable and analytically characterize the conditions under which this strategy is optimal.

From an economist's perspective, the treatment of infectious diseases is fundamentally different from the treatment of noninfectious conditions such as arthritis, cardiovascular disease, or cancer. Unlike the case of noninfectious or chronic diseases, two social externalities—one positive and the other negative—inherently characterize the treatment of infectious diseases. Take the case of antibiotics (although the situation can be generalized to antivirals and antimalarials as well). On the one hand, antibiotic treatment cures the patient, thereby preventing the disease from being transmitted to other individuals. On the other hand, drug treatment selects in favor of harmful mutations or organisms that are resistant to the drug, increasing the likelihood that the drug will be less effective in the future. Because the individual patient fails to take into account either of these externalities when deciding to seek treatment, a Pigovian tax or subsidy of treatment could in principle correct for

externality (depending on whether its impact on overall social welfare is negative or positive).[1]

The externality problem implicit in the decision on whether to seek treatment for infectious diseases has been well documented in earlier work (Philipson 2000). In this chapter, we extend this literature to look at externalities arising from the *choice* of drug treatment once the decision to treat has been made. The degree of homogeneity in the choice of drug treatments for infectious diseases is remarkable. For instance, in 1997, nearly 60 percent of all cases of acute ear infections (a common condition in young children) in the United States were treated with amoxicillin. In fact, amoxicillin accounted for 35 percent of *all* antibiotics used by physicians, and the five most commonly used antibiotics used accounted for 72% of *all* antibiotics used by physicians in this country. This degree of homogeneity has been witnessed even in the developing world. In most African countries, chloroquine was the most commonly used drug to treat malaria for many years. In fact, in some malaria-endemic countries, it was even mixed in with common salt to ensure widespread and uniform malarial prophylaxis.

There are reasons why uniformity of treatment is frequently encountered. In many developing countries, all drug procurement is centralized and controlled by the government. Therefore, the government determines which drugs should be even allowed into the country, thereby influencing the choice of treatment. In developed countries such as the United States, clinical treatment guidelines for community-level infections are typically issued by national public health bodies, such as the American Association of Pediatrics and the Centers for Disease Control and Prevention. In addition, individual hospitals both set and follow treatment guidelines on the basis of the advice of the hospital's infection-control committee. The choice of drug treatment is, all other things being equal, often made on the basis of the principle known as cost-effectiveness (Weinstein and Fineberg 1980). In simple terms, the drug with the smallest ratio of treatment cost to effectiveness is the drug of first choice for *all* patients. In addition, individual patients acting in their own self-interest tend to prefer the most cost-effective drug option.

There are good reasons why such homogeneity is actively promoted among the medical profession. Clinical guidelines and national treatment policy recommendations provide guidance to individual physicians on which drugs are to be used for first-line treatment, which drugs are to be used for second-line treatment should the first-line treatment fail, and which drugs are to be used in case of complications. By specifying treatment in the form of simple uniform decision rules, national policies are particularly useful in ensuring safe and accurate medical treatment while relieving the physician of some of the burden of medical decisionmaking.[2] Following uniform guidelines reduces the liability associated with medical error for physicians.

However, as this chapter demonstrates, significant disadvantages may be associated with promoting a single drug as the first-line treatment for a given condition.

The starting point of this chapter is the observation that, to the extent that most patients in a region or country are treated with the same drug for a given infectious disease, the use of a single drug places "excessively" high selection pressure on organisms that are susceptible to that particular drug and increases the likelihood that a resistant strain will evolve and proliferate. As resistance to the recommended first-line drug builds up, that drug is replaced by an alternative that is used until resistance to this second drug also increases, and so on in succession. The main message of this chapter is that the optimal solution may therefore be to use not just a *single* drug throughout the population as first-line agent, but to prescribe a variety of drugs, randomized over patients, to ensure inordinate selection pressure is not placed on any single drug or class of drugs.

This chapter also indirectly addresses the question of whether more expensive, highly effective drugs should be kept on the sidelines for use in the event of serious, resistant infections or whether they should be deployed alongside less effective agents on the frontlines against infectious diseases. The benefit of having an effective drug available as backup should all else fail cannot be disregarded, nor should the more effective drug's ability to relieve selection pressure on the first-line drug be ignored (when the effective drug is also used).[3] What, then, is the optimal solution? The answer, as it often happens, lies somewhere between the black and the white—it may be to use more effective drugs in both roles.

The problem is not defining the extent to which the more effective drug should be used but rather describing a standard policy based on guidelines for first-line treatment in this situation. As the simple model of this chapter will show, it is generally more desirable to use less expensive agents on a greater fraction of patients and more expensive agents on a smaller fraction—right from the beginning—all else being equal. In this sense, the concept of uniform guidelines may be fundamentally flawed in the presence of endogenously generated resistance. Of course, it is difficult to specify these "mixed-policy" fractions in the form of a standard, uniform, guidelines-based policy. For instance, in a geographically isolated area, it may be optimal for the single family practitioner serving these areas to prescribe a wide array of antibiotics (only in cases in which they are required, of course) so that selection pressure on no single antibiotic is allowed to build up. Clearly, there are practical difficulties of doing so. The single most important difficulty is that it makes sense from the individual physician's perspective to do what everyone else in the country is doing and to prescribe the most commonly used antibiotic. Herein lies the intrinsic externality issue related to drug prescribing. Physicians have

an incentive to prescribe in concordance with the rest of the medical community in the interest of, if nothing else, reducing their liability in malpractice claims. Of course patients are put at ease with a single, unified decisive choice. But such concordance increases the selective pressure on the drug of common choice. Guidelines that pick out a single drug for such targeted, nationwide use may therefore be exacerbating selection pressure on that single drug to a degree that is socially undesirable.

The emphasis placed on using a single drug may occur even in the absence of uniform treatment guidelines. Decentralized decisionmakers (i.e., individual physicians or patients) may not take into account the risk involved in prescribing a single drug repeatedly for a common condition such as an ear infection. The individual physician's encouragement of the development of resistant organisms globally whenever he or she decides to use that drug represents a negative externality. This externality remains uncorrected because the individual physician bears only a negligible fraction of the total burden of resistance that he or she may be placing on others with every treatment decision.

The problem of excessive selection pressure arising from the use of a single drug occurs not only in countries where physicians are the primary source of treatments but also in countries where the disease is home treated, as is the case with malaria in Africa. Here too, patients would prefer to be treated with the most cost-effective drug available to them. However, from a societal perspective, it may be optimal to use other drugs that are not cost-effective from the individual patient's perspective. The question then is how patients might be persuaded to use these other drugs even if it is not in their self-interest to do so.

One might argue that the logical extension of the strategy to treat different patients with different drugs is to treat individual patients with more than one drug. Such a strategy is already standard practice for the treatment of human immunodeficiency virus and tuberculosis. In each of these cases, the underlying principle is that the probability of a multigenic resistance in a microbe is much lower than the probability of a genetic mutation conferring resistance to one drug. Using two drugs ensures that each drug exercises a protective effect over the other. However, with the argument used for drug combinations, resistance to a drug is *exogenous*. The reasoning we follow in developing the argument for using a wider variety of drugs as first-line agents runs along similar lines, as do our policy prescriptions, except that in our case, drug resistance develops *endogenously* as an evolutionary reaction to excessive usage. The solution to the problem of endogenously growing drug resistance then may be to extend the combination treatment concept to a community level. Further, routinely using two antibiotics on a single patient may be undesirable for medical reasons.[4] The alternative is to treat different patients suffering from the same infectious disease with different drugs, a prescription that is difficult to implement using a guidelines-based policy. In any event,

this chapter concentrates on this case—of treating different patients with the same infectious disease with different drugs—in the context of endogenously induced disease.

We fully recognize that to simplify the complex task of medical decision-making to fit into the boundaries of theoretical economic analysis is to issue an open invitation for criticism. The constraints imposed by the degree of abstraction in developing the arguments in this chapter—or the specific applicability of the results—cannot be overstated. This chapter addresses only problems associated with guidelines that recommend one kind of drug per patient as first-line therapy and does not refer to guidelines that promote judicious drug use, safe doses, overall safety, and so forth. We are not suggesting the use of combinations of drugs on individual patients but rather a strategy of treating different patients with different drugs. This principle, known as antibiotic heterogeneity, is beginning to enter the set of options being considered by medical professionals. However, it runs fundamentally counter to the long-held belief in the medical profession of the existence of a "best treatment" for a disease and the deeply felt need for uniformity in drug treatment.

Guidelines that promote uniformity in the choice of drug for treating infectious diseases may be inherently self-defeating because using the greatest variety of drugs decreases the likelihood that microbes will acquire and maintain resistance to any single class of drugs. The single most important message of this chapter is that, from a societal perspective, it may even be desirable to treat some patients with more expensive drugs even while it is individually suboptimal to do so. The precise fraction of patients that should be treated with these more effective drugs can be determined using fairly straightforward criteria, which we demonstrate in the sections that follow.

Model of Endogenous Resistance

This section presents our "core model" of endogenously generated resistance to drug therapies. It goes without saying that such a model must of necessity be formulated at a very high level of abstraction. Nevertheless, as will become clear, it is little short of amazing how much analytical insight emerges from even such a simple formulation.

Let there be available m possible drug therapies (indexed $i = 1,2,\ldots,m$), each of which may be used to counter some particular disease. For analytical simplicity, we imagine that everyone in the population is treated with exactly one complete treatment dose of one of the drugs. Critical to our analysis are the ideas that we are allowing a "mixed strategy" of different drugs to be used on different people and that the model should tell us when this strategy is optimal rather than excluding it a priori. Let x_i represent the fraction of the population treated with drug i, where

$$\sum_{i=1}^{m} x_i = 1 \tag{1}$$

and

$$0 \leq x_i \leq 1 \tag{2}$$

Let the cost (inclusive of c non-drug treatment costs) of drug i (per unit of population) be given by

$$c_i > 0 \tag{3}$$

Resistance to drug i by the underlying pathogen is assumed to be a Poisson process with intensity parameter

$$\theta_i > 0 \tag{4}$$

where θ_i is a (very small-valued) parameter representing the probability that resistance to drug i will develop endogenously (presumably by mutation) in the pathogen in any one person treated by that drug. (Here we refer to *endogenously acquired* resistance, which develops spontaneously by Poisson mutation in the pathogen in a patients being treated using the drug, as opposed to *epidemic* resistance by the pathogen, which results from infection by a drug-resistant pathogen from another person treated by that same drug.)

When a fraction x_i are treated by drug i, the probability that a resistant strain emerges is (to a first-order approximation)

$$\theta_i x_i \tag{5}$$

If such a resistant strain emerges, it will put at risk of epidemic resistance all x_i people treated by drug i. Let the social loss *per person* of being placed "at risk" by resistance developing in the drug by which they are being treated be denoted

$$L > 0 \tag{6}$$

Then, combining Equations 5 with 6, the expected social loss *per person* of being put "at risk" by drug i is

$$L[\theta_i x_i] \tag{7}$$

whereas the *total* expected social loss from being put "at risk" by being exposed to pathogens that are resistant to drug i is

$$[Lx_i][\theta_i x_i] \tag{8}$$

In other words, we assume that it takes time to change these treatment fractions and that individuals who continue to be treated with drug i after a resistant strain has emerged are at risk for treatment failure.

Let

$$N_i \equiv \frac{1}{\theta_i} \tag{9}$$

be the average number of people that can be expected to use drug i before resistance sets in. Then the total expected social loss Expression 8 can be rewritten as

$$\frac{Lx_i^2}{N_i} \tag{10}$$

The *optimal drug combination problem* in this model is one of minimizing

$$\sum_{i=1}^{m}\left[c_i x_i + \frac{L}{N_i} x_i^2\right] \tag{11}$$

subject to

$$\sum_{i=1}^{m} x_i = 1 \tag{12}$$

and

$$0 \leq x_i \leq 1 \tag{13}$$

Characterizing the Optimal Drug Combination

The effectiveness of all drugs is assumed to be identical. Without loss of generality, suppose the drugs are arrayed from least to most expensive, so that

$$c_1 \leq c_2 \leq \ldots \leq c_n \tag{14}$$

It is quite obvious that it will never be optimal to use (to prescribe positive amounts of) a *more* expensive drug while not using (prescribe zero amount of) a *less* expensive drug. To see beyond this what is the form of an optimal policy, and what it depends on, let us begin by analyzing in full detail the situation for *two* drugs ($m = 2$).

There are two possible solutions—an interior solution and a corner solution of the form $x_1 = 1$, $x_2 = 0$. The latter corresponds to the necessary and sufficient first-order corner condition

$$c_2 \geq c_1 + \frac{2L}{N_1} \tag{15}$$

From Equation 15, we may say that a "mixed" interior solution using both drugs is optimal if and only if the following condition is met:

$$(c_2 - c_1)N_1 \leq 2L \tag{16}$$

What is the intuition behind Condition 16? The precise economic condition under which it is optimal to include drug 2 in our menu is that the increase in cost associated with treating with the more expensive drug in place of the cheaper drug is less than or equal to the expected benefit from using two drugs in place of one. The term on the right-hand side, $2L/N_1$, represents the marginal expected social cost per person associated with treating another patient with drug 1. As long as the increased treatment cost of using drug 2 in place of drug 1 is less than the expected increase in cost associated with endogenously generated resistance if drug 1 were to be used, it makes economic sense to use drug 2.

Next, consider the more general case in which m is an arbitrary positive integer (larger than two). The first-order condition for a fully interior solution is the existence of a positive multiplier λ, which is dual to Equation 12, that satisfies for positive x_i the conditions

$$c_i + \frac{2Lx_i}{N_i} = \lambda \tag{17}$$

The multiplier λ can therefore be interpreted as the "user cost" of any drug being used in the menu. Therefore, for any drug i that is being used, the total user cost equals the sum of the treatment cost c_i and the resistance cost $(2x_i/N_i)L$, in which the resistance cost equals the marginal probability of inducing a resistant infection with another treatment multiplied by L, the associated social cost of inducing resistance in the population. Although the treatment costs of drugs in our optimal menu can vary greatly, their user cost is identical. In other words, if two drugs are included in our optimal menu and one costs less than the other, then the resistance cost of the cheaper drug must exceed that of the more expensive drug so that the user cost of the two drugs is identical. The resistance cost of a drug is, of course, a function of the fraction of the population being treated with that drug, and a high treatment fraction implies a larger resistance cost.

The astute reader may have guessed where we are headed. The optimal decision rule is to use the lowest cost drug(s) first, as standard economic intuition would dictate. What is not so standard, however, is the form in which these costs arise. In addition to the treatment cost that the individual patient faces, there is an additional cost associated with the increased probability of drug resistance associated with each use of the drug. This resistance cost is endogenously determined by the fraction of the infected population that is

administered the drug in question. Therefore, the optimal menu design is such that the sum of treatment and resistance costs of *all* drugs on the menu is identical, thus ensuring that some drugs may find their way into this menu even if they are not the least expensive from a treatment cost perspective. Making use of Equation 17, Condition 12 can be rewritten as

$$\lambda = \frac{\sum_{k=1}^{m} c_k N_k}{\sum_{k=1}^{m} N_k} + \frac{2L}{\sum_{k=1}^{m} N_k} \tag{18}$$

The next step is to determine the optimal user cost for a given set of drugs that are available to the social planner (not just those that will be included on the menu). The optimal user cost can be expressed as the sum of the resistance probability weighted *average* cost of *all* available drugs and the expected marginal cost of treatment failure associated with any single treatment when *all* available drugs are being used.

Combining Equations 18 and 17, the "interiorness" condition $x_i > 0$ is equivalent to the condition $c_i \leq \lambda$, or equivalently,

$$x_i = (\lambda - c_i)\frac{N_i}{L} \tag{19}$$

which is the appropriate generalization of Equation 16. From an economic perspective, it is optimal to include any drug i in the menu of the drugs so long as the cost of the drug is less than or equal to the benchmark user cost λ. It is now intuitively clear what is an easy-to-apply myopic algorithm for determining optimal drug use. Suppose by induction it is known that an optimal solution includes a positive use of all drugs j where $j < i$ for some i. The next question to ask is whether it is additionally optimal to use drug i at a positive level. The answer is "yes" if and only if

$$\sum_{j<i} (c_i - c_j) N_j < 2L \tag{20}$$

By repeatedly asking the induction question in the form of Equation 20, it is possible to build up an optimal solution inductively by using a simple recursive algorithm based on the easily interpretable economic Equation 20.

We can rewrite the "interiorness" Equation 20 as follows:

$$c_i \leq \frac{\Sigma c_j N_j}{\Sigma N_j} + \frac{2L}{\Sigma N_j} \tag{21}$$

Equation 21 means that drug i will be used as long as the marginal cost of using this drug is less than the weighted average cost of all drugs that are

already in use. The weights are expected life span, measured in number of treatments before resistance evolves, plus the average cost of treatment failure associated with each additional treatment.

Let us assume that the second term on the right-hand side of Equation 21 is small. Even so, we could still want to use more than one drug, which is quite unlike the standard cost-effectiveness criterion in which drugs are used strictly in the order of lowest to highest cost, and only one drug is used at any given time. When this second term is large, then we may want to use drugs whose marginal cost *exceeds* the average cost of all drugs weighted by their expected life spans. Note, however, that we would *never* want to use a more costly drug j while excluding a cheaper drug k, even if

$$N_j > N_k \tag{22}$$

In other words, the value of N_i by itself does not determine whether a drug will be included in an optimal program. However, N_i does determine the fraction of patients who should be treated with drug i, as is demonstrated by rewriting Equation 17 as

$$x_i = (\lambda - c_i)\frac{N_i}{L} \tag{23}$$

From Equations 17 and 18, for any two drugs j and k being used in positive amounts, we can write

$$\frac{x_j}{x_k} = \frac{(2L + \Sigma c_i N_i - c_j \Sigma N_i) N_j}{(2L + \Sigma c_i N_i - c_k \Sigma N_i) N_k} \tag{24}$$

We have already noted that the parameter N_i does not ever invert the order in which a drug i is included in the overall drug menu. However, from Equation 24, the average useful lifetime parameters $\{N_i\}$ could result in a relatively less cost-effective drug being used on a larger fraction of patients, such that $x_k > x_j$ even while $c_k > c_j$, so long as N_j is sufficiently larger than N_k.

Referring back to Expression 14, if one were to follow the traditional medical cost-effectiveness criterion, one would first use only drug 1, then later switch to drug 2 when resistance evolved to drug 1, and so on. However, moving sequentially in strict order of increasing cost-effectiveness ratios and treating all patients with the same drug at the same time can be myopically ineffective whenever account is taken of the inescapable fact that immunity is endogenous—as we have just shown. In fact, it is not even optimal to use the most cost-effective drug on the largest number of patients. When resistance evolves endogenously, a parameter representing the average number of patients who must be treated before resistance appears determines (along with drug costs) the optimal intensity of drug usage.

Discussion

The externality problem associated with the treatment of infectious diseases—one that is related to a treatment's dual properties of reducing contagion and limiting drug resistance—has a reduced-form structure that is extremely familiar to any economist. Externalities are a common problem, whether they are related to highway congestion or air pollution, and copious economics papers have dealt with these issues. Always, a negative externality calls for using less of the privately optimal good and more of the privately more expensive alternatives. What is unusual about drug resistance is that this problem has not been widely recognized as a social externality—possibly of enormous consequence. Following this line of thinking, we arrived at simple criteria for choosing an optimal antibiotic policy, which contrasts sharply with the conclusion of the standard conventional health economists' individualistic cost-effectiveness analysis.

Under the standard cost-effectiveness approach, the economic criteria most commonly used in offering an economic perspective on the optimal choice of first-line treatment is that the drug with the lowest ratio of cost to effectiveness is selected as the primary or first-line drug. When this criterion is followed, it ignores the possibly large negative externality of overusing a particular drug. A large number of papers in the medical literature use the private-cost approach to determine the "optimal" treatment for a communicable disease. But the very nature of a communicable disease means there is a potentially large externality associated with drug treatments. The standard medical approach fails to recognize the externality problem associated with the uniform use of a single drug. The externality here is similar to the one encountered in agriculture in which all farmers decide to grow a single "optimal" crop, thereby encouraging the evolution of pests that can wipe out the entire monoculture. Although in the agricultural context, the solution is to grow different varieties dispersed spatially, in the medical context, the true optimal solution is analogously to use a "mixed" variety of drugs in fractions that are proportional to their individual cost-effectiveness.

There are many instances in which we could move from a policy of a nationally recommended treatment to a policy in which local doctors have more control over the drug prescribed. So the recommended policy change is from one of active promotion of treatment heterogeneity to a more decentralized approach to decisionmaking. Such a strategy would raise much concern over the lack of a "national strategy" to combat a disease such as malaria even if such a coordinated strategy would hasten the day when the prescribed guideline treatment would become ineffective.

Without a doubt, there may be practical problems with using a variety of drugs at the health care setting for a single infectious condition. For instance, a physician may have to explain to individual patients why they are getting

different drugs. The specific treatment given by different doctors will differ depending on their (different) assessments of probability weights. This is potentially problematic because patients typically look to doctors to resolve uncertainty by prescribing the "single best" treatment.

Herein lies the dilemma. We have boxed ourselves into a particular way of reasoning that there is a "best" treatment for an ailment, one that is attributable to the fact that we are not used to having any externalities in medicine. The single best treatment approach works well for noninfectious conditions but breaks down badly for infectious diseases, in which significant negative externalities are likely to be present in the form of endogenously generated drug resistance. Once we become aware of the nature of this particular externality as one that requires the physician to also consider society's best interests, while determining what is in the best interest of the patient, then an optimal strategy may well involve a mixture of less expensive and more-expensive drug therapies.

Drug resistance is endogenous. The current strategy has been to wait for resistance to evolve before being surprised each time it appears, as if it were an ad hoc problem requiring some quick fix. Economists can contribute to the formulation of strategies that would internalize the cost of endogenously generated resistance into the process of treatment decisionmaking. This chapter tries to take a first step in such a direction.

References

Philipson, T. 2000. Economic Epidemiology. In *Handbook of Health Economics*, edited by A.J. Culyer and J.P. Newhouse. New York: Elsevier, 1762–99.

Weinstein, M.C., and H.V. Fineberg. 1980. *Clinical Decision Analysis*. Philadelphia, PA: W.B. Saunders Company.

Notes

1. It is useful to contrast *appropriate* drug treatment (or treatment for a bacterial infection that is likely to be cured faster because of that treatment) with *inappropriate* drug treatment (which does not cure the patient any faster than if that treatment were not used). An example of appropriate drug treatment is the use of antibiotics to cure bacterial infections; an example of inappropriate drug treatment is prescribing antibiotics for viral infections. Appropriate drug treatment benefits both the individual patient and society, whereas inappropriate drug treatment benefits neither the patient nor society. Although inappropriate drug treatment is a significant factor in the growing resistance of microbes to drugs, this chapter focuses exclusively on optimal policies related to appropriate drug treatment. In practice, appropriate drug treatment often is linked to a guidelines-type policy under which physicians are expected to adhere uniformly to a predetermined sequence of drugs to be used for treatment.

2. National policies are especially useful in countries in which the primary health care provider is typically a health care worker with limited training. In countries in which government-run public health facilities are the primary sources of drugs, national policies determine which drugs are available at different levels of the health care system. For instance, a second-line drug may only be available at a hospital and not at a primary care clinic.

3. An important argument against keeping newer, more effective drugs on the sidelines as backups is that such a policy tends to lower the incentive for drug firms to develop new drugs that may not be used extensively during the life of their patent protection.

4. These medical reasons may include undesirable side-effects from using two drugs, more complicated dosage regimens, and economic costs.