CHAPTER 7

Discounting health effects for medical decisions

W. Kip Viscusi

The set of virtually unanswerable questions includes the discount rate used for future health status.

Lester Lave, *The Strategy of Social Regulation*

In any decision-making context in which effects of decisions are generated over time, it is necessary to establish some relative weight on deferred outcomes as opposed to immediate impacts. The task of converting future effects into present dollars by discounting them at some rate of interest is not a process unique to medical decision making. Indeed, all economic decisions in which resource allocations involve an intertemporal element entail some weighting of the effects at different points in time. The main economic question is what weight impacts should have at different times. In particular, how much more highly should current, as opposed to future, impacts be weighted?

Intertemporal aspects are inherent in the medical decision making context as well. The entire notion of investment in human health implies some concern with future well-being. Although strategies for treating headaches and the common cold involve remedies with imminent payoffs, most major medical decisions involve a long-run element (see, among others, Edelson, Tosteson, and Sax 1990; Weinstein and Stason 1977). The communicable nature of AIDS, for example, means potential future beneficiaries. Pharmaceutical products, such as those for treating hypertension and cancer, have long-run effects on future earnings and individual health that require some discounting process (Shibley et al. 1990; Kawachi and Malcolm 1991). Choice of diet, exercise patterns, and smoking behavior clearly have long-run effects on physical well being. Ailments such as heart disease and cancer are often the result of a long
period of individual decisions. Some of these effects are cumulative, and others occur with a lag.

Intertemporal elements also are intrinsically involved in almost all major surgical decisions. Patients contemplating back surgery must decide whether relief of immediate pain warrants back surgery. An operation offers the promise of improved health in the long run but at the cost of increased risk of complications at the time of surgery. Thousands of back patients annually must confront the decision to undergo back surgery and make some trade-off involving expected future health effects, current costs, and potentially substantial losses associated with immediate surgery.

Intertemporal aspects are also an inherent part of societal decisions. In the case of the drug Taxol, the U.S. government had to trade off the present benefits of saving lives against long-run environmental consequences of destroying the Pacific yew trees used to make the drug (Box 7.1). To what extent, then, should the government allocate health sector resources to promote the welfare of the citizens presently alive as opposed to future generations? Future generations may be more affluent and value health more than those now alive, but transferring resources to them without some means of compensating those now alive would be regressive. Moreover, because good health is an economic good, it may be more desirable to confer good health status and other in-kind transfers, such as a clean environment, than simply to raise future welfare levels.

Substantial pressures are often brought upon the medical research establishment because of the temporal dimension of research breakthroughs. The AIDS lobby has exerted a strong influence on AIDS research because a large identifiable population who might benefit immediately from research breakthroughs will not survive to benefit later. Since this research has highly uncertain prospects, and since time may be an important part of the research process (allowing for dissemination of early research results, etc.), the degree to which this research should be accelerated is not clear.

Because of the central role of research and development in the medical industry, companies involved in producing products related to medical decisions necessarily must place weights on outcomes over a period of time. The classic case in which a substantial research effort is expended is that of the pharmaceutical industry, where companies invest years in research and supervised testing of drugs before bringing them to market.

If companies were only concerned with immediate rewards and acted, in effect, as if the rate of interest used to discount future benefits were infinite, then they would never undertake research involving any lag time. The only concern would be how to achieve the greatest possible profit from products currently on the market. At the other extreme, if companies acted as if they had a zero rate of interest, then effects far into the future would receive the same weight as current profits. Considering the greater affluence and larger market
Box 7.1 Intertemporal trade-offs: the case of Taxol

A classic example of the intertemporal choice trade-off arose in 1991 with respect to the drug Taxol. The 1992 Food and Drug Administration (FDA) approval of this drug for use in treatment of cancer of the ovary, breast, and lungs has heightened the controversy. Using the bark from the Pacific yew tree, scientists were able to develop a drug that appeared to be effective in treating several kinds of cancer, such as breast cancer. The main problem is that obtaining the bark involves destroying 100-year-old trees, which cannot be readily replaced, at least in the near term. To produce enough Taxol to treat a single cancer patient requires the bark of as many as six trees of diameter of at least ten inches.

From a societal standpoint, a question arose as to how fast the trees should be chopped down for use in production of this drug. Exploitation of the bark now offered the promise of immediate rewards, but a slower depletion of the stock of Pacific yew trees would enable scientists to better understand the effects of the drug and perhaps perfect it so that more lives would be saved in the future. Should we deplete the stock of trees now or do so over time, and at what rate?

This decision was complicated even further by the prospect of developing synthetic substitutes for the drug that, in the long run, might make it unnecessary to cut down the trees. Moreover, environmentalists raised the issue of trade-offs other than those involving health. Society was also sacrificing a scarce resource, Pacific yew trees, so that it could promote individual health. The ability to use hybrid yews cultivated by Weyerhaeuser eases these environmental concerns but does not eliminate them.

The trees soon had their advocates. A group known as the Native Yew Conservation Council formed a lobby to protect them. Rather than fending off lumberjacks eager to make boards, these environmentalists sought to preserve trees that were being used to save lives. The stakes had escalated enormously, particularly considering that the tree involved is one that few Americans knew even existed.

Almost all of the issues involved in this policy debate have a strong intertemporal aspect. Preservation of natural resources necessarily entails that weight be given to the effects on future generations, and the assessment of the merits of using the bark from the Pacific yew tree for medical purposes hinges on one’s assessment of the likely developments with respect to the use of this bark as well as the development of substitute drugs. Unless one is willing to place weights on the effects that occur at different points in time, one cannot even begin to think about such decisions, much less make them. The FDA ultimately decided to save lives rather than trees, but the development of semisynthetic Taxol from the wild yew trees promises to ease the long-term harvesting of the Pacific yew.
of future generations, companies would have little incentive to develop pharmaceutical products for those now alive. Their efforts would shift to likely consumers of their products in the distant future.

In practice, pharmaceutical companies have not gone to either extreme. The evidence suggests that in their research and development decision-making process, pharmaceutical companies use a real rate of interest of 9 percent in converting future effects into present value.\(^2\) This is not, however, a riskless rate of return, since it embodies a return for some of the uncertainty in pharmaceutical research. This rate will consequently exceed the rate of discount that should be used for discounting health effects. Pharmaceutical companies have undertaken an appropriate response to the wishes of the current stockholders. The same kinds of intertemporal concerns that stockholders evidence in their other economic decisions will be reflected in their stock purchases as well.

This chapter focuses primarily on why one should choose to discount in the health context, and what discount rate one should use. From an economic standpoint, raising this question may be belaboring the obvious. However, government policy frequently prevails in this arena, by no means recognizing the need for universal discounting. In assessing the merits of its health-enhancing regulations, the U.S. Environmental Protection Agency (EPA) frequently assumes that deferred effects, such as reduced cases of cancer resulting from diminished asbestos exposure, need not be discounted. This chapter will be concerned not only with the need to discount but how one should approach discounting in the context of medical decision making.

The next section provides an overview of the rationale for discounting in medical contexts. Why do we discount in the health area, not simply for costs but also for nonfinancial effects? The third section motivates the rationale for discounting even further by indicating some of the problems that can arise by failing to discount. The fourth section explores analytically some of the ramifications of discounting for optimal experimentation strategies with respect to new drugs. The fifth section presents the multiperiod choice problem that involves discounting health effects influencing mortality and it summarizes the existing evidence on the magnitude of these discount rates. The sixth section concludes the chapter.

The rationale for discounting

*Health as an investment*

Discounting deferred impacts is an essential concern in almost any context where there is investment activity. The basic task is to convert future effects into terms that can be compared with present impacts. Using an interest rate of \(r\), the value of $1 received \(n\) years from now is \(1/(1 + r)^n\).
Discounting health effects for medical decisions

Present-value calculation used in financial contexts can be applied to other contexts as well. For example, a standard practice in policy analysis is to discount both the benefits and the costs using such a discount rate. At the present time, the U.S. Office of Management and Budget (OMB) requires that all federal agencies evaluate government regulations using a real (i.e., net of inflation) discount rate of 10 percent, although agencies are free to show the present value of their policies using other discount rates as well.

To abstract from the role of inflation, all the discount rates referred to in this chapter will be in real terms.\(^3\) One rationale for discounting future dollar amounts is that the purchasing power of this money will be less in the future than it is today because of inflation. However, one can account for the role of inflation directly in the assessment of benefits and costs. This approach is generally preferable, because there is no reason to believe that inflation rate patterns over time will differ from those of interest rates. Inflation rates for medical care, for example, have been much higher than those for the economy as a whole, with 6–9 percent inflation for the past decade. One would want to incorporate the effect of such inflation directly into any analysis, rather than assuming that the same discount rate including inflation will be pertinent in medical contexts as in other situations.

Even in an inflation-free world, there is a preference for having resources now rather than later. Money can be invested and earn a rate of interest. This real rate of interest available to investors establishes the opportunity cost of capital. In addition, economists such as Marglin (1967) have suggested that an important factor is the social rate of time preference. Capital market interest rates may not be a perfect guide, because these markets are not perfect. Society’s collective interests in transferring resources across generations may not be fully expressed in private actions, whereas social decision making may yield quite different results. For example, if society wished collectively to transfer certain kinds of resources to the future in greater measure than is reflected in private decisions, then private levels of savings would be too low and observed interest rates too high. A lower rate of interest could be used to reflect this social rate of time preference in situations in which there was a desire to foster a greater shifting of resources to the future than is provided by private market forces.

Although there is a substantial literature on the selection of the rate of discount, the choice of the discount rate continues to be a matter of substantial debate.\(^4\) Most estimates of the current real risk-free rate of return to capital range from 1 to 3 percent. What is clear is that the rate of discount is not zero. Moreover, it appears to be substantially below the real rate of interest of 10 percent that the OMB requires for policy assessments.

In situations in which individuals advocate a substantial rate of time preference, one should also take into account the fact that these high rates of time
preference may affect benefit values as well. In particular, the principal factor
driving the real rate of interest is the rate of productivity in the economy. Higher
rates of productivity in turn will raise per capita income. Since there is a
positive income elasticity of demand for health, higher income levels imply
higher valuations of health status.5

Estimates reported in Viscusi and Evans (1990) suggest that the income
elasticity of the implicit value of job injuries is approximately 1.0. The value
of job injuries is the value workers require to bear the risk of one statistical
injury. Thus, if the annual wage premium for risk were $30 for an injury risk
of 1/10,000, the value per statistical injury would be $30,000.

An income elasticity of 1.0 implies that the $30 compensation amount rises
proportionally with one’s income. If this value is applied to other valuations of
health impacts as well, and if income grows at some percentage growth rate $g$
and an implied interest rate of $r$, then the appropriate net rate of discount that
is applicable is not $1/(1 + r)$. Rather, one should net out the growth and the
benefit values so that the net rate of discount becomes $1/(1 + r - g)$. The
underlying rationale for incorporating income growth in the discounting pro-
cess is that the value of income $Y$ after $n$ years of growth is $Y(1 + g)^n$. When
brought to present value, this amount is $Y(1 + g)^n/(1 + r)^n$, which is approxi-
mately $Y/(1 + r - g)^n$. In effect, one should reflect the effect of growth in income
on unit benefit values in the analysis.

Although the cost–effectiveness literature for medical decisions has devoted
far too little attention to the role of discounting, in other contexts economists
have made discounting a central matter of concern. As one would expect, a
substantial debate exists over the extent to which society should transfer re-
sources to future generations. A higher discount rate reflects a greater present
orientation and a low discount rate reflects a greater orientation toward future
benefits and rewards.6 One can expect continuing debate over the choice of the
parameter that in effect drives the relative weight on present and future out-
comes. However, it is important to recognize that whereas the debate over the
magnitude of the discount rate is legitimate, the concept of discounting at a
nonzero discount rate is not controversial.

What is being discounted

Although there may be debate over the correct rate that should be used,
discounting costs and other dollar expenditures is straightforward. However,
bigger problems arise in discounting health effects. For example, are we dis-
counting statistical lives, years with different morbidity effects, or similar
health effects? Quality-adjusted life years (QALYs) certainly should take into
account the discounted value of the life years saved and not treat them sym-
metrically. When discounting such health effects, what is the rationale for using
the same discount rate as that for monetary effects?
Discounting health effects for medical decisions

This controversy can be muted by establishing a financial bridge between the health effects and their financial counterparts. In particular, one can first convert these health effects into a monetary equivalent and then discount these valuations.

For example, consider mortality-reducing efforts. Although some medical contexts involve the likelihood of certain life and certain death, what is usually at risk is a statistical life rather than a certain life. The issue then becomes the dollar value of a statistical life (see Chap. 5 for further elaboration). Based on evidence using labor market data, the value of a statistical life to the workers at risk is in the range of $3–$7 million. Workers who face an annual mortality risk of 1/10,000 receive extra wage compensation for risk ranging from $300 to $700.

The value of a statistical life reflects individual attitudes toward life extension. Individuals should not be expected to have the same attitudes toward life and death any more than they should be expected to have the same valuations of standard economic commodities, such as food and housing. Some of these differences are systematic. More-affluent individuals generally place a higher value on a statistical life.

In social decision contexts concerning life and death, it is useful to assume that what is being discounted is the implicit value of the statistical lives that will be saved in different years. The discount rate is applied to a monetary willingness-to-pay figure rather than to a number-of-lives-saved amount. We continue to discount money, as in the case of costs, and there is no need to discount health effects per se.

More generally, what is being discounted is the utility stream associated with certain health effects rather than the health effects themselves. It is no more controversial to discount utility streams in which health status is an argument of the utility function than it is to discount utility streams more generally. The fundamental source of all discounting stems from an economic model in which an individual maximizes the present value of a stream of utility over time, subject to an intertemporal budget constraint.

Health as a commodity

Although one can convert health into willingness-to-pay amounts, as well as subsume it within utility functions, health does differ from money in a number of ways. Most important, one cannot trade health either across time or across individuals. I am limited in terms of how I can exchange my health. I can purchase medical care and other inputs that may have a probabilistic effect on my health. But if my health were to deteriorate significantly, it might be difficult to restore it. Many health outcomes are irreversible in character.

This irreversibility also makes it difficult to trade in health status across time. If we value our health at forty-five but do not at twenty-five, then we
cannot simply shift health status across time in the same way that we would shift monetary resources. Deterioration of one's health at twenty-five may prevent one from reaching forty-five, or if one does live that long, health deterioration may have a permanent effect on one's well-being at that age.

Although these different aspects of health status do make it a very distinctive commodity, health nevertheless enters utility functions of individuals in much the same way as do other objects of choice.⁹

**Present versus terminal value**

Even if one is reluctant to discount health effects or the monetary values associated with health effects, the rationale for the discounting process is clear-cut.¹⁰ Moreover, one can establish the validity of discounting even without converting deferred health effects into their present value by converting the immediate expenditure of costs into their terminal value. Thus, one can compare the terminal value of the health effects with the value of the costs after they are converted to their future value, taking into account the interest that can be earned (Box 7.2).

**Social versus private rates of discount**

Selecting the appropriate discount rate depends in large part upon the decision maker. As a general rule, the appropriate discount rate for decisions is the intertemporal rate of trade-off that reflects the value to the decision maker of the effects being considered. Attitudes toward risk should be handled separately from discounting through appropriate valuation of the payoffs, since the premium one demands for risk may not have the same time pattern as would be imposed by incorporating it into the discount rate. In the case of private decisions, the appropriate reference point is the private decision maker's rate of trade-off and his or her relative valuation of the future effects. In the case of social decisions, one would want to use the social rate of time preference, although many economists have suggested that private opportunity cost of capital provides the most reliable guideline for this social rate of discount.

As a practical matter, these discounting decisions should be made either using the private rates of interest as the reference point or on a decentralized basis by an individual decision maker who will apply whatever discount rate he or she believes is appropriate. For example, government decisions to invest in different kinds of medical research will be governed by social rates of discount.

Private decision makers within pharmaceutical companies and other businesses will be influenced by market forces. Consequently, private decision makers' decisions will reflect a cognizance of prevailing private rates of interest.
Box 7.2 Reluctance to discount health effects not a valid reason for not discounting

Consider the following example. Suppose that in year 10 we will save two statistical lives through the introduction of a new drug. The cost of doing so is an expenditure in year 0 of $8 million. Let \( V \) be the value of life in year 10. Under what circumstances will the benefits of introducing the new drug exceed the costs? First, instead of converting the benefit values into their current amount, let us assess the attractiveness of this policy ten years from now. The benefit of saving two statistical lives in year 10 will be \( 2V \). The costs associated with this policy in year 10 will be $8 million plus the accumulated interest that we would have been able to earn over the ten-year period had we invested it. Thus, the pertinent cost figure in terms of its terminal value is \( 8(1 + r)^{10} \). In this situation, one will conclude that the benefits of introducing the drug will exceed the costs provided that

\[
2V > 8(1 + r)^{10}.
\]

Note that the health effects are never discounted.

Suppose that instead of converting the amounts into the terminal values we assess their present value. The present value of the $8 million cost allocation is simply $8 million. The present value of the lifesaving effects with a ten-year delay is \( 2V/(1 + r)^{10} \). We will conclude that the benefits of introducing the drug exceed the costs, based on a present value calculation, provided that

\[
2V/(1 + r)^{10} > 8.
\]

Straightforward rearrangement of terms shows that the terminal value requirement and the present value requirement are mathematically equivalent. One can achieve the same results as would occur under a present value calculation in which health effects are discounted by converting the monetary expenditures into their terminal value.

Although this calculation becomes more complicated in situations with multiple health effects in different time periods, the underlying economic rationale is the same. Shifting the reference point in this manner does not alter the relative attractiveness of the policies. The real issue is not whether health effects will be discounted. The fundamental question is whether one will appropriately recognize that economic effects at various points in time should be weighted differently to reflect the opportunity cost of capital.
and the private cost of capital. Private decision makers also will be affected by intertemporal resource constraints. Because of capital market imperfections, the market borrowing and lending rates may not be identical. Moreover, the private rate of interest may not always be the same as the social rate of discount. These differences are not irrational. They simply reflect the fact that opportunities may vary across individuals for transferring resources across time, and these differences in opportunities may be manifested in the rates of time preference that are applied.

**Problems with the failure to discount**

Notwithstanding the compelling economic rationales for discounting, discounting in health contexts remains controversial. In some cases, the failure to discount is an oversight. In others, the lack of any discounting of future effects reflects a conscious decision to treat all health effects symmetrically across time. For example, in the controversial case of the EPA’s asbestos regulation, the effects of the regulation on reducing cancer cases occurred with a substantial lag of two decades or more. A required discount rate of 10 percent implies that a dollar in benefits received in twenty years has a present value of only fifteen cents. The long time lag, coupled with such a high rate of discount, greatly depresses the attractiveness of this regulatory proposal. Rather than seeking a more rational discount rate, such as 2 percent, which would have led the present value of $1 in twenty years to be sixty-seven cents, the EPA objected more generally to the discounting process.

Although discounting at a zero rate frequently makes health investments appear more attractive, this is often misleading, as examples in the following section indicate.

**Policy deferral**

Suppose that we have a situation in which there are two policy options. Policy A costs $1 million and will save ten lives next year. Policy B also costs $1 million but will not begin until next year and will save ten lives two years from now. In a world without discounting of health effects, each of these policies will save ten lives and the fact that one policy saves ten lives one year from now and the other saves ten lives two years from now is a matter of indifference. Although the $1 million monetary cost of the policies is identical, Policy B will always be more attractive from a financial standpoint, since at a 5 percent rate of interest an investment of approximately $950,000 will yield $1 million a year from now. A current outlay of $950,000 will generate the saving of ten lives in two years, whereas a current outlay of $1 million is required to generate a saving of ten lives next year. The current allocations required to save ten lives next year will be higher than those for saving ten lives the following year. When
policies that save the same number of lives at different time periods are available and costs do not change over time, it will always be desirable to defer these policies when health effects are not discounted.

The role of affluence

Over time, society has become richer. Because of the positive income elasticity of the valuation of health status, this increase in income over time will increase the value placed on health outcomes. Recognition of this increased valuation over time when assessing Policy A and Policy B will simply enhance the desire to defer expenditures and to focus on efforts like Policy B. The farther in the future the lives are saved, the more highly these lives will be valued, since future generations will be more affluent. Wholly apart from the fact that Policy B will entail lower cost allocations than Policy A, there will be an advantage to deferral because the health effects generated by the policy will be more highly valued if they are generated farther into the future than if they are generated today.

Although the impetus in a zero discount rate world to defer policies will be quite strong, it is questionable whether this will be advocated as a rational policy. Clearly, the transfer of some resources to future generations is desirable. For example, there is some societal value to decreasing genetic damage and decreasing the rate of communicable diseases that will affect future health status. However, our generosity toward the future does have limits, as the public’s resistance to incurring an extra nickel per gallon gas tax to promote energy conservation has demonstrated. Most public demand for pharmaceutical research is for products related to ailments that affect those now alive or that can potentially affect the current population. This emphasis is true of government-sponsored research as well.

If future generations will be more affluent and live longer, one would ideally like these more affluent future generations to compensate their poorer counterparts who must undertake actions now to protect the future. Unfortunately, such transactions are not feasible. Moreover, if intergenerational markets like this did exist, the prevailing rate of interest for such intertemporal rates of transfer would not be zero. Weighting the present value of $1 in health effects equally with $1 in health effects a century from now will lead to a substantial income redistribution to the more affluent future generations. It is doubtful that once the full implications of such a policy were understood, any of the advocates of the zero discounting of future health effects would pursue it.

Technological change

An analogous influence arises within the context of technological change. Suppose that Policy C cost $1 million in 1993 and will save ten lives in 1994.
This policy may appear attractive when viewed in isolation. However, other potentially more attractive alternatives may emerge in the future. Medical technologies, pharmaceutical products, and other mechanisms for enhancing individual health have become increasingly effective in enhancing mortality and reducing morbidities and will continue to do so.

Suppose that Policy D takes advantage of a modest technological improvement in how this $1 million could be spent so that an allocation of $1 million in 1993 would lead to the saving of eleven lives in 2094. In a situation in which there is no discounting, Policy D will be preferable to Policy C.

Current resources can be parlayed into substantial future gains if we simply invest the money and then wait for the benefits of technological change to accrue. Spending money to save lives at the current time simply makes no sense whatsoever in a world in which the value of health effects is not discounted. More productive means for allocating medical resources will make it desirable to shift resource expenditures from the present to the future, where greater dividends will be reaped.

This is one of the trade-offs embodied in the Taxol example already mentioned. Using the Pacific yew tree bark now may save lives next year, but if we defer use of the bark until the properties of Taxol are better understood, then we may be able to save more lives utilizing these trees in the future than if we deplete this resource now.

*The permanent cost slam dunk*

One of the most fundamental problems arising from a failure to discount is that it places one in a situation in which one would never accept a risk of permanent harm in return for present gains. Suppose that society could defer by a decade all deaths that would take place in the world this year for an expenditure of $1 per year forever. Every person who would have died this year will have an extra ten years added to their lives. In a situation in which there is no discounting, one would never undertake such an effort, because an infinite $1 expenditure stream would impose an infinite loss that would dominate any finite benefits associated with the current lifesaving gains.

*Discounting and medical decisions under uncertainty*

The classic problem surrounding uncertain medical decisions involving a sequence of trials has come to be known as the two-armed bandit problem. Consider a situation in which a sequence of trials involves two different drugs. Drug A offers a probability $p$ of leading to a successful outcome in each period. The value of $p$ is, however, uncertain, since the properties of this drug are not yet known. In contrast, drug B offers a known probability $q$ of success with each trial that is known with precision. If a patient is engaged in a sequence of
Discounting health effects for medical decisions

trials, what is the optimal selection of drugs in each period? If all we are concerned with is the patient's own welfare and not how this experimentation will affect our ability to treat other patients, how should we proceed? Is drug B always superior to drug A; is \( q > p \)?

What drives this particular class of problems is its multiperiod aspect. If one were only concerned with a single period choice problem, the only issue would be whether or not \( p \) was greater than \( q \). However, because of the potential gains from experimentation, the benefits of which are necessarily deferred, the problem assumes an intertemporal nature. Discounting consequently plays a fundamental role, since the discount rate determines the weight that will be placed on the gains from experimentation and how much one can learn about the properties of the drug with uncertain effects.

This section explores different models of testing drugs, focusing on two different classes of models. The first class consists of those in which there are lotteries on life and death. In this model, success means that the patient survives, and failure means that the patient dies. In such a situation, adverse experimental outcomes with a drug that has unfavorable properties will have dire consequences. In the second class of models, which is more akin to the classic two-armed bandit model structure, continued experimentation after an adverse outcome can occur. To keep the models amenable to analysis and to obtain closed form solutions, this discussion will focus on two-period models. Within the context of two periods, one can analyze the role of learning and the potential gains from experimentation, while at the same time not overly complicating the model in a manner that would make it unwieldy.

Experimentation with lotteries on life and death

A cancer patient might be given a drug with known properties or a drug with uncertain properties. The probability that the patient will die in the initial period is higher for the unknown drug. Which drug should the patient pick? This class of experimentation models can be solved, yielding some potentially surprising conclusions.

Consider a situation of experimentation in which two outcomes are possible in each period. The first outcome is that the patient survives, which we will assume has a payoff of 1. Alternatively, the patient may die, which has a payoff of 0. For von Neumann and Morgenstern utility functions, one can establish a metric of this type with no loss of generalizability. In the second period, should the patient survive, the initial lottery is repeated, with a payoff structure that is the same as before. Rewards received in period 2 must be discounted to put them on the same basis as the expected payoffs in period 1. For purposes of this calculation, it is assumed that the patient uses a discount factor \( \beta \), equal to \( 1/(1 + r) \), where \( r \) is the rate of interest.

Trials with the uncertain drug A offer an initial probability of success of \( p \).
However, if the patient survives the trial in period 1, then after a Bayesian updating process, the assessed probability of success in the second period becomes \( p^* > p \). The key element of structure of the problem is that it terminates after an unfavorable outcome in the first period so that there is no need to take into consideration the downward revision of the probability following an unfavorable outcome.

Uncertain drug A offers an expected utility \( EU_A \) over the two periods given by

\[
EU_A = p + pp^* \beta, \tag{7.1}
\]

and drug B, which has known properties, offers the discounted expected rewards given by

\[
EU_B = q + q^2 \beta, \tag{7.2}
\]

Because of the learning process and the influence of the upward revision of \( p \) in equation (7.1) but with no effect of a downward revision following an unfavorable trial, experimentation with an uncertain drug A will generally be preferable to selection of a drug B whose properties are more precisely understood and that offers an expected immediate probability of success of \( q \). Viewed somewhat differently, the uncertain drug A offers the prospect of long-term survival. If the drug proves to have a favorable effect, then the patient can continue with it, with an expected probability of continued success in excess of the initial value \( p \).

The key concern here is the effect of the discount factor \( \beta \) on the relative attractiveness of experimentation. Clearly, if the value of \( \beta \) is zero or the interest rate is infinite, one will be concerned only with the immediate payoffs in the first period so that the decision will turn solely on whether or not \( p \) is in excess of \( q \). For finite values of \( \beta \), the value of experimentation assumes a critical role. One measure of the effect of \( \beta \) on the relative attractiveness of experimentation is to assess the effect of \( \beta \) on \( EU_A - EU_B \), yielding the following equation:

\[
EU_A - EU_B = p(p^*)(1 + \beta) + p(1 - p^*) - q^2(1 + \beta) - q(1 - q) = \beta(pp^* - q^2) + p - q. \tag{7.3}
\]

Hence,

\[
pp^* - q^2 \leq 0 \quad \text{implies} \quad p < q. \tag{7.4}
\]

In that case, differentiating with respect to \( \beta \) to find the effect of the discount factor on the value of the experiment, one obtains

\[
\frac{\partial V}{\partial \beta} = pp^* - q^2. \tag{7.5}
\]
Discounting health effects for medical decisions

Increasing $\beta$ raises the relative attractiveness of experimentation. This result is true even if the decision maker has biased risk perceptions, as shown in this chapter's Appendix.

Experimentation with nonfatal outcomes: case 1

In situations in which the lotteries facing the patient do not involve life and death, there is a chance of survival after an unfavorable trial in the first period that has a payoff of 0. Two possible situations can prevail. First, it may be optimal to continue experimentation with the uncertain drug after an unfavorable outcome in the first period. This would be the case if the drug's properties are such that it offers an expected probability of success in excess of $q$ even after an adverse outcome on the first trial. Alternatively, it may be desirable to switch to drug B following an adverse outcome with drug A if the assessed probability of success with drug A is lowered sufficiently by an adverse trial in the first period. In this section, I will consider the strategy $X$, in which the patient switches to drug B after an unfavorable trial with drug A in the first period. I will then address the situation in which there is no drug switching.

Under this scenario, the individual begins experimentation with drug A in the first period and either switches to drug B after an adverse trial in the first period or continues with drug A after a successful trial in the first period. Assuming a value of failure of 0, this leads to a discounted expected utility for strategy $X$ of

$$EU_X = p + (1 - p)q\beta + pp^*\beta.$$  \hfill (7.6)

The discounted expected utility associated with drug B is simply given by

$$EU_B = q + q\beta.$$  \hfill (7.7)

It will never be optimal to switch from drug B to drug A once started on drug B, since the gains from experimentation are necessarily deferred and the attractiveness of drug B never decreases over time.

The main issue of interest is the relative attractiveness of strategy $X$ as compared with taking drug B in the first trial and this value is given by the difference between $EU_X$ and $EU_B$, or

$$V = (p - q) + \beta p(p^* - q).$$  \hfill (7.8)

Again, if $p^* \leq q$, then $p < p^* \leq q$; strategy X can be optimal only if $p^* > q$.

Differentiating with respect to the discount factor $\beta$ to determine the effect of discounting, one finds that

$$\frac{\partial V}{\partial \beta} = p(p^* - q) > 0.$$  \hfill (7.9)

Increasing the discount factor $\beta$ will increase the value of experimentation.
Experimentation with nonfatal outcomes: case 2

If drug A offers a sufficiently high probability that even after an unfavorable trial in the initial period, the assessed probability of success exceeds that associated with drug B, then it will always be optimal to utilize drug A. Thus, in this situation, the discounted expected utility for drug A is given by

\[ EU_A = p + p\beta, \]  
(7.10)

and the discounted expected utility for drug B is given by

\[ EU_B = q + q\beta, \]  
(7.11)

The difference in the discounted expected utility between the two drugs is

\[ V = (p - q)(1 + \beta). \]  
(7.12)

Increasing the value of the discount factor \( \beta \) will necessarily boost the value of experimentation, since

\[ \frac{\partial V}{\partial \beta} = p - q > 0. \]  
(7.13)

Modeling and estimating rates of discount for health

One's immediate inclination in framing policy analyses of long-term health decisions is simply to employ financial rates of discount within the context of health decisions. Although this approach may be correct, a useful exercise is to determine whether or not implicit discount rates for health are in fact comparable to the implicit rates of discount used in financial contexts.

In a series of papers, Michael J. Moore and I have used labor market data in which workers are making long-term decisions with respect to their well-being, which is influenced by the fatality risks they face on the job (Moore and Viscusi 1988, 1990a; Viscusi and Moore 1989). Based on the wages workers receive to bear greater fatality risks, one can impute the rate of discount that they display with respect to the years of life that are at risk (Box 7.3).

The main implication of three quite different empirical models estimated using a variety of econometric assumptions is that the implicit rate of discount for life is in the range of 1.0 to 14.2 percent. Although double-digit discount rates are clearly too high, the confidence intervals for these estimates are sufficiently broad that prevailing market rates of interest are generally included within them. Thus, generally one cannot reject the hypothesis that the implicit rate of discount that individuals place on years of life equals prevailing financial rates of return. Given such evidence, it would seem that the presumption should be that the same rates of discount used to discount financial impacts should be applied within the context of medical decision making as well.
Box 7.3 Deriving implicit rates of discount from wage data

The most simplistic approach is to assess the discounted expected years of life that will be gained or lost. Within the context of the labor market models that have been used to estimate risk-dollar trade-offs, one could consequently estimate the wage premiums workers receive for the expected life years lost because of a job.

To calculate this amount, let \( R \) be the remaining years of life, \( r \) be the rate of interest, and \( p \) the annual probability of death. The expected years of life lost based on working on a risky job for one year is given by

\[
\frac{1}{r} (1 - e^{-rR})p. \quad (7.14)
\]

This approach neglects the fact that if one were to repeat this job choice over time, then the lifetime at risk would not be \( R \), but would be something less than \( R \) which would accommodate the future fatality risks the worker faces.

Using this simple framework, if one includes the discounted expected years-of-life-lost term in a wage equation and estimates the implicit rate of discount that workers display for years of life, one obtains an estimate of the rate of discount that the workers have of between 9.6 and 12.2 percent, depending on the particular empirical specification used.

A more detailed approach that reflects the fact that a repetition of risky job choices affects one’s current, as well as future, well-being is to formulate a Markov decision model to capture this choice process. In particular, suppose that the worker must select a fatality risk \( p \) in each period, where there is a probability \( p \) of death and a probability \( 1 - p \) of survival. The worker is paid a wage rate \( w(p) \) for the risky job, from which he or she derives a utility \( U[w(p)] \). For time-invariant probabilities and wages, the worker’s choice problem takes the form

\[
\max_p V = (1 - p)U[w(p)] = \beta(1 - p^2)U[w(p)]
+ \cdots + \beta^{r-1}(1 - p)U[w(p)] + \cdots \quad (7.15)
\]

or

\[
\max_p V = U[w(p)](1 - p)\sum_{i=1}^{\infty} [\beta(1 - p)]^{r-1}. \quad (7.16)
\]

One can solve this optimization problem leading to a wage equation based on the structural decision problem facing the worker. This approach yields an implicit rate of discount of 10.7 percent (Viscusi and
Moore 1989). In addition, one could estimate variants of this model in which one places restrictions on the market opportunity locus based on estimated wage equations, makes different simplifying assumptions when going from the Markov decision model to the estimating equation, or adds a bequest term explicitly to the analysis. These variants of the model yield implicit rates of discount of 1.0 percent, 1.0 percent, and 14.2 percent, respectively (Moore and Viscusi 1990b).

A final variant of the fatality risk problem is to assume that instead of leading to a period-by-period risk of death, the fatality risk simply shortens the path of life one has. Some common descriptions of smoking, for example, suggest that each cigarette shortens your life by a certain number of minutes. In reality, the risk does not work in that fashion. The years of life lost may not simply come from the tail but instead may involve premature death at, for example, fifty years of age. Thus, the life-cycle approach has the virtue of simplicity, but it sacrifices realism to the extent that one could lose years of life in the interim that would not be reflected by simply shortening the path of life in response to the risks that are faced. The main advantage is that it leads to a simple functional form. In particular, if an individual has \( T \) years of life left that are a function of the fatality risk \( p \) on the job, the worker’s objective is to select the job risk that will maximize

\[
\max_p V = \int_0^{T(p)} U[w(p)]e^{-\gamma t} dt.
\]  

(7.17)

Based on this approach, Moore and Viscusi (1990a) have derived estimates of the discount rate of between 1.6 and 2.0 percent.

The range involved in the estimates of the discount rates for health may appear to be substantial, but when compared to other discount rate estimates that have appeared in the literature, this range is actually quite tight. Consider, for example, the implicit rate of discount that individuals have revealed through their decisions to purchase household appliances. If one has a low rate of time preference, the value of the deferred energy savings associated with energy-efficient appliances will appear to be substantial when compared to the initial outlay associated with the energy-efficient appliances. Thus, one can use consumers’ choice of the degree of energy efficiency and the associated price that they pay for their appliances to infer the implicit rate of discount that they have with respect to appliance-related financial savings.

Studies of these decisions have yielded perhaps surprising estimates of the
Discounting health effects for medical decisions

implicit rate of discount. Hausman (1979) reports a series of estimates for refrigerator purchases, where these discount rates are close to 20 percent and, in some cases, much higher. Gately (1980) performed a similar study of appliance efficiency and found that the implicit rates of discount range from 45 to 300 percent. Thus, at least in this context, the estimates of the rates of discount are implausibly high. Indeed, if these estimates do in fact reflect actual behavior, they suggest that individuals may display a kind of temporal myopia with respect to energy-efficiency choices. In contrast, there appears to be no evidence of substantial temporal myopia with respect to fatality risk decisions stemming from job hazards.

The estimates based on worker decisions also appear to be in a more reasonable range than those based on survey evidence. Fuchs (1986) undertook an exploratory study of discount rates in health contexts and found, based on survey responses, that individuals often reported discount rates around 30 percent. If one actually encountered a medical decision-making context in which the patient indicated an implicit rate of discount of 30 percent, presumably the first task would be to ascertain whether the patient understood the particular choice that was being made. In particular, rather than accepting an extremely high discount rate at face value, one should determine whether the respondent understood the consequences of these intertemporal decisions for his or her own well-being.

Corporate executives appear to display discount rates in a much more reasonable range, since survey evidence suggests that their discount rates are on the order of 15 percent. Moreover, as indicated, pharmaceutical companies display a rate of discount of 9 percent. These corporate rates may embody a premium for risk that should not be included when computing the discount rate, which should pertain to riskless decisions. In particular, attitudes toward risk should be addressed through proper valuation of the payoffs.

Conclusion

An examination of the intertemporal choices in the medical decision-making context suggests that discounting is important. Many health decisions involve a substantial investment component. The choices we make now, both in our daily lives and with respect to medical care decisions, have a long-term effect on our future welfare. Indeed, the irreversible nature of many health-related decisions ensures that discounting will remain a prominent concern.

Discounting of monetary effects is an essential and widely accepted procedure for reflecting the terms of trade for individual welfare across different periods of time. The discounting of health effects has been more problematic. A failure to discount health effects altogether by employing a zero rate of discount may appear to be more farsighted in terms of its emphasis on the
future, but in practice this no-discounting approach may have the opposite result. In particular, it may lead one to defer decisions in a manner that will enhance the well-being of future generations rather than those now alive. Similarly, discount rates that are excessively high are not ideal.

Discount rates have a fundamental effect on the character of decisions in terms of their temporal orientation, particularly within the context of experimentation. Situations in which there is learning and acquisition of risk information, as in most sequential trials, are strongly affected by the rate of discount. In particular, higher rates of time preference imply a lower weight on future benefits and will reduce the gains from experimentation. Experimentation is similar to prevention in that it is a form of investment.

Because there are no markets for explicitly trading health status across time, the choice of the appropriate discount rate for health status has remained a substantial subject of debate. Estimates based on workers' job risk decisions with respect to fatality risks indicate that the implicit rate of discount individuals apply to welfare at different years in their lives is not significantly different from prevailing financial rates of return. Indeed, the estimates of the discount rates displayed in this context appear in many respects to be more reasonable than the implicit rates of time preference that analysts have found with respect to other individual decisions, such as appliance efficiency choices.

As a practical matter, many cost–effectiveness studies currently use a real rate of discount of 5 percent. Although this approach is not unreasonable, real rates of return of 3 percent, or even less, appear more in line with U.S. economic performance in the past decade. In most contexts, the medical decision will not be sensitive to reasonable variations in the discount rate. Sensitivity analyses using discount rates ranging from 1 percent to 7 percent would be of assistance in identifying the degree to which the discount rate estimate must be refined within that range.

There is no question that health is special as an economic commodity. Individual health status has many unique characteristics, not the least of which is its ability to derive welfare from any other expenditures we might make. Nevertheless, the appropriate way to recognize the special status of health is through appropriate benefit valuation of one’s health status in different periods of time. One should not distort intertemporal rates of time preference in an effort to boost the intertemporal rate emphasis a particular policy might have on health.

Appendix: Experimentation contexts with biased risk perceptions

It is useful to contrast the properties of rational experimentation with the outcomes that would prevail if there were biased risk perceptions. This bias
Discounting health effects for medical decisions

pertains to the situation in which the decision maker (possibly the physician, not the patient) has biased probabilistic beliefs. A large literature in psychology and economics has documented a variety of such biases, such as the tendency to overassess low probability risks and underassess large risks.

Consider the case where an unfavorable trial leads to the patient's death. For concreteness, assume that individuals' probabilistic beliefs can be characterized by a beta distribution, where \( \gamma \) corresponds to the precision of the individual's prior beliefs. In particular, the decision maker acts as if he or she has observed trials from a Bernoulli urn, of which a fraction \( p \) are successful and a fraction \( 1 - p \) are failures. With this formulation, the assessed probability \( p^* \) following a favorable trial of the drug in period 1 will equal \( (\gamma p + 1)/(\gamma + 1) \). With a rational Bayesian formulation such as this, the probability that the decision maker will survive for both periods, which is the key component determining the value of the experiment in equation (7.3), is given by

\[
pp^* = p \left( \frac{\gamma p + 1}{\gamma + 1} \right) = \frac{\gamma p^2 + p}{\gamma + 1}.
\]

(7.18)

It is useful to contrast this probability of survival over both periods with the perceived probability given biased risk perceptions. Although individuals may act in a rational Bayesian fashion, a variety of studies have suggested that there are systematic biases in risk perceptions. In Viscusi (1989) I formulate a model that I call "prospective reference theory" in which individuals, in effect, act as if they have some underlying prior information regarding the decision context. Instead of treating the new information as if it had an implied probability \( p \) with an associated precision \( \gamma \), they also act as if they have prior information that the probability is \( s \) with precision \( \xi \).

This formulation is in many respects a linearized counterpart of Kahneman and Tversky's (1979) prospect theory model, in which the risk perceptions flatten out. The main difference is that the prospective reference theory formulation predicts a wide variety of forms of deviations from standard expected utility maximization. In particular, this model predicts the Allais Paradox, the overweighting of low-probability events, the existence of premiums for certain elimination of risks, and the representability heuristic. In short, a large class of the paradoxes that have been identified in the choice-under-uncertainty literature can be predicted using this quasi-Bayesian formulation. As a result, I will use it here as a reference point for assessing the impact of inadequacies in risk perceptions on experimental behavior.

More specifically, under a prospective reference theory model, individuals act as if the perceived probability \( \pi(p) \) associated with any stated probability \( p \) is somewhat different in that it takes on the functional form

\[
\pi(p) = \frac{\xi s + \gamma p}{\xi + \gamma}
\]

(7.19)
and the assessed probability after a favorable trial in period 1 assumes the form

\[ \pi(p^*) = \frac{\xi s + \gamma p + 1}{\xi + \gamma + 1}. \] (7.20)

The perceived probability of survival over both periods is no longer given by equation (7.19). Instead, it is the product of the perceived probabilities, and it appears as

\[ \pi(p)\pi(p^*) = \frac{\xi s + \gamma p}{\xi + \gamma} \frac{\xi s + \gamma p + 1}{\xi + \gamma + 1}. \] (7.21)

The main issue from the standpoint of optimal experimentation, judged from the perspective of the risk perceptions of the decision maker, is whether the perceived probability of survival over both periods is greater given the biases in risk perceptions than when the probabilities are assessed as in the standard Bayesian learning case. In particular, is

\[ \pi(p)\pi(p^*) >, <, \text{ or } = pp^*? \] (7.22)

One can rewrite this condition substituting from equations (7.20) and (7.18) as follows:

\[ \frac{\xi s + \gamma p}{\xi + \gamma} >, <, \text{ or } = \frac{\gamma p^2 + p}{\gamma + 1}. \] (7.23)

As can be seen by inspection, the effect of the biases in risk perceptions hinges on two sets of parameters. First, it matters whether the individuals act as if they have some prior information implying a chance of success \( s \) before being told that the associated chance of success with drug A is \( p \). Higher values of \( s \), which imply greater chances of success, will necessarily enhance the attractiveness of the experimental drug. The second key parameter is \( \xi \). Even if the value of \( s = p \), it is of consequence that individuals act as if they have more information about drug A than is simply dictated by the precision \( \gamma \). In effect, this additional prior information tightens the probability assessments in a manner that will reduce the value of information and the informational content provided by a successful trial on the first experiment. Reducing this value in turn will decrease the importance of the deferred effects influenced by the discounting process.

The role of biased risk perceptions for the nonfatal outcome experimentation model case 1 is fairly similar to that for fatal outcomes. The condition that \( p^* > q \) can be rewritten as

\[ \frac{\xi s + \gamma p + 1}{\xi + \gamma + 1} > q. \] (7.24)
Discounting health effects for medical decisions

The key issue here is whether this condition is more or less demanding and hinges on the value of $p$ and the associated precision $\xi$. Increasing the value of $p$ makes experimentation more attractive. The effect of increasing the precision of the reference probability $p$ depends to some extent on the value of $s$ that enters in equation (7.24). If the value of $s = p$, then the net effect is to make the probabilities tighter, thus reducing the potential gains to experimentation in terms of how much the experiment will alter the probabilistic beliefs.

From the standpoint of biased perceptions in the nonfatal outcome experimentation model case 2, what matters is not the effect on perceived probabilities following success but rather the assessed probability in the initial period. Using the prospective reference theory model, the assessed probability $\pi$ of $p$ is given by

$$\pi(p) = \frac{\xi s + \gamma p}{\xi + \gamma}.$$  \hspace{1cm} (7.25)

As before, the main parameters of interest are $\xi$ and $s$, and their influence follows the same patterns.