2009

Treatment Uncertainty and Irreversibility in Medical Care: Implications for Cost-Effectiveness Analysis

Joshua Graff Zivin, Columbia University
Matthew Neidell, Columbia University
Lauri Feldman, Columbia University

Available at: https://works.bepress.com/josh_graffzivin/40/
Chapter 12

Treatment uncertainty and irreversibility in medical care: implications for cost-effectiveness analysis

Joshua Graff Zivin, Matthew Neidell and Lauri Feldman

Introduction

Over the past 20 years, the global policy environment that governs the approval and price-setting process for new pharmaceuticals has shifted its orientation from an almost exclusive focus on safety to one that is increasingly preoccupied with the ‘value’ of new products. As a result, we have witnessed the rise of cost-effectiveness analysis (CEA) and other economic valuation techniques as important tools in the evaluation of new medical interventions. In resource-constrained environments, this emphasis on value is an important step for achieving more efficient outcomes.

However, the metrics employed for determining the value of new medical technologies often produce measures that do not accurately capture its full costs or benefits, and thus its true contribution to social welfare. For example, some relatively inexpensive and effective interventions are neither covered by insurance nor provided by national health care systems, while other more expensive, but less effective treatments are regularly covered and provided. These inconsistencies may reflect the fact that economic evaluation methods in health care, as they are currently implemented, do not include all the dimensions germane to making a welfare-improving decision. The divergences between cost-effectiveness criteria and actual decision-making are often attributed to ignorance, politics, financial constraints, and equity concerns about the distribution of outcomes (Birch and Donaldson 1987; Gold et al. 1996). This divergence, however, may also be a consequence of important elements missing from the analysis as currently performed.

1 Joshua Graff Zivin is an Associate Professor in the Department of Health Policy and Management, Columbia University. Matthew Neidell is an Assistant Professor in the Department of Health Policy and Management, Columbia University. Lauri Feldman is an undergraduate student in Columbia College.
In determining the value of any new medical technology, it is essential to weigh the treatment’s immediate benefits against any potential impact on the set of future treatment options. The influence of current medical technology on the effectiveness of future potential interventions is often overlooked in adoption and coverage decisions.\(^2\) There are often large degrees of uncertainty about both the current costs and benefits of technology adoption, and/or coverage. Moreover, some health interventions, once exercised, restrict future potential interventions for both related and unrelated medical conditions.

Since actions taken today may involve some irreversible transformation of the set of available interventions in the future, this phenomenon has become known as the irreversibility problem. The benefit associated with actions that preserve treatment choices in the future, above and beyond the direct value associated with those actions, is referred to as the option value of the intervention. Investment rules that ignore this option value can be grossly in error.\(^3\)

Incorporating option values in medical technology evaluations is potentially important for several reasons. First, this setting is one where there is tremendous uncertainty about the demand for future products. When we begin treating a population of individuals, we do not know what additional conditions they will develop in the future. Since new diseases are constantly emerging, we do not even necessarily know the nature of these future conditions. Moreover, recent improvements in life expectancy, which increase the opportunity for new conditions to arise—especially those associated with aging such as cancer and dementia—make option values in this context an especially important piece of the valuation equation. Secondly, despite brisk growth over the past several decades in the number of treatments available for a wide range of conditions, treatments generally share a fairly small set of common mechanisms of action, making inter-dependencies especially likely. Since the choice of a given present treatment would probably preclude some future treatment option, this implies particularly high option values. Finally, unlike many private investment decisions, those taken by national governments may be effectively irreversible for political reasons. Once medical technologies have been authorized for public consumption, it is extremely difficult to limit their use, such as ongoing concerns regarding antibiotic over-use and the resulting development of antibiotic-resistant bacteria.

---

\(^2\) The general point that current treatment decisions affect the likelihood of developing future medical conditions has received considerable attention in the health technology assessment literature (see, for example, Weinstein et al. 1980; Meltzer 1987). Here we focus on the ‘irreversible’ influence of current treatment decisions on the effectiveness of treatment for future diseases, and thus the value associated with preserving treatment options in the future.

\(^3\) See McDonald and Siegel (1985) and Pindyck (1988), for empirical evidence about firm’s operating decisions.
Irreversibility, intertemporal uncertainty, and decision making

Certain interventions, once exercised, may restrict the set of future potential interventions available to treat future medical conditions that may arise. For example, treatment of certain types of cancer patients with a bone marrow transplant and massive doses of chemotherapy will reduce the patient’s ability to tolerate and respond to chemotherapy in the future, should some form of cancer recur (Messori et al. 1997; Schouten et al. 2000). Coverage and adoption of this course of treatment may be the best decision, but clearly its influence on the availability of future treatment protocols and the effectiveness of alternate, option-preserving treatments should be incorporated in the treatment decision.

Drugs that are subject to resistance are another example. Indeed, these concerns may help explain why there is still no consensus about when to start therapy in HIV patients (Cohen 2000; Harrington and Carpenter 2000). Some advocate the ‘hit hard and hit early’ approach, which suggests the initiation of complete treatment at the time of diagnosis in order to prevent the disease from progressing. Others are concerned that starting therapy at early stages, when T-cell counts are high and viral loads are low, may lead to the development of viral resistance to these drugs and related compounds. These clinicians advocate waiting until the disease reaches a more advanced stage to initiate therapy so that future therapeutic options can be preserved, although the disease may progress to an advanced stage more rapidly. While either approach may be appropriate, clinical and coverage decisions about when to initiate therapy should clearly weigh the benefits and costs of starting early versus waiting over the entire potential treatment time horizon.

This problem of current actions influencing the availability of future potential actions has received considerable attention in the environmental economics literature (Arrow and Fisher 1974; Hanneman 1989; Kolstad 1996) and, more recently, in the economic investment literature (for a good review, see Pindyck 1991). This literature suggests that if more information about the costs and benefits of these future potential interventions will become available over time, it may be optimal to delay investment in the ‘irreversible’ technology. The value of preserving options by delaying decisions until a future time when more information is available is called the quasi-option value of the decision.

Several recent studies suggest that investment and environmental development rules that ignore this value can be grossly in error (e.g. Brennan and Schwartz, 1985; McDonald R and D Siegel 1985; Pindyck 1991). For example, Fisher and Hanneman (1986) analysed a land development decision, where the land being considered for development might contain valuable genetic material for commercial maize production. While the benefits were more than triple the costs of development, the quasi-option value to waiting 1 more year for additional information was almost half the expected benefit. When the costs of waiting are less than the benefits, current development would clearly be sub-optimal. Given the similarities between environmental protection decisions and healthcare coverage decisions, similar effects may undermine the value of current CEA practice.
Let us further examine the role that uncertainty and irreversibility can play in determining the value of a particular medical intervention. For simplicity, we suppose that people live for only two periods, the discount rate is zero, and all treatment costs are equal and negligible. Consider a chronic disease $X$, which first occurs in period 1 (and always lasts two periods). Suppose that there are two choices in treating disease $X$, denoted $T_1$ and $T_2$, respectively. Patients with disease $X$ treated with $T_1$ gain 6 health units in each period. Patients treated with $T_2$ gain 4 health units each period.

A traditional economic analysis would indicate that patients should be treated with $T_1$ in both periods, as illustrated in Fig. 12.1 (with purely dominated strategies demarcated by short parallel lines).

Now consider an acute disease, $Y$, which may occur in period 2. There is only one treatment ($T_Y$) for disease $Y$. Patients who contract $Y$ lose 8 health units in the period in which the disease occurs. Patients who are treated with $T_Y$ recover all of these units. Clearly, treatment with $T_Y$ is strongly preferred to non-treatment for those patients with condition $Y$.

The problem of irreversibility occurs in circumstances when these two situations are combined and the treatments are inter-related. Suppose that patients who, in period 1, contracted disease $X$ and were treated with $T_1$ cannot tolerate $T_Y$ (i.e. $T_Y$ is ineffective in treating $Y$) if they later develop disease $Y$. Let $p$ denote the probability of developing disease $Y$, which we will assume is independent of developing (or treating) disease $X$. In this case, $T_2$ has two values associated with it—one due to its effectiveness in treating $X$ and another due to it preserving the option to effectively treat $Y$. The treatment decision in this case is illustrated in Fig. 12.2. When patients are offered $T_1$ in the first period, it will always be optimal to offer $T_1$ again in the second period, although the overall benefit of this treatment strategy is clearly diminished for cases where disease $Y$ develops. When patients are offered $T_2$ in the first period, the arrival of disease $Y$ will dictate the second period protocol. If $Y$ does not arrive, patients will

---

4 Note that we are being intentionally vague about our quality-of-life measure so as to abstract away from the specific assumptions associated with conventional outcome measures, such as the quality-adjusted-life-year.
MODELLING TREATMENT VALUES WHEN ACTIONS ARE IRREVERSIBLE

199

switch to T₁ in the second period, since there is no benefit from option preservation at this point. If Y does arrive, then patients are treated with T₂ and T₁.

The incremental value of T₂ (relative to the value of T₁) can be expressed as the difference in expected effectiveness between starting with T₂ and providing patients with T₁ in both periods:

\[ p \cdot (8) + (1 - p) \cdot (10) \] – \[ p \cdot (4) + (1 - p) \cdot (12) \]

This value reflects both the direct treatment benefits of T₂ for condition X, as well as the value of preserving the option to use T₁ in period 2. This latter value is known as the option value associated with T₂. While this option value will always increase the value of T₂ relative to T₁, it will not always be the case that this ‘extra’ value is enough to justify its use. In this simple example, T₂, which is an inferior treatment for disease X alone, would only be the preferred treatment option if the probability of developing disease Y is greater than one-third. In general then, we see that the optimal treatment plan is contingent upon future expectations.

**Modelling treatment values when actions are irreversible**

Option values can be quantified and incorporated in analyses by explicitly formulating a multi-period, dynamic decision model that attends to uncertainty and the potential role of irreversibility. Generally speaking, there are three cases we must deal with complete irreversibility, partial irreversibility, and partial irreversibility with the
possibility of learning more about the probability of future disease arrival (hereafter partial irreversibility with learning).

To illustrate these three cases, we again consider diseases X and Y, and the same set of treatment options as in our earlier model. We begin assuming that disease X is present in the first period. Disease Y will then arrive in the second period with some probability equal to $p$. Remember from the previous illustration that $T_2$ is an inferior treatment for condition X, but yields better results should it need to be combined with $T_Y$. Thus, there is an option value associated with using $T_2$, but that option comes at a cost.

First consider the case of complete irreversibility. Since the downstream impacts from ever receiving a given treatment are permanent, it will never make sense to switch from initial treatment regimens. Therefore, patients initiated on $T_1$ will receive that treatment in perpetuity and the same will be true for patients initiated on $T_2$.

As such, the decision-maker will use the forecast about the arrival rate of disease Y to calculate the expected present discounted value (PDV) of each treatment regime and select the one with the highest return.

The second case in need of consideration is that of partial irreversibility without learning. In this version of the model, we relax the assumption of complete irreversibility by assuming that the negative impacts of $T_1$ on the effectiveness of $T_Y$ are temporary. In particular, we assume that after some time or cooling off period of duration $n$ following treatment with $T_1$, $T_Y$ can be fully effective once again. Since agents still cannot perform tests to update their knowledge about the arrival of Y, the least sophisticated treatment protocol would start patients on $T_1$ and switch them to $T_2$ when disease Y arrives. Under such a protocol, patients will pay a ‘penalty’ equal to $n$ times the value of $T_Y$, which they forgo while they are waiting for the cooling off period to end.

Of course, decision-makers can use the information that they have about the expected arrival rate of disease Y to do better than simply waiting for it to arrive before switching to $T_2$. In this case, the optimal treatment strategy requires switching when the expected marginal benefit of one more period of treatment with $T_1$ is exactly equal to the expected present discounted value of adopting treatment with $T_2$.

Now compare the value of $T_2$ under complete irreversibility to its value under partial irreversibility. When the cost of the cooling-off period is smaller than the relative value of the advantage of $T_1$ over $T_2$, the value of $T_2$ for those patients that would not receive it under complete irreversibility increases in a world of partial irreversibility. If the cooling off cost is larger than the benefits that would be received from introducing $T_2$, then the value of $T_2$ again decreases under partial irreversibility.

5 Unlike the two period example given in the previous section, the last period of the disease is unknown, so individuals who begin on $T_2$ will not switch to $T_1$.

6 Note this characterizes the optimal interior solution where both treatments are used. A corner solution where patients exclusively receive $T_1$ or $T_2$ could also arise when the probability-adjusted costs associated with the inferior treatment of Y is extremely high or low, respectively.

7 In this last case, the value of $T_2$ is going from negative under complete irreversibility to more negative under partial irreversibility. If we view the value of $T_2$ as bounded from below at zero, then the value of this treatment will be the same for these types of patients under both scenarios.
For those patients who would not have received $T_2$ under complete irreversibility, the move to partial irreversibility has an ambiguous effect on its value. The intuition for this result relies on the recognition that the introduction of partial irreversibility has two distinct effects. First, it reduces the amount of ‘wasteful’ $T_2$ that needs to be offered to patients to ensure that they can be effectively treated for condition Y when it arrives. This yields changes on the intensive margin, since under partial irreversibility these individuals no longer need to initiate treatment ($T_2$) at time zero. Secondly, by reducing this ‘wasteful’ spending, the use of $T_2$ may become attractive for some patients who did not find it attractive when costs were higher. This yields changes on the extensive margin, since a subset of patients that would not have received any $T_2$ under complete irreversibility will receive some when irreversibility is partial. The first effect leads to reductions in the value of $T_2$, while the second effect leads to increases. The overall impact of a move to partial irreversibility on the value of $T_2$ will depend on the distribution of patients in each of these groups. Moreover, the value of $T_2$ largely depends on underlying economic and biologic parameter values, and the incident rate of disease Y.

Finally, we must consider the situation of complete irreversibility with possibility of learning more about the probability of disease Y occurring. Up to this point we have assumed that decision makers have basic knowledge about the population-level incidence of disease Y, but that they have no means of updating that knowledge with information about individual patients until the disease actually arrives. Now, consider the possibility of a test that can be performed to measure levels of the biologic indicator for Y within patients. The decision maker will perform the test precisely at the moment where the net benefits from waiting one more period to switch equal the costs of the test. If the test reveals that the biologic indicator is higher than expected, the decision-maker will switch earlier. If the test reveals that the biologic indicator is lower than expected, the decision-maker will wait. This process can be repeated indefinitely—thus, always ensuring that the treatment is optimal for the given period. The ability to increase knowledge about the probability of disease incidence in future greatly assists the present decision-maker in determining the optimal present treatment.

This test also significantly impacts on the value of $T_2$. The introduction of learning will lower the overall option value of $T_2$ for those that would have received it under partial irreversibility absent learning. Moreover, increases in the cost of the test make deviations from the expected protocol under partial reversibility without learning less likely and thus increase the value of $T_2$. The responsiveness of the value of $T_2$ with respect to underlying parameters remains qualitatively the same.

**Conclusions**

For many physicians, the observation that current medical treatment decisions have repercussions for the treatment of health conditions in the future, is an obvious one that often factors into their clinical decision-making, albeit heuristically. Yet, at present, such considerations form no part of health care technology assessment calculations at societal or sub-societal levels, leading to potentially significant mischaracterizations of treatment value. Growth in the availability of treatments for
chronic diseases that require long-term interventions, along with general increases in life expectancy, suggest that the importance of this omission will only become larger.

It is thus essential to use a valuation model that attends to these inter-temporal dependencies—one that considers the uncertain arrivals of future diseases and the varying degrees of effectiveness with which they can be treated. More generally, there are several key insights that we can all consider. Irreversibility raises the value of the option-preserving treatment. The existence of an option value—that value above and beyond the direct value of treatment for the current condition—means that a treatment that is inferior to an alternative for its indicated use may be the superior choice when lifetime welfare is considered. Optimal decision-making requires a careful comparison of the ‘costs’ of a less effective treatment for a condition today with the ‘benefits’ of more effective treatments for conditions in the future. The size of the option value and thus the degree to which valuations that ignore it are miscalculated, depends critically on the relative effectiveness of treatments, the likelihood of diseases arriving in the future, the extent to which current interventions limit the ability to treat these future conditions, and the relevant discount rate.

Comparing treatment values under complete irreversibility to a scenario where irreversibility is only partial reveals a more nuanced role for option values. For those patients at high-risk for the future disease the value of the option-preserving treatment falls when irreversibility is incomplete, while the value increases for those at low-risk. In a world with patients that are heterogeneous in their risk for future diseases, the distribution of patient types will determine whether the option-preserving treatment is more valuable in an environment of partial or complete irreversibility. Each of these impacts is compounded when a test for future disease progression is introduced.

The intuition for these results is deepened when we recognize that one of the principal features driving our results is that health risks for future diseases increase with age, a characteristic that accurately describes the incidence profile for many diseases throughout the world. When patients are heterogeneous in early environmental exposures or their genetic predisposition to disease, older low-risk patients are conceptually akin to younger high-risk ones, since both have similar risks of developing disease in the coming years. The key distinction is that under complete or near-complete irreversibility, low-risk patients must begin the option-preserving treatment when they are young and it is precisely at this point that the treatment is least valuable. Only in the cases where future disease is both very likely and of significant consequence, i.e. when the option value is very large, will it make sense to place low-risk patients on this option-preserving regimen. As the ‘costs’ of reversing current treatment impacts fall, it becomes more feasible to provide the option-preserving treatment to these low-risk individuals at a later age, even if the impacts of future disease are modest.

This basic framework of option-valuation also highlights potentially important macro-level strategies to improve social welfare through medical technologies. Research investments that focus on transforming irreversibility from complete to partial could generate large social benefits. Clearly, investments in the development of alternative treatments for future diseases are also important, but the return to such
investments will hinge critically on the degree of irreversibility in the system. Insofar as irreversibilities are a function of shared modes of actions among interventions, investment strategies that focus on a diversified portfolio of interventions in terms of these modes will also create social benefits. At a broader policy level, this argument points toward support for research that strengthens our understanding of disease and its evolution over time, which stands in contrast to approaches that focus on investing in disease treatments for the sole purpose of reducing the onset and severity of symptoms. Assessing the value of such strategies is an important area for future research.

Also essential is a revision of the CEA methodology itself. CEA is used to determine optimal funding and resource allocation for medical technologies, research, and treatment. As long as CEA fails to incorporate uncertainty and irreversibility, it will continue to miscalculate medical valuations with potentially severe consequences. Revision will take time and will come at a cost, but the option value of future policy gains is high.

References


