

Where Is the Good Stuff?
Drug Quality Measures and Economic Research*

Josh Feng
(Harvard University)

Luca Maini
(Harvard University)

December 5, 2016

Abstract

Prescription drug quality is an important variable in several strands of the health economics literature, including research on the value of medical R&D, the structure of the pharmaceutical markets, and the impact of innovation policy. In this paper, we provide a starting point for researchers interested in these topics by first summarizing the data, methodology, and literature surrounding the measurement of drug quality, and then highlighting areas of economic research that use quality measures. We conclude with a discussion of possible areas for future research.

*Research reported in this publication was supported by the National Institute on Aging under Award Number R24AG048059 to the National Bureau of Economic Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or NBER.

Contents

1	Introduction	3
2	Drug Quality Measurement in Economics	4
2.1	Revealed Preference Approaches	4
2.2	Inference from Outcomes	5
2.3	Direct Measures of Drug Quality	6
2.3.1	Limitation of clinical trial data	6
3	Data Sources on Drug Quality	7
3.1	ClinicalTrials.gov	7
3.2	FDA Drug Trial Snapshots	8
3.3	Evidex by Advera Health	8
3.4	Oncology Drugs: ASCO and DrugAbacus	9
3.5	Tufts Medical Center Cost-Effectiveness Analysis Registry	9
4	Pharmacoeconomic Approaches to Measuring Drug Quality	10
4.1	Models for the Evaluations of Health Technologies	10
4.1.1	Decision tree models	10
4.1.2	Markov models	12
4.1.3	Individual level models	13
4.1.4	Why are these models useful?	14
4.2	A Markov Model of Hepatitis-C	14
5	Usage of Drug Quality in Economics Research	17
5.1	Value of Medical Research	17
5.2	Quality and Market Outcomes	18
6	Future Work	19

1 Introduction

Prescription drugs play a key role in the healthcare system, with retail sales representing about 10% of US health expenditures, and non-retail sales representing an additional 4%.¹ Several actors play key roles in the market for prescription drugs. Pharmaceutical companies and university researchers discover and develop molecular compounds, government agencies such as the US Food and Drug Administration (FDA) evaluate prospective drugs, payers decide which ones to cover in their formularies, and doctors and patients choose between available treatments for various ailments.

In order to evaluate how well this overall system or any of its parts are functioning, economists need some measure of the value generated by prescription drugs, which depends on their quality or characteristics. Having reliable measures of quality can help researchers ascertain whether policies related to the development and prescription of drugs improve overall health outcomes in a cost-effective manner. This motivates our overview of drug quality measures in this paper. We do not claim to provide the ultimate, or even the most complete overview of this topic. Rather, our hope is that this paper can serve as a starting point for researchers interested in these questions.

To focus the discussion, we define drug quality as the *effect of a drug on patient health outcomes*. This includes individual efficacy and safety attributes (as measured by clinical trial endpoints and side-effect frequency and severity), as well as summary measures such as gain of quality-adjusted life years (QALYs).² This definition of quality is closely tied to the notion of cost-effectiveness, which is often measured in terms of dollars spent per QALY, thus incorporating information about a drug's cost to society.

We exclude additional considerations that are sometimes part of drug evaluations, like budgetary impact, scientific novelty, R&D costs, rarity of the disease, and number of existing treatments available. These factors are included in quality reviews by some organizations, such as the Institute for Clinical and Economic Review (ICER) and the Evidence Driven Drug Pricing Project. However, they account for social preferences that are generally less objective than clinical measures of safety and efficacy. Due to the limited scope of our paper, we will therefore only touch on these additional factors in passing.

The rest of the paper is structured as follows. In Section 2, we survey existing approaches to measuring prescription drug quality in economics. We distinguish three strategies used in the literature: revealed preferences, inference from longitudinal health gains, and direct measurements. Each has its advantages and drawbacks, which highlight some of the difficulties in measuring drug quality. Then, in Sections 3 and 4, we discuss data sources and more advanced methodologies that scientists outside of economics use to model the impact of drugs. Next, in Section 5, we discuss the importance of these measures to a variety of topics in health economics. These

¹The first figure comes from the 2015 National Center for Health Statistics Report (<http://www.cdc.gov/nchs/data/abus/abus15.pdf>), and the second from Altarum Institute (2014).

²A count of years alive, weighted by a quality score between 0 and 1 for different health states, with 1 being perfect health.

questions include the value or productivity of medical research, the role of quality in formulary construction, and the relationship of quality to market outcomes such as price, patient choice, and advertising decisions. Some of these questions are tied to important public policy decisions. Section 6 concludes with a discussion of areas of future research surrounding the measurement and usage of quality measures.

2 Drug Quality Measurement in Economics

Economics and health policy researchers estimate drug quality to measure the impact on longevity and health, the efficiency of government spending, or the effectiveness of policies that are meant to stimulate innovation.

Much of the existing literature has relied on more traditional economic approaches to calculate the welfare generated by drugs, such as revealed preferences. However, some researchers have taken alternative approaches, such as using longitudinal health gains to measure welfare changes directly, or incorporating quality measures from clinical studies in economic models. We review these approaches and their relevance in this section.

2.1 Revealed Preference Approaches

The revealed preference approach consists of deriving the quality of a drug from patients choices through a demand system, and then measuring the welfare drugs generate by tracing the estimated demand curve. This approach usually requires some structural assumptions in the form of a demand system (for an example, see [Arcidiacono et al., 2013](#)), unless the researcher can identify an exogenous source of price variation that fully identifies the demand curve (see for example, [Goldman et al., 2010](#)).³

Such an approach is predicated on the assumption that a patient's welfare is reflected through her choices, which is standard in economics. However, there are reasons to believe that it might not hold in the pharmaceutical market. Patients usually choose therapies together with a physician. Both parties have imperfect information that might prevent them from reaching the optimal decision: patients might not be familiar with the effects of drugs, while doctors might not know exactly the symptoms experienced by the patient. Moreover, drugs do not have identical effects on all patients. In many cases, different drugs might be appropriate for different patients with identical diagnosis. Because of this idiosyncratic component, drugs are sometimes modeled as experience goods - that is, goods for which the quality is uncertain prior to consumption. [Crawford and Shum \(2005\)](#) develop one such model to study the market for anti-ulcerants. While they find evidence of heterogeneous effects, the overwhelming majority of patients in their data only ever tries one drug, and no patient tries more than three. On the one hand, this result might suggest

³A demand system is a discrete choice model estimated using either aggregate or individual level data. A demand system of pharmaceuticals will usually include a price sensitivity parameter and indicators for each drug molecule, which capture the average quality of a given drug.

that doctor and patients are remarkably effective at identifying the right drug from the onset. On the other hand, it might mean that learning about drugs is difficult and takes a long time, or that patients exhibit a high degree of inertia.

Another piece of evidence suggesting that revealed preferences might not reflect true welfare is the fact that quality estimates from individual choice models are usually much larger for brand drugs than for their generic counterparts (for example, see [Dunn, 2012](#)). But generic drugs are engineered to exactly replicate the active principle contained in brand medications, and research shows that more informed consumers tend to prefer generics to brands at much higher rates than less informed ones ([Bronnenberg et al., 2015](#)).

Finally, even if one assumes that patients and physicians have perfect knowledge of drug quality, the fact that neither party bears the full financial cost of the prescription and that they sometimes are not aware of the costs, can lead to distortions in inferred quality when using a revealed preference approach.

Despite these drawbacks, revealed preference estimation has two main advantages. First, it considers the decisions of the individuals who are actually going to be using the products, helping to capture idiosyncratic match quality between patient and drug; and second, it provides a comprehensive measure that synthesizes all drug characteristics, from efficacy to side effects and more.

2.2 Inference from Outcomes

Another strategy to capture the quality and welfare impact of innovative pharmaceutical products is to look at how they affect population health. This approach exploits heterogeneity in the availability of drugs across geographical areas and over time. However, since population health is the result of many different inputs, identifying credible sources of variation is sometimes difficult.

Researchers have tried to solve the variation issue in two ways. One way is to focus on a specific disease. This requires data on disease-specific mortality. Oncology is a particularly appropriate area of study for this strategy because health gains in oncology are almost entirely due to medical innovations, and because several countries break down mortality by type of cancer. For example, [Dubois and Kyle \(2016\)](#) exploit variation in the introduction of oncology drugs in Europe to look at changes in site-specific cancer mortality. [Budish et al. \(2015\)](#) - who study underinvestment in early-stage cancer R&D - use differential progress in five-year survival rates across cancer types to estimate the quality-adjusted life years gained from alternative R&D policies. Another strategy is to try and look at the overall contribution of all drugs to longevity, as in [Lichtenberg \(2005\)](#).

Similarly to revealed preferences, this method provides a relatively complete assessment of the impact of a drug. However, it usually leaves out a few important aspects of quality, such as the impact of side-effects on the quality of life. It also requires a significant amount of time to track health improvements, making it difficult to evaluate new drugs. Its advantages over demand systems are that it uses actual health improvement outcomes, and that it is simpler to implement,

provided that one can identify a credible source of variation.

2.3 Direct Measures of Drug Quality

Finally, several papers use “direct” measures of drug quality. Direct measures include quality ratings and labels from various agencies and organizations, as well as estimates of drug quality from clinical evidence.

The simplest way to directly measure the quality of a drug is to use FDA classifications. One simple classification is new molecular entity (NME), which the FDA assigns to drugs based on a previously unapproved molecule.⁴ Researchers have usually used NME counts as a measure of the quality of R&D output, and this method can be a quick and effective way to estimate the impact of innovation policy or market size on R&D outcomes, as in [Acemoglu and Linn \(2004\)](#) or [Blume-Kohout and Sood \(2013\)](#). A slightly more sophisticated approach is to classify NMEs in terms of their degree of innovativeness. [Lanthier et al. \(2013\)](#) provide one such classification by dividing NMEs in three categories: first-in-class, advance-in-class, and addition-to-class. First-in-class drugs represent new ways to treat a specific disease, while addition-to-class drugs are drugs that, despite not being first-in-class, received priority review designation from the FDA.⁵

However, these counts only represent proxies for quality. A more precise approach used by some researchers is to use data on efficacy and safety profiles from clinical trial evidence. Researchers can pull data on several characteristics (e.g. endpoints, side effects, drop-out rates, etc.) and use them as inputs in demand systems or hedonic price indices ([Berndt et al., 1996](#); [Cockburn and Anis, 2001](#); [Lucarelli and Nicholson, 2009](#); [Suslow, 1992](#)). Another approach is to use estimates of QALYs gained from taking each drug, which are based on clinical trial results, to estimate health gains under different market outcomes and policies. Several databases, including some recently established ones, collect data on these measures. We cover some of them in Section 3.

The key advantage of direct measures is that they are typically available by the time a drug enters the market.⁶ A further advantage of the clinical trial measures is that they can be used to estimate health gains and welfare.

2.3.1 Limitation of clinical trial data

The inclusion of direct measures of quality is probably the approach that holds the most future promise, because of the increasing availability of new datasets (see Section 3 of this paper). How-

⁴As opposed to drugs that are modified versions of previously approved ones, such as reformulations and changes in dosage.

⁵The FDA assigns priority review designation to drugs that have the potential to substantially improve the safety, effectiveness, diagnosis or prevention of serious conditions. While [Lanthier et al. \(2013\)](#) only uses priority review, the FDA also assigns several other designations, which could in principle be used to create alternative classifications: Fast-Track, Breakthrough Therapy, Accelerated Approval, and orphan status. More information on how these designations are assigned can be found on the FDA’s website: <http://www.fda.gov/ForPatients/Approvals/Fast/default.htm>

⁶However, there are also clinical trials run after drug approval, the results of which could be incorporated in choices made by payers, doctors, and patients.

ever, a word of caution about using this data is necessary. Clinical trials can often only offer a noisy signal of a drug's performance in the real world. Rothwell (2005; 2006) provides a detailed review of the factors that can affect external validity. He mentions that patients who enroll in these trials are not randomly selected, so they might not be representative of the population who will use the drug once it is approved. Enrollees are then followed very carefully by a team of physician and nurses throughout the whole duration of the trial, thus minimizing issues of non-adherence and inappropriate drug use. In addition, differences in trial protocol and setting have made comparisons of drugs based on separate studies difficult, to the point that the American Society of Clinical Oncology (ASCO) now limits its comparative analysis to head-to-head trials (Schnipper et al., 2015). Finally, because clinical trials enroll small cohorts of patients and follow them for a relatively short time, their ability to capture heterogeneity in response or long-term outcomes is limited by design. Side effects in particular have been singled out to be rather noisy in trials, which makes the safety profile of a drug hard to evaluate (Ahmad, 2003), and empirical research on how well clinical trials can predict real world outcomes is sparse (Panayidou et al., 2016 review a few such studies).

3 Data Sources on Drug Quality

In this section, we describe several databases that contain direct measures of drug quality. For each database, we will discuss the types of data they offer, and possible benefits for researchers. Many of these contain disaggregated measures of quality, separating out clinical benefits and side-effects into several dimensions. The exception is Tufts Medical Center Cost-Effectiveness Analysis Registry, which contains QALY estimates for each drug. These databases are all based on clinical trial evidence, so the caveats mentioned in 2.3.1 apply. For reference purposes, we summarize the contents of this section in Table 1.

3.1 ClinicalTrials.gov

The most commonly used database of clinical evidence is the one available from ClinicalTrials.gov. The website was established as part of the Food and Drug Administration Modernization Act of 1997 (FDAMA), in order to collect and share information on clinical trials conducted on experimental drugs with “serious or life-threatening diseases or conditions.”⁷ Over time, the requirements have expanded to include the registration of more clinical trials and the timely posting of trial results. This means that ClinicalTrials.gov should contain information on all approved drugs, as well as all unsuccessful drugs that reach the latter stages of clinical testing. However, studies such as Saito and Gill (2014) have found that close to 30% of trial results are not reported within 4 years of trial completion. In addition, the results posted are not always complete and may be difficult to parse for researchers.

⁷<https://clinicaltrials.gov/ct2/about-site/history>

In terms of specifics, the database contains information on over 220,000 clinical trials. For each trial, it lists characteristics such as the drug(s) being tested, the condition treated, and the study phase. Trials with posted results will have a page displaying primary and secondary endpoint measurements as well as side-effect rates, broken down by treatment and control groups. Researchers can search trials by attributes such as the drug(s) being tested, the condition treated, and the study phase, and then download information on the set of relevant trials in XML format.

3.2 FDA Drug Trial Snapshots

The FDA has been publishing Drug Trial Snapshots of approved drugs since May of 2014, putting together information based on the outcomes of clinical trials that led to approval.⁸ Each approved drug has its own snapshot page, which contains tables on efficacy and side-effects, broken down by age, gender, and race - provided enough data is available.⁹ These tables summarize the results of pivotal trials that led to FDA approval, and drugs in the same indication usually have the same set of clinical outcomes.

The main advantage of Drug Trial Snapshots is the clarity with which the information is presented. It is also an accurate reflection of what the FDA knew at the time of approval, which may help researchers studying drug approval policies. It only contains data on drugs approved in the last two years, but it will expand over time, and it is possible that in the future the FDA might decide to provide Snapshots for older drugs.

3.3 Evidex by Advera Health

In 2016, Advera Health began collecting all clinical trial results for drugs in several major indications, as part of its proprietary Evidex service.¹⁰ Currently, they have completed data collection for Hepatitis C, diabetes, and multiple sclerosis. The limited coverage of the database is its main drawback, though coverage should expand in the future.

Evidex collects data from ClinicalTrials.gov and FDA Drug Snapshots, as well as from all other pre- and post-approval clinical trials published in medical journals. For each indication, they collect the data on the same endpoints and side effect across trials of all drugs, allowing for simple aggregation and comparisons. In addition, they collect characteristics of each trial, including the sponsor, the phase, and the number of participants, allowing researchers to test robustness through filtering and/or weighting of trial results. Finally, Evidex also collects evidence on real-world side-effects for all FDA-approved drugs, based on data from the FDA Adverse Event Reporting System (FAERS).

There are a few key advantages to this dataset. First, the dataset is collected and organized by medical experts, removing the need to parse raw results data. Second, the dataset is longitudinal

⁸<http://www.fda.gov/Drugs/InformationOnDrugs/ucm412998.htm>

⁹For example, <http://www.fda.gov/Drugs/InformationOnDrugs/ucm457839.htm> contains information on Hepatitis C treatment Daklinza.

¹⁰<http://www.adverahealth.com/solutions/overview>

in nature, which allows tracking of scientific evidence changes over time for each drug, akin to the approach taken by [Azoulay \(2002\)](#).¹¹ Third, it contains all drugs that reached at least Phase II of development, even if they did not receive FDA approval. This can be helpful for studying the role of quality in drug development decisions. Finally, it offers evidence on real-world side-effect rates, which may be a more accurate reflection of a drug's safety profile.

3.4 Oncology Drugs: ASCO and DrugAbacus

Next, we will cover the efforts of two organizations in collecting detailed information on the quality of oncology drugs.

The first organization is the American Society of Clinical Oncology (ASCO), which has recently developed a rating system for assessing the value of cancer treatments. As detailed in ([Schnipper et al., 2015; 2016](#)), the goal of ASCO is “to provide a standardized approach to assist physicians and patients in assessing the value of a new drug” in relation to existing standards of care. Specifically, the ASCO value framework combines measures of clinical efficacy (median length of overall survival or progression-free survival) and toxicity (based on the frequency of serious side-effects) from comparative clinical trials relative to the existing standards of care into a net health benefit (NHB) points measure. This effort remains in its infancy, as ASCO has only released preliminary reports on a few drugs.

The second organization is the Evidence Driven Drug Pricing Project, which has created the DrugAbacus tool.¹² The tool is based on data collected for “52 cancer drugs approved between 2001 and 2015 by the US FDA for the treatment of cancer.”¹³ The quality measures are gathered from FDA labels and include median survival and progression-free survival, as well as side-effects. The tool also collects data on several other dimensions that we have not covered in this paper, including drug novelty, and R&D costs.

3.5 Tufts Medical Center Cost-Effectiveness Analysis Registry

The Tufts Medical Center Cost-Effectiveness Analysis Registry (CEAR) is slightly different with respect to previous datasets.¹⁴ It focuses on studies that evaluate cost-effectiveness rather than direct clinical evidence. The database not only covers drugs, but also medical interventions such as immunizations, medical devices, and surgical procedures.

The database was created by searching MEDLINE using keywords related to cost-effectiveness. Relevant articles were then manually processed, recording the results of the study, which includes the target population, the relevant intervention (drug name), any comparator therapies used in the control arm, all methodological assumptions, the QALYs gained through the treatment, and

¹¹Azoulay looks at the impact of new clinical information on firm advertising and market shares in the anti-ulcer market.

¹²www.drugabacus.org

¹³The underlying dataset is used in [Howard et al. \(2015\)](#), which we discuss in Section 5.2.

¹⁴<https://research.tufts-nemc.org/cear4/AboutUs/WhatistheCEARRegistry.aspx>

also an estimate of the cost-effectiveness ratio value.¹⁵

This database has the widest coverage of all the ones we presented, going back to 1976, and provides ready-to-use QALY measures, which means economic researchers don't need to create a welfare measure from raw data on efficacy and safety. However, it has incomplete coverage of approved prescription drugs.

4 Pharmacoeconomic Approaches to Measuring Drug Quality

Many of the datasets we presented in the previous section contain raw data about a drug's efficacy and safety, which can be hard to distill into a single measure of welfare. Moreover, individual trials tend to report average or median effects, which can sometimes mask the fact that drugs affect patients in different ways, or that they can have important dynamic effects on the progression of a disease.

To fully illustrate these effects, a different approach is required. Pharmacoeconomics is a field dedicated to the evaluation of drugs in more complex and structured way. Researchers in this area have developed a series of disease models that simulate how drugs affect the progression of disease in the population.

In this section we provide a theoretical overview of these models, and discuss how they can contribute to economic research. We also provide an example of how these models are implemented in practice, drawn from the assessment of Sovaldi done by the National Institute of Health and Clinical Excellence (NICE) in the UK.

4.1 Models for the Evaluations of Health Technologies

Broadly speaking, models for the evaluation of health technologies fall into three classes: decision tree models, Markov models, and individual level models. Below we provide a brief description of the main characteristics of each model.¹⁶

4.1.1 Decision tree models

Decision tree models have a finite horizon, and are therefore most appropriate for interventions that are meant to cure a disease, rather than manage it. They consist of one or more decisions whose outcomes can be stochastic. Figure 1 provides an example of one such model to evaluate when and if a course of antibiotics should be provided after a tick bite as a prevention measure against Lyme disease (Magid et al., 1992). In the tree, the square represents the decision, while the circles indicate stochastic outcomes. In this case, the researchers are comparing three alternatives:

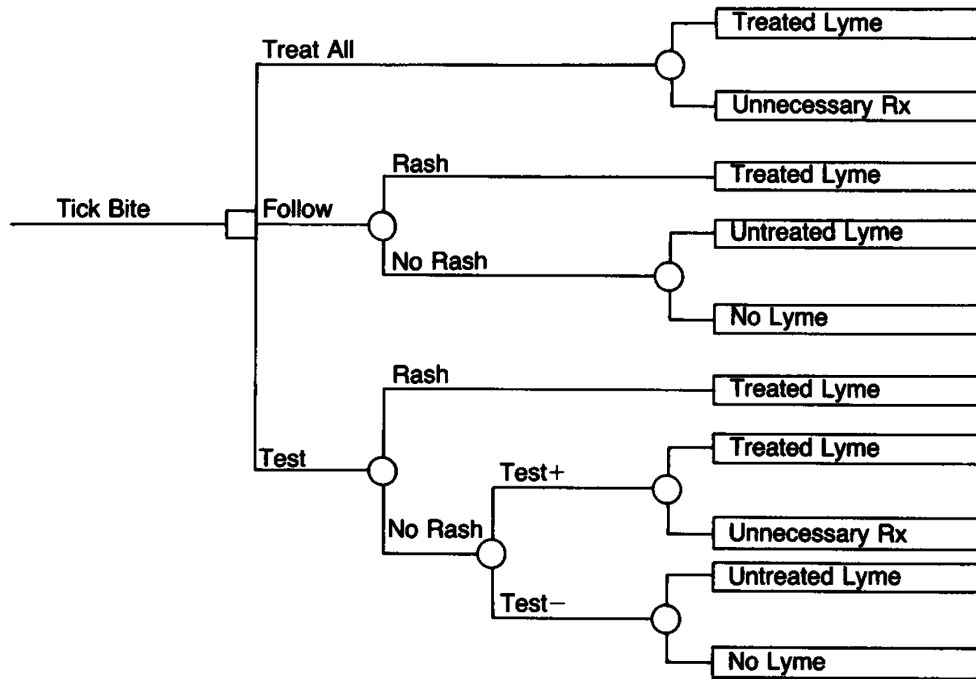
¹⁵See Section 4 for a more detailed discussion of methodologies. Many of these studies use Markov models, and the CEAR contains data on the welfare weights and discount rates they use to compute QALYs gained over a given period.

¹⁶Because the focus of our discussion is the application of these models in economics, we present a simplified taxonomy of these models loosely based on Brennan et al. (2006). Readers who wish to learn more about these models should consult that paper and Barton et al. (2004).

Table 1: Drug Quality Data Sources Summary

Data Source	Description	Coverage
ClinicalTrials.gov	<ul style="list-style-type: none"> Publicly available dataset on over 220,000 clinical trials. Contains trial characteristics and clinical trial outcomes (endpoints and side-effect rates) for trials with posted results. 	Established in 1997. More complete data after 2007.
FDA Drug Trial Snapshots	<ul style="list-style-type: none"> Publicly available dataset covering clinical trial results for FDA-approved drugs. Contains information on endpoint measurements and <i>side-effect</i> rates, and gives breakdowns by demographics whenever available. 	Drugs approved by FDA after May 2014.
Advera Health Evidex	<ul style="list-style-type: none"> Proprietary database that collects clinical trial results based on ClinicalTrials.gov, FDA Drug Trial Snapshots, and clinical trial results published in peer-reviewed journals. Also contains data on real-world side-effect rates based on FAERS data. 	Complete data for drugs in three indications: Hepatitis C, diabetes, and multiple sclerosis
DrugAbacus	<ul style="list-style-type: none"> Survival measures based on median survival and progression-free survival and tolerability in terms of side-effects, gathered from FDA label data. 	Oncology drugs approved between 2001 and 2015 by the FDA
Tufts CEAR	<ul style="list-style-type: none"> Contains data on QALYs gained from a variety of medical treatments, based on cost-effective analyses published in peer-reviewed journals. Partially available to the public with premium options offered through subscription. 	Established in 1976. Database contains more data on post-2000 studies.

Figure 1: Decision Tree Model for the Prevention of Lyme Disease (Magid et al., 1992)



treating everyone (Treat All), treating everyone who develops a rash (Follow), and treating only individuals who develop a rash *or* tested positive to a diagnostic test (Test). The model is evaluated by assigning probabilities and utility weights for all possible outcomes. The optimal strategy is then determined by maximizing expected utility.

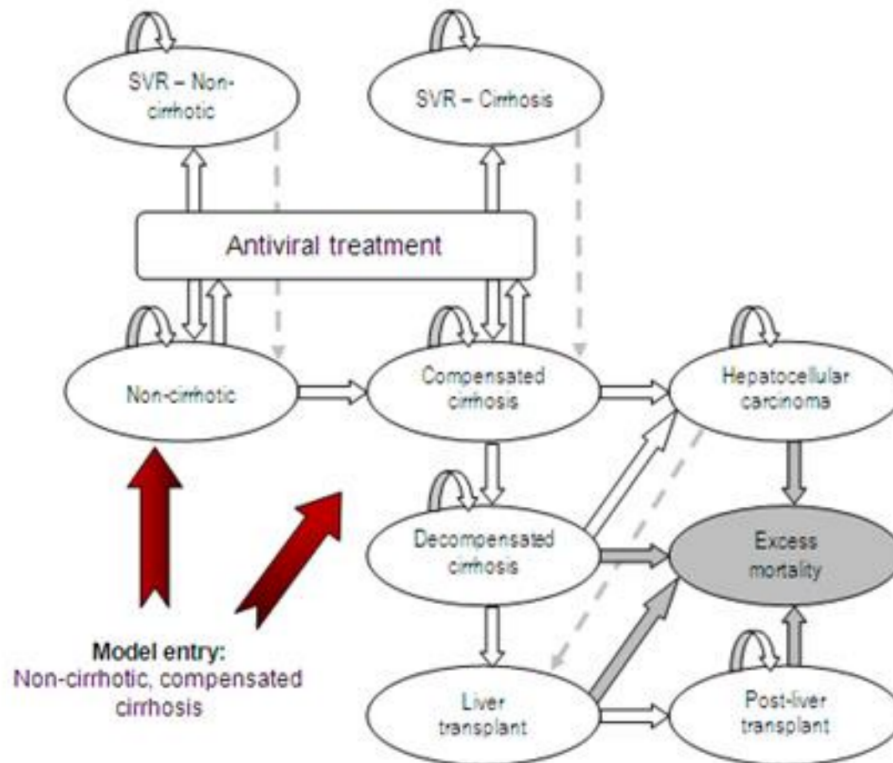
Decision tree models generally describe decisions that do not repeat over time, and resolve within a short time span (so that discounting usually is not an important factor). Even though they can be used to describe more complicated set of decisions, their complexity quickly increases with the time horizon encompassed.

4.1.2 Markov models

Markov models are similar to decision tree models, but have an infinite horizon. They consist of a series of states (with associated utility weights), and a matrix of transition probabilities regulating the likelihood of moving between these states. Their key assumption is the *memoryless property*, which requires that the probability of moving from one state to another can only depend on the current state, and not on past history. In general, this restriction can be circumvented by changing the definition of a state. For example, since the risk of having an acute episode increases with age for most chronic diseases, age could be included as a state variable. It is important to remember however, that the estimation of Markov models becomes increasingly complicated with the number of states.

Because Markov models have an infinite horizon and allow for the repetition of states, they

Figure 2: Markov Model Diagram for Hepatitis-C



are most appropriate for modeling chronic diseases. As an example, consider Figure 2, which shows the diagram for a Markov model of Chronic Hepatitis C, taken from a NICE assessment report (National Institute for Health and Clinical Excellence, 2015).¹⁷ We will use this model as an example in the next section and describe it in more detail.

Markov models can be estimated by simulating a cohort of patients and cycling it through the model for a certain number of periods. After each cycle, patients will be distributed across the various states, and will generate a certain amount of utility, which is then discounted and summed over time to obtain an overall number. The impact of a new drug can be estimated in this model by simulating how it would affect transition probabilities. Alternatively, the path of several individual patients can be simulated separately, allowing the researcher to construct a distribution of outcomes.

4.1.3 Individual level models

Decision tree models and Markov models are both examples of so-called aggregate models, which means that they can be simulated rather easily by letting a full cohort of individuals go through the model. Conversely, individual level models require the researcher to simulate the outcomes

¹⁷The figure is not part of the NICE report, but can be found on page 169 of the manufacturer submission, available for download at <https://www.nice.org.uk/guidance/ta330/evidence>

of individual patients and are therefore more computationally intensive. These models are more flexible, which is useful in situations where it is hard to write transition probabilities in a way that preserves the Markov property. For example, individual level models are often used to simulate the outbreaks of infectious diseases, where the transition probability is a function of the number of individuals in a given state (see for example [Kaplan et al., 2002](#)).

4.1.4 Why are these models useful?

Pharmacoeconomic models are best seen as complementary tools for studies that want to incorporate raw measures of drug quality from clinical trials. Usually, economic papers using this type of data will focus on a single dimension (generally efficacy, disregarding side effects), or include a few drug attributes in linear form in demand systems. By using established clinical models, economists will be able to incorporate a lot more information in their analysis, thus improving the precision of their estimates. For example, a paper analyzing the relationship between the quality of drugs and market outcomes within a disease area could use a quality measure based on simulating the total health gains generated by each drug in a cohort of individuals.

A second advantage is that these models provide a framework to understand the effects of drugs over the life-cycle, which might actually help researchers formulate economic questions, and create economic models. For example, the Markov model of Hepatitis-C might show that a lot of the cost savings generated by Sovaldi occurs later in life, in the form of reduced liver transplants, gains that might not be captured by other approaches.

4.2 A Markov Model of Hepatitis-C

To highlight the features of Markov models, we discuss the model of Hepatitis-C summarized in Figure 2. This model was created by Gilead Sciences and submitted for review to NICE as part of a pricing and reimbursement application for sofosbuvir (Sovaldi) in the UK.¹⁸ NICE is a UK government organization whose main role is to provide evidence-based guidelines and recommendations for public health practitioners and the National Health Service (NHS).¹⁹

The structure of this particular model includes nine states (represented by the nine ellipses), with patients entering at the non-cirrhotic or compensated cirrhosis stage, reflecting the population that is eligible for treatment. Patients then undergo antiviral treatment and either achieve a sustained virological response (SVR) and remain in that state forever, or they fail to achieve SVR and remain in their initial state, with a chance to progress towards more advanced stages of the disease: liver cancer (hepatocellular carcinoma), and decompensated cirrhosis. Note that the model makes a few simplifying assumptions about transition patterns: SVR is assumed to be an absorbing state (i.e. a state from which the patient cannot escape), and if a patient entered the

¹⁸The full report ([National Institute for Health and Clinical Excellence, 2015](#)) is available on NICE's website at <https://www.nice.org.uk/guidance/ta330>

¹⁹[Sorenson et al. \(2008\)](#) provide a good background on the role of NICE and its history.

Table 2: NICE approach to Cost-Effectiveness Analysis

Phase	Goal	Details
Scoping	Establishes parameters for the analysis	<ul style="list-style-type: none"> • the disease and population for which the technology is being appraised, • the relevant comparator technologies, • the health outcome relevant for the analysis, • the time horizon for the estimation of the health effects and costs.
Assessment	Generates evidence to inform the appraisal process	<ul style="list-style-type: none"> • Evidence is solicited from a variety of sources, including the manufacturer, patient groups, independent academic groups, and healthcare organizations • Guidelines for generation of evidence are pre-established, and presented in Table 3
Appraisal	Determines whether the drug is approved or not.	<ul style="list-style-type: none"> • conducted by an independent advisory body called the Appraisal Committee • decision can incorporate considerations outside cost-effectiveness, such as promoting innovation

decompensated cirrhosis state, the model assumes that she cannot recover.²⁰

The model was simulated using a series of competing antiviral treatments. The transition probabilities from the entry state to the SVR state for each possible antiviral were obtained from clinical trials that pitted sofosbuvir against competing treatments, while the probabilities of disease progression came from studies specific to the UK population. This last point is an important one: since the technology was being evaluated for reimbursement in the UK, NICE required probabilities specific to the population that was going to be affected.

Finally, utility weights associated to each disease state were obtained through a combination of patient surveys administered during the trial, and literature estimates. NICE requires utility weights to be expressed in QALYs (see Table 3 for more details). QALYs are utility weights

²⁰White arrows represent transitions to other states. Grey arrows represent *excess mortality* associated with having decompensated cirrhosis, liver transplant or hepatocellular carcinoma (patients can die in each health state). Dashed grey arrows represent health state transitions only investigated in sensitivity analysis.

Table 3: Summary of NICE Reference Case Requirements

Element of HTA	Guideline	Notes
Type of economic evaluation	Cost-utility analysis with full incremental evaluation	The output is supposed to be an Incremental Cost-Effectiveness Ratio (ICER), sometimes supplemented by an overall net monetary and health benefits, where 1 QALYs is valued at £20,000 or £30,000
Perspective	Patient's perspective for health gains, and National Health Service (NHS) perspective for costs	
Time horizon	Long enough to reflect important differences in cost or outcomes	The main decision is whether to follow patients over their lifetime or to use a shorter horizon (appropriate for treatment options that don't have differential mortality effects)
Evidence	RCTs comparing the technology against the relevant comparator technologies	
Measure of health effects	QALYs (EQ-5D are the preferred weights).	EQ-5D is a system of weights developed by EuroQol. For more information, see http://www.euroqol.org/about-eq-5d.html
Equity considerations	An additional QALY is given the same weight regardless of the characteristics of the individual receiving the benefit	
Discounting	Same for health benefits and costs (currently 3.5% yearly)	

scaled from 0 to 1, with 1 being perfect health, and 0 meaning death. For example, in the model considered, the two entry states carried a utility weight of 0.74 and 0.55 for non-cirrhotic and compensated cirrhosis respectively. Utility would then increase to 0.79 and 0.60 if SVR was achieved, or drop, if the disease progressed further.

To conclude, we encourage researchers interested in learning more about how these assessments are conducted to refer to NICE methodological guidelines, which are useful summaries of the considerations that go into a technology assessment.²¹ We also provide a summary of these considerations in the forms of two tables. Table 2 is a summary of the strategy NICE uses to conduct cost-effectiveness analysis. Table 3 lists some of the main requirements for “reference cases”, which is a set of rules that manufacturers should follow when they submit a cost-effectiveness analysis like the one we presented. NICE requires that any submission include an analysis of results generated using these reference case methods.

5 Usage of Drug Quality in Economics Research

In this section, we highlight a few main topics in economic research that require or could benefit from the use of the drug quality data and methodology discussed in the preceding sections.

As mentioned earlier, there are several types of agents in the market for prescription drugs. Universities and biopharmaceutical companies are involved in the discovery and development of new drugs, while further downstream, companies, payers, doctors, and patients interact to determine which drugs are prescribed and at what price. Several strands of the existing literature touch on the relationship between drug quality and the behavior of these agents, in order to construct welfare-enhancing policies. Our goal is to provide insights into the types of quality measures researchers might need to collect or construct in order to improve on our existing knowledge.

5.1 Value of Medical Research

One overarching question on the upstream part of the system is the value of medical research, or put another way, the productivity of research and development. Governments provide direct funding for basic and clinical research, and also provide a variety of subsidies for private R&D done by pharmaceutical companies. The main question here is how efficient public and private R&D spending is in terms of generating new drugs to improve health outcomes, which can have implications for biomedical R&D policy.

In a series of papers dating back to the early 1990s, DiMasi and coauthors (1991; 2003; 2016) have attempted to estimate the productivity of private R&D of biopharmaceutical companies, using internal data obtained from various companies. They show that the average development cost per NME approved by the FDA has risen dramatically over the past two decades, at an annual rate of 7-8% above inflation. The shortcoming of this set of studies is that NMEs are a crude

²¹NICE's *Guide to the methods of technology appraisal* is available at <http://publications.nice.org.uk/pmg9>.

measure of R&D output, as approved drugs have very heterogeneous attributes. It's possible that drug development has become less efficient, but one cannot rule out that newer drugs are equally effective per R&D dollar.

Using a different approach, [Hult \(2014\)](#) collects detailed outcomes data from the CEAR database, in order to estimate R&D productivity, with a focus on including gains from incremental drugs (i.e. new FDA-approved drugs that are reformulations of existing drugs).²² Each NME and incremental FDA approval is matched to CEAR data, in order to come up with a QALY value for each treatment.²³ The paper claims that R&D productivity remained steady in the 2000s relative to the 1990s, in contrast to the NME-based estimates from DiMasi, Hansen, and Grabowski.

The two studies highlighted here make useful contributions towards our understanding of R&D productivity trends in the pharmaceutical industry, but have some shortcomings in terms of measurement. Better drug quality data can allow for the estimation of accurate productivity measures at finer levels (e.g. by disease area), opening the door for micro-level studies on the determinants of R&D productivity and the value of different types of medical research.

5.2 Quality and Market Outcomes

Further downstream, drug quality is also relevant for thinking about prescription drugs market outcomes such as price, market shares, advertising, and formulary construction. In this section, we highlight a few papers that touch on these topics.

As mentioned above, there has been a growing debate over pharmaceutical drug pricing and cost-effectiveness. One simple way to isolate areas of inefficient spending is to compare drug quality measures to equilibrium price outcomes. [Howard et al. \(2015\)](#) do this for all oncology treatments currently on the market. They use median survival time as their measure of quality and compare it to launch price. They find that in the cross-section, higher launch prices reflect greater quality, but that over time the effectiveness of new drugs has not tracked the accompanying increase in launch prices, suggesting an overall decrease in cost-effectiveness.

In terms of drug choice and quality, several papers in the industrial organization literature have estimated models of demand for prescription drugs that incorporate quality measures. An example is [Lucarelli and Nicholson \(2009\)](#), which uses clinical trial measures of efficacy and side effects to estimate a demand system for colorectal cancer drugs. Significant coefficients on these quality variables would suggest that consumers and doctors do factor in quality into their drug choices. Other papers, such as [Dickstein \(2014\)](#), also use side-effects as a determinant of successful treatment or compliance in a learning model in the market for anti-depressants. The paper incorporates outside measures from the medical literature to evaluate the welfare generated by alternative policies such as value-based insurance design and informational interventions. This

²²We discuss the CEAR database in detail in Section 3.

²³The author has to make several assumptions here, because the CEAR data is somewhat sparse. Also, in contrast to the DiMasi approach, the cost-effectiveness of each drug is weighted by the total number of users, in order to calculate the aggregate welfare contribution of each drug.

represents an interesting approach in terms of integrating traditional demand models and measures of drug quality and welfare from the medical literature.

Drug quality also affects other firm behaviors such as advertising. [Azoulay \(2002\)](#) measures changes in perceived quality of drugs over time in the anti-ulcer market, based on post-approval clinical trials. He shows that demand responds directly to these quality changes, but firms also respond to this information by investing more in advertising, which creates a second channel for increasing demand for higher quality drugs.

Finally, there is a growing literature surrounding the effects of downstream market policies on the development of drugs. In particular, the Medicare Modernization Act of 2003, which added prescription drug coverage for seniors (Medicare Part D), has been the focus of recent research. [Blume-Kohout and Sood \(2013\)](#) show that Medicare Part D expansion has led to an increase in NMEs approved by the FDA. However, [Dranove et al. \(2014\)](#) build on their analysis and find that incremental innovation concentrated largely in areas where a large number of therapies was already available, suggesting a more moderate increase in welfare. Their results shows that more nuanced measures of quality seem to matter for evaluating policy, at least in certain contexts.

6 Future Work

In this paper, we have provided a sketch of the key topics, data sources, and methodologies surrounding the evaluation of prescription drug quality. We conclude with some thoughts on fruitful areas for future research, both on the measurement of quality discussed in Sections 3 and 4 and on the usage of quality measures discussed in Section 5.

The first issue in this literature, is the lack of a unifying framework for the evaluation of prescription drugs. Papers use a variety of different quality measures and approaches to measuring welfare, sometimes reaching different conclusions on the same research question. We believe that economists should create a consensus approach (or set of approaches) for evaluating drugs, which future papers can adopt. This will make it easier to compare findings across papers.

Even without a unifying framework, we think that a mix of the quality measures outlined in this paper can help improve our understanding of topics outlined in Section 5. As mentioned above, using quality measures in modeling demand can help researchers identify areas where prescription choices and drug quality are not closely linked and the possible policy remedies. Further upstream, we can use clinical trial outcomes on candidate drugs to project downstream demand, which can help us understand the link between the downstream market and firm development incentives and the usefulness of various R&D subsidies.

Finally, we think that the data and methodologies we presented will prove useful in assessing some of the big changes looming on the horizon for the prescription drugs market. As the public discourse focuses more and more on high spending and prices, payers will feel increased pressure to cut costs while maintaining efficiency. Several proposals have been advanced to this end, some of which have been in place for a few years now: formulary exclusions and tiers, Value-Based In-

insurance Design, pay-for-performance agreements with manufacturers, and focused interventions to foster innovation in specific disease areas.²⁴ Indicative of this trend, Medicare Part-B recently announced a pilot program to test new ways to reimburse prescription drugs to physicians and hospitals.²⁵ Having quality measures, particularly ones used by payers, can help us model and uncover the impacts of these activities on market outcomes.

²⁴The first example of a pay-for-performance contract was reached this year between CIGNA and the PCSK9 class of cholesterol drugs: <http://www.wsj.com/articles/health-insurers-push-to-tie-drug-prices-to-outcomes-1462939262>

²⁵Part B covers drugs administered in outpatient settings. See the announcement from the Centers for Medicare and Medicaid Services: <https://innovation.cms.gov/initiatives/part-b-drugs>

References

- Acemoglu, Daron and Joshua Linn**, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *The Quarterly Journal of Economics*, aug 2004, 119 (3), 1049–1090.
- Ahmad, Syed Rizwanuddin**, "Adverse drug event monitoring at the Food and Drug Administration.," *Journal of general internal medicine*, 2003, 18 (1), 57–60.
- Altarum Institute**, "The Prescription Drug Share of National Health Expenditures," Technical Report June 2014.
- Arcidiacono, Peter, Paul B. Ellickson, Peter Landry, and David B. Ridley**, "Pharmaceutical followers," *International Journal of Industrial Organization*, 2013, 31 (5), 538–553.
- Azoulay, Pierre**, "Do Pharmaceutical Sales Respond to Scientific Evidence?," *Journal of Economics & Management Strategy*, 2002, 11 (4), 551–594.
- Barton, Pelham, Stirling Bryan, and Suzanne Robinson**, "Modelling in the economic evaluation of health care: selecting the appropriate approach," *Journal of Health Services Research & Policy*, 2004, 9 (2), 110–118.
- Berndt, Ernst R., Iain M. Cockburn, Zvi Griliches, Theodore E. Keeler, and Martin Neil Baily**, "Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs," *Brookings papers on Economic Activity. Microeconomics*, 1996, pp. 133–199.
- Blume-Kohout, Margaret E. and Neeraj Sood**, "Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development," *Journal of Public Economics*, jan 2013, 97 (1), 327–336.
- Brennan, Alan, Stephen E. Chick, and Ruth Davies**, "A taxonomy of model structures for economic evaluation of health technologies," *Health Economics*, dec 2006, 15 (12), 1295–1310.
- Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro**, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," *The Quarterly Journal of Economics*, nov 2015, 130 (4), 1669–1726.
- Budish, Eric, Benjamin N. Roin, and Heidi Williams**, "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials," *American Economic Review*, jul 2015, 105 (7), 2044–2085.
- Cockburn, Iain M. and Aslam H. Anis**, "Hedonistic analysis of Arthritis Drugs," *Medical Care Output and Productivity*, 2001, 1 (January), 439–462.
- Crawford, Gregory S. and Matthew Shum**, "Uncertainty and learning in pharmaceutical demand," *Econometrica*, 2005, 73 (4), 1137–1173.

- Dickstein, Michael**, “Efficient provision of experience goods: Evidence from antidepressant choice,” *Harvard, unpublished*, 2014.
- DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen**, “Innovation in the pharmaceutical industry: New estimates of R&D costs,” *Journal of Health Economics*, may 2016, 47, 20–33.
- , **Ronald W. Hansen, and Henry G. Grabowski**, “The price of innovation: New estimates of drug development costs,” *Journal of Health Economics*, 2003, 22 (2), 151–185.
- , – , – , and **Louis Lasagna**, “Cost of innovation in the pharmaceutical industry,” *Journal of Health Economics*, 1991, 10 (2), 107–142.
- Dranove, David, Craig Garthwaite, and Manuel Hermosilla**, “Pharmaceutical Profits and the Social Value of Innovation,” 2014.
- Dubois, Pierre and Margaret Kyle**, “The Effects of Pharmaceutical Innovation on Cancer Mortality Rates,” 2016.
- Dunn, Abe**, “Drug innovations and welfare measures computed from market demand: The case of anti-cholesterol drugs,” *American Economic Journal: Applied Economics*, 2012, 4 (3), 167–189.
- Goldman, Dana P., Anupam B. Jena, Darius N. Lakdawalla, Jennifer L. Malin, Jesse D. Malkin, and Eric Sun**, “The value of specialty oncology drugs,” *Health Services Research*, 2010, 45 (1), 115–132.
- Howard, David H., Peter B. Bach, Ernst R. Berndt, and Rena M. Conti**, “Pricing in the Market for Anticancer Drugs,” *NBER Working Paper Series*, feb 2015, 53 (9), 1689–1699.
- Hult, Kristopher J.**, “Incremental Innovation and Pharmaceutical Productivity,” 2014.
- Kaplan, Edward H., David L. Craft, and Lawrence M. Wein**, “Emergency response to a smallpox attack: the case for mass vaccination.,” *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99 (16), 10935–10940.
- Lanthier, Michael, Kathleen L. Miller, Clark Nardinelli, and Janet Woodcock**, “An improved approach to measuring drug innovation finds steady rates of first-in-class pharmaceuticals, 1987-2011,” *Health Affairs*, 2013, 32 (8), 1433–1439.
- Lichtenberg, Frank R.**, “The impact of new drug launches on longevity: Evidence from longitudinal, disease-level data from 52 countries, 1982-2001,” *International Journal of Health Care Finance and Economics*, 2005, 5 (1), 47–73.
- Lucarelli, Claudio and Sean Nicholson**, “A quality-adjusted price index for colorectal cancer drugs,” 2009.
- Magid, David, Brian Schwartz, Joseph Craft, and J. Sanford Schwartz**, “Prevention of Lyme Disease after Tick Bites,” *New England Journal of Medicine*, aug 1992, 327 (8), 534–541.

National Institute for Health and Clinical Excellence, “Sofosbuvir for treating chronic hepatitis C,” Technical Report 2015.

Panayidou, Klea, Sandro Gsteiger, Matthias Egger, Gablu Kilcher, Máximo Carreras, Orestis Efthimiou, Thomas P. A. Debray, Sven Trelle, and Noemi Hummel, “GetReal in mathematical modelling: a review of studies predicting drug effectiveness in the real world,” *Research Synthesis Methods*, 2016, *in press* (November 2014).

Rothwell, Peter M., “External Validity of Randomised Controlled Trials: “To Whom Do the Results of This Trial Apply?”,” *Lancet*, 2005, 365 (9453), 82–93.

– , “Factors That Can Affect the External Validity of Randomised Controlled Trials,” *PLOS Hub for Clinical Trials*, 2006, 1 (1), e9.

Saito, Hiroki and Christopher J. Gill, “How frequently do the results from completed US clinical trials enter the public domain? - A statistical analysis of the ClinicalTrials.gov database,” *PLoS ONE*, 2014, 9 (7), 1–9.

Schnipper, Lowell E., Nancy E. Davidson, Dana S. Wollins, Courtney Tyne, Douglas W. Blayney, Diane Blum, Adam P. Dicker, Patricia A. Ganz, J. Russell Hoverman, Robert Langdon, Gary H. Lyman, Neal J. Meropol, Therese Mulvey, Lee Newcomer, Jeffrey Peppercorn, Blase Polite, Derek Raghavan, Gregory Rossi, Leonard Saltz, Deborah Schrag, Thomas J. Smith, Peter P. Yu, Clifford A. Hudis, and Richard L. Schilsky, “American Society of Clinical Oncology statement: A conceptual framework to assess the value of cancer treatment options,” *Journal of Clinical Oncology*, 2015, 33 (23), 2563–2577.

– , – , – , – , **Douglas W. Blayney, Adam P. Dicker, Patricia A. Ganz, J. Russel Hoverman, Robert Langdon, Gary H. Lyman, Neal J. Meropol, Therese Mulvey, Lee Newcomer, Jeffrey Peppercorn, Blase Polite, Derek Raghavan, Gregory Rossi, Leonard Saltz, Deborah Schrag, Thomas J. Smith, Peter P. Yu, Clifford A. Hudis, Julie M. Vose, and Richard L. Schilsky**, “Updating the American Society of Clinical Oncology Value Framework: Revisions and Reflections in Response to Comments Received,” *Journal of Clinical Oncology*, 2016.

Sorenson, Corinna, Michael Drummond, Panos Kanavos, and Alistair McGuire, *The National Institute for Health and Clinical Excellence (NICE): how does it work and what are the implications for the U.S.?* number April 2008.

Suslow, Valerie, “Are There Better Ways to Spell Relief? A Hedonic Pricing Analysis of Ulcer Drugs,” 1992.