I acknowledge funding through the National Institutes of Health (NIH) Agreements OT3HL152448. Data access and analyses were supported through the National Institute of Child Health and Human Development research infrastructure grant, P2C HD042828. I thank Chris Adams and numerous seminar participants for their helpful comments on the paper. All errors are mine. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Anirban Basu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
ABSTRACT

I demonstrate that to achieve dynamic efficiency, the optimal share of total surplus that a social payer should transfer to an innovating industry for a current asset depends on the marginal product of investment and the share of profits invested by the industry on the current asset and not on returns from future innovations. This insight arises from using a dynamic multi-period model of optimal transfers rather than a static two-period model with one optimal transfer, as used in the literature. I delve into the implications for alternative pricing of healthcare innovations - value-based prices using cost-effectiveness analysis, monopoly prices under the social demand curve, and monopoly profit preserving prices under insurance – for surplus appropriation by the innovating industry. I also explore how alternative financing mechanisms used by social payers and the demand uncertainty that innovators face impact this appropriation share. I illustrate these concepts with a substantive example of pricing gene therapy for sickle cell disease in the United States.

Anirban Basu
The CHOICE Institute
Departments of Pharmacy, Health Services, and Economics
University of Washington, Seattle
1959 NE Pacific St., Box - 357660
Seattle, WA 98195
and NBER
basua@uw.edu
I. INTRODUCTION

How do we determine prices for healthcare innovations without a market mechanism to resolve them? This continues to be a key question in healthcare economics. A fundamental challenge in biomedical innovation is the significant capital cost of entry into the market. Recent estimates show that it takes about a billion dollars on average to bring a drug to the market. Naturally, this creates a dilemma for investors who would be reluctant to take on such a significant risk without some guarantee of returns from the market. In the 1980s, the United States passed laws to offer pharmaceutical patent holders certain regulatory exclusivity. Exclusivity provides specific delays and prohibitions on the approval of competitor drugs to the patent holders whose products are approved for marketing by the Food and Drug Administration (FDA) so that the patent holder can market their drug exclusively and generate a stream of rents that suffices to generate enough profits so that future investments in research and development are encouraged. The laws, however, never specify how much profit is enough or how much of the welfare generated by the innovation should be appropriated by the manufacturer during this exclusivity period. In essence, the laws assign specific monopoly powers to the patent holder, otherwise illegal in the marketplace, for a fixed period.

That higher returns usually lead to more innovation is well established. For example, the literature establishes that larger market sizes, either through a larger indication size of current innovation (Acemoglu and Linn, 2004; Dubois et al., 2015), insurance expansion (Blume-Kohout and Sood, 2013; Bennette et al., 2019), or higher utilization (Finkelstein, 2004), lead to additional future innovations. Finkelstein further shows that not all induced innovations necessarily enhance welfare. However, this literature does not answer the question about the optimal private appropriation share of social surplus from current innovations.

Another large body of literature has used principles of economic evaluations to obtain estimates for what is known as "value-based price" (VBP) for healthcare innovations.
These works rely on an ever-evolving discussion on the definition of value and from whose perspective. The Second Panel on Cost-Effectiveness Analysis in Health and Medicine argued for using a complete societal perspective (Sanders et al., 2016; Meltzer et al., 2016), especially when decisions such as pricing are made. The ISPOR Taskforce on Value added additional items to this discussion (adding petals to the infamous "value flower"). These discussions are essential in the context of the optimal appropriation share of surplus because previous literature seems to assert that such evaluations do not fully represent the marginal benefits to consumers and, therefore, misrepresent the actual demand curve based on which social surplus should be calculated (Jena and Philipson, 2008). Indeed, in certain jurisdictions, economic evaluations are carried out from narrow institutional perspectives, where such misrepresentations are possible. However, in the US, there is a consistent movement toward taking a more comprehensive societal perspective reflecting innovation's full social marginal benefits.

None of this literature, however, establishes what optimal share of surplus generated by an innovation should be captured by the innovator. Economists have asserted the need to account for dynamic efficiencies in this market to fully maximize the long-term returns of innovation (Finkelstein, 2004; Jena and Philipson, 2008; Moreno and Ray, 2016). However, debates remain about the level of surplus (at the competitive equilibrium quantity level) that should be appropriated by the manufacturer so that the innovation remains affordable today without sacrificing incentives for future research by private parties and the generation of future surplus (Jena and Philipson, 2008; Woods et al., 2024). Such estimates are a necessary first step to calculating optimal prices and quantity to reach such an appropriation share. Jena and Philipson (2008) claim that this appropriation share should be 100%. Woods et al. (2024) estimate this share fraction based on empirical estimates of market productivity for pharmaceuticals on quantity and quality of innovation and find it to be about 22%. Both account for the current welfare and future benefits of R&D investments but use a static model, where this calculation is determined once and for all for a single decision. Neither uses a fully dynamic model, where the
optimal appropriation share decisions are made repeatedly over time between a social payer and the innovators, with dynamic links between current investments, current payments, profits, and future welfare, a structure necessary to establish conditions for dynamic efficiency (Abel et al., 1989). Adams (2021) is an exception that employed a fully dynamic model but only assesses firms' investment decisions.

Finally, irrespective of the appropriation share, there is genuine concern about the affordability and financing of these innovations by payers. Typically, analysts perform a budget impact analysis, assuming some reasonable estimate of the innovation's uptake to highlight the scale of the affordability issue. More recently, researchers have discussed alternative mechanisms of financing this budget impact, most of which are designed to postpone and delay payment to the innovator (Mattke and Hoch, 2015).

From the innovator's perspective, the appropriation share is only one piece of the puzzle. For innovators, total revenues (and profits) matter, determined by market size and the rate at which consumers adopt the innovations before their patent period runs out. For example, Goldman et al. (2011) argue that innovators of small-molecule innovations should get more extended exclusivity to achieve dynamic efficiency.

In this paper, I combine many of these debates using a dynamic model for a social payer (who subsidizes the innovation through cost-sharing so that private demand equals the social marginal benefit curve) to answer the question of optimal appropriation share from the social payer's perspective. I then offer standard economic principles for correctly interpreting alternate estimates of prices in the context of appropriation share, the role of competition on the demand and supply side in further shaping the VBP, and mitigating innovators' risk on adoption rates. Finally, I illustrate alternate pricing and financing of innovations incorporating these principles using a case study of gene therapy in sickle cell disease and implications for the Centers for Medicaid and Medicare Services (CMS).

The following section establishes the optimal appropriation share using a dynamic behavior model from the social payer. In Section III, I set up the theoretical pricing
arguments based on the market demand and supply curves for innovation. Section IV discusses alternative financing mechanisms for the social payer. Section V applies these principles to the case study. Section VI concludes with a discussion.

II. OPTIMAL APPROPRIATION SHARE FOR INNOVATORS

Let \( D(p) \) denote the demand for the innovation as a function of its price \( p \). Usually, the market demand curve can only be determined in a competitive market with random price fluctuations. Even estimating the price elasticity of demand requires strict natural experiments, which are challenging to come by after the product is marketed and are, by design, non-existent before marketing. We rely on the alternate structural approach where the marginal social benefits curve is estimated based on data from clinical trials and other epidemiological and economic evidence. This principle is followed in cost-effectiveness studies from a societal perspective that aims to establish the marginal social benefits of an intervention and compare it to its price. Throughout this paper, I will keep discounting implicit for the ease of exposition. The marginal health benefits are valued at some social marginal willingness to pay for a unit of health. Let the marginal social benefit function \( B \) be given as:

\[
B = H \cdot \theta - M - \tilde{M},
\]

where \( H = \) Health Benefit, \( \theta = \) the marginal willingness to pay for a unit of health, \( M = \) healthcare costs offsets other than the price of innovation, \( \tilde{M} = \) non-healthcare costs offsets in the society, and the distribution of \( B \), \( F_B(.) \), is absolutely continuous with respect to a Lebesgue measure. The derived social demand function is given as follows:

\[
\tilde{D}(p) = N \cdot \Pr(B - p > 0) = N \cdot (1 - F_B(p)),
\]
where \( N \) = the target population size. The cost of paying for the innovation is not included in (1) and is considered below.

The following holds:

(1) \( E(B) \) is finite

\[
E(B) = E_P(\bar{D}(p)/N) = \int_{p \in P} \left(1 - F_B(P = p)\right) \cdot f(p) dp = \int_0^1 \left(1 - F_B(P = p)\right) dF_B. \tag{3}
\]

Thus, the expected value of \( B \) is derived by integrating the area under the entire social demand function. The social surplus (\( S \)) produced by the innovation is therefore represented as

\[
S = \int_{y_C}^{\infty} B(y) - MC(y) dy \tag{4}
\]

where \( y_C \) is the competitive quantity demanded, and \( MC(y) \) is the marginal cost curve.

**A Social Payer’s Optimization Problem**

Let a social payer wish to maximize an infinite stream of consumer surplus, \( h() \), sequentially over assets \( (j)^\dagger \) such that:

\[
U = \text{Max} \sum_{j=0}^{\infty} h(j, S^j, Y^j) \quad \text{where} \quad S^{j+1} = k(j; S^j, Y^j), \tag{5}
\]

\( S^j = S^j(I^j) \): The net present value\(^\ddagger\) of the total surplus generated in the society from the \( j^{th} \) asset, which is a function of the total investment, \( I_j \), made by the industry. Assume, \( S^j_i \geq 0 \ \forall j \), where partial derivatives are denoted with subscripts.\(^\ddagger\)

---

\(^\dagger\) Or groups of assets over time \((j)\).

\(^\ddagger\) Discounting is kept implicit for clarity.

\(^\ddagger\) Note this assumption is only made on average, at the industry-level.
\( Y^j = \beta^j \cdot S^j(i^j) \): Total payment made by the social payer on the \( j \)th asset to the industry that is also the revenue for the industry for that asset, and

\( \beta^j \): Proportion of the total surplus society pays the industry for the \( j \)th asset. \( \beta^j \geq 0 \).

Here, \( k() \) represents the social payer's surplus-generating function for the next asset as a function of surplus and medical expenditures for the current asset. Putting some structure around \( h() \) and \( k() \), let

\[
\begin{align*}
    h(j, S^j, Y^j) &= S^j - Y^j = (1 - \beta^j) \cdot S^j \\
    S^{j+1} &= k(j, S^j, Y^j) = S^{j+1}(i^{j+1}) = S^{j+1}(Y^{j+1} \cdot (Y^j - I^j)),
\end{align*}
\]

\( \gamma^{j+1} \): Proportion of industry profits (= \((Y^j - I^j)\)) from the \( j \)th asset that the industry invests in R&D for the \((j+1)\)th asset. Assume, without loss of generality, \( \gamma^{j+1} > 0 \).

Equation (7) shows that current payment by the social payer does impact the industry's bottom line, which in turn influences the total amount reinvested by the industry in R&D. However, the portion of profits reinvested in R&D is assumed to be independent of the profit levels and is determined by other factors like technological progress and capital constraints.†† The dynamic budget constraints for the social payer and the industry are implicit in the social surplus produced and the rents paid to the industry.

Setting the social payer's objective function as a recursive Bellman equation (Bellman, 1971),

\[
U(j, S^j) = \max_{Y^j} \{ h(j, S^j, Y^j) + U(j + 1, S^{j+1}) \}
\]

The first-order condition of (8) with respect to \( Y^j \),

†† It is easy to relax this assumption and make \( \gamma^{j+1} \) a function of \( Y^j \), i.e., \( \gamma^{j+1}(Y^j) \). In this case, the term \( \gamma^{j+1} \) in (9) and (12) would be replace by \( \gamma^{j+1} + \gamma^{j+1}(Y^j - I^j) \). The qualitative implications for optimal appropriation remains the same.
\[ h_Y(j, S^j, Y^j) + U_S(j + 1, S^{j+1}) \cdot (S^{j+1})_Y = 0, \text{ or} \]

\[ -1 + U_S(j + 1, S^{j+1}) \cdot (S^{j+1})_Y = 0 \quad (9) \]

The left-hand side of (9) says that the marginal dollar of medical expenditures (-1) the \( j \)th asset will fetch \((S^{j+1})_Y\) dollars of surplus from the next innovation through the investment of additional \( Y \) dollars by the industry. This will translate to the long-term consumer surplus through the marginal \( U_S(j + 1, S^{j+1}) \) function.

Next, I differentiate both sides of the maximum operator in (8) with respect to \( S \) and use the star superscript for optimal quantities:

\[ U_S^*(j, S^j) = h_S(j, S^j, Y^j) + U_S^*(j + 1, S^{j+1}), \text{ or} \]

\[ U_S^*(j, S^j) = (1 - \beta^*) + \frac{\partial U^*(j + 1, S^{j+1})}{\partial Y^{j+1}} \cdot \frac{\partial Y^{j+1}}{\partial S^j} \quad (10) \]

Following the envelope principles, \( \frac{\partial U^*(j + 1, S^{j+1})}{\partial Y^{j+1}} = 0 \), as decision makers will reoptimize payment decisions every period. (10) indicates that at the optimal level of medical expenditures, a marginal dollar of surplus from the current asset will change the long-term total consumer surplus through the marginal \((1 - \beta^*)\) dollar of consumer surplus gained on the current asset. Therefore, focusing only on the direct effects, we have

\[ U_S^*(j, S^j) = (1 - \beta^*), \text{ or} \]

\[ U_S^*(j + 1, S^{j+1}) = (1 - \beta^{j+1*}) \quad (11) \]

Substituting (11) in (9), we have our main result:

\[ \beta^{j+1*} = 1 - \frac{1}{S_Y^{j+1}}, \text{ or} \quad \beta^* = 1 - \frac{1}{S_Y} \quad (12) \]

The result in (12) identifies the relationship between the optimal appropriation of surplus by the industry and the marginal product of investment in the current asset. The appropriation share for the current asset increases when the marginal product of
investment from the current asset is high. It also increases when the industry has invested more of its previous profits in the asset. Note that the appropriation share asymptotes to 100% as these quantities increase. However, in realistic scenarios, it should be less than 100%.

Perhaps the most surprising implication of these results is that the optimal appropriation share for the current asset does not depend on expected dynamic returns, which show up in two-period models (Jena and Philipson, 2008; Woods et al., 2024). For example, Woods et al. use a two-period model and find that the optimal share of surplus depends on the ratio of today's benefits (static) versus tomorrow's benefits (dynamic) from innovations. This is not surprising when one is trying to solve for the optimal appropriation share that the social payer must pay private innovator only once. In a dynamic model, when this optimization is happening repeatedly, current results show that the optimal share has no direct dependence on tomorrow's benefits (or the notion of elasticity of innovation due to revenue). This is because maximizing locally for each asset will lead to the global maximum for the social payer. These local maxima imply an appropriation share that does not depend on future returns, as the social payer will also optimize future appropriation shares and adjust to those future returns.

These results also contrast with the usual discourse in the literature about how a reduction in industry revenue leads to a reduction in investments in R&D (Philipson And Durie, 2021). While this may very well be true, our results show that the reverse effect is also true: optimal appropriation share depends on the share of profits invested by the industry.

The Pharmaceutical Research and Manufacturers of America (PhRMA), an industry trade association, reports that the pharmaceutical companies that PhRMA represents,‡‡ invested about 25% of their revenues in R&D efforts in 2017-2018 (PhRMA, 2019). This would imply that the share of profits invested is higher. Figure 1 illustrates the relationship

‡‡ Some smaller biotech companies are not represented by PhRMA.
between the optimal appropriation share of the surplus, the marginal product of investment, and the share of profits reinvested by industry. Since we don't expect the industry to invest the marginal dollar if they don't get back an expected return of at least a dollar, and since the appropriation share of the total surplus is less than 100%, we only consider marginal products greater than 1. At a 50% profit reinvestment rate by the industry, even a smaller marginal product of 6 to 9 would indicate an appropriation share of 70 to 80% of the total surplus. Although more research is needed to obtain a better estimate of such marginal product of investment for the industry, a very rough calculation is presented here. US life expectancy increased (GBD US Health Disparities Collaborators, 2022), on average, by 2.3 years from 2010 to 2019, which amounts to a total surplus of about $69 trillion.*** During this time, pharmaceutical companies spent an average of $80 billion annually, totaling $800 billion (PhRMA, 2019). Assuming conservatively that only 10% of the total surplus was due to the pharma R&D spending (Lichtenberg, 2014), the marginal product could be estimated to be about 8.6.

Gene therapies are likely to represent the scenario where the marginal product of investment is quite high due to the curative nature of therapies. Novel gene editing techniques are only beginning to emerge and are likely to be more efficient in the future and applied to a much broader set of disease areas (Li et al., 2023). Moreover, the share of profits invested in R&D for gene therapy, especially by small biotech firms, is also likely to be high. For example, the cost of producing gene-related products is exorbitantly high today, and there are expectations for a reduction in the marginal costs of producing these therapies by an order of magnitude in the next decades (Harris, 2019). This would indicate bigger investments by the industry. Consequently, expecting a social payer offering to pay closer to the full social surplus for current gene therapies is not unreasonable.

\[ \frac{dS}{dj} > \frac{dP}{dj} > 1, \text{ at optimal investments by the industry.} \]

*** Calculated for a population size of 300 million at $100,000/life year.
III. ALTERNATE PRICING PRINCIPLES FOR INNOVATIONS AND IMPLICATIONS FOR SURPLUS APPROPRIATION

Total surplus appropriation depends on the price set by the innovator or negotiated between the innovator and payer, as well as the actual quantity demanded in the market (which can deviate from the competitive equilibrium quantity even with full insurance). Discussions around appropriate surplus appropriation assume that the competitive equilibrium quantity will be sold during the patent period. However, this is often not the case for various reasons, exposing innovators to revenue risks. Innovators may be willing to sacrifice a fraction of the total surplus appropriation to mitigate quantity risks. This section discusses alternate principles for price setting, assuming competitive equilibrium quantity is demanded under full insurance. In the next section, we discuss alternative financing mechanisms, some of which can mitigate the risk of under-demand for innovators.

Value-based Pricing with Cost-effectiveness Analysis

Following (3), cost-effectiveness analysis conducted from a societal perspective can be used to derive a "value-based price", $p_V$, where,

$$p_V = E(B).$$

This approach is illustrated in Figure 2(a). Suppose the value-based price derived using this cost-effectiveness approach is used to pay for the innovation. In that case, it implies that society is transferring the total surplus from the innovation to the manufacturer. Essentially, this is equivalent to allowing the manufacturer to perfectly (first-degree) price-discriminate for each consumer in society, even without identifying who may have a higher $B$ versus a lower $B$ at the individual level. The monopolist manufacturer will produce an output level for anyone with $B > 0$. Therefore, from a complete societal

---

††† Ideally, this should be $p_V = E(B | B > MC)$. However, since physicians and patients cannot often target those who get less versus more benefits from an intervention the average is used.
perspective, a CEA-based value-based price can be considered the upper bound for the price of an innovation as long as $\beta^{l_1} \leq 1$.

**Direct Monopoly Pricing without Insurance**

In contrast, if the manufacturer was allowed to take away only monopoly profits instead of the entire surplus, one can calculate it using the marginal costs curve, $MC(y)$, of manufacturing the innovation and the marginal revenue ($MR(y)$) function at any level of output $y$. The marginal revenue for a monopolist, based on the social benefits curve, is given as the rate of change of total revenue with respect to the quantity supplied:

$$MR(y) = \frac{d(p(y)\cdot y)}{dy} = p(y) \left[1 - \frac{1}{|\epsilon(y)|}\right]$$

where $\epsilon(y)$ is the price elasticity of the demand curve, and $p(y)$ is the inverse demand function. The optimality condition for a monopolist suggests that:

$$MR(y) = MC(y) \left[1 - \frac{1}{|\epsilon(y)|}\right] = MC(y)$$

This optimality condition is only met over the elastic part ($|\epsilon(y)| > 1$) of the demand curve. If the demand curve is inelastic ($|\epsilon(y)| < 1$), a monopoly will always have the incentive to reduce output to maximize profit, since otherwise, the marginal revenue is negative and cannot be equal to any positive marginal cost. As such, the optimal output from a monopolist ($= y_M$) is always smaller than a fully competitive equilibrium. Therefore, the monopolistic price estimate would be given as:

$$p(y_M) = MC(y_M) \cdot \left[1 - \frac{1}{|\epsilon(y_M)|}\right]^{-1}$$

And the maximum monopoly profit that the manufacturer can expect is:

$$Profit_M = p(y_M) \cdot y_M - \int_{y_M}^{\infty} MC(y)dy$$

This is illustrated in Figure 2(b).
**Volume-based Contracts for the Monopolist with an Insurance Plan**

Consequently, if an insurance plan wishes the monopoly to increase its output to a competitive equilibrium quantity, $y_c$, where $p(y_c) = MC(y_c)$, it must guarantee at least the monopoly profits to the manufacturer. Therefore,

$$p_{eff}(y_c) \cdot y_c - \int_{y_m}^{y_c} MC(y)dy = p(y_M) \cdot y_M - \int_{y_m}^{y_M} MC(y)dy \quad (18)$$

In such a case, the effective monopoly price that the insurer must offer the manufacturer is given as:

$$p_{eff}(y_c) = [(p(y_M) \cdot y_M - \int_{y_m}^{y_M} MC(y)dy) + \int_{y_m}^{y_c} MC(y)dy] \cdot (y_c)^{-1}$$

$$= [p(y_M) \cdot y_M + \int_{y_m+1}^{y_c} MC(y)dy] \cdot (y_c)^{-1} \quad (19)$$

The effective monopoly price offered would be the total monopoly revenue plus the total costs of producing the extra unit of the product beyond the monopoly quantity, all together divided by the total quantity output requested by the insurer (Figure 2c). This price expression also captures the basis for any volume-based price contract that an insurer may try executing with a monopolist in the market.

**Volume-based Contracts for the Monopolist with Multiple Insurance Plans**

A case may arise when multiple insurance plans intend to execute volume-based contracts applicable to only their beneficiary population. If the distribution of B is the same across each beneficiary pool (even when the sizes of the pools are different), the main results from the last two sections do not change. The monopolist would arrive at the same
monopoly price as in (16), and the effective monopoly price offered by plans to preserve monopoly profits would be the same as in (19).

However, if the distribution of $B$ differs in each beneficiary pool, the manufacturer may be able to carry out third-degree price differentiation. To illustrate this, let's say there are two plans with the same target population size in each plan (i.e., $N_1 + N_2 = N$). However, the cumulative distribution of $B$ in plan 1, $F_B^1(b)$, first-order stochastically dominates the cumulative distribution of $B$ in plan 2, $F_B^2(b)$. i.e. $(1 - F_B^1(b)) \geq (1 - F_B^2(b)), \forall b$ (Figure 3 (a)). Figure 3 (b) and (c) show the corresponding demand functions for each plan. Plan 1 has a more inelastic demand function overall compared to Plan 2. Hence, barring any arbitrage between plans, the monopolist's optimal output for Plan 1 will be curtailed even further, and the equilibrium price will be higher than that with a single plan. Volume-based contract negotiations with the monopolist would lead to Plan 1 transferring a larger proportion of the total surplus to the manufacturer than Plan 2. This represents third-degree price discrimination, implying that the manufacturer's total profit across both plans will be higher than when pricing is based on pooled demand. Hence, any volume-based contracts done separately by plans will fetch a higher surplus for the manufacturer than when such contracts are negotiated centrally. Consequently, such third-degree price discrimination would also occur when state Medicaid programs engage in negotiations in silo instead of a federal body negotiating for all states, especially when the distribution of benefits varies substantially across states.

**Competition in Oligopolistic Market**

Sometimes, more than one innovation for the same indication comes to the market simultaneously, creating an oligopolistic market. In most cases, there is only a duopoly. If the products offered by these manufacturers are similar enough so that the distribution of $B$ for each product is very similar to each other, then following a simple Bertrand competition model, one would expect the prices will reduce to the point where the
marginal costs equal the marginal benefits of the product. Hence, insurance plans should be able to negotiate an equilibrium price close to the marginal costs of supplying to the entire target population. This will be true whether one or multiple insurance plans are in the market.

Consequently, it is often of interest for competing oligopolistic manufacturers to differentiate their products even when they appear similar. With such horizontally differentiated products that are imperfect substitutes, typically, we would expect these manufacturers to compete on quantity following a Cournot-type competition, as that strategy will maximize their profits (Singh, 1984). However, insurance changes the dynamics of the typical oligopolistic market equilibrium. The manufacturers cannot dictate the quantity they will produce. They will instead compete on prices following a Bertrand-type competition. The manufacturers' equilibrium prices will differ depending on the extent of the product differentiation.

Once the manufacturers set the Bertrand equilibrium prices, plan(s) can follow the volume-based contracts to negotiate the prices down, following the same principles as dealing with a monopolist. Specifically, suppose plans can identify differentiated distributions of B for each manufacturer. In that case, they can calculate the manufacturer's expected profit at the Bertrand prices and negotiate to supply a larger quantity if the plans can guarantee expected profits under those Bertrand prices.

It is interesting to note that sometimes plans may perceive products as perfect substitutes, even when the manufacturers believe they have differentiation. If manufacturers cannot convince the plans regarding such differentiation, the manufacturer with the lower Bertrand price will be more likely to have a bigger volume contract. This situation will more likely arise when plan(s) engage in subscription-based pricing, where closed bid tenders on prices determine which manufacturer wins the full market share. However, while such a strategy can attain static efficiency, it will likely sacrifice dynamic
efficiency by removing adequate appropriation of rents by innovators that would impact future R&D decisions.

IV. ALTERNATE FINANCING STRATEGIES FOR INNOVATIONS AND IMPLICATIONS FOR SURPLUS APPROPRIATION AND RISK

Given the typically significant budget impacts of paying for curative therapies, a range of alternative financing options has been proposed in the literature that can help mitigate the upfront cost impact of paying for these cures by the insurer and mitigate risk due to the current uncertainty about the effectiveness of therapies. However, these alternative models have not been evaluated in the context of surplus appropriation. These financing options include:

1. **Traditional Upfront**: This is the standard mechanism where insurers pay the product's price when consumed. Because of liquidity constraints, insurers may borrow from a bank to pay these upfront for a one-time gene therapy. The insurer waits for competition to enter the market for the prices to decrease.

2. **Annuity**: This is a variation on the traditional upfront mechanism, where the insurer is still liable to pay the price of the gene therapy as they are consumed, but instead of paying the full price upfront, they negotiate with the manufacturer to pay them on an annuity basis with some interest rate so that the net present value (NPV) of the annuity stream equals the upfront price. Such a mechanism may allow insurers to make payments with their liquidity constraints and prevent them from having to borrow money from the bank to pay for the upfront costs.

3. **Outcomes-based payment or risk-sharing agreements**: Here, the insurer waits to ascertain a specific positive outcome for the patients, ensuring that the new therapy works for them, and then issues a payment to the manufacturer. For gene therapy, such
ascertainment will likely be a few years after the patient receives the therapy. Therefore, the patient must survive until then and meet the ascertainment criterion.

None of the above mechanisms guarantees that competitive equilibrium quantity will be demanded over the innovation's patent period. Such a situation presents a risk to the innovator regarding the total surplus appropriated and also leads to static and dynamic inefficiency for the social payer. These inefficiencies can be overcome using:

(4) Patent-buyout: This is a kind of subscription model where the insurers anticipate the potential competitive equilibrium demand in the market over a specific time during which the manufacturer holds the intellectual property rights and estimates the costs of buying out the patent so that the manufacturer will continue to supply the product either at marginal costs or freely. Negotiations between the payer and the innovator will determine the price at which the buyout happens. The buyout price will likely lie between the benchmark price for optimal total surplus appropriation and a reduced price for risk mitigation for the innovator. The payer would typically pay the buyout amount in annuity over this period.

V. PRICING AND FINANCING FOR GENE THERAPY FOR SICKLE CELL DISEASE

Sickle cell disease (SCD) is characterized by recurrent vaso-occlusion and hemolysis that contribute to acute episodes of pain, tissue ischemia, inflammation, and progressive organ damage (Ware et al. 2017). Approximately 100,000 people in the United States are living with SCD. More than 80% of those affected are of African heritage. In December of 2023, two gene therapies were approved, promising potential cures for this disease.

Value-based Price

A recent cost-effectiveness study comprehensively compared gene therapy to the common care these patients receive (Basu et al., 2024). Common Care included
hydroxyurea and transfusions but excluded use of other disease-modifying therapies or hematopoietic transplants. Assuming a $2 million price for gene therapy, the incremental cost-effectiveness ratio from a societal perspective was estimated to be $126,000/QALY\footnote{The corresponding ICER in terms of health years in total (HYT) was $76,000/HYT.}. At an equity-informed threshold of $150K/QALY, the value-based price would be estimated at $2.29 million. For our purposes, the components of E(B) are estimated as: E(H) = 11.9 quality-adjusted life years (QALYs). \( \theta = $150K/QALY; E(M) = $298,780, E(\bar{M}) = - $799,492 \). This leads to a value-based price \((p_v)\) of $2.29 Million. If society were to pay this price for every eligible patient, it would be equivalent to transferring the full surplus from gene therapy to the manufacturer.

**Monopoly and Effective Monopoly Pricing**

We could also derive the individual-level distribution of B in the eligible population using the microsimulation model that generated these estimates. Based on that distribution, Figure 4 shows the downward-sloping estimate of \( \bar{D}(p) \) (assuming the total eligible population is 1). Figure 4 also shows the marginal revenue curve.

Next, we focus on the marginal cost curve. There remains tremendous uncertainty around the marginal costs of manufacturing gene therapy. There are no empirical estimates available in the peer-reviewed literature. Expert opinion suggests that currently, it can be as high as $500,000 to $1 million (Harris, 2019). We assume a current estimate of marginal costs of $800,000 with a standard deviation of $100,000. There is also substantial evidence and discussion showing that manufacturing costs will decline by orders of magnitude over the next decade.

Given that the supply to the whole target population of gene therapy-eligible sickle cell disease would take ten years or more, we allow the marginal costs of gene therapy production to decline with the amount supplied. It is estimated that in 10 years, manufacturing costs for the marginal dose of gene therapy could go down to as low as...
$35,000 (Harris, 2019). We assume that marginal costs would decline by 20% per year. We allow for uncertainty in this estimate with a standard deviation of 5% points. This ensures that we are conservative about the decline in marginal costs, so the $35,000 estimate would only be achieved if we used the 95% higher bound of the decline rate. The marginal cost curve and its uncertainty are shown in Figure 4.

Table 1 shows the resulting market-clearing monopoly price and conditions and an effective monopoly price by a social payer that could guarantee monopoly profits for supplying the competitive equilibrium quantity.

**Volume-based Contracts with Multiple State Medicaid Programs**

Figure 5 shows the distribution of eligible patient population by state and the distribution of eligible patients enrolled in Medicaid only, Medicare only, and dual eligible within a state. Corresponding state-specific distribution of benefits is also shown. While these distributions of benefits are quite similar across states, states with a higher proportion of dual-eligible patients have lower benefits, which indicates that the demand curve in these states may be shifted inward and flatter than in other states (Figure A1). Consequently, based on our discussions above, allowing individual states to carry out negotiations on their own may provide an opportunity for the innovator to carry out third-degree price discrimination. These features of the potential demand function indicate that it will be optimal for CMS to conduct these negotiations centrally.

**Alternate Financing Mechanisms**

Table 2 reports the NPV of the 10-year budget impact for CMS under alternate pricing and financing mechanisms, assuming that the expected eligible population of 5000 patients will be treated with gene therapy over 10 years with an uptake by 500 patients per year. This assumes that the competitive equilibrium quantity of 98% of the eligible 5000 patients will be demanded in 10 years. This time frame also aligns with CMMI's scope of evaluating alternate financing of gene therapy in SCD. The net present value of the total
costs of manufacturing the gene therapy for 5000 patients over 10 years amounts to about $2.28 Billion. Table 2 also reports the NPV of the 10-year revenue stream and profits for the manufacturer under these alternative scenarios. Annuities (in patent buyout and the Annuity model) were evaluated at a 3% interest rate.

As discussed in the theory sections, the traditional upfront payment of the gene therapy at the value-based price, as consumed over the year, represents the transfer of the total surplus for this gene therapy to the manufacturer. This approach amounts to an NPV of $7.82 Billion, which the manufacturer appropriates as profits.

The patent buyout is the only option presented under the effective monopoly price where CMS guarantees a volume contract to preserve monopoly profits. However, instead of paying $1.3 million times 5000, CMS will use annuities to spread this payment over the 10-year period, whose NPV to the manufacturer would still be $6.5 Billion (Revenue). The overall NPV of the impact on the CMS budget would be $7.76 Billion, evenly distributed over the years. The NPV of the manufacturer's 10-year profits is $4.22 Billion.

Not surprisingly, if CMS attempts to do a patent buyout at the value-based or monopoly price, the budget impact will be much higher, leading to higher profits for the manufacturer. Under the value-based price, the 10-year profit of $9.22 Billion for the manufacturer represents the upfront transfer of the total expected surplus for this gene therapy to the manufacturer, which amounts to 118% of the total surplus in NPV. The fact that manufacturer profits are higher ($14 Billion) on a patent buyout at the monopoly price suggests that this mechanism could be a welfare-decreasing proposition unless there is absolute reason to believe that the innovator should appropriate more than the total surplus of the current asset.

Since patent buyout mitigates the risk of quantity demanded for the manufacturer, an effective negotiation could bring payout prices closer to the effective monopoly prices rather than the other ones. Among the other financing mechanisms, three themes stand out (Table 2):
1. Annuity and outcomes-based payments imply a similar 10-year budget impact for CMS and revenues and profit for the manufacturer. They are all lower than the traditional upfront payment of the gene therapy at any given price.

2. Patent buyout indicates that CMS won't owe any money to the manufacturer for treating these 500 patients beyond the 10-year mark. For other mechanisms, CMS will continue to owe money to the manufacturers.

3. All the other financial mechanisms have a higher 10-year budget impact, manufacturer revenue, and profits compared to patent buyout at the effective monopoly price.

For SCD gene therapy, a patent buyout at an effective monopoly price would still transfer 54% of the surplus to the manufacturer. Such a mechanism also shifts the risk due to the uncertainty in the future uptake of gene therapy complete to CMS. Given the risk-adjusted estimates for the investment costs of bringing new gene therapy to the market of $2 billion (Wouters et al. 2022; Sabatini and Chalmers, 2023), the patent buyout at the effective monopoly price still projects to deliver over a 200% net return to the manufacturer.

One issue that will come up with a patent buyout option is who bears the dynamic risk of quantity. For example, what happens if the buyout is based on an expected total demand of 5000 cases, but ultimately, only 2500 or 10000 adopt the technology? In the first case, CMS loses out as it has paid way more than it should have paid for the buyout. In the second case, the manufacturer loses out as it did not receive enough in the payout. This dynamic issue can be solved with annuity payment with an upfront rider of changing the annuity scheme, say in 5 years, if better estimates of total demand become available. It is important to point out here that since the payout option is a mechanism to mitigate the risk of under-demand for the manufacturer, the initial contract should be based on an estimated total demand based on the social benefits curve at the outset but would only be updated if private demand, under insurance, outpaces such estimates. The social payer assumes that risk if private demand falls below the contracted amount.
For example, Figure A5 shows the upfront annuity scheme for payout with different estimates of total demands for SCD gene therapy. Suppose in year 5 it is revealed that the current annuity scheme is incorrect. In that case, there will be scope for corrections to switch to a new annuity scheme over the remaining years so that effective net present value represents the overall annuity stream for the correct total demand.

**Duopolistic Competition**

In December of 2023, two separate gene therapies were approved by US FDA for sickle cell disease in the United States. These approvals set up a stage for potential duopolistic competition. The two products are technically different, one using gene editing (innovator#1) and the other gene addition (innovator#2). Both showed curative-level outcomes in the short run (up to 3-8 years of trial evidence). US FDA gave Innovator#2 product a black box warning for hematologic malignancy (blood cancer) because two patients developed such malignancy in the trial, triggering life-long monitoring. However, there is a growing clinical opinion that these malignancies cannot be attributed to gene therapy. Evidence on Innovator#2 product is more mature than innovator#1 product. Overall, the evidence base for the benefits of gene therapy has not differentiated between the two products.

Interestingly, innovator#1 set a launch price of $2.2 million for their product, close to the value-based prices derived from the economic evaluation analysis. Innovator#1 set a launch price of $3.1 Million, close to the monopoly price derived here. There is no information or evidence of differential marginal production costs between the two therapies. Without a differentiated market, if CMS effectively negotiates without allowing full Bertrand-level competition, an effective monopoly price-based payout could be attained for both products that can achieve static and dynamic efficiencies and mitigate demand risk for the innovators.
VI. CONCLUSIONS

There is an ongoing debate about the level of reimbursement necessary to achieve dynamic efficiency in the innovation markets. Although the discussions center around the tradeoff between static and dynamic efficiency, which is the exact tradeoff a social payer faces when deciding whether to transfer today's consumer surplus (static efficiency) to the innovating industry for future gains through future innovations (dynamic efficiency), no models have approached this problem through solving a full dynamic model. In this article, I show that when such a dynamic model is solved, the optimal share of the total surplus from the current innovation that should be given to the innovator industry does not directly depend on the expected dynamic surplus. Instead, it depends on the portion of profits invested by the industry and the marginal product of that investment for the current asset. The dynamic flow of surpluses are internalized within these two parameters. When either of these two parameter increase, optimal appropriate share of total surplus increases for the current asset. I argue that in the case of gene therapy, both these parameters could be high and, therefore, imply a high appropriation share of total surplus.

In the literature, calculating the appropriation share of the total surplus for the industry is usually estimated through varying prices. However, the total demand is assumed to reach the competitive equilibrium quantity under insurance. Under these assumptions, I study the implications for alternative derivation of value-based prices for the appropriation share for the industry. However, the total quantity demanded often fails to reach the competitive equilibrium quantity, as private demand often lags the socially optimal demand (Pauly and Held, 1990; Finkelstein, 2004; Newhouse, 2006). Consequently, derived estimates of prices combined with private sub-optimal demand reduce the effective appropriation share for the industry. A social payer could mitigate this risk for the innovating industry by instituting optimal payout contracts that guarantee the anticipated competitive equilibrium quantity but, perhaps, at a price that lowers the optimal appropriation share under full demand. This can be a win-win for both the industry and the social payer.
I study the implications of these theories, alternative pricing approaches, and alternative financing contracts for gene therapies in sickle cell disease. CMS may be better off instituting a patent buyout for sickle cell gene therapy based on the anticipated socially optimal market size, allowing for interim contract updates should private demand outpace the anticipated demand.

I hope these results and discussions shed new light and continue to shape the debate about the level of reimbursement necessary to achieve dynamic efficiency in the innovation markets.
REFERENCES


Pharmaceutical Research and Manufacturers of America, 2019 PhRMA Annual Membership Survey (PhRMA, 2019), Table 2, https://tinyurl.com/ycvneve7, accessed April 19, 2024.


Table 1: Monopoly price and effective monopoly price estimates for gene therapy in sickle cell disease.

<table>
<thead>
<tr>
<th></th>
<th>Monopoly Price, $p(y^{Mo})$, eq (7) (in millions)</th>
<th>Monopoly quantity*, $y^{Mo}$ (in millions)</th>
<th>Marginal Cost at $y^{Mo}$ (in millions)</th>
<th>Price Elasticity at $y^{Mo}$ (in millions)</th>
<th>Monopoly profits /person, eq (8) (in millions)</th>
<th>Effective Monopoly price, $p_{eff}(y^c)$, ++ eq (10) (in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>$2.8</td>
<td>0.355</td>
<td>$0.56</td>
<td>1.25</td>
<td>$2.11</td>
<td>$1.28</td>
</tr>
<tr>
<td>Low</td>
<td>$2.7</td>
<td>0.370</td>
<td>$0.40</td>
<td>1.17</td>
<td>$2.06</td>
<td>$1.29</td>
</tr>
<tr>
<td>High</td>
<td>$2.9</td>
<td>0.340</td>
<td>$0.80</td>
<td>1.38</td>
<td>$2.21</td>
<td>$1.29</td>
</tr>
</tbody>
</table>

* $y^{Mo}$ : Monopoly quantity expressed as a fraction of the total target population.

++ $y^c$ : Competitive equilibrium quantity as a fraction of the total target population (= 0.98 in the case of sickle cell disease gene therapy). $p_{eff}(y^c)$ : Effective monopoly price that would maintain monopoly-level profits for manufacturers but all supply to $y^c$. 
<table>
<thead>
<tr>
<th>Alternate Prices</th>
<th>Budget Impact &amp; Profits</th>
<th>Patent Buyout*</th>
<th>Traditional Upfront</th>
<th>Annuity**</th>
<th>Outcomes-based: 5-year survival†</th>
<th>Outcomes-based: 3-year severe VOEs††</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-based Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($2.3M)</td>
<td>CMS 10-year NPV BI</td>
<td>$12.77 B</td>
<td>$11.36 B</td>
<td>$8.01 B</td>
<td>$7.88 B</td>
<td>$8.61 B</td>
</tr>
<tr>
<td></td>
<td>CMS Year 1 NPV BI</td>
<td>$1.48 B</td>
<td>$1.32 B</td>
<td>$0.31 B</td>
<td>$0.17 B</td>
<td>$0.17 B</td>
</tr>
<tr>
<td></td>
<td>CMS Year 5 NPV BI</td>
<td>$1.29 B</td>
<td>$1.14 B</td>
<td>$0.76 B</td>
<td>$1.31 B</td>
<td>$1.08 B</td>
</tr>
<tr>
<td></td>
<td>CMS Year 10 NPV BI</td>
<td>$1.10 B</td>
<td>$0.98 B</td>
<td>$1.28 B</td>
<td>$1.12 B</td>
<td>$0.92 B</td>
</tr>
<tr>
<td></td>
<td>10-year Revenue NPV</td>
<td>$11.50 B</td>
<td>$10.10 B</td>
<td>$6.80 B</td>
<td>$6.62 B</td>
<td>$7.35 B</td>
</tr>
<tr>
<td></td>
<td>10-year Profit NPV</td>
<td>$9.22 B</td>
<td>$7.82 B</td>
<td>$4.52 B</td>
<td>$4.33 B</td>
<td>$5.06 B</td>
</tr>
<tr>
<td></td>
<td>% Surplus Appropriated</td>
<td>118%</td>
<td>100%</td>
<td>58%</td>
<td>55%</td>
<td>65%</td>
</tr>
<tr>
<td>Monopoly Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($2.8M)</td>
<td>CMS 10-year NPV BI</td>
<td>$15.23 B</td>
<td>$13.56 B</td>
<td>$9.54 B</td>
<td>$9.32 B</td>
<td>$10.20 B</td>
</tr>
<tr>
<td></td>
<td>CMS Year 1 NPV BI</td>
<td>$1.76 B</td>
<td>$1.57 B</td>
<td>$0.34 B</td>
<td>$0.17 B</td>
<td>$0.17 B</td>
</tr>
<tr>
<td></td>
<td>CMS Year 5 NPV BI</td>
<td>$1.54 B</td>
<td>$1.37 B</td>
<td>$0.90 B</td>
<td>$1.57 B</td>
<td>$1.29 B</td>
</tr>
<tr>
<td></td>
<td>CMS Year 10 NPV BI</td>
<td>$1.32 B</td>
<td>$1.17 B</td>
<td>$1.54 B</td>
<td>$1.34 B</td>
<td>$1.10 B</td>
</tr>
<tr>
<td></td>
<td>10-year Revenue NPV</td>
<td>$14.00 B</td>
<td>$12.30 B</td>
<td>$8.28 B</td>
<td>$8.06 B</td>
<td>$8.94 B</td>
</tr>
<tr>
<td></td>
<td>10-year Profit NPV</td>
<td>$11.71 B</td>
<td>$10.01 B</td>
<td>$5.99 B</td>
<td>$5.77 B</td>
<td>$6.65 B</td>
</tr>
<tr>
<td></td>
<td>% Surplus Appropriated</td>
<td>150%</td>
<td>128%</td>
<td>77%</td>
<td>74%</td>
<td>85%</td>
</tr>
<tr>
<td>Effective Monopoly Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($1.3M)</td>
<td>CMS 10-year NPV BI</td>
<td>$7.76 B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CMS Year 1 NPV BI</td>
<td>$0.91 B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CMS Year 5 NPV BI</td>
<td>$0.78 B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CMS Year 10 NPV BI</td>
<td>$0.67 B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10-year Revenue NPV</td>
<td>$6.50 B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10-year Profit NPV</td>
<td>$4.22 B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>% Surplus Appropriated</td>
<td>54%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Total amount paid to manufacturer via 10-year annuity. ** Each price of gene therapy is paid through a 10 year annuity starting from when the gene therapy was used. † If patient survived till the end of 5 years. †† If patient experienced less than 1 (on average) serious VOE and survived till the end of 3 years.
Figure 1: Relationship between optimal surplus share appropriation, Marginal product of investment, and share of profits reinvested by industry.
Figure 2: Alternate scenarios for (a) value-based pricing, (b) Monopoly pricing, and (c) Effective monopoly pricing.
Figure 3: Benefit distribution and differential demand across two plans.

(a) Benefits Distribution in Two Plans

(b) Monopoly Pricing in Plan 1

(c) Monopoly Pricing in Plan 2

$F_{i_B}(b) = \Pr(B \leq b) \quad \forall b, j = 1, 2$

MC = Marginal Costs
MR = Marginal Revenue
$p_{1H} = \text{Monopoly price for plan 1}$
$y_{1H} = \text{Monopoly quantity for plan 1}$

$abcd = \text{Manufacturer Profits from plan 1}$

$abcd = \text{Manufacturer Profits from plan 2}$
Figure 4: Derived demand, marginal revenue, marginal costs for gene therapy in sickle cell disease.
Figure 5: Size of eligible population enrolled in plans funded by the Centers for Medicaid and Medicare Services (CMS) and the distribution of marginal benefits for gene therapy in sickle cell disease across states in the United States.
APPENDIX
Figure A1:

[Diagram showing discounted costs over years since treatment for different medical cost categories]
Figure A2:

(a) Under Traditional Upfront Payment

(b) Under Annuity Model

(c) Under 5-year survival-based contracting

(d) Under 3-year severe VoE-based-contracting
Figure A3:

(a) Under Traditional Upfront Payment

(b) Under Annuity Model

(c) Under 5-year survival-based contracting

(d) Under 3-year severe VoE-based-contracting
Figure A4

The graph illustrates the manufacturer revenue annually under patent buyout with effective monopoly price, plotted against years since the start of GT coverage. The target pop size is indicated for three different scenarios: 2,500, 5,000, and 10,000. The revenue values decrease over time, with each target pop size represented by a different line type and color.