http://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 24, No. 4, July–August 2022, pp. 2029–2047
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Outcome-Based Reimbursement: The Solution to High Drug Spending?

Liang Xu,[a] Hongmin Li,[b] Hui Zhao[c]

[a] University of Nebraska–Lincoln, Lincoln, Nebraska 68588; [b] Arizona State University, Tempe, Arizona 85281; [c] The Pennsylvania State University, State College, Pennsylvania 16801

Contact: liang.xu@unl.edu, https://orcid.org/0000-0003-2218-9177 (LX); hongmin.li@asu.edu, https://orcid.org/0000-0002-1107-6348 (HL); huz10@psu.edu, https://orcid.org/0000-0002-8995-6644 (HZ)

Received: October 20, 2019
Revised: October 9, 2020; April 8, 2021; July 26, 2021
Accepted: September 7, 2021
Published Online in Articles in Advance: January 19, 2022

https://doi.org/10.1287/msom.2021.1051

Abstract. ***Problem definition*:** The continuously soaring prices of new drugs and their uncertain effectiveness in clinical practice have put substantial risks on insurers/payers. To induce insurer coverage of their new drugs, manufacturers start to propose an innovative outcome-based reimbursement (OBR) scheme under which manufacturers refund insurers (and possibly patients) if the drugs fail to achieve a prespecified treatment target. We investigate the impact of OBR on insurers, manufacturers, and patients. ***Academic/practical relevance*:** Although OBR sounds intuitively appealing, its true impact is under much debate and depends particularly on the design of OBR. Our study sheds light on the optimal design of OBR and the debate around OBR, considering key trade-offs and key elements not covered in prior literature. ***Methodology*:** We develop a Stackelberg game under which the manufacturer designs a rebate scheme for its drug, either non-OBR or OBR, considering the trade-off between a favorable formulary position and the rebate provided. The insurer subsequently determines its formulary for the drug as well as other alternative drugs within the same disease category considering the trade-off between its spending and patient health benefits. Using data on 14 drugs treating a common disease, hyperlipidemia, we estimate through a multinomial logit model the demand of the 14 drugs and conduct counterfactual analyses on the impact of OBR. ***Results*:** Under the optimal OBR, the manufacturer lowers the insurer's risk but inflates the wholesale price (hence, may not reduce insurer spending). OBR also induces a better formulary position for the manufacturer, which, hence, improves patient access to new drugs. Further, rebates to the insurer and patients affect demand through different mechanisms. Including patient rebates in OBR lowers patient expenses and increases drug demand but further increases insurer spending. ***Managerial implications*:** We demonstrate the structure of an optimal formulary and its application in practice. We caution insurers/payers who are seeking OBR to reduce their spending.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2021.1051.

Keywords: outcome-based reimbursement • risk sharing • drug pricing • formulary design • risk aversion • multinomial logit models

1. Introduction

Drug prices, especially those of the newly approved drugs, have been continuously soaring, casting significant financial burdens on the U.S. healthcare system. In 2017, total U.S. spending on retail prescription drugs reached $333.4 billion, or $1,074 per capita, which accounts for 10% of total U.S. health expenditure (Centers for Medicare and Medicaid Services 2017b). At the same time, 49.5% of prescription drug spending is attributed to the most expensive drugs, which account for only about 2.2% of the prescriptions (IQVIA 2019). For example, Herceptin, a drug for breast cancer, has an annual treatment cost of $70,000 per patient (Nordqvist 2012). Given the high prices, insurers' decisions of whether and how to cover these drugs impact both their spending and patients' out-of-pocket expenses and access to these drugs.

Risks of lackluster performance of expensive new drugs in clinical *practice* (as opposed to clinical *trials*) further compounds insurers' coverage decisions. The actual health benefit of a drug in clinical practice (defined as *effectiveness*) is typically uncertain and lower than that observed in clinical trials (defined as *efficacy*) for several reasons. First, clinical trials have strict eligibility criteria for participants, often excluding patients with multiple complications and diseases. For example, only 6% of patients with asthma, a common disease in the United States, could meet the eligibility criteria for asthma-related clinical trials (Travers et al. 2007). Such strict eligibility criteria make it nearly impossible to extrapolate the clinical trial results to the general patient population in clinical practice. Second, unlike participants in clinical trials who are placed under rigorous protocols to take their medication as instructed, patients

in clinical practice up to 50% of the time do not take their medications as prescribed unintentionally or intentionally (Brown and Bussell 2011). Such nonadherence behavior especially affects high-priced drugs, for which patients may skip or split dosage to reduce cost, thus compromising the effectiveness of these drugs. Third, some drugs approved under the FDA's accelerated approval process may later prove to be ineffective in postmarket studies (Xu et al. 2021).

High drug prices combined with uncertain effectiveness pose tremendous risk for insurers as they decide the coverage and formulary of these new drugs. Although individual patients may respond to a drug differently because of patient heterogeneity, the proportion of patients with successful treatment outcomes is typically established and known for an existing drug; thus, an insurer can estimate its health benefit. However, this is not the case for a new drug because of the aforementioned reasons. As a result, many insurers, particularly small ones, such as employer-sponsored insurance plans, are hesitant to include these new drugs in their formulary (Reddy 2017) because of this "ambiguous risk" of failure (Kunreuther et al. 1993). Therefore, insurers may be reluctant to provide insurance coverage because they are unable to precisely estimate the effectiveness of the new drug, impeding patients' access to the new drug. The insurer formulary, which determines whether to include a drug and, if included, the copayment of the drug, directly affects patients' access, out-of-pocket expenses, and consequently demand of the drug.

To encourage coverage and a favorable copayment (i.e., formulary position) for their drugs, manufacturers recently have proposed outcome-based reimbursement (OBR), under which the manufacturer provides rebates to the insurer (and possibly patients) if the drug does not achieve a prespecified treatment target. OBR has received wide attention in the pharmaceutical industry. For example, Harvard Pilgrim, an insurer, has recently signed eight OBR schemes. The Centers for Medicare and Medicaid Services (CMS) also plan to extend OBR for its Medicare Part B prescription drugs (Centers for Medicare and Medicaid Services 2016). Refer to Online Table 1A in the online appendix for a summary of OBR schemes recently implemented in the United States. Although most manufacturers provide OBR rebates to insurers only, some provide rebates to both insurers and patients.

Intuitively, OBR could reduce insurer spending because the insurer only pays for patients who respond positively to the drug. Indeed, Novartis' CEO predicts that OBR, if widely implemented, could potentially reduce U.S. healthcare expenditure by 25% (Mukherjee 2017). CMS is also exploring the possibility of using OBR to lower spending (Centers for Medicare and Medicaid Services 2017a). However, many are skeptical of OBR's impact because the manufacturer may provide a rebate but inflate the base price (Thomas and Ornstein 2017). Since 2006, the Italian National Health System has implemented OBR for several drugs, but the rebate from manufacturers is trifling (only 3.27% out of the total 3.7 billion sales) (Navarria et al. 2015). Indeed, it remains debatable whether OBR truly lowers insurer spending.

It is also not clear whether the manufacturer benefits from OBR or how the manufacturer should design the OBR scheme. On the one hand, the manufacturer has a strong incentive to exploit the most profit from a new drug to recoup its R&D cost before patent expiration. This may drive the manufacturer to design an exploitative OBR that places the insurer at a disadvantage. On the other hand, the manufacturer must also provide a generous OBR in exchange for a more favorable formulary position for its drug. As mentioned, insurer formulary, to a large extent, determines the drug demand and its profit. Therefore, the impact of OBR boils down to how the manufacturer weighs the trade-off between a favorable formulary position and the cost of a rebate as well as how the insurer leverages its formulary to influence the manufacturer's offer of OBR.

In this paper, we investigate, for a new drug with uncertain effectiveness, the manufacturer's optimal design of the OBR scheme considering the insurer's optimal formulary design and how the OBR scheme would correspondingly impact the insurer, the manufacturer and the patients, as compared with a non-OBR scheme. Specifically, considering the uncertainty in treatment outcomes resulting from a new drug's uncertain effectiveness in addition to patient heterogeneity, how should the manufacturer design its OBR, for example, how to set the optimal wholesale price, whether to offer outcome-based rebates also to patients, and if so, how to optimally allocate/split the rebates between the insurer and patients? Further, how should the insurer design its optimal formulary as a leverage to obtain better OBR terms from the manufacturer? Will and how will the rebates affect insurer spending, manufacturer profit, and patient access to a new drug? How do the risk attitudes of the insurer and patients influence OBR and its impact? To address these questions, we develop a Stackelberg game under which the profit-maximizing manufacturer first sets a price and rebate scheme for its new drug, either non-OBR or OBR, considering the trade-off between a favorable formulary position and the rebate provided to the insurer (and possibly patients). The insurer subsequently determines its formulary for the drug as well as other alternative drugs within the same disease category, considering the trade-off between patients' health benefits and its spending.

We develop an analytical model whose parameters can be estimated using real-world data to provide

answers to these questions. To this end, we calibrate our model using data of 14 drugs used to treat a common disease, hyperlipidemia, and conduct a counterfactual analysis accordingly. Specifically, we first estimate through a multinomial logit (MNL) model the demand of the 14 drugs as a function of drug effectiveness and copayments. The estimation results are then used as input to our analytical model to obtain the optimal OBR scheme and the corresponding insurer formulary. We then quantify the impact of OBR on the insurer, manufacturer, and patients through counterfactual analyses of one hyperlipidemia drug, Simcor, which has evident uncertain effectiveness.

Our analysis brings forth a few interesting insights: First, manufacturers can leverage OBR to increase the drug price and induce a better formulary position, hence improving patient access to the drug at the same time. Although the former has been documented in the literature (e.g., Adida 2021), the latter has not been shown previously, but is a primary driver for manufacturers' adoption of OBR. Second, although OBR can lower the insurer's risk, it does not lower the insurer's expected spending because of the manufacturer's inflated price. Specifically, based on our counterfactual analyses for the case of Simcor, the insurer would spend about 2.4 ~ 8.9% more under OBR than under non-OBR. Third, compared with the rebate to the insurer, which increases drug demand indirectly through a favorable formulary tier, it is often more efficient for the manufacturer to allocate a rebate to patients to directly improve drug demand. However, the increased drug demand further increases insurer spending. In particular, our counterfactual analyses show that insurer spending increases by about 53.8% ~ 64.1% under OBR with a patient rebate as compared with under non-OBR. Fourth, we demonstrate that the optimal formulary has a price-nested structure in which drugs with a lower effective price should be assigned to a more favorable tier (i.e., a tier with lower copayment), thus incentivizing the manufacturer to increase its rebate or reduce its wholesale price. Finally, if the insurer cannot adjust the formulary of existing drugs when deciding the formulary tier of the new drug, we show with an example that the insurer has to pay a 38.5% higher wholesale price for the new drug. This demonstrates how the insurer may leverage its formulary design to influence the effective price of a new drug.

Our paper makes several important contributions. First, from a public policy perspective, current high drug prices and spending are a contentious issue for the government. Compared with the limited previous literature on the impact of OBR, our paper further sheds light on this topic by considering additional key components, including the insurer's optimal formulary design under OBR, the uncertainty in treatment outcomes resulting from a drug's uncertain effectiveness in addition to patient heterogeneity, and whether the manufacturer should provide OBR to patients and how to optimally split OBR between patients and the insurer. Second, from a modeling perspective, (1) we provide a framework for the optimal design of the key parameters of OBR, the lacking of which has been one of the most notable barriers for implementing OBR (Nazareth et al. 2017); (2) we model both patients' and the insurer's risk attitude toward uncertainty in the new drug's effectiveness and their implications on the impact of OBR, which has not been done by prior literature; and (3) we collect and leverage data from the 14 drugs that treat hyperlipidemia to calibrate the model and quantify the impact of OBR in this case, demonstrating how our model can be operationalized and implemented in practice. Third, from a theoretical perspective, our paper extends assortment planning analysis to the formulary design setting. The formulary design problem resembles assortment planning but possesses special features that require extension of the state-of-the-art techniques in this literature. We establish structural properties of the optimal formulary and demonstrate how the special features of formulary design affect the optimal assortment solution.

The remainder of the paper is organized as follows. In Section 2, we position our paper in regard to the related literature. In Section 3, we depict the model setup, and in Section 4, we solve the optimal formulary without uncertainty in drug effectiveness. In Sections 5, we optimize the design of OBR and the insurer formulary with uncertain drug effectiveness. In Section 6, we calibrate our model to apply data on drugs treating hyperlipidemia to assess the impact of OBR. Section 7 concludes the paper with managerial implications. Proofs of analytical results are in the online appendix.

2. Literature Review

Because of ever-increasing drug prices and the pressure to ensure value for drug spending, OBR has received wide attention from payers, manufacturers, and government agencies. Studies in the health policy field discuss qualitatively the good practices for design, implementation, and evaluation of OBR. For example, Garrison et al. (2013) summarize the current practices of OBR in France, Italy, Netherlands, the United Kingdom, and the United States, and Carlson et al. (2010) introduce the emerging OBR schemes in the United States. Coulton et al. (2010) provide a comprehensive taxonomy of different variants of OBR schemes implemented worldwide. In contrast, our paper provides an analytical framework for the optimal design of the key parameters of OBR. The absence of such framework is one of the most notable barriers for implementing OBR (Nazareth et al. 2017). We focus

on OBR practices in the United States because of its unique regulatory context and insurance market.

Studies in the health economics field examine OBR in various settings analytically to assess its impact on the pharmaceutical industry (e.g., Zaric and Xie 2009, Antonanzas et al. 2011, Mahjoub et al. 2018). These studies do not capture the relationship between drug demand and insurer formulary or other alternative treatments on the insurer formulary, which are some key considerations of the manufacturer when providing OBR in the U.S. market.

The operations management field has some works on OBR. So and Tang (2000) examine how the insurer should design OBR to discourage unnecessary or excessive prescriptions but do not consider the risk associated with uncertain drug effectiveness. A working paper by Truong and Yao (2013) models the manufacturer's OBR decision but assumes that copayments of drugs can be any continuous amount instead of the tiered formulary commonly used in practice. As a result, they conclude that the insurer subsidizes a fixed amount across different drugs, and patients pay the remaining amount although a typical insurance policy calls for patients paying a fixed amount as copayment and the insurer paying the remaining amount. In addition, they conclude that OBR is equivalent to non-OBR with a risk-adjusted price, which is a special case of the risk-neutral insurer in our study. Adida (2021) compares the manufacturer's optimal pricing decision under non-OBR and OBR (under which a fixed fraction of drug cost is refunded to the insurer and patients). She examines the insurer's coverage decision of a single drug without considering other alternative drugs in the same disease category, nor does she consider the insurer's formulary decision.

Our paper differs from the cited OBR literature in the following ways: (1) A drug's uncertain treatment outcomes can be attributed to two different sources: patients' heterogeneity and the drug's uncertain effectiveness. Adida (2021) considers only the former by assuming the proportion of patients successfully treated by the drug (ρ) to be a known constant; hence, the insurer essentially bears no risk because of risk pooling. In contrast, we model both sources of uncertainty (ρ is uncertain) because an insurer faces much uncertainty in the effectiveness of a newly released drug. We examine how the manufacturer leverages OBR to alleviate the risk of the insurer (and the patients if OBR includes patient rebate) in exchange for better insurance coverage, which is a key incentive for the manufacturer to provide OBR. Our study shows that the manufacturer can provide OBR to induce a better formulary position for its new drug, whereas such an advantage of OBR on the insurer formulary is not studied in Adida (2021). (2) Although it is common practice for insurers to leverage their power in formulary design to negotiate rebates and prices with manufacturers (Kouvelis et al. 2018), most OBR studies do not consider the insurer's formulary design. We are the first to evaluate the impact of OBR considering the insurer's adjustment of its tiered formulary as a response to the manufacturer's OBR. The design of the tiered formulary leads to a nontrivial, nonlinear, binary mathematical program, for which we propose a novel solution. Modeling the tiered formulary and the risk of drug effectiveness in clinical practices also allows us to parameterize the model using real-world data to provide realistic evaluation on the impact of OBR. (3) By considering alternative drugs in a disease category and allowing the insurer to adjust its formulary as a new drug enters the market, we capture the insurer's influence on the manufacturer's design of OBR: a lower effective price of the new drug warrants a formulary tier with a lower copayment than existing drugs, incentivizing the manufacturer to increase the rebate or to reduce the wholesale price. (4) We optimize both the manufacturer's wholesale price and rebate decisions under OBR, and previous works (e.g., Adida 2021) only consider the manufacturer's pricing decision when the rebate is a fixed fraction of the drug price. Additionally, we also consider rebates to both insurers and patients under OBR and the optimal split between them. Compared with rebates to the insurer, which increases drug demand indirectly through a favorable formulary tier, rebates to patients increases drug demand directly and is often more effective for the manufacturer. (5) We capture the insurer's risk aversion as well as patients' risk preference (risk averse or risk seeking) toward the uncertain drug effectiveness, depending on patients' disease severity and available treatment options.

The insurer's formulary design is studied in contexts unrelated to OBR. For example, Kouvelis et al. (2015) study the design of tiered formulary from the perspective of pharmacy benefit managers (PBMs) who offer such a formulary to the patronage of a client (such as employers and insurers). In Kouvelis et al. (2015), the PBM determines jointly the formulary and its resell price to clients to maximize profit. They conclude that, in a special case in which all drugs have the same effectiveness, PBM places the most cost-effective drugs on the preferred tier, which is, in spirit, similar to our result about the optimal formulary. However, their decision setting is rather different from ours in which the insurer decides the formulary and the manufacturer decides the price. In addition, they consider neither the uncertainty of drug effectiveness nor OBR.

The mathematical formulation of the insurer formulary design problem is, to some extent, similar to the assortment-planning problem in retail (e.g., Talluri and Ryzin 2004, Davis et al. 2013) in that the insurer

needs to decide which drugs to include in its formulary. However, the formulary design problem has two distinct features that add to its complexity. First, in addition to deciding whether to cover a drug (i.e., include the drug in the formulary), the insurer also decides the formulary tier for the drug. Second, the assortment-planning problems typically maximize the profit from an assortment, and in our context, the insurer must balance the trade-off between spending and patients' health benefits. In this paper, we extend the state-of-the-art technique for assortment-planning problems (e.g., see Davis et al. 2013) to accommodate these added complexities and establish properties of the optimal formulary. As a result, we generalize the analytical method of assortment planning and broaden its applicability.

OBR is related to the broad class of risk-sharing contracts, which are studied in various contexts. For example, Zhang et al. (2011) and Zaric and O'Brien (2005) investigate how to design risk-sharing contracts to mitigate the risk of uncertain demand for new drugs. These studies investigate a risk-sharing contract to contain the insurer spending, under which the drug price is discounted if demand of the drug exceeds a threshold. In contrast, OBR addresses the risk derived from the quality uncertainty of the drug. This is somewhat similar to a manufacturer warranty, which guarantees a level of performance of products or services to countervail the ex post quality uncertainty. A warranty is a type of risk-sharing contract that is widely used and studied (e.g., Thomas and Rao 1999). Our context is different because, when designing OBR, the manufacturer has to consider how the insurer will respond through its formulary decision, which is unique in the pharmaceutical context.

Finally, we remark that, OBR for prescription drugs is different from pay-for-performance for medical services adopted by CMS and other private insurers, under which insurers reimburse medical service providers based on the *quality* of care rather than *quantity* of care provided. Studies on pay-for-performance mostly focus on how to incentivize providers to improve their service quality (e.g., Gupta and Mehrotra 2015, Zorc et al. 2017). In comparison, the effectiveness of drugs is exogenous and not fully revealed when the drugs enter the market; the decision and challenges are, thus, quite different as we describe.

3. Model Setups

Consider a disease category with $J-1$ existing drugs and a newly approved drug J in the market. Each drug is intended to deliver a prespecified health benefit q_j, determined by its distinct chemical or biological structure and informed by its premarket clinical trial results. For example, for drugs treating hyperlipidemia, the health benefit q_j is measured by the improvement on patients' hyperlipidemia level. However, patients do not uniformly achieve the prespecified treatment benefit because of their heterogeneity in terms of disease status, underlying conditions, etc. For a new drug that moves from clinical trial to practice, the uncertainty in treatment outcome originates from not only patient heterogeneity, but also uncertainty in drug effectiveness. We capture both sources of uncertainty. Specifically, we denote $\tilde{\rho}_j$ as the proportion of patients treated successfully (i.e., achieving the predefined treatment benefit of drug j). For the new drug J, its effectiveness in clinical practice is still uncertain; hence, the true value of $\tilde{\rho}_J$ is unknown (Nallamothu et al. 2008). We assume the insurer has a prior belief about $\tilde{\rho}_J$ based on the clinical trial results and characterized by a probability distribution function (pdf) f with $E(\tilde{\rho}_J) = \rho_J$. Thus, ρ_J estimates the proportion of patients treated successfully by drug J. This contrasts with Adida (2021) in which $\tilde{\rho}_J$ is modeled as a deterministic constant; hence, the insurer bears no risk. For existing drugs $j = 1..J-1$, their effectiveness has been long established; thus, we assume $\tilde{\rho}_j$ equals a constant ρ_j, $\forall j = 1..J-1$. This assumption reflects a normalized difference in risk between a new drug and an established drug, a key driver for the OBR arrangement for new drugs. We assume that the manufacturer and the insurer have the same information about ρ_j, $\forall j = 1,..J-1$, and $\tilde{\rho}_J$, which is also shared with physicians. Online Figure A2 summarizes patients' possible responses to new and existing drugs and their corresponding health benefits, which we elaborate later.

A pharmaceutical manufacturer typically sets a list price for its new drug based on many strategic factors beyond the scope of this paper. Insurers seldom pay the list price but instead pay a lower wholesale price to the manufacturer by leveraging the formulary decisions. If a drug j is excluded from the insurer formulary, the manufacturer determines a direct-sell cash price to patients instead. We assume that the wholesale price p_j and cash price p_{j0} for all existing drugs are exogenously given as these drugs have been long established in the market. We focus on manufacturer J's decisions of wholesale price, cash price, and rebates.

The manufacturer of drug J can choose either a non-OBR scheme (i.e., no rebate based on patients' realized health outcome) or an OBR scheme, which provides a rebate to the insurer (and potentially patients as well) based on the realized health outcome. For example, for Simvastatin (a drug treating hyperlipidemia), Merck offered an OBR that refunds both insurers and patients if Simvastatin does not lower patients' blood cholesterol below a target level (Carlson et al. 2009). Such a "no cure, no pay" campaign has seen increasing popularity in the pharmaceutical industry. To capture these features, we model the manufacturer's

pricing scheme as follows: the manufacturer decides a wholesale price p_J to the insurer; if a patient does not achieve the predefined treatment benefits after taking the drug, the manufacturer provides a rebate R_1 to the insurer and R_2 to the patient. Thus, $R_1 = 0$ indicates the case of a non-OBR scheme, $R_1 > 0$ indicates the case of an OBR scheme, and $R_2 > 0$ indicates that the OBR scheme includes patient rebates. Thus, the effective price for drug J (i.e., price after insurer rebate) is $p_J - R_1(1 - \tilde{\rho}_J)$.

To solve the manufacturer's optimal rebate scheme (i.e., p_J, R_1, and R_2), we develop a Stackelberg game under which the manufacturer first decides its rebate scheme, either non-OBR or OBR, to maximize its profit, and the insurer subsequently decides its optimal formulary considering the trade-off between patients' health benefits and its spending. Given the manufacturer's rebate scheme and the insurer formulary, which determines the copayments of different drugs, patients/physicians choose the drug that benefits them the most. Once the effectiveness of the new drug among patients is revealed, the manufacturer pays the outcome-based rebate to the insurer (and possibly patients) retrospectively. We next describe the key elements of the model.

3.1. The Insurer's Tiered Formulary

Under a tiered formulary, the insurer assigns drugs within a particular disease category to different tiers, each tier corresponding to a different copayment level. The tiered formulary is widely used by insurers to direct patients toward less expensive drugs. In 2011, 77% of private insurance plans (Claxton et al. 2011) and 91% of Medicare Part D plans (Hoadley et al. 2011) had a formulary with three or more tiers. Online Table A2 illustrates the 2015 tiered formulary for hyperlipidemia from Cigna, the second largest private insurer in the United States. As Online Table A2 shows, tiers 1–4 correspond to copayments of \$0, \$10, \$45, and \$95, respectively. Insurers may also choose to not cover a drug, a strategy increasingly used by insurers to reduce spending.

Without loss of generality, we assume that the insurer has a K-tier formulary with a_k being the copayment for the kth tier, $k = 1, 2, \ldots, K$, and $a_1 < a_2 < \ldots < a_K$. In addition, $k = 0$ indicates that the drug is excluded from the formulary. The insurer's formulary decision is to assign each drug to a specific tier,for which $x_{jk} = 1$ if drug j is assigned to the kth tier and $x_{jk} = 0$ otherwise, $k = 1, 2, \ldots, K$. In addition, $x_{j0} = 1$ indicates that the insurer excludes drug j from the formulary, and $x_{j0} = 0$ indicates that the insurer includes drug j in the formulary. Thus, a feasible formulary assignment $\boldsymbol{x}$ is characterized by $\sum_{k=0}^{K} x_{jk} = 1$, $\forall j$, that is, a drug is either assigned to one specific tier or excluded from the formulary. If the insurer decides to cover drug j and assigns it to the kth tier, then patients are responsible for the copayment a_k, and the insurer pays $p_j - a_k$ to the manufacturer for each unit of drug j purchased. If, however, the insurer decides to exclude drug j from its formulary, then the patients have to purchase the drug on their own at the cash price p_{j0}. Accordingly, given the copayment value for each tier $\boldsymbol{a} = (a_1, \ldots, a_K)$ and the insurer's formulary assignment decision x_{jk}, $j = 1, \ldots, J$, $k = 0, 1, \ldots, K$, the corresponding copayment for drug j is $c_j(x) = \sum_{k=1}^{K} x_{jk} a_k + x_{j0} p_{j0}$, and we denote $\boldsymbol{c}(x) = (c_1, \ldots, c_J)$ as the drug copayment vector (denoted as $\boldsymbol{c}$ for simplicity).

3.2. The Demand of Drugs

We assume patients and their physicians to be one agent in choosing drugs because physicians, in principle, make prescription decisions according to patients' best interests. Although an individual patient's drug choice is complex and involves many factors, the aggregated demand can be reasonably estimated. We model the aggregated demand of drugs using an MNL model, which is widely used in the literature to capture patients' demand of drugs (see, e.g., Kouvelis et al. 2015, Ching and Lim 2020). Specifically, under the MNL model, patients choose the drug that maximizes their utility, considering the effectiveness of the drug and its uncertainty and out-of-pocket copayments as well as other random factors, such as the characteristics and preferences of the patients and their physicians. We capture patients' risk preference toward the uncertainty in drug effectiveness through the mean-variance approach, following the literature on consumer choice of products with quality uncertainty (e.g., Erdem and Keane 1996, Ching and Lim 2020).

Suppose a patient purchases the new drug J: if the patient achieves the treatment target after taking the drug, the patient gains the health benefits q_J from the treatment and pays the copayment c_J. Otherwise, the patient receives a rebate R_2 from the manufacturer. Let $\mathbb{I}_{ij}$ be the binary indicator of whether patient i achieves the treatment target after taking drug j and μ_{ij} be the corresponding realized utility. Thus,

$$\mu_{iJ} = q_J \mathrm{I}_{iJ} - \theta(c_J - R_2(1 - \mathrm{I}_{iJ})) + \epsilon_{iJ},$$

where θ measures patients' sensitivity to spending (i.e., copayment) and ϵ_{iJ} represents patients'/physicians' idiosyncratic preference toward drug J because of factors such as patients' disease status, age, income, and other related characteristics of the patients and their physicians. We assume ϵ_{ij} follows a standard type I extreme value distribution.

Because of patient heterogeneity and uncertain drug effectiveness, patients face the risk of treatment failure when taking the new drug. To model patients' aversion to such risk, we adopt the mean-variance

approach under which patients' risk-adjusted utility from taking the new drug J is

$$\begin{aligned} u_{iJ} &= E_{\tilde{\rho}_J}[E_{\mathrm{I}_{ij}}(\mu_{iJ}|\tilde{\rho}_J)] - \vartheta\sigma(\mu_{iJ}) \\ &= E_{\tilde{\rho}_J}[\tilde{\rho}_j(q_J - \theta c_J + \epsilon_{iJ}) + (1-\tilde{\rho}_J)(-\theta(c_J - R_2) + \epsilon_{iJ})] \\ &\quad -\vartheta|q_J - \theta R_2|\sqrt{\rho_J(1-\rho_J)} \\ &= q_J\rho_J - \vartheta q_J\sigma_J + \vartheta\theta R_2\sigma_J - \theta(c_J - (1-\rho_J)R_2) + \epsilon_{iJ}, \end{aligned}$$

where $\vartheta\sigma(\mu_{iJ})$ captures patients' disutility from uncertain treatment outcome, and we can derive $\sigma(\mu_{iJ}) = |q_J - \theta R_2|\sqrt{\rho_J(1-\rho_J)}$. Let $\sigma_J = \sqrt{\rho_J(1-\rho_J)}$ be the standard deviation of the Bernoulli random variable representing the treatment outcome. ϑ indicates patients' risk preference. Depending on the severity of the disease and patients' alternative treatment options, patients may be either risk averse (i.e., $\vartheta > 0$) or risk seeking (i.e., $\vartheta < 0$) in trying a new drug (even participating in clinical trials). $q_J\rho_J$ measures the ex ante expected health benefit from the drug, and $\vartheta q_J\sigma_J$ measures the drug-specific disutility from the risk of treatment failure. Note that the manufacturer's rebate to the patient (i.e., R_2) affects patients' risk-adjusted utility through both the expected utility and the risk disutility term. We focus on the case in which $R_2 \le q_J/\theta$ because it is never optimal for the manufacturer to refund a patient more than the prespecified health benefit.

Similarly, patient i's realized utility from purchasing an existing drug j, $j = 1,..,J-1$, is

$$\mu_{ij} = q_j\,\mathrm{I}_{ij} - \theta c_j + \epsilon_{ij}.$$

Define $\sigma_j = \sqrt{\rho_j(1-\rho_j)}$. Then, the risk-adjusted utility from drug j is

$$\begin{aligned} u_{ij} &= E_{\mathrm{I}_{ij}}[\mu_{ij}] - \vartheta\sigma(\mu_{ij}) = q_j\rho_j - \vartheta q_j\sigma_j - \theta c_j + \epsilon_{ij},\ \forall j \\ &= 1,\ldots,J-1. \end{aligned}$$

Note that the disutility associated with the risk of treatment failure applies to all drugs $j = 1...J$ because no drug is guaranteed to treat every patient successfully. In addition, if patient i does not purchase any drug from the list of drugs, we assume that the patient acquires a normalized utility $u_{i0} = \epsilon_{i0}$ from the outside option.

Given the insurer's formulary assignment $\boldsymbol{x}$ and the manufacturer's rebate to patients R_2, patients choose the drug that yields the highest utility. Accordingly, the market share for drug j is

$$d_j(\boldsymbol{x}, R_2) = \frac{\sum_{k=0}^{K} x_{jk}v_{jk}}{1 + \sum_{t=1}^{J}\sum_{k=0}^{K} x_{tk}v_{tk}},\ \forall j = 1,\ldots,J,$$

where $v_{jk} = \exp(q_j\rho_j - \vartheta q_j\sigma_j - \theta a_k),\ \forall k \ge 1$, and $v_{j0} = \exp(q_j\rho_j - \vartheta q_j\sigma_j - \theta p_{j0}),\ \forall j = 1,..,J-1$ represent the attractiveness of drug j to patients if placed on the kth tier and if excluded from the formulary, respectively; $v_{Jk} = \exp(q_J\rho_J - \vartheta q_J\sigma_J + \vartheta\theta R_2\sigma_J - \theta(a_k - (1-\rho_J)R_2))$ and $v_{J0} = \exp(q_J\rho_J - \vartheta q_J\sigma_J - \theta p_{J0})$ represent the attractiveness of drug J to patients if placed on the kth tier and if excluded from the formulary, respectively.

4. Formulary Design Without Uncertain Drug Effectiveness

We first develop a base model for the insurer's formulary design given drug J's effectiveness (i.e., $\tilde{\rho}_J$ equals a constant ρ_J). This base model lays the groundwork for analyzing the insurer's formulary design and serves as a benchmark for the case with uncertain effectiveness in Section 5.

The insurer determines its formulary by weighing the trade-off between patients' health benefits and its spending: on the one hand, the insurer needs to provide sufficient coverage and an affordable copayment to deliver health benefits to its insured patients; on the other hand, the insurer needs to reduce its spending to maintain profitability. Given a formulary assignment $\boldsymbol{x}$, a patient's expected health benefit can be written as

$$B(\boldsymbol{x}) = E_{\epsilon=(\epsilon_{ij})'s}(\max_{j=0,1..J}(u_{ij})) = \ln(1 + \textstyle\sum_{j=1}^{J}\sum_{k=0}^{K} x_{jk}v_{jk}),$$

where $\ln(1 + \sum_{j=1}^{J}\sum_{k=0}^{K} x_{jk}v_{jk})$ represents the expected health benefit derived from the formulary based on the probability distributions of ϵ_{ij} (Anderson et al. 1992). Correspondingly, the insurer spending associated with the formulary assignment $\boldsymbol{x}$ is

$$S(\boldsymbol{x}) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(p_j - a_k)x_{jk}v_{jk}}{1 + \sum_{j=1}^{J}\sum_{k=0}^{K} x_{jk}v_{jk}}.$$

When the effectiveness of the drugs is certain, patients' heterogeneous response to drugs are averaged out in aggregate. Therefore, the insurer knows the proportion of successful treatments and, thus, can easily evaluate the health benefits and spending associated with a formulary. Thus, the insurer solves the following nonlinear binary programming problem to maximize the spending-adjusted health benefits $u(\boldsymbol{x})$:

$$\begin{aligned} \max_{x}\ & u(\boldsymbol{x}) = wB(\boldsymbol{x}) - S(\boldsymbol{x}) \\ \text{s.t.}\ & \sum_{k=0}^{K} x_{jk} = 1,\ \forall j = 1,\ldots,J, \\ & x_{jk} \in \{0,1\},\ \forall j = 1,\ldots,J,\ \forall k = 0,1,\ldots,K. \end{aligned} \tag{1}$$

Here, w represents the insurer's weight on patients' health benefits relative to its spending. Note that we do not impose that patients and the insurer have the same weight between health benefits and monetary

spending (i.e., w is not necessarily equal to $1/\theta$) because, depending on the specific disease, patients may place more or less weight on spending compared with the insurer. The first constraint ensures feasibility of an assignment such that each drug is either assigned to one specific tier or excluded from the formulary. Note that our model can also accommodate other constraints that the insurer may encounter in practice. For example, if the insurer designates a drug j to or not to a specific tier k, the resulting constraint would be $x_{jk} = 1$ or $x_{jk} = 0$; if drugs i and j cannot be both covered, the resulting constraint would be $x_{i0} + x_{j0} > 0$.

When the number of alternative drugs in a disease category is relatively small, we can solve the optimal formulary through enumeration. However, as J increases, such a brute-force procedure soon becomes computationally prohibitive because the number of enumerations $(K+1)^J$ increases exponentially. For example, given the number of drugs available for hyperlipidemia as shown in Online Table A2, the number of enumerations to search for the optimal formulary is more than 10^9. In fact, many diseases have more alternative drugs than hyperlipidemia: diabetes and hypertension both have more than 20 alternative drugs, resulting in more than 10^{14} enumerations. Therefore, it is imperative to derive the structural properties of the optimal formulary to expedite the search procedure. Such properties can also provide useful insights and heuristics for the insurer to design and refine its formulary. Toward this end, we transform Problem (1) as follows. Let $z_{jk} = \frac{x_{jk}v_{jk}}{1+\sum_{t=1}^{J}\sum_{k=0}^{K} x_{tk}v_{tk}}$, $\forall j,k$ represent the market share of drug j if assigned to the kth tier and let $z_0 = \frac{1}{1+\sum_{t=1}^{J}\sum_{k=0}^{K} x_{tk}v_{tk}}$ represent the market share of the outside option. The demand for drug j is $d_j = \sum_{k=0}^{K} z_{jk}$, $\forall\, j = 1,\ldots,J$. Accordingly, Problem (1) can be transformed into the following relaxed Problem (2):

$$\begin{aligned} \max_{z:z_0,z_{jk},\forall j,k} \; & -w\ln z_0 - \sum_{j=1}^{J}\sum_{k=1}^{K}(p_j - a_k)z_{jk} \\ \text{s.t.} \quad & z_0 + \sum_{j=1}^{J}\sum_{k=0}^{K} z_{jk} = 1, \\ & \sum_{k=0}^{K}\frac{z_{jk}}{v_{jk}} = z_0, \;\forall j, \\ & z_0 \geq 0, z_{jk} \geq 0, \;\forall j,k. \end{aligned} \tag{2}$$

The assortment-planning literature uses a similar relaxation technique to reformulate the assortment-planning problem into a linear programming problem (e.g., Davis et al. 2013). Our formulary design problem has two distinct features that add to its complexity. First, in addition to deciding whether to cover a drug, which is similar to the assortment decision, the insurer must also decide the formulary tier to which to assign the drug. Second, the assortment-planning literature typically maximizes the profit of an assortment. However, in our context, the insurer has to balance its spending with patients' health benefits, resulting in a nonlinear objective function in Problem (2). We are able to demonstrate the equivalency between Problems (1) and (2) even with the added complexity.

Lemma 1. *Problems* (1) *and* (2) *are equivalent in solving the optimal formulary.*

Lemma 1 ensures that we can construct the optimal formulary for Problem (1) by solving Problem (2), through which we can derive useful structural properties of the optimal formulary.

Definition 1. A formulary assignment $\tilde{x}$ is efficient if and only if, for some value $\gamma \geq 0$, $\tilde{x}$ solves the following problem:

$$\max_{x} \; \gamma B(\boldsymbol{x}) - S(\boldsymbol{x}),$$

where $B(\boldsymbol{x})$ and $S(\boldsymbol{x})$ represent the patients' health benefits and insurer spending, respectively, as defined previously. According to this definition, an efficient formulary optimally balances patients' health benefits $B(\boldsymbol{x})$ and insurer spending $S(\boldsymbol{x})$ for some weight γ. Therefore, it follows directly that the optimal formulary for Problem (1) must be efficient.

Definition 2. A formulary is ordered by effective price if any drug with a lower wholesale price than drug j is assigned to a tier with a lower copayment than drug j, that is, if $p_l \leq p_j$, then $c_l \leq c_j$, $\forall\, l \neq j$.

Proposition 1 characterizes the efficient formulary.

Proposition 1. *An efficient formulary must be ordered by wholesale price of drugs.*

Proposition 1 suggests that the efficient formulary has a similar structure as the efficient assortment. Talluri and Ryzin (2004) demonstrate that the efficient assortment follows a nested structure under which higher margin products are included first, followed by lower margin products. Likewise, in our context, an efficient formulary is nested according to the wholesale price such that drugs with lower wholesale prices are assigned first to a lower tier (with a lower copayment), followed by drugs with higher wholesale prices assigned to a higher tier (with a higher copayment). Therefore, despite the distinct features and added complexity of our problem, we establish a similar structure of the optimal formulary as in Talluri and Ryzin (2004).

The property of the efficient formulary in Proposition 1 has several implications on how the insurer should design its formulary. First, if a drug is included in the formulary, all other drugs with lower wholesale

prices should also be included. Therefore, when designing its formulary, the insurer can include drugs sequentially in the order of their wholesale prices such that less expensive drugs are on the tiers with a lower copayment. Such a formulary structure allows the insurer to direct patients toward less expensive drugs. This is consistent with practice. Second, price rather than effectiveness is the qualifier for a drug to be included in the formulary. This is because each drug, regardless of its expected effectiveness level, may best suit certain patients because of patients' idiosyncratic preferences characterized by ϵ_{ij}. Therefore, barring the prohibitive administrative cost for including more drugs, it is beneficial for the insurer to cover all FDA-approved drugs as long as their prices are reasonable because some subpopulation of patients benefit from this drug more than other drug alternatives. Nevertheless, the drug effectiveness does impact the overall demand for each drug: drugs with lower q_j have a lower demand as the subpopulation of patients who find this drug to be their utility maximizer is smaller.

To find the optimal formulary, we rank the drugs according to their wholesale prices and evaluate all wholesale price–ordered formulary assignments to identify the optimal one. Without such an optimal formulary structure, one has to enumerate all $(K+1)^J$ possible formulary assignments. With such a structure, we only need to evaluate the $\binom{J+K}{K}$ efficient formulary assignments, which substantially alleviates the computational burden. Please see Online Figure A1 for a numerical example of the efficient formulary.

5. Formulary and OBR Design Under Uncertain Drug Effectiveness

In the base model, we characterize the optimal formulary without uncertainty in drug effectiveness. Next, we solve the insurer's formulary design under uncertain drug effectiveness given the manufacturer's price and rebate scheme, followed by the manufacturer's optimal price and rebate decision.

5.1. Formulary Design Under Uncertain Drug Effectiveness

In this section, we examine the insurer's formulary design given the price and rebate scheme when the effectiveness of drug J is uncertain. In this case, the insurer does not know the proportion of successful treatments by drug J, and we explore the impact of the insurer's risk attitude on its formulary design. Consistent with the practice in which an insurer routinely updates its formulary in response to new drugs and new contracts, our model allows the insurer to determine the formulary tier for the new drug as well as adjust tiers of other drugs in the formulary.

Specifically, for a given realization of $\tilde{\rho}_J$, patients' *realized* health benefit associated with the insurer formulary decision z is

$$\tilde{B}(z) = -\ln(z_0) - (\rho_J - \tilde{\rho}_J)q_J \sum_{k=1}^{K} z_{Jk} + (\rho_J - \tilde{\rho}_J)\theta R_2 \sum_{k=1K} z_{Jk},$$

where $-\ln(z_0)$ represents patients' *expected* health benefits prior to the realization of $\tilde{\rho}_J$. The term $(\rho_J - \tilde{\rho}_J)q_J\sum_{k=1}^{K} z_{Jk}$ indicates the lost health benefits because of the proportion of patients not achieving the health benefit beyond the initial expectation, and $(\rho_J - \tilde{\rho}_J)\theta R_2\sum_{k=1}^{K} z_{Jk}$ indicates the manufacturer's rebate to the patients. Please see the online appendix for a detailed derivation of $\tilde{B}(z)$.

Correspondingly, the insurer's realized spending associated with formulary decision z for a given realization of $\tilde{\rho}_J$ is

$$\tilde{S}(z) = \sum_{k=1}^{K}(p_J - a_k)z_{Jk} - R_1(1-\tilde{\rho}_J)\sum_{k=1}^{K} z_{Jk} + \sum_{j=1}^{J-1}\sum_{k=1}^{K}(p_j - a_k)z_{jk},$$

where $R_1(1-\tilde{\rho}_J)\sum_{k=1}^{K} z_{Jk}$ represents the manufacturer's rebate to the insurer for patients who do not achieve the treatment benefit.

Let $V(u)$ be the insurer's utility given its spending-adjusted health benefits u. To capture the insurer's risk aversion, we assume $V(u)$ is a concave and increasing function of u. Therefore, the insurer determines the optimal formulary to maximize its expected utility from the spending-adjusted health benefits $V(u(\boldsymbol{x}))$ as follows:

$$\begin{aligned} \max_{z} \int V(w\tilde{B}(z) - \tilde{S}(z))f(\tilde{\rho}_J)d\tilde{\rho}_J \\ \text{s.t.} \quad z_0 + \sum_{j=1}^{J}\sum_{k=0}^{K} z_{jk} = 1, \\ \sum_{k=0}^{K}\frac{z_{jk}}{v_{jk}} = z_0, \ \forall j, \\ z_0 \geq 0, z_{jk} \geq 0, \ \forall j,k. \end{aligned} \tag{3}$$

Note that Problem (3) cannot be directly reduced to the base model because the insurer considers not only the expected value of patients' health benefits, but also the uncertainty of its realized value. Therefore, to solve Problem (3) directly, we have to assume certain functional forms of $V(u)$ and the pdf of $\tilde{\rho}_J, f(\tilde{\rho}_J)$. To circumvent this, we first characterize the structure of the optimal rebate scheme, which allows us to convert Problem (3) into the base model and then solve the corresponding insurer formulary without any functional assumptions of $V(u)$ and $f(\tilde{\rho}_J)$.

5.2. Optimal OBR Design Under Uncertain Drug Effectiveness

The manufacturer has two alternatives for selling its drug: first, it could provide a rebate to induce the insurer to place its drug on the formulary such that patients pay only the copayment; second, it could sell directly to patients without insurance coverage. The manufacturer decides the best alternative by comparing the expected profit of the two alternatives. The latter is rarely optimal for an expensive new drug, but we include it in our analysis for completeness.

If the manufacturer seeks the insurer coverage, it decides its optimal price and rebate scheme (i.e., p_J, R_1, and R_2) by weighing the trade-off among the expected cost of the rebate provided to the insurer and patients, the obtained formulary position, and the corresponding drug demand. Note that the manufacturer's price and rebate scheme can influence the demand of its drug in two ways: (1) by influencing the insurer's formulary decision and (2) through the patient rebate R_2. Given the manufacturer's price and rebate scheme (p_J, R_1, R_2), the insurer then decides whether to include the drug in its formulary and the corresponding formulary tier. Let $z_{Jk}(p_J, R_1, R_2)$ denote the demand of drug J if it is placed on the kth tier. The manufacturer solves its optimal price and rebate scheme to maximize its expected profit as follows:

$$\Pi = \max\left[\max_{p_{J0}} p_{J0} z_{J0}(p_{J0}),\ \max_{p_J, R_1, R_2} \int ((p_J - (R_1 + R_2)(1 - \tilde{\rho}_J)) \sum_{k=1}^{K} z_{Jk}(p_J, R_1, R_2) f(\tilde{\rho}_J) d\tilde{\rho}_J\right],$$

where $z_{J0}(p_{J0})$ represents the demand for drug J if it is excluded from the formulary and $p_{J0} z_{J0}(p_{J0})$ represents the corresponding manufacturer's profit under which the manufacturer incurs no rebate cost, but may have only a slender demand of $z_{J0}(p_{J0})$. If the drug is included in the formulary, its demand is then $\sum_{k=1}^{K} z_{Jk}(p_J, R_1, R_2)$ with a margin of $p_J - (R_1 + R_2)(1 - \tilde{\rho}_J)$, where $(R_1 + R_2)(1 - \tilde{\rho}_J)$ represents the manufacturer's total expected rebate to the insurer and patients.

If the insurer can adjust the formulary tiers of all existing drugs $j = 1 \ldots J - 1$ as a response to the manufacturer's direct cash price p_{J0}, we can use a grid search to find the optimal p_{J0}. Otherwise, Lemma 2 describes the optimal p_{J0} when fixing the formulary tiers of existing drugs.

Lemma 2. *Given the insurer formulary assignment for existing drug $j = 1 \ldots J - 1$, there exists a unique solution to the optimal p_{J0}, which can be solved through $\theta p_{J0}(1 - z_{J0}(p_{J0})) = 1$.*

However, solving the optimal wholesale price and rebate scheme directly is challenging because, for each combination of p_J, R_1, and R_2, we need to solve the corresponding insurer formulary through Problem (3), whose solution, as mentioned, depends on assumptions of the functional forms of $V(u)$ and $f(\tilde{\rho}_J)$. Instead, we first characterize the structure of the optimal rebate scheme.

Proposition 2. *Under the optimal OBR, for a patient not achieving the treatment benefit, the manufacturer should provide to the insurer and the patient a rebate with $\frac{1}{w} R_1 + \theta R_2$ completely offsetting the amount of lost health benefits q_J, that is, $\frac{1}{w} R_1 + \theta R_2 = q_J$.*

Proposition 2 implies that, given the rebate to patients R_2, the optimal rebate to the insurer is $R_1 = w(q_J - \theta R_2)$, indicating that the lost health benefits are partly offset by the manufacturer's direct rebate to patients, and the remainder is offset by the rebate to the insurer. As a result, the insurer's objective, the spending-adjusted health benefits, remains the same regardless of the realized effectiveness of the drug. Therefore, under such an OBR scheme, the insurer's risk is reduced to none. This is, in spirit, consistent with most OBR practices by which manufacturers fully refund its drug cost if patients do not obtain the intended effectiveness, thus eliminating risk exposure for the insurer. Therefore, under the optimal OBR scheme, the insurer decides its formulary as if there were no uncertainty in the drug's effectiveness regardless of the specific form of utility function $V(u)$ and the pdf of $\tilde{\rho}_J, f(\tilde{\rho}_J)$. Accordingly, Problem (3) can be reduced to the base model under such a rebate scheme (Please see the online appendix for details).

Although Proposition 2 prescribes the total rebate to the insurer and patients, the allocation of rebate between the insurer (R_1) and patients (R_2) is still at the manufacturer's discretion. Recall that there are two ways the manufacturer can affect demand: through an increased rebate to patients to directly increase demand or through an increased rebate to the insurer to improve its formulary position, which indirectly increases its demand. Intuitively, given the insurer formulary, the manufacturer should allocate more rebate to patients (R_2) because it directly increases patients' preference for the drug and, thus, increases the demand of the drug. Thus, if the manufacturer offsets the drug's uncertain effectiveness through patient rebate only, then $R_2 = q_J/\theta$. However, there is one caveat. According to Proposition 2, a higher R_2 leads to a lower R_1. To retain the formulary position of its drug, the manufacturer has to lower its wholesale price to the insurer (i.e., p_J) to offset the decrease in R_1. When patients are less sensitive toward monetary spending than the insurer is (i.e., $\theta < 1/w$), for example, for certain life-threatening diseases such as cancer, increasing the patient rebate may only increase drug demand incrementally but lower the wholesale price considerably.

In this case, the manufacturer should allocate more rebate to the insurer. We set c_J as the upper bound of R_2 because, otherwise, the manufacturer subsidizes patients more than their out-of-pocket copayment and undercuts the rebate to the insurer. Formally, Corollary 1 depicts the optimal allocation between R_1 and R_2.

Corollary 1. *Given the insurer formulary*

$$R_2 = \begin{cases} \min(c_J, q_J/\theta), & \text{if } (w\theta - 1)(1-\rho_J) \geq -\dfrac{\partial p_J}{\partial R_2}, \\ \min(\hat{R}_2^+, c_J, q_J/\theta) & \text{otherwise.} \end{cases}$$

Here, $\hat{R}_2^+ = \max(\hat{R}_2, 0)$ *and* $\hat{R}_2$ *solves the manufacturer's first-order condition* $w\theta(1-\rho_J) - (1-\rho_J) + \frac{\partial p_J}{\partial R_2} + (p_J - (R_1+R_2)(1-\rho_J))\theta(1-\rho_J+\vartheta\sigma_J)(1-d_J) = 0$, *and* $-\frac{\partial p_J}{\partial R_2}$ *represents the marginal rate of substitution between* p_J *and* R_2 *to retain the formulary position of the drug.*

Corollary 1 suggests that the allocation between R_1 and R_2 depends on patients' and insurer's weights on health benefits versus spending as well as the need to maintain an appropriate rebate to the insurer to retain the drug's formulary position. If patients are more sensitive to spending than the insurer (i.e., $\theta \gg 1/w$), then the manufacturer should refund patients' full copayment. Otherwise, the manufacturer balances the trade-off between allocating more rebate R_1 to the insurer to retain its formulary position and increasing drug demand through rebate R_2 to patients, which is captured by $\hat{R}_2$, derived from the manufacturer's first-order condition with respect to R_2.

It is noteworthy that the choice of R_1 and R_2 depends on the manufacturer's wholesale price to the insurer (p_J), another decision variable for the manufacturer. However, the relationship between R_1 and R_2 identified in Proposition 2 and Corollary 1 enables us to simplify the search for the optimal wholesale price: for each p_J, we calculate the corresponding R_1 and R_2 based on which we calculate the insurer formulary and the manufacturer's expected profit. The manufacturer should choose p_J that rewards the highest expected profit. A detailed description of the algorithm for solving the optimal (p_J, R_1, R_2) is in the online appendix.

Corollary 2. *Compared with the optimal non-OBR, under the optimal OBR, the manufacturer can charge a higher effective price and, at the same time, induce the insurer to place the drug on a formulary tier with a lower or at most the same copayment. Hence, the manufacturer can earn a higher expected profit, and the insurer incurs a higher expected spending under OBR.*

Corollary 2 shows that, with OBR, the manufacturer can induce the insurer to place the new drug on a more favorable tier and broaden the coverage of the drug, hence improving patients' access to the drug. Meanwhile, under the optimally designed OBR, the manufacturer raises its wholesale price by more than the amount of the expected rebate provided to the insurer such that the insurer pays a higher effective price. Hence, although OBR shields the insurer from the risk of uncertain drug effectiveness, it does not lower insurer spending as compared with non-OBR. This is because OBR transfers the risk from the insurer to the manufacturer but does not eliminate such risk from the supply chain. As a result, the insurer pays a risk premium to the manufacturer, which then earns a higher expected profit under OBR than under non-OBR.

Although our study focuses on OBR in the format of the ex post rebate, it can be shown that other formats of OBR, such as deferred payment that allows the insurer to pay a partial price up-front and defer additional payment based on the realized drug effectiveness, also have similar impact. We also compare OBR with non-OBR under a special case in which the insurer is risk neutral. (Please see the online appendix for details).

6. Model Calibration and Data Analysis for Hyperlipidemia Drugs

In this section, we apply our analytical model to drugs treating a common disease, hyperlipidemia, to illustrate the procedure for implementing our solution method and to provide additional insights that complement our analytical results. We focus on three questions. First, we quantify the impact of OBR on the insurer, manufacturer, and patients as compared with non-OBR. In particular, although our analytical results establish that OBR does not lower the insurer's expected spending, it is important to quantify the scale of risk premium that the insurer has to pay under OBR. Second, we demonstrate how the comparison between OBR and non-OBR depends on model parameters, such as the insurer's weight on health benefits against spending and its risk preference. Third, we explore numerically the difference between OBR with and without the patient rebate and examine the impact of including the patient rebate in OBR on different parties.

To provide realistic answers to these questions, we calibrate our analytical model based on data from 14 drugs treating hyperlipidemia. We choose hyperlipidemia because it has a clear and easily measurable metric of patient health outcome (i.e., blood cholesterol level), and OBR has been implemented for some hyperlipidemia drugs, such as Simvastatin (Carlson et al. 2009). Using the data on patients' purchase record of hyperlipidemia drugs, we estimate through an MNL model the demand of the 14 drugs based on the drug effectiveness (defined later) and copayments. The

estimation results are then used as input for our analytical model to optimize the design of OBR and insurer formulary. We then provide calibrated answers to the earlier questions through counterfactual analyses of a hyperlipidemia drug Simcor, which has evident uncertain effectiveness. Simcor was approved in 2008 for treating hyperlipidemia. However, several postmarket studies suggest that Simcor does not provide sufficient health benefits to patients in clinical practice. Consequently, the FDA withdrew the drug from the market in 2016 (Food and Drug Administration 2016).

6.1. Data Description

As mentioned, hyperlipidemia refers to an abnormal level of lipids (fat particles) in the blood, including elevated levels of LDL-C (the so-called bad cholesterol) and triglyceride (TG) and decreased levels of HDL-C (the so-called good cholesterol). Medical literature establishes that abnormal levels of lipids may result in heart attacks, strokes, and peripheral arterial diseases, all of which are the leading causes of mortality in the United States (Wadhera et al. 2016).

According to the 2015 Medical Expenditure Panel Survey (MEPS), 14 drugs are used by patients for treating hyperlipidemia, differing in their effectiveness, prices, and copayments, which depend on their formulary positions. In general, the demand for each of these 14 drugs depends on their effectiveness (measures of which are described in Section 7.2), patients' out-of-pocket copayments, and other random factors. Accordingly, we collate the information about each drug's treatment effects, adverse effects, and copayment from several data sources, including

- The 2015 MEPS: Patients' purchase records of hyperlipidemia drugs.
- Various medical literature: The treatment effects of different hyperlipidemia drugs.
- FDA adverse event reporting system: Adverse effect reports associated with hyperlipidemia drugs.

Online Table A3 summarizes the 14 drugs and their information detailed as follows. Specifically, the MEPS randomly surveys U.S. households for their prescription drug purchases and insurance coverage. This data set is considered quite representative of the general population in the United States (Medical Expenditure Panel Survey 2019). The MEPS data set records the drugs that patients purchase, the corresponding copayments, and the amount paid by their insurers. For our study, we extracted from the 2015 MEPS all purchase records of hyperlipidemia drugs. Online Table A3 summarizes the average patient copayment and total price paid by patients and their insurers. For example, as Online Table A3 shows, 8,342 patients purchased Atorvastatin (i.e., the generic counterpart of the blockbuster drug Lipitor), their average copayment for atorvastatin is \$4.6, and the total price paid by patients and insurers is \$37.95. All copayments and prices are normalized for a unit of 30-day supply. Typically, once the generic drug enters the market, it takes most, if not all, of the demand from the corresponding brand drug. Hence, the only generic–brand pair present in the MEPS records is atorvastatin/Lipitor, and the demand for Lipitor is about 2% of that of atorvastatin. Note that the copayment for the same drug varies across patients because they have different insurance plans, which may have different formularies. Such variation in copayments helps us identify patients' sensitivity to their copayments.

From various medical literature, including Sabatine et al. (2015), Mohiuddin et al. (2009), Birjmohun et al. (2005), and Hou and Goldberg (2009), we collected information on the treatment effects of different hyperlipidemia drugs, summarized in Online Table A3. The treatment effects are evaluated in three standard metrics: percentage of LDL-C decrease, percentage of triglyceride decrease, and percentage of HDL-C increase as compared with the corresponding baseline level (i.e., the lipid level of the untreated patient group). For example, atorvastatin can lower LDL-C on average by 43%, lower triglyceride by 28%, and increase HDL-C by 7%.

In addition to the treatment effects in these metrics, potential adverse effects are another critical consideration in patients' choice of drugs. To construct a measure of potential adverse effects associated with each drug, we obtained the adverse effect reports of these hyperlipidemia drugs from the 2015 FDA adverse event reporting system data set. For example, statins, a class of drugs including Lovastatin, Simvastatin, Pravastatin, atorvastatin, and Crestor (Simcor and Vytorin are also considered statins but mixed with other medical ingredients) may cause severe adverse effects, such as muscle pain and cramps. We define $AdvEff_j$ of drug j as the total number of reported incidences of adverse effects associated with the drug divided by its total quantity sold from the MEPS data set to measure its risk of adverse effects. The resulting risk measure of each drug is shown in the last column of Online Table A3.

6.2. Demand Estimation and Optimal Formulary

Recall that, in our analytical model, patients' risk-adjusted utility from a drug $u_{ij} = q_j\rho_j - \vartheta q_j\sigma_j - \theta c_{ij} + \epsilon_{ij}$ consists of the expected health benefit, the risk of treatment failure, the cost of copayment, and the idiosyncratic shock. In our application, we construct and estimate an MNL model to quantify the impact of the expected health benefit (measured by the three treatment effect metrics and adverse effects), risk of treatment failure, and copayment on the demand of drugs. However, the treatment outcome data for the 14 drugs

is not readily available. Therefore, for our numerical analysis, we normalize the risk of treatment failure for the existing drugs to zero and include a binary indicator to capture the risk of treatment failure for the new drug, Simcor, which, as mentioned, has more evident risk of treatment failure than the existing drugs.

Specifically, under the MNL model, patients choose the drug that maximizes their expected utility, which is given by

$$u_{ij} = \gamma_1 LDLC_j + \gamma_2 HDLC_j + \gamma_3 TG_j + \gamma_4 AdvEff_j + \gamma_5 Simcor_j - \theta c_{ij} + \epsilon_{ij},$$

where $LDLC_j$, $HDLC_j$, and TG_j represent the three treatment effect metrics: improvement of LDL-C, HDL-C, and triglyceride associated with drug j and $AdvEff_j$ captures the impact of a drug's potential adverse effects on its demand. $Simcor_j$ is a dummy variable indicating whether the drug is Simcor, and γ_5 is the coefficient signifying the additional disutility resulting from the risk of treatment failure for Simcor compared with existing drugs. c_{ij} captures the impact of copayment of a drug on its demand, and ϵ_{ij} captures patients' idiosyncratic preferences. Note that we do not include the no-purchase option because patients with abnormal LDL-C level are usually recommended to undertake hyperlipidemia drugs to prevent potential heart attacks and strokes, and the MEPS data set only includes the purchase records of hyperlipidemia drugs.

To estimate patients' sensitivity to copayment (θ), we need data on the copayments of all alternative drugs for each patient. However, the MEPS data set only provides the copayment of the drug purchased by a patient. Fortunately, the MEPS data set also provides patients' insurance types, for example, private, Medicare, Medicaid, Tricare, and their combinations or being uninsured. Because the formulary from the same type of insurers is generally similar, we approximate the copayment of drugs that are not purchased by a patient through the average copayment paid by all other patients with the same type of insurers. In addition, because patients with the same type of insurers may share similar characteristics (e.g., patients with Medicare all have age above 65), their idiosyncratic preference ϵ_{ij} may be correlated. Thus, we clustered the standard errors by insurer types to account for the potential correlation of ϵ_{ij} among patients with the same insurer types.

Table 1. MNL Model Estimates for Demand Function of Hyperlipidemia Drugs

Variables	(1)	(2)	(3)
Copayment	−0.0410***	−0.0416***	−0.0415***
	(0.0083)	(0.0089)	(0.0088)
LDL-C	7.014***	7.006***	6.948***
	(0.195)	(0.189)	(0.271)
TG	2.690***	2.601***	2.585**
	(0.758)	(0.793)	(0.814)
AdvEff	−1.146***	−1.100**	−1.047*
	(0.397)	(0.458)	(0.510)
HDL-C		0.798	1.267
		(1.268)	(1.992)
Simcor			−0.338
			(0.696)
Pseudo R^2, %	20.23	20.24	20.25
Observations	309,183	309,183	309,183

Note. Cluster (on insurer type) standard errors in parentheses.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Table 1 shows the estimation results of the MNL model. We include the variables sequentially to show the robustness of the estimates, and the results are consistent across these model specifications. Column (3) of Table 1 represents the results of the full model. We refer to the results from column (3) for our statistical inferences and counterfactual analyses hereafter.

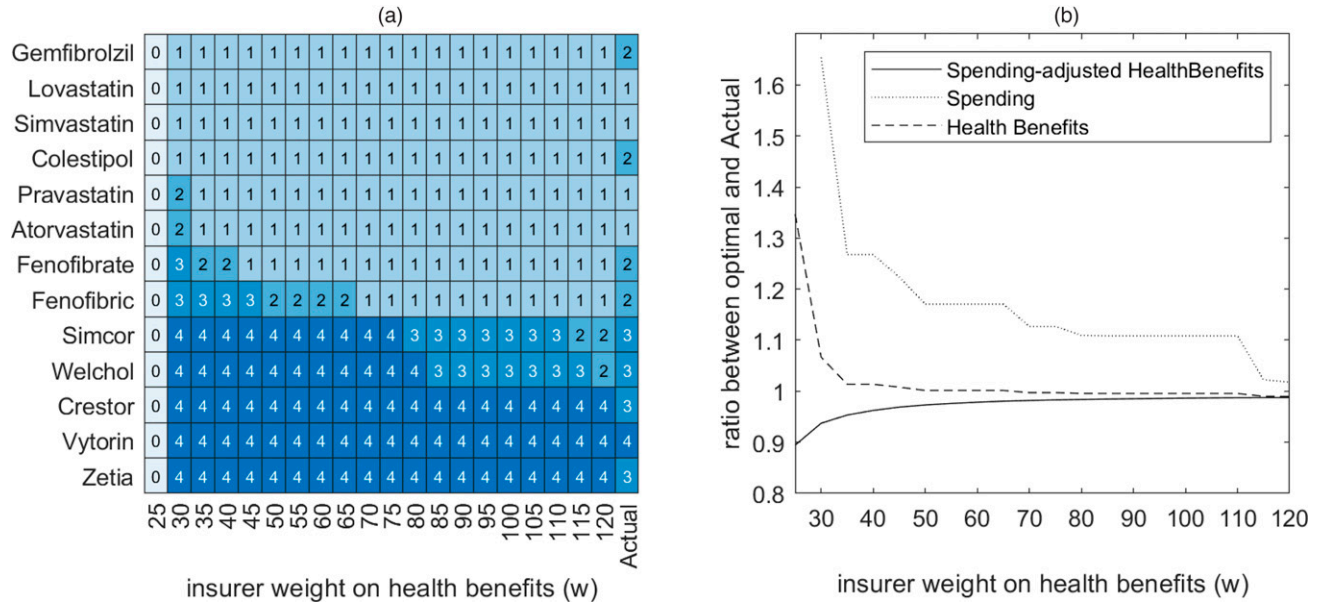
Column (3) of Table 1 highlights a few observations. First, among the three treatment effect metrics, the ability of a drug in lowering LDL-C has the strongest impact on the demand of a drug. This is consistent with the current treatment guideline for hyperlipidemia, which sets the primary goal of hyperlipidemia treatment as LDL-C reduction (Hou and Goldberg 2009). Second, the potential adverse effects of a drug also significantly (with $p < 0.05$) influence its demand. This is consistent with the fact that patients may discontinue their hyperlipidemia drugs because of potential serious adverse effects. Third, the coefficient of copayment is significant (with $p < 0.001$), suggesting that the insurer can leverage its formulary to influence the demand of different drugs. Note that the coefficient of *Simcor* is not significant, potentially because the risk of uncertain treatment outcome has been captured by some of the drug attributes such as *AdvEff*.

Using these estimation results as input for our model, we can optimize the insurer formulary accordingly. Recall, as Problem (2) suggests, the insurer decides the formulary to balance patients' health benefits and its spending. The patients' risk-adjusted health benefits can be calibrated as

$$q_j\rho_j - \vartheta q_j\sigma_j = \gamma_1 LDLC_j + \gamma_2 HDLC_j + \gamma_3 TG_j + \gamma_4 AdvEff_j + \gamma_5 Simcor_j. \quad (4)$$

Note that q_j and ρ_j cannot be separately identified because we do not have the data on patients' treatment outcome. Nevertheless, we only need to approximate the drug-specific risk-adjusted health benefit $q_j\rho_j - \vartheta q_j\sigma_j$ for the purpose of our analysis. The insurer spending depends on the demand of different drugs, which can be calibrated by the MNL model with the estimated risk-adjusted health benefit $q_j\rho_j - \vartheta q_j\sigma_j$ and

Figure 1. (Color online) Cigna's Optimal Formulary for Hyperlipidemia Under Different Weight of Health Benefits w and Cigna's Actual Formulary in 2015



patients' sensitivity to copayment θ. Applying these estimates in Problem (2), we next solve the optimal insurer formulary.

Take the insurer, Cigna, as an example. In 2015, Cigna adopted a four-tier formulary with copayments for tier 1–4 drugs as \$0, \$10, \$45, and \$95, respectively. Because the estimation of the insurer's weight on patients' health benefits versus spending w is not straightforward, we vary w relative to $1/\theta$, the patients' weight between health benefits and copayment spending, to examine the impact of w.

Figure 1(a) demonstrates Cigna's optimal formulary solved through Problem (2) (i.e., base model without uncertain effectiveness) under different w and its actual formulary for hyperlipidemia in 2015. The y-axis of Figure 1(a) indicates the 13 drugs (excluding Lipitor), ordered by their effective prices, as seen in Online Table A3. Note that Lipitor is not included here as well as the following analysis because Cigna excludes Lipitor from its formulary because of the existence of the generic counterpart atorvastatin. The x-axis of Figure 1(a) represents different values of w, and the last column represents Cigna's actual formulary. As Figure 1(a) shows, the optimal formularies under different w, as demonstrated in Proposition 1, follows a nested structure in which drugs with lower effective prices are assigned to a lower tier (with a lower copayment), followed by drugs with higher effective prices assigned to a higher tier (with a higher copayment). As the insurer puts more weight on patients' health benefits over its monetary spending, it places more drugs on tiers with lower copayments. In an extreme case in which the insurer's weight on patients' health benefits is really small, that is, $w = 25$, it excludes all drugs from its formulary.

Figure 1(a) illustrates that Cigna's actual formulary is approximately ordered by the effective prices of drugs. In Figure 1(b), we benchmark Cigna's actual formulary with the optimal formulary in terms of insurer spending, patients' health benefits, and the spending-adjusted health benefits. The spending-adjusted health benefits associated with Cigna's actual formulary is comparatively close to that of the optimal formulary, achieving mostly more than 90% of the optimal spending-adjusted health benefits under the different w we tried. This result highlights that the "effective price ordering" principle works well in designing insurer formulary over a broad range of w.

The example illustrates how our analytical model, together with data, can be used to guide the design of insurer formulary. Next, we assess the impact of OBR on the insurer, manufacturer, and patients through counterfactual analyses on the case of Simcor, which is featured with uncertain effectiveness.

6.3. Counterfactual Analysis with Simcor

As mentioned, Simcor has evident uncertain effectiveness, which eventually led to its withdrawal from the market. Under the conventional non-OBR scheme, the insurer spends a great amount of money on Simcor but likely does not obtain the anticipated health benefit. In this section, we explore how OBR, if used for Simcor, would have impacted the insurer, the manufacturer, and patients.

6.3.1. Model Parameters. We first parameterize patients' risk preference toward the uncertain drug effectiveness. According to Erdem and Keane (1996), for typical consumer products with uncertain quality, consumers' risk coefficient (i.e., ϑ in our model) is between 0.14 and 7.11. We choose $\vartheta = 1$ as an illustrative example for this analysis. We model the insurer's risk aversion through a smooth function $V(u) = -e^{-Au}$ with a constant risk-aversion coefficient A. The risk-aversion coefficient A is insurer-specific and not readily available. Thus, we set $A = 0.3$ as an illustrative example and later vary A for sensitivity analysis. The insurer's weight on health benefits w is also insurer-specific. We choose $w = 115$ for the analysis because, as Figure 1 shows, it leads to a formulary closest to Cigna's actual formulary, indicating that the true w might be in its vicinity. We later replicate the analysis for different w to examine its impact. To capture the uncertain effectiveness of Simcor, we assume that the proportion of patients who take Simcor and achieve the treatment target $\tilde{\rho}_J$ follows a uniform distribution between zero and one, that is, $\tilde{\rho}_J \sim U(0,1)$.

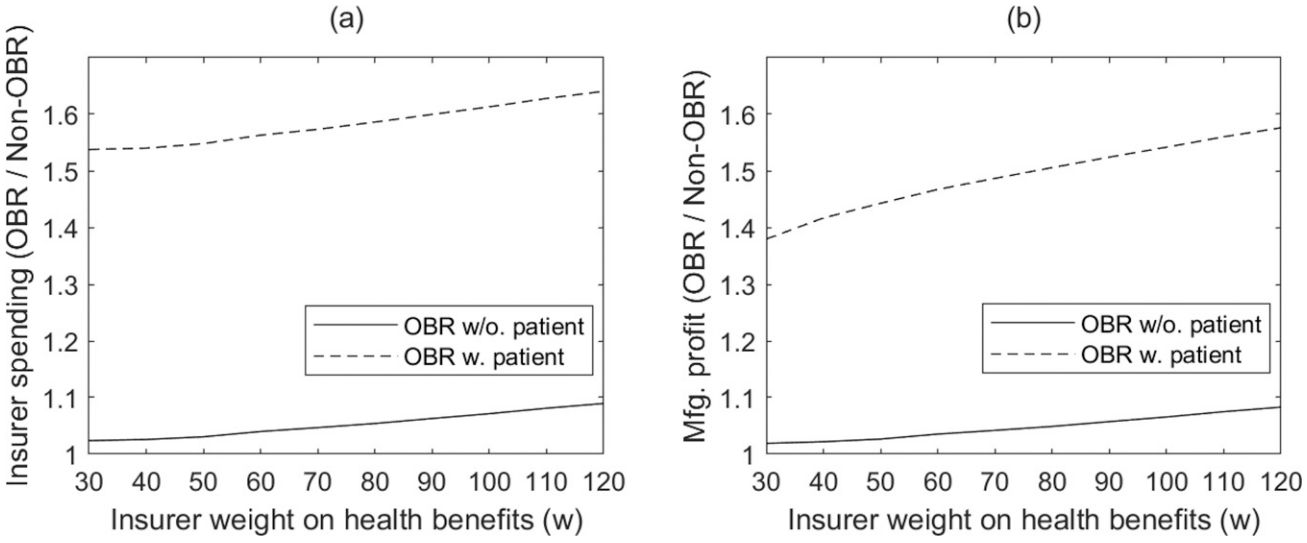
6.3.2. Main Results. If the manufacturer sells its drug directly to patients without insurance coverage, it sets a cash price of \$24.8 and earns a profit of \$0.5616, much lower than if the manufacturer sets a rebate scheme to induce insurance coverage. Table 2 highlights the three optimal rebate schemes (i.e., non-OBR, OBR without/with patient rebate) offered by the manufacturer and the corresponding optimal insurer formulary. As Table 2 shows, under non-OBR, the manufacturer decides a wholesale price of \$135.08 to the insurer. Accordingly, the insurer places Simcor on tier 2 of the formulary with a copayment of \$10. Hence, patients pay a \$10 copayment, and the insurer pays the remaining \$125.08 regardless of the realized health outcomes among patients. In contrast, under OBR without the patient rebate, the manufacturer provides a rebate of \$154.04 to the insurer for any patient who does not achieve the treatment target but, meanwhile, raises the wholesale price to \$223.46. Accordingly, the insurer still places Simcor on tier 2 of the formulary. This supports the speculation that the manufacturer may inflate the drug price under OBR (Thomas and Ornstein 2017). Under such an OBR scheme, the expected effective price of Simcor becomes \$146.44 ($= 223.46 - 154 * 0.5$), and patients pay a \$10 copayment, and the insurer pays the remaining \$136.44, which is 9.1% higher than that under non-OBR. This is consistent with Corollary 2 that the insurer needs to pay a risk premium under OBR.

The manufacturer could also include a patient rebate in OBR, under which the manufacturer refunds the full copayment \$10 to patients and \$106.20 to the insurer for any patient who does not achieve the treatment target. Because of a lower rebate to the insurer, to retain its formulary position, the manufacturer sets a lower wholesale price of \$200.49 (a smaller price inflation as compared with \$223.46) to offset the reduced outcome-based rebate, R_1, to the insurer. Accordingly, the expected effective price of the drug becomes \$147.39 ($= 200.49 - 106.2 * 0.5$), and patients pay a \$10 copayment, and the insurer pays the remaining \$137.39. Because of the direct patient rebate, the patients' expected out-of-pocket payment is, thus, \$5 ($= 10 - 10 * 0.5$). Thus, including a patient rebate in OBR has a similar effect as a manufacturer's coupon

Table 2. Cigna's Formulary for Hyperlipidemia Treatments with Simcor

				OBR	
Drugs	Price	Effectiveness	Non-OBR	Without patient	With patient
Gemfibrozil	8.36	0.810	1	1	1
Lovastatin	11.81	1.782	1	1	1
Simvastatin	17.27	2.903	1	1	1
Colestipol	21.41	0.333	1	1	1
Pravastatin	32.25	1.599	1	1	1
Atorvastatin	37.95	2.936	1	1	1
Fenofibrate	51.48	1.289	1	1	1
Fenofibric	75.80	1.087	1	1	1
Simcor	**143.77**	**1.002**	**2**	**2**	**2**
Welchol	149.36	−1.838	3	3	3
Crestor	200.57	2.792	4	4	4
Vytorin	211.62	2.152	4	4	4
Zetia	226.97	0.031	4	4	4
(p_J, R_1, R_2)(\$)			(135.08, 0, 0)	(223.46, 154.04, 0)	(200.49, 106.20, 10)
Effective price, \$			135.08	146.44	147.39
Insurer pay, \$			125.08	136.44	137.39
Patient pay, \$			10	10	5
Market share, %			2.97	2.97	4.43
Mfg. Profit, \$			4.0089	4.3257	6.2783

Figure 2. The Risk Premium and Manufacturer Profit Under Different Weight of Health Benefits w



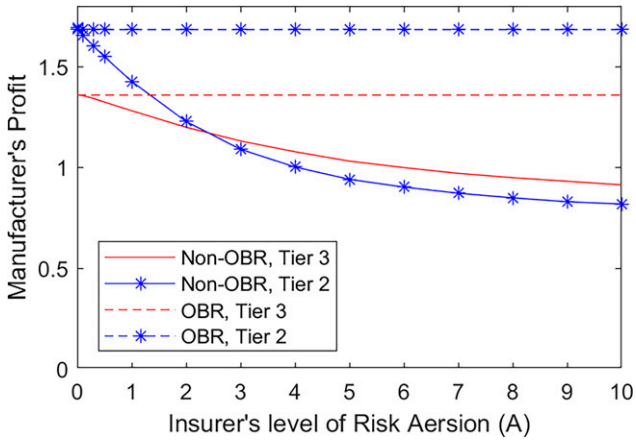
(except that the rebate is only redeemable if the drug is not effective): it could lower patients' out-of-pocket payment, thus increasing the market share of the drug without changing the insurer's formulary, and improve the manufacturer's profit (King et al. 2019). As for the insurer, providing a rebate to patients further escalates its spending because of the increased demand for the drug.

Our model allows the insurer to adjust formulary tiers of the existing alternative drugs upon the entry of the new drug. Such flexibility is especially important when the number of alternative drugs is small. See a numerical example in the online appendix, which indicates that the insurer may have to pay a 38.5% higher wholesale price if the insurer cannot adjust the formulary of existing drugs.

6.3.3. Sensitivity Analysis on *w*. As Figure 1 shows, the insurer's weight on health benefits w has a significant impact on how an insurer decides its formulary. Thus, we perform the same analysis to compare non-OBR and OBR with different w. Figure 2 highlights the ratio of the insurer's spending and manufacturer's profit for Simcor under OBR to that under non-OBR. Under OBR without the patient rebate, the insurer's spending increases by 2.4% ~ 8.9% compared with non-OBR because of the inflated price as a risk premium. This risk premium increases as w increases, that is, the insurer puts more weight on patients' health benefits. In contrast, under OBR with the patient rebate, the insurer's spending increases by 53.8% ~ 64.1% compared with non-OBR because of the inflated drug price and, more importantly, the increased drug demand. Additionally, such spending increase is prominent even when w is low.

Reciprocally, as Figure 2(b) shows, the manufacturer can exploit a higher profit by providing OBR (especially OBR with a patient rebate) in lieu of non-OBR, and the advantage of OBR is more prominent when the insurer's weight on health benefits is higher. Note that, under OBR with the patient rebate, the increase of manufacturer profit is not as high as the increase of insurer spending because the manufacturer has to provide not only an additional incentive (in the form of a lower wholesale price) to the insurer to maintain its formulary position, but also an outcome-based rebate to both the insurer and patients.

Figure 3. (Color online) Manufacturer's Rebate and Corresponding Formulary Position Under Different Risk Aversion A



6.3.4. Sensitivity Analysis on *A*. Because the insurer's level of risk aversion A plays a critical role in the comparison between OBR and non-OBR, we conduct the same analysis with different values of A. It is noteworthy that the copayment for drugs on tiers 2 and 3 are \$10 and \$45, respectively. Given such a substantial difference in copayment, when θ is comparatively large, it is better for the manufacturer to induce the insurer to place its drugs on the tier 2 formulary for higher demand regardless of OBR or non-OBR. To highlight the distinct formulary position of the drug under OBR and non-OBR, in this sensitivity analysis, we choose a smaller $\theta = 0.015$ (i.e., the lower bound of the 99.5% confidence interval of θ). Figure 3 summarizes the resulting manufacturer profit as a function of A if the drug were placed on tier 2 and 3. We do not include OBR with patient rebate in Figure 3 in order to focus on the impact of the insurer's risk preference.

As demonstrated in Proposition 3 and illustrated in Figure 3, the optimal non-OBR is equivalent to the optimal OBR when the insurer is risk neutral (i.e., $A = 0$). In this case, the manufacturer can provide either non-OBR or OBR to induce the insurer to place its drug on tier 2. However, under non-OBR, as A increases, the manufacturer has to set a lower wholesale price to keep such a favorable formulary tier to an extent that such a favorable tier becomes unappealing because of the much reduced wholesale price. As a result, when A is sufficiently large, the manufacturer chooses to induce the insurer to place its drug on tier 3 under non-OBR. Accordingly, patients have to pay a higher copayment for the drug. In contrast, as Figure 3 illustrates, with OBR, the manufacturer can induce the insurer to place its drug on tier 2 regardless of the insurer's level of risk aversion. This is because the optimal OBR scheme provided by the manufacturer takes away the insurer's risk of uncertain drug effectiveness. Thus, compared with non-OBR, OBR indeed can induce the insurer to place a new drug on the same, if not better, tier of the formulary, improving patients' access to the drug. Such an advantage of OBR over non-OBR becomes more evident with a more risk-averse insurer. In addition, as the insurer becomes more risk averse, the risk premium the insurer has to pay under OBR increases (given that the drug's formulary tier stays unchanged).

7. Conclusions

Exorbitant prices for new drugs in combination with the typical uncertainty in new drugs' effectiveness in clinical practice result in substantial risks for a payer/insurer. This has driven the pharmaceutical industry to explore new reimbursement schemes (Reddy 2017). OBR, proposed by the manufacturer to tie an insurers' payment to the realized drug effectiveness, appears promising for lowering insurer spending because manufacturers promise to refund insurers (and possibly patients) if their drugs fail to deliver the expected health benefits. However, the true impact of OBR is under much debate and depends particularly on the design of OBR. Our study sheds light on the optimal design of OBR and the debate around OBR, considering key trade-offs and elements not covered in prior literature, such as the insurer's formulary design and the optimal split of rebates between patients and the insurer, among others.

We develop a Stackelberg game under which a manufacturer designs a rebate scheme for its drug, either non-OBR or OBR, considering the trade-off between a favorable formulary position and the rebate provided. The insurer subsequently determines its formulary for the drug as well as other alternative drugs within the same disease category considering the trade-off between its spending and patient health benefits. We also include the insurer's and patients' risk attitudes, which are not investigated by previous literature. Despite the distinct complexities, we solve the optimal insurer formulary and establish the structure of the optimal OBR scheme and eventually OBR's impact on different parties. We further calibrate our model with data of 14 drugs used to treat the disease hyperlipidemia and conduct a counterfactual analysis to complement our analytical results.

We show that the optimal OBR reduces the insurer's risk but not the insurer's spending because of the manufacturer's inflated wholesale price as a price premium. Our counterfactual analyses for the drug Simcor show that the insurer would need to pay 2.4% ~ 8.9% more under OBR compared with non-OBR. The more risk averse the insurer is, or the more weight the insurer puts on patients' health benefits over its own spending, the higher risk premium the manufacturer could exploit. In addition, OBR can induce the insurer to place a new drug on a better formulary position, hence improving patient access to new drugs. Such a benefit of OBR becomes more evident with a more risk-averse insurer.

The manufacturer may also provide an outcome-based rebate to patients in addition to the insurer. Such a rebate to patients assures the value of patients' out-of-pocket spending, thus increasing the demand of the drug. To some extent, it works as a manufacturer's coupon does (except that it is only redeemable if the drug is not effective). In allocating between the outcome-based rebate to the insurer (R_1) and the patients (R_2), the manufacturer considers the direct impact on its demand from R_2 and R_1 through the insurer's formulary. The increased drug demand resulting from the outcome-based rebate to patients further increases insurer spending. In particular, our counterfactual analyses show that insurer spending increases

by 53.8% ~ 64.1% under outcome-based rebates to both insurer and patients as compared with that under non-OBR. Therefore, including a patient rebate in OBR would potentially encounter resistance from the insurer. Thus, the manufacturer may have to lead the implementation of any patient refund scheme because of the lack of incentive from the insurer.

Our study captures two sources for uncertain treatment outcomes: uncertainty in drug effectiveness and patient heterogeneity. For the insurer, because the latter can be averaged out because of risk pooling, the former is particularly important, especially when the insurer is risk averse. We assume in this paper that the insurers are risk averse and the manufacturer is risk neutral. This corresponds to the fact that the manufacturer already incurred a huge R&D cost developing the drug, and its main objective for a newly approved drug is to generate the highest demand and profit. Our results still hold qualitatively even with a risk-averse manufacturer as long as it is not as risk averse as the insurer. That is, although the optimal OBR may not have the simple form as we currently have, the manufacturer will still offer to partially offset the insurer's risk, and the discussion on how to allocate the rebate between the insurer and patients is still valid but with much more technical complexity. We further note that we do not consider the administrative cost associated with OBR in this study. Instead, our goal is to establish the benefits of OBR, which could then be benchmarked with its administrative cost.

Our study uses the MNL model to approximate the aggregated demand of drugs. This inevitably omits certain nuances in patients' choice of prescription drugs. For example, for an expensive new drug, an insurer may impose step therapy and prior authorization (insurer approves the use of the new drug only if other cheaper alternative drugs failed). Such additional constraints may curb the demand of the new drug and, thus, affect the manufacturer's pricing scheme. This can be an interesting direction for future study.

A recent survey of 14 U.S. payers and five European payers suggests that 26% of the payers perceive the value of OBR to be spending reduction, and 13% perceive the value to be an increased rebate from the manufacturer (Nazareth et al. 2017). However, our results caution insurers/payers who seek OBR as the path to spending reduction. As Harvard Pilgrim, an insurer who has signed several OBR schemes, stated, "What we have done so far is not going to solve the high drug price crisis" (Loftus 2017). As shown in this paper, OBR shifts the risk from the insurer to the manufacturer, who, in turn, uses the design of its pricing scheme to compensate for its risk. Although OBR can potentially induce a better formulary position for a new drug, hence improving access to the drug, it does not curb insurers' spending. Looking ahead, it is imperative to reform the payment schemes for new drugs because of the continually increasing prices. However, the current format of OBR does not seem to be the ultimate solution. More promising payment innovations should consider ways to reduce the uncertainty of drug effectiveness in clinical practices through, for example, enhanced medication adherence and better treatment recommendations.

Acknowledgments

The authors thank the review and editorial team for constructive and useful comments.

References

Adida E (2021) Outcome-based pricing new pharmaceuticals via rebates. *Management Sci.* 67(2):892–913.

Anderson SP, De Palma A, Thisse JF (1992) *Discrete Choice Theory of Product Differentiation* (MIT Press, Cambridge, MA).

Antonanzas F, Juarez-Castello C, Rodriguez-Ibeas R (2011) Should health authorities offer risk-sharing contracts to pharmaceutical firms? A theoretical approach. *Health Econ. Policy Law* 6(3):391–403.

Birjmohun RS, Hutten BA, Kastelein JJ, Stroes ES (2005) Efficacy and safety of high-density lipoprotein cholesterol-increasing compounds: A meta-analysis of randomized controlled trials. *J. Amer. College Cardiology* 45(2):185–197.

Brown MT, Bussell JK (2011) Medication adherence: WHO cares? *Mayo Clinic Proc.* 86(4):304–314.

Carlson J, Garrison LP Jr, Sullivan SD (2009) Paying for outcomes: Innovative coverage and reimbursement schemes for pharmaceuticals. *J. Managed Care Pharmacy* 15(8):683–687.

Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ and Veenstra DL (2010) Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. *Health Policy* 96(3):179–190.

Centers for Medicare and Medicaid Services (2016) CMS proposes to test new Medicare Part B prescription drug models to improve quality of care and deliver better value for Medicare beneficiaries. Accessed April 2021, https://www.cms.gov/newsroom/press-releases/cms-proposes-test-new-medicare-part-b-prescription-drug-models-improve-quality-care-and-deliver.

Centers for Medicare and Medicaid Services (2017a) Innovative treatments call for innovative payment models and arrangements. Accessed April 2021, https://www.cms.gov/newsroom/press-releases/cms-innovative-treatments-call-innovative-payment-models-and-arrangements.

Centers for Medicare and Medicaid Services (2017b) National Health Expenditures 2017 Highlights. Accessed April 2018, https://www.cms.gov/newsroom/press-releases/cms-office-actuary-releases-2017-national-health-expenditures.

Ching AT, Lim H (2020) A structural model of correlated learning and late-mover advantages: The case of statins. *Management Sci.* 66(3):1095–1123.

Claxton G, Rae M, Panchal N, Lundy J, Damico A, Osei-Anto A, Kenward K, Whitmore H, Pickreign J (2011) *Employer Health Benefits: 2011 Annual Survey* (Henry J. Kaiser Family Foundation, Menlo Park, CA; Health Research & Educational Trust, Chicago). Accessed December 2021, https://www.msbo.org/sites/default/files/8225.pdf.

Coulton L, Annemans L, Javier J, Brown R, Keskinaslan A (2010) Risk-sharing schemes worldwide: A landscape analysis of health outcomes-based reimbursement agreements. *ISPOR 4th Asia-Pacific Conf.*, Abstracts, Phuket, Thailand.

Davis J, Gallego G, Topaloglu H (2013) Assortment planning under the multinomial logit model with totally unimodular constraint structures. Working paper, Columbia University, New York, NY.

Erdem T, Keane MP (1996) Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* 15(1):1–20.

FiercePharma (2016a) Cigna inks results-based deals on pricey Amgen, Sanofi PCSK9 meds. Accessed April 2018, https://www.fiercepharma.com/pharma/cigna-inks-results-based-deals-pricey-amgen-sanofi-pcsk9-meds.

FiercePharma (2016b) Lilly's Trulicity joins pay-for-performance trend with Harvard Pilgrim deal. Accessed April 2018, https://www.fiercepharma.com/pharma/lilly-s-trulicity-joins-pay-for-performance-trend-harvard-pilgrim-deal.

Food and Drug Administration (2016) Withdrawal of approval of new drug applications for Advicor and Simcor. Accessed April 2018, https://www.federalregister.gov/documents/2016/04/18/2016-08894/abbvie-inc-withdrawal-of-approval-of-new-drug-applications-for-advicor-and-simcor.

Garrison LP, Towse A, Briggs A, De Pouvourville G, Grueger J, Mohr PE, Severens JL, Siviero P, Sleeper M (2013) Performance-based risk-sharing arrangements—good practices for design, implementation, and evaluation. *Value Health* 16(5):703–719.

Gupta D, Mehrotra M (2015) Bundled payments for healthcare services: Proposer selection and information sharing. *Oper. Res.* 63 (4):772–788.

Harvard Pilgrim (2017) Harvard Pilgrim signs second groundbreaking contract with Amgen for Repatha. Accessed September 2019, https://www.harvardpilgrim.org/public/news-detail?nt=HPH_News_C&nid=1471912937208.

Hoadley J, Hargrave E, Merrell K (2011) Medicare Part D formularies, 2006–2011: Update to chartbook. Report, MedPAC, Washington, DC. Accessed April 2018, http://67.59.137.244/documents/Aug11_PartDFormulariesChartbook_Updated_CONTRACTOR_ss.pdf.

Hou R, Goldberg AC (2009) Lowering low-density lipoprotein cholesterol: Statins, ezetimibe, bile acid sequestrants, and combinations: Comparative efficacy and safety. *Endocrinology Metabolism Clinics* 38(1):79–97.

IQVIA (2019) Medicine use and spending in the U.S. Accessed December 2021, https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/medicine-use-and-spending-in-the-us—a-review-of-2018-outlook-to-2023.pdf?_=1640885282192.

King GJ, Chao X, Duenyas I (2019) Who benefits when prescription drug manufacturers offer copay coupons? *Management Sci.* 65 (8):3758–3775.

Kouvelis P, Xiao Y, Yang N (2015) PBM competition in pharmaceutical supply chain: Formulary design and drug pricing. *Manufacturing Service Oper. Management* 17(4):511–526.

Kouvelis P, Xiao Y, Yang N (2018) Drug pricing for competing pharmaceutical manufacturers distributing through a common PBM. *Production Oper. Management* 27(8):1399–1419.

Kunreuther H, Hogarth R, Meszaros J (1993) Insurer ambiguity and market failure. *J. Risk Uncertainty* 7(1):71–87.

Loftus P (2017) Drug companies tie costs to outcomes. *Wall Street Journal* (September 12, 2017), accessed December 2021, https://www.wsj.com/articles/drug-companies-tie-costs-to-outcomes-1505268421.

Mahjoub R, Ødegaard F, Zaric GS (2018) Evaluation of a pharmaceutical risk-sharing agreement when patients are screened for the probability of success. *Health Econom.* 27(1):15–25.

Medical Expenditure Panel Survey (2019) Survey background. Accessed April 2021, https://meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp.

Mohiuddin SM, Pepine CJ, Kelly MT, Buttler SM, Setze CM, Sleep DJ, Stolzenbach JC (2009) Efficacy and safety of ABT-335 (fenofibric acid) in combination with simvastatin in patients with mixed dyslipidemia: A phase 3, randomized, controlled study. *Amer. Heart J.* 157(1):195–203.

Mukherjee S (2017) Is $475,000 too high a price for Novartis's historic cancer gene therapy? *Fortune Online* (August 31). Accessed December 2021, http://fortune.com/2017/08/31/novartis-kymriah-car-t-cms-price/.

Nallamothu BK, Hayward RA, Bates ER (2008) Beyond the randomized clinical trial: The role of effectiveness studies in evaluating cardiovascular therapies. *Circulation* 118(12):1294–1303.

Navarria A, Drago V, Gozzo L, Longo L, Mansueto S, Pignataro G, Drago F (2015) Do the current performance-based schemes in Italy really work? "Success Fee": A novel measure for cost-containment of drug expenditure. *Value Health* 18(1):131–136.

Nazareth T, Ko JJ, Sasane R, Frois C, Carpenter S, Demean S, Vegesna A, Wu E, Navarro RP (2017) Outcomes-based contracting experience: Research findings from US and European stakeholders. *J. Managed Care Specialty Pharmacy* 23(10):1018–1026.

Nordqvist C (2012) One year on herceptin for breast cancer ideal. Accessed April 2021, https://www.medicalnewstoday.com/articles/250912.

Reddy S (2017) When cholesterol drugs cost $14,000, an insurance tug-of-war. *The Wall Street Journal Online* (June 19), accessed April 2021, https://www.wsj.com/articles/when-cholesterol-drugs-cost-14-000-an-insurance-tug-of-war-1497889667.

Sabatine MS, Giugliano RP, Wiviott SD, Raal FJ, Blom DJ, Robinson J, Ballantyne CM, et al. (2015) Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *New England J. Medicine* 372(16):1500–1509.

So KC, Tang CS (2000) Modeling the impact of an outcome-oriented reimbursement policy on clinic, patients, and pharmaceutical firms. *Management Sci.* 46(7):875–892.

Talluri K, Van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* 50(1):15–33.

Thomas K, Ornstein C (2017) Considering the side effects of drugmakers' money-back guarantees. *The New York Times Online* (July 10), accessed April 2018, https://www.nytimes.com/2017/07/10/health/prescription-drugs-cost.html.

Thomas MU, Rao SS (1999) Warranty economic decision models: A summary and some suggested directions for future research. *Oper. Res.* 47(6):807–820.

Travers J, Marsh S, Williams M, Weatherall M, Caldwell B, Shirtcliffe P, Aldington S, Beasley R (2007) External validity of randomised controlled trials in asthma: To whom do the results of the trials apply? *Thorax* 62(3):219–223.

Truong VA, Yao D (2013) Analytical models for designing pharmaceutical contracts. Working paper, Columbia University, New York, NY.

Wadhera RK, Steen DL, Khan I, Giugliano RP, Foody JM (2016) A review of low-density lipoprotein cholesterol, treatment strategies, and its impact on cardiovascular disease morbidity and mortality. *J. Clinical Lipidology* 10(3):472–489.

Xu L, Zhao H, Petruzzi NC (2021) Inducing compliance with post-market studies for drugs under FDA's accelerated approval pathway. *Manufacturing Service Oper. Management* 23(1):170–190.

Zaric GS, O'Brien BJ (2005) Analysis of a pharmaceutical risk sharing agreement based on the purchaser's total budget. *Health Econom.* 14(8):793–803.

Zaric GS, Xie B (2009) The impact of two pharmaceutical risk sharing agreements on pricing, promotion, and net health benefits. *Value Health* 12(5):838–845.

Zhang H, Zaric GS, Huang T (2011) Optimal design of a pharmaceutical price-volume agreement under asymmetric information about expected market size. *Production Oper. Management* 20(3): 334–346.

Zorc S, Chick SE, Hasija S (2017) Outcomes-based reimbursement policies for chronic care pathways. Working paper, INSEAD.