Formulary Design and Consumer Choices in Medicare Part D*

Meagan Madden†

November 9, 2026

[Latest Version Here]

Abstract

Health insurers use formulary tiering and exclusion to steer patients across competing prescription drugs, and manufacturers pay rebates for formularies that favor their products. This paper studies vertical contracting between Medicare Part D plans and insulin manufacturers and the implications for consumers. I model formulary design as a multi-principal, multi-agent game in which manufacturers offer rebate contracts to insurers that are contingent on formulary choices. I first estimate consumer demand for Part D plans and long-acting insulins and construct insurer and manufacturer payoffs under alternative formularies. I then solve for equilibrium formulary choices and rebate contracts. I evaluate Medicare Part D policies that expand access to insulin: (1) prohibiting exclusion, (2) requiring coverage of biosimilar insulin, and (3) capping patient out-of-pocket (OOP) costs. These policies generate small gains in consumer surplus through broader formulary coverage or lower OOP costs. However, these restrictions weaken insurers' leverage in the contracting game, resulting in lower manufacturer rebates and higher net drug costs for insurers—average rebates fall from 29% under the status quo to 6% when exclusion is prohibited. These results highlight a tradeoff between consumer access and the cost of prescription drugs.

Keywords: Prescription drug insurance, formulary, drug rebates, vertical contracts, Medicare Part D, insulin

JEL Codes: I11, L14

*I am grateful for the guidance and support from Brian McManus, Christopher Adams, Donna Gilleskie, Jonathan Williams, Marco Duarte, Alexander Olssen, and Qing Gong. I thank seminar participants at the Triangle Health Economics Workshop, UNC IO Workshop, and UNC Applied Microeconomics Workshop for their valuable comments.

†Department of Economics, University of North Carolina at Chapel Hill. Email: mkmadden@unc.edu

1 Introduction

In the United States, 91% of prescription drugs are purchased with insurance.[1] Insurers play a central role in determining patients' access to drug treatment—they decide which drugs to cover on the formulary and how much patients pay in out-of-pocket (OOP) costs.[2] Insurers can use formulary tiering and exclusion to steer patients across competing drugs within a therapeutic class. Consequently, manufacturers are willing to pay transfers ("drug rebates") to insurers in exchange for formulary choices that increase demand for their drugs.[3] Manufacturers routinely pay larger rebates if their drug is on a tier with low OOP costs and if rival drugs are on a tier with high OOP costs or excluded from coverage altogether (US GAO 2023). These post-sale rebates are substantial and reduce insurers' net cost of drug coverage. In 2021, drug rebates paid to Medicare Part D plans totaled $49 billion, or 23% of gross Part D spending. Tiering and exclusion give insurers leverage to negotiate lower net prices with drug manufacturers, but they also affect patient OOP costs and drug choices. My paper evaluates this trade-off: What are the effects on patient access and net-of-rebate drug prices when drug coverage and rebates are jointly determined by insurers and manufacturers?

I study this question in Medicare Part D and the market for long-acting insulins. In this setting, competing insurers offer prescription drug plans and determine formulary coverage of competing insulin products. Existing and proposed Part D policies restrict insurers' formulary choices in order to expand patient access. For example, Medicare Part D plans cannot exclude drugs in Medicare's six protected classes.[4] Recent congressional proposals would require coverage of generic and biosimilar drugs,[5] or require preferential tiering of generics and biosimilars relative to reference drugs.[6] Starting in 2023, Part D beneficiaries pay no more than $35 per month for covered insulin products; insurers decide which insulin products are included and excluded from coverage.[7]

Competition among insurers and drug manufacturers is a key feature of the Part D mar-

[1] Cash pay prescriptions represented approximately 9% of drug volume in 2024: https://www.iqvia.com/locations/united-states/blogs/2024/02/top-10-us-market-access-trends-for-2024.

[2] Insurers are ultimately responsible for their plan formularies, but they may hire pharmacy benefit managers (PBMs) for certain services. For example, PBMs maintain committees of medical professionals that evaluate drugs' clinical effectiveness and may advise insurers on formulary design. PBMs may also conduct rebate negotiations with drug manufacturers on behalf of insurers. Large insurers often perform these services in-house or through vertically integrated PBMs.

[3] In 2019, Medicare Part D plans excluded 42% of brand drugs, on average, and assigned 41% to non-preferred tiers with high OOP costs (Kyle et al. 2026).

[4] Anticonvulsants, antidepressants, antineoplastics, antipsychotics, antiretrovirals, immunosuppressants.

[5] *Ensuring Access to Lower-Cost Medicines for Seniors Act* (S. 4323): https://www.congress.gov/bill/119th-congress/senate-bill/4323

[6] *Ensuring Access to Lower-Cost Medicines for Seniors Act of 2026* (H.R. 8143): https://www.congress.gov/bill/119th-congress/house-bill/8143

[7] My paper studies formulary coverage of insulins in 2014–2019 before the OOP cap was introduced.

ket that shapes the effects of formulary regulation. Insurers may differentiate themselves through the formularies they offer, while manufacturers' payoffs depend on substitution both within and across plans. Consumer switching costs determine how strongly formulary changes redirect demand across drugs and plans, and therefore how much contracting leverage tiering and exclusion create. Exclusion by one insurer does not necessarily eliminate patient access when competing plans cover the drug, but plan switching costs may limit consumers' responses. As a result, policies that restrict exclusion may have modest effects on consumer access while eliminating an important source of insurers' contracting leverage.

To evaluate formulary regulations, I develop a model with multiple insurers competing in the health insurance market and multiple manufacturers competing in the prescription drug market. Manufacturers offer insurers formulary-contingent rebate schedules, insurers choose the tier placement of each drug, and consumers choose both an insurance plan and a prescription drug. I estimate consumer demand for plans and prescription drugs and construct insurer and manufacturer payoffs under observed and alternative formularies. I then solve for equilibrium formularies and rebate schedules under status quo rules and counterfactual policies. I find that policies requiring broader formulary coverage limit insurers' ability to steer patients across competing drugs, weakening their contracting leverage and reducing rebates. The effects on consumers depend on how much equilibrium formularies—and the resulting OOP costs—change relative to the status quo.

To construct payoffs for the contracting game, I first estimate consumer demand for Part D plans and long-acting insulins using Medicare claims data. During the sample period (2014–2019), up to five insulin brands were available in two delivery forms.[8] Consumers choose a Part D plan taking into account their expected value from the plan's insulin formulary, as well as annual premiums and OOP costs for non-insulin drugs. They choose an insulin product (brand-form combination) based on plan-specific OOP costs and preferences for brand and form. I estimate both plan and drug switching costs using cohort-level initial choice probabilities as instruments for consumers' lagged plan and drug choices. Consumers are willing to pay $82 per month to avoid switching insulin brands, $106 per month to avoid switching delivery form, and $208 per month to avoid switching insurance plans. Brand preferences, price sensitivity, and switching costs determine how consumers respond to formulary changes through plan and drug choices, and therefore govern firms' payoffs under alternative formularies.

I model drug manufacturer–insurer negotiations over insulin formularies and rebates as a multi-principal, multi-agent contracting game. Before insurers choose their insulin for-

[8]The insulin brands are Lantus, Levemir, Toujeo, Tresiba, and Basaglar. Basaglar is a follow-on version of Lantus. All brands are available in prefilled pens; Lantus, Levemir, and Tresiba are also available in vial form.

mularies, each manufacturer offers each insurer a *schedule* of rebates paid under different outcomes.[9] Specifically, each bilateral schedule specifies the amount that the manufacturer pays to the insurer contingent on the formulary choices of *all insurers*.[10] The rebate schedules offered by manufacturers in the first stage alter payoffs in the insurer subgame. Insurers observe all offers and choose the tier placement of each insulin; consumers then observe plan formularies and choose an insurance plan and a long-acting insulin. I solve for the joint formulary profile that maximizes aggregate insurer and manufacturer payoffs, and I construct candidate rebate schedules to support the contract-efficient outcome as a weakly truthful equilibrium (WTE) following Weber and Xiong (2006, 2026). In equilibrium, manufacturers weakly prefer the equilibrium outcome over alternatives given their rebate schedules, insurers have no incentive to choose an alternative formulary, and manufacturers have no incentive to alter their offers. Importantly, equilibrium rebates depend on the equilibrium formulary choices *and* the payoffs associated with off-equilibrium choices. Policies that restrict insurers' feasible formularies therefore remove credible off-equilibrium threats and may affect equilibrium rebates even when the equilibrium formularies themselves do not change.

In game simulations, I model formulary choices and rebate contracts for a subset of Part D insurers. The main specification includes three insurers and three insulin brands.[11] Each insurer decides the formulary position of each drug—preferred, non-preferred, or excluded. I simulate plan and drug choices under every formulary profile to construct insurer and manufacturer payoffs. I identify the contract-efficient formulary profile that maximizes joint firm payoffs and calculate the payoff gains relative to each alternative outcome. I impose sharing rules to allocate these gains across firms and construct bilateral transfers that align firms' incentives at the efficient outcome. I then map these transfers into nonnegative lump-sum rebates. Following Weber and Xiong (2026), I check whether any manufacturer could earn higher net profits by revising its offers to one or more insurers in the first stage; I hold rival manufacturers' offers fixed and consider every pure-strategy equilibrium induced in the insurer subgame.

Under status quo rules, average rebates are 29% of gross insulin spending. I evaluate counterfactuals motivated by existing and proposed Part D policies: (1) prohibiting exclusion of any insulin, (2) requiring coverage of biosimilar insulin, and (3) capping insulin OOP

[9] This differs from bargaining models, where the bilateral contracting object is a price or transfer. Here, the bilateral contracting object is a schedule of transfers contingent on insurers' subsequent formulary choices.

[10] In practice, bilateral rebate schedules may specify transfers contingent on (1) the tier of the manufacturer's own drug in the insurer's formulary, (2) the tiers of rival drugs, and (3) volume or market-share thresholds (US GAO 2023). Because realized market shares depend on all insurers' formulary choices, I capture these contingencies by allowing rebates to depend on the joint formulary profile.

[11] Lantus, Levemir, and Basaglar comprise 77% of long-acting insulin utilization in 2019. Humana, SilverScript (CVS), and UnitedHealthcare comprise 65% of enrollment in 2019, ranging from 54% to 80% across Part D regions.

costs at $35 per month. For each policy, I use the contracting model to solve for a new equilibrium, allowing both formularies and rebate schedules to adjust endogenously. I quantify impacts on consumer surplus, firm profits, and net-of-rebate drug prices. Prohibiting exclusion and requiring biosimilar coverage generate the same equilibrium formularies and increase consumer surplus by 2%. However, average rebates fall to 6% when no insulin can be excluded, compared with 25% when only Basaglar cannot be excluded. These policies have different effects on rebates because they leave insurers with different off-equilibrium contracting threats.

The incidence of these policies extends beyond beneficiaries taking insulin. Part D plan bids reflect their expected net-of-rebate drug costs and determine both enrollee premiums and federal subsidies. Policies that lower rebates and raise insurers' net costs will increase government spending and beneficiary premiums. In future analyses, I will re-estimate plan demand for both insulin users and non-users, update firm payoffs, and re-solve the contracting game. For each counterfactual policy, I will quantify changes in consumer surplus and government spending under the assumption that 25% of changes in drug costs are passed through to beneficiary premiums and 75% to federal subsidies.

This paper is broadly related to the literature on selective contracting and network design in health care. A central finding is that selective networks can lower negotiated prices but restrict consumer access. In Medicare Part D, preferred pharmacy networks lower retail drug prices, and the average enrollee benefits because lower OOP costs at preferred pharmacies outweigh reduced access to non-preferred pharmacies (Starc and Swanson 2021). In hospital markets, narrow and tiered networks similarly reduce spending by strengthening insurers' bargaining leverage and steering patients toward lower-cost providers (Prager 2015). Policies that prohibit exclusion or impose network-adequacy requirements can broaden provider choice but raise negotiated prices and premiums, illustrating the tradeoff between access and bargaining leverage (Ghili 2022; Ho and Lee 2019). Prager and Tilipman (2026) show that this tradeoff also depends on disagreement payoffs: lowering out-of-network payments reduces negotiated in-network prices but can lead to narrower networks. In much of this literature, insurer–provider relationships are characterized by a single negotiated price.[12] In contrast, the contracts in my model are schedules specifying manufacturer rebates that depend on the formulary placement of both the manufacturer's own drug and competing drugs, consistent with observed rebate contracts.

My paper is therefore more closely related to the growing literature on formulary design and pharmaceutical rebates. Existing work shows that formulary regulations can substantially affect negotiated rebates. Ho and Lee (2026) and Feng and Maini (2024) both

[12]Or a single baseline price that determines prices for different patient types and hospital services.

study contracting between a single PBM and competing manufacturers of statins, a class of cholesterol-lowering drugs. Ho and Lee (2026) develop a contracting model in which a single PBM offers schedules of formulary-contingent rebates to competing manufacturers, and they find higher rebates when the PBM can use tiered formularies and exclusion. Feng and Maini (2024) allow drug switching costs in consumer demand for statins, and they estimate the role of consumer inertia in a dynamic model of PBM–manufacturer contracting over formularies and net drug prices. Although formularies rarely exclude multiple drugs in their setting, the PBM uses the threat of exclusion to extract higher rebates from manufacturers. Related empirical work on Medicare's protected-class policy provides suggestive evidence of lower rebates for drugs that cannot be excluded from the formulary.[13] Conti et al. (2026) develop a theoretical model of formulary design as an all-pay contest, and they show that competition for preferred formulary tiers can improve efficiency relative to manufacturers setting prices to consumers directly. Unlike the rebate contracts in my model, these payments are committed ex ante and do not vary with the realized tier placement of the manufacturer's own drug or rival drugs. In this literature, Olssen and Demirer (2024) present the only model of formulary choices by competing insurers. They use a moment-inequalities approach to infer rebates for statins from observed formulary tiers in Medicare Part D, and they examine insurers' formulary responses to an exogenous decrease in rebates; however, they do not model manufacturers or rebate negotiations. I build on this existing literature by solving for both equilibrium formularies and rebate schedules in a model with competing insurers and competing drug manufacturers. My empirical application studies long-acting insulins and a distinct set of policy interventions, including biosimilar coverage and the $35 cap on insulin OOP costs.

My paper builds on theoretical models of bilateral contracting with contingent transfers. Early work studied one-to-many contracting environments, including the common-agency *bidding game*, in which multiple parties offer contingent payments to influence a common decision maker, and the complementary *offer game*, in which a single principal makes bilateral contract offers to multiple agents in the presence of contracting externalities (Bernheim and Whinston 1986; Segal 1999; Segal and Whinston 2003). More closely related, Prat and Rustichini (2003) study a multi-principal, multi-agent game with bilateral contracts that are contingent on each agent's individual action, and they characterize weakly truthful equilibria (WTE) in which all principals weakly prefer the same action profile, given the contracts offered in the first stage. WTE outcomes are efficient when agents' payoffs do not depend on the actions of other agents; Strulovici and Weber (2009) provide a constructive characteriza-

[13] Adams et al. (2025) compare the difference between Part D and commercial-market rebates for protected and non-protected drug classes, and Kakani et al. (2024) compare rebate trends for protected and non-protected classes.

tion of equilibrium contracts in this setting. Jackson and Wilkie (2005) show more generally that the continuation game is endogenous when players offer strategy-contingent side payments before game play; deviations in offers to a single player may affect the payoffs and choices of multiple players. Weber and Xiong (2006) extend the multi-principal, multi-agent environment to allow payoff externalities across agents. There exist contracts satisfying the WTE conditions at the efficient outcome when transfers are contingent on the action profile of *all* agents. Their updated analysis in Weber and Xiong (2026) distinguishes principals' incentives over outcomes in the continuation game from principals' contract deviations that alter the continuation game itself. They characterize equilibrium conditions, including conditions so that no principal has an incentive to deviate in the first stage, and construct contracts that are robust to such deviations. The framework developed in Weber and Xiong (2006, 2026) is well suited to my empirical application, where multiple drug manufacturers negotiate with multiple insurers over formulary-contingent rebate contracts.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting. Section 3 describes the data and variable construction. Section 4 presents the model of manufacturer–insurer contracting and consumer demand for insurance plans and prescription drugs. Section 5 details demand estimation, identification, and empirical results. Section 6 describes implementation of the contracting game and the simulated equilibrium. Section 7 analyzes equilibrium and payoffs under counterfactual policies, and Section 8 concludes.

2 Institutional Setting

2.1 Medicare Part D

Medicare is a federal health insurance program for Americans ages 65 and older and individuals with certain long-term disabilities. The Medicare Part D program provides outpatient prescription drug coverage through private insurers. Eligible beneficiaries can enroll in standalone Part D plans (SAPD) alongside Traditional Medicare, or Medicare Advantage plans that combine medical and drug benefits (MAPD).

The SAPD market is divided into 34 geographic regions covering the 50 states and Washington, D.C., and the MAPD market is defined at the county level. In 2019, the typical beneficiary could choose from 27 SAPD and 21 MAPD plans.

Part D enrollees pay monthly premiums and out-of-pocket (OOP) costs for prescription drugs. In 2019, average premiums for SAPD plans were $40 per month, ranging from $28 to $76 in the 10 largest SAPD plans.[14] OOP costs are determined by plan cost-sharing

[14] https://www.kff.org/medicare/issue-brief/medicare-part-d-a-first-look-at-prescription-drug-plans-in-2019/

rules. In 2019, unsubsidized beneficiaries paid an average of $39 per month in OOP costs, representing 17% of their total drug expenditures (average of $232 per month).[15] Federal subsidies to plans cover approximately 75% of the cost of standard Part D benefits, with additional premium and cost-sharing subsidies for qualifying low-income beneficiaries.[16]

Beneficiary OOP costs depend on their plan's benefit design and drug formulary. Part D has four coverage phases: deductible, initial coverage, coverage gap, and catastrophic. Under the standard benefit design, beneficiaries pay 100% of drug costs up to the deductible and 25% of costs during the initial coverage phase. In 2019, beneficiaries paid 25% cost-sharing for brand drugs and 37% for generic drugs in the coverage gap.[17] In 2019, beneficiaries paid 5% of drug costs after the catastrophic threshold, and there was no cap on OOP expenses until 2024. Figure A.1 illustrates the evolution of total drug costs and beneficiary OOP costs under the standard benefit design in 2019. Most plans offer alternative benefit designs but must provide coverage that is actuarially equivalent to or more generous than the standard benefit. For instance, 30% of SAPD plans offered zero-deductible coverage in 2019.

The drug formulary is the set of drugs covered by the plan and the associated cost sharing in each benefit phase. Most Part D plans use tiered formularies, so drugs in different tiers have different cost-sharing requirements. Part D formularies have up to six tiers with different cost-sharing rules, generally with lower OOP costs on lower tiers. For example, most plans have distinct tiers for generic drugs, preferred and non-preferred branded drugs, and high-cost specialty drugs. Beneficiaries pay lower OOP costs for brand drugs on "preferred" tiers and higher OOP costs for brand drugs on "non-preferred" tiers; they face the full pharmacy cost for any excluded drugs. Part D formularies must include at least two drugs in every therapeutic class. In addition, they must cover substantially all drugs in Medicare's six protected drug classes: anticonvulsants, antidepressants, antineoplastics, antipsychotics, antiretrovirals, and immunosuppressants.[18]

During annual open enrollment (October 15–December 7), existing beneficiaries can switch or drop plans for the next contract year beginning January 1. Plan premiums, deductibles, drug formularies, and pharmacy networks are posted on October 1, so beneficiaries can evaluate their plan options.[19] If an enrollee does not take action during the annual en-

[15] https://aspe.hhs.gov/sites/default/files/documents/1b652899fb99dd7e6e0edebbcc917cc8/aspe-part-d-oop.pdf

[16] https://www.medpac.gov/wp-content/uploads/2024/10/MedPAC_Payment_Basics_24_PartD_FINAL_SEC.pdf

[17] Beneficiaries previously paid 100% of drug costs in the coverage gap ("donut hole") until it was phased out over 2011–2020. In 2019, plans covered the remaining 63% of generic drug costs in the gap. For brand drugs, manufacturers provided a 70% discount at the point of sale, and plans covered the remaining 5%. These mandatory discounts were separate from manufacturer rebates paid to Part D plans for formulary design. For details on the phaseout, see Table 4: https://www.congress.gov/crs_external_products/R/PDF/R40611/R40611.15.pdf.

[18] Brand drugs with generic equivalents can be excluded if the generic is included.

[19] Beneficiaries can also access this information via the Medicare Plan Finder tool. Individuals enter their current medications and preferred pharmacies and the tool provides an estimate of total annual costs for all available plans in their geographic area.

rollment period, they are re-enrolled in their current plan.[20] Table 1 provides summary statistics for the set of SAPD plans available each year (2014–2019).

Table 1: Standalone Part D Market

	2014	2015	2016	2017	2018	2019
Annual Premium	\$646	\$637	\$645	\$624	\$632	\$567
Deductible	\$157	\$163	\$222	\$226	\$245	\$270
Zero Deductible	0.473	0.420	0.327	0.375	0.372	0.292
Number of Insulin Brands Covered	1.77	2.30	3.27	3.13	3.22	3.17
Share of Available Brands Covered	0.883	0.767	0.818	0.625	0.643	0.635
Number of Plans	1,169	1,001	886	746	782	901
Average Number of Plans Per Region	34.4	29.4	26.1	21.9	23	26.5

Notes: The top panel reports average plan characteristics across the set of SAPD plans each year. I report the average number of long-acting insulin brands included on the formulary. New brands enter the market during the sample period, so I also report the average share of available insulins included on the formulary. The bottom panel reports the total number of SAPD plans across regions and the average number of plans in the choice set by region.

2.2 Long-Acting Insulin

Insulin is a naturally occurring hormone that regulates blood glucose levels. Type 1 diabetes is an autoimmune condition in which the pancreas makes little or no insulin, and Type 2 diabetes is a condition in which cells stop responding efficiently to insulin. Type 2 diabetes accounts for approximately 90–95% of diabetes diagnoses in the U.S.[21] Long-acting insulin is injected once or twice per day to manage baseline blood glucose levels; some patients supplement with rapid-acting insulin before meals to control peaks in blood sugar.[22]

Long-acting insulin is an important drug class in Medicare Part D, used by approximately 4% of beneficiaries. Their annual OOP spending on long-acting insulins was \$520 on average in 2019, accounting for 40% of their total OOP drug costs. Diabetes is a chronic condition requiring ongoing treatment, so beneficiaries can anticipate their future insulin needs. Therefore, formulary coverage of long-acting insulins may be an important factor in their annual plan choice.

By the end of the sample period, five long-acting insulins were available: Lantus, Levemir, Toujeo, Tresiba, and Basaglar. Brands are differentiated by their insulin molecules,

[20] If plans are discontinued or consolidated, current enrollees will be automatically enrolled in a default plan if they do not make an active selection. They are notified by mail before open enrollment.

[21] https://www.cdc.gov/diabetes/about/about-type-2-diabetes.html

[22] The American Diabetes Association guidelines recommended GLP-1 receptor agonists for certain types of patients starting in 2019: https://diabetesjournals.org/clinical/article/37/1/11/32671/Standards-of-Medical-Care-in-Diabetes-2019.

concentrations, and delivery forms (prefilled pens or vials). Basaglar is a follow-on version of Lantus; both contain the same insulin analog (glargine U-100), while Toujeo contains a more concentrated formulation (glargine U-300).[23] Unlike generic drugs, follow-on biologics and biosimilars are not necessarily exact copies of their reference products.[24] Consumers may therefore have different preferences across brands, even between original and follow-on products containing the same insulin analog. Consumers may also differ in how they value the convenience and flexibility of each delivery form. Vial insulin must be drawn up and administered with separate syringes, but patients with specialized regimens or fluctuating needs may prefer the flexibility to tailor doses; prefilled pens are more convenient and easier to administer, particularly for patients with limited dexterity. Table 2 summarizes the long-acting insulin market in 2019.[25]

Table 2: Long-Acting Insulins

Brand	Molecule	Entry	Delivery	Gross Cost	OOP Cost	Plan Coverage	Market Share
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lantus	Glargine U-100	2000	Pen, vial	$388	$70	0.67	0.44
Levemir	Detemir	2005	Pen, vial	$490	$73	0.84	0.22
Toujeo	Glargine U-300	2015	Pen	$527	$81	0.67	0.08
Tresiba	Degludec	2016	Pen, vial	$560	$75	0.69	0.14
Basaglar	Glargine U-100	2016	Pen	$315	$51	0.31	0.11

Notes: Column (3) indicates the year of U.S. market launch; Toujeo launched in April 2015, Tresiba launched in January 2016, and Basaglar launched in December 2016. Columns (5) and (6) report the average gross cost and beneficiary OOP cost, respectively, for a 30-day supply of the pen form in 2019. Column (7) reports the share of SAPD plans covering each insulin brand in 2019. Column (8) reports each brand's share of long-acting insulin utilization in standalone Part D in 2019.

Part D plans vary in their coverage of long-acting insulin. Figure 1 shows the observed formulary tiers for each insulin brand across SAPD plans in 2019. Formulary exclusion is common—each long-acting insulin brand is excluded by some plans, and 60% of plans excluded at least two brands in 2019. Plans also differ in which drugs occupy the preferred tier. Nearly all SAPD plans placed at least one of Lantus and Levemir on the preferred tier in 2019—across all plans, 40% preferred both, 27% preferred only Lantus, and 33% preferred

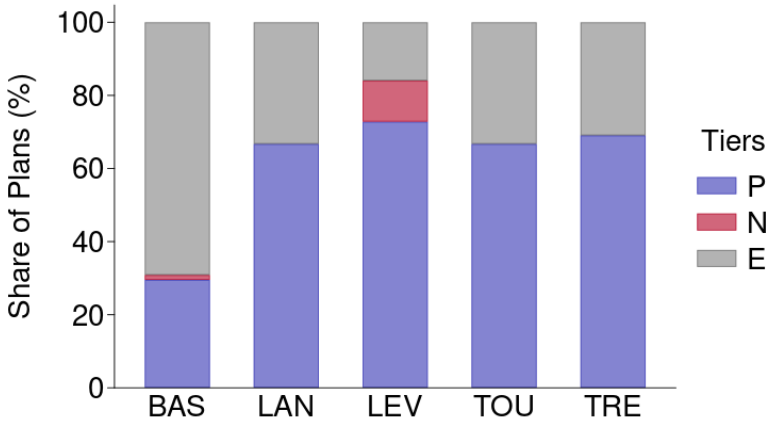
[23] Basaglar was approved through the FDA's 505(b)(2) "follow-on" pathway, relying in part on the demonstrated safety and effectiveness of Lantus. See https://diabetesjournals.org/clinical/article/35/3/181/35381/Basaglar. Additional follow-on insulins have been approved through the FDA's 351(k) biosimilar pathway which requires the comparison trials to show no clinically meaningful differences in safety or effectiveness relative to the reference product.

[24] https://www.fda.gov/drugs/biosimilars/overview-health-care-professionals

[25] Two additional products, Soliqua and Xultophy, combine long-acting insulin with a GLP-1 receptor agonist. These are excluded from the analysis since each represents less than 1% of Part D utilization during the sample period.

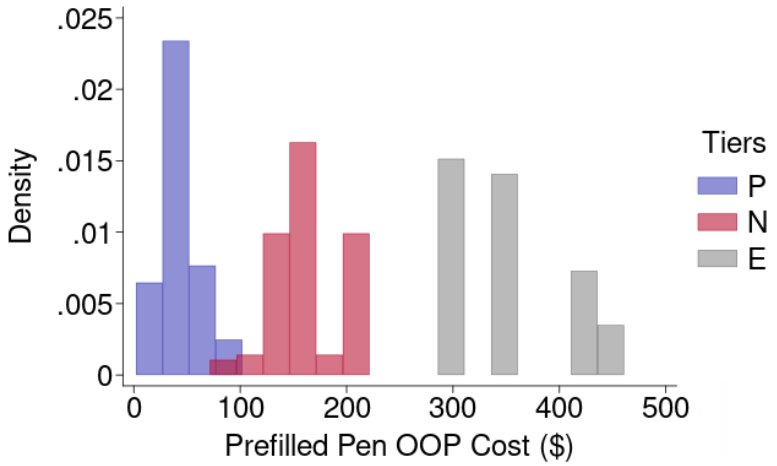
only Levemir. These formulary tiering decisions generate substantial variation in OOP costs across insulin products and across plans. Figure 2 shows the distribution of costs in 2019 conditional on observed tier. Appendix Figure A.2 shows the distribution of costs separately for each insulin brand.

Figure 1: Observed Formulary Tiers (2019)



Notes: Stacked bars represent the distribution of formulary tiers of five long-acting insulin brands in 2019: Basaglar (BAS), Lantus (LAN), Levemir (LEV), Toujeo (TOU), and Tresiba (TRE). The proportions within each bar represent the percentage of SAPD plans in 2019 assigning each insulin brand to the respective formulary tier: Preferred (P), Non-Preferred (N), or Excluded (E).

Figure 2: Observed Formulary Tiers (2019)



Notes: Figure shows the distribution of OOP costs ($) in the initial coverage phase for a 30-day supply of prefilled insulin pen. Bars represent the density of OOP costs conditional on observed formulary tier, across all SAPD plans and insulin brands. Appendix Figure A.2 shows the distribution of costs separately for each insulin brand.

Insurers can offer multiple Part D plans with different formularies but often use the same formulary design across plans. In 2019, 79% of insurers assigned Lantus, Levemir, and Basaglar to the same tiers in every SAPD plan they offered. The remaining 21% typically used two distinct insulin formularies (and at most three). For example, SilverScript offered two SAPD plans in each region—one placed Levemir and Basaglar on the preferred tier and excluded Lantus (95% of its enrollment), and another placed Levemir and Lantus on the preferred tier and excluded Basaglar (5% of its enrollment). In the contracting game, I assume each insurer chooses a single insulin formulary to use for all its SAPD plans. Beneficiary OOP costs can differ across plans with identical tiering due to differences in benefit design. For example, Humana offered three national SAPD plans in 2019 that all placed Lantus on the preferred tier; OOP costs were \$39, \$53, or \$70 for a 30-day supply of Lantus pen in the initial coverage phase.

2.3 Drug Rebates

Drug manufacturers negotiate rebates with Part D insurers in exchange for favorable formulary treatment. In 2019, Part D rebates totaled \$49 billion (23% of gross Part D drug spending). Just three therapeutic classes accounted for 73% of rebates: antidiabetic drugs, blood modifiers, and respiratory agents (US GAO 2023).[26] Average rebates exceeded 50% of gross spending for the class of antidiabetic drugs, which includes insulins. Competition within the class, particularly among insulin brands, contributes to these high manufacturer rebates (MedPAC 2023).

Contracts between manufacturers and insurers are confidential, and the terms and realized rebate payments are generally unobserved. The U.S. Government Accountability Office (GAO) reviewed contracts with drug manufacturers for a sample of Part D plans and documented common contractual terms. For all reviewed contracts, drug rebates were contingent on the drug's formulary tier; most also contained provisions tying rebates to the formulary placement of competing drugs. In addition, contracts frequently included thresholds for drug market shares or sales volumes and specified rebates for meeting those targets (US GAO 2023).[27] Manufacturer rebates are paid to insurers after the point of sale, reducing insurers' net drug costs. Beneficiaries do not directly receive these rebates; instead, they pay a fixed copayment or coinsurance rate applied to the point-of-sale pharmacy price before manufacturer rebates.

[26] These rebates are separate from mandatory manufacturer discounts in the coverage gap described in Section 2.1, which totaled approximately \$9 billion in 2019: https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/jun20_ch5_reporttocongress_sec.pdf.

[27] Rebates may also depend on utilization management practices, such as prior authorization and step therapy. However, there are very few cases of utilization management for insulins in Part D.

Insurers may utilize pharmacy benefit managers (PBMs) to conduct negotiations on their behalf. PBMs may also provide clinical guidance on formulary design through their Pharmacy and Therapeutics (P&T) committee of medical professionals. Major Part D insurers often perform these services internally or through vertically integrated PBMs. Humana operates its own pharmacy benefit services, while UnitedHealthcare and SilverScript use affiliated PBMs (OptumRx and CVS Caremark, respectively). In Sections 6–7, I model formulary choices and drug rebates for these three insurers, each of which operates through a distinct pharmacy benefit organization.

3 Data and Sample

3.1 Data Sources

I use administrative data for a 20% sample of Medicare Part D beneficiaries enrolled in standalone Part D plans (2007–2019). Data include plan choices, medical and prescription drug claims, beneficiary demographics (region, age, sex, race), and diagnosed chronic conditions. For each prescription drug claim, I observe the National Drug Code (NDC11), drug quantity, total pharmacy cost, and the amounts paid by the beneficiary and the plan.

For each of the 34 Part D regions, I construct the choice set of SAPD plans available each year. I link plan-level characteristics including premiums, deductibles, and drug formularies. The formulary files are available beginning in 2010; they include the list of covered drugs and their assigned formulary tiers for each plan, and the associated cost sharing for each tier and benefit phase.[28] I also link data on insurers' use of pharmacy benefit managers (PBMs) to administer Part D benefits, obtained through a Freedom of Information Act (FOIA) request.[29] I use these data to identify cohorts of beneficiaries enrolled in plans with the same PBM at the time of diagnosis. These cohorts are used to construct instruments based on initial drug choice probabilities in order to identify drug switching costs from persistent tastes (see Section 4.2).

[28] Cost sharing may vary by pharmacy type (retail or mail-order), pharmacy network (preferred in-network or standard), and drug quantity (30-day or 90-day supply). I use the lowest-cost option per day supplied across pharmacy types and networks to construct the plan-tier-phase coinsurance rates and copay amounts.

[29] Part D insurers must report any subcontracted organizations performing essential pharmacy benefits: (1) enrollment processing, (2) enrollee appeals & grievances, (3) customer service, (4) management of pharmacy & therapeutics committee, (5) coordination with drug benefits programs, (6) rebate negotiations, (7) pharmacy network development, (8) pharmacy technical assistance, (9) drug benefit administration, and (10) claims adjudication.

3.2 Beneficiary Sample

The sample includes Medicare beneficiaries ages 65 and older enrolled in standalone Part D (SAPD) plans. I exclude beneficiaries who are dually eligible for Medicaid or receive the Part D Low-Income Subsidy (LIS) since these programs affect beneficiaries' plan choice sets and OOP drug costs in ways not captured in my model. I also exclude beneficiaries enrolled in employer-sponsored Part D plans, since they have a different set of plans available to them. Finally, I exclude beneficiaries who switch plans during the contract year, since they face different formularies and cost-sharing rules over the contract year.

My analysis focuses on SAPD beneficiaries who use long-acting insulin. I restrict attention to beneficiaries who appear to initiate insulin treatment after enrolling in Medicare Part D, to help identify their initial insulin choices and corresponding choice sets.[30] To identify likely new insulin users, I exclude beneficiaries with insulin claims in the first six months of enrolling in Medicare Part D for the first time. The final sample used to estimate drug demand includes 269,381 beneficiary-year observations for 101,729 unique beneficiaries.

Table 3 describes the sample of beneficiaries enrolled in SAPD plans during 2014–2019 (Column 1), those taking long-acting insulin (Column 2), and the sample of insulin users enrolled in SAPD plans whose initial treatment choices are observed (Column 3). Long-acting insulin users are more likely to be male and non-Hispanic Black and have substantially higher gross drug expenditures and OOP costs, compared with beneficiaries not using long-acting insulin. Among insulin users, beneficiaries with observed initial choices are similar to those without observed initial choices on most dimensions; however, those with initial treatment choices observed in Part D are 1.3 years older and have 0.1 additional chronic conditions, on average.

Nearly all beneficiaries use a single insulin brand (94%) and a single delivery form (96%) throughout the contract year. [31] In the model of insulin demand in Section 4.2.1, I assume beneficiaries choose a single long-acting insulin each year. For beneficiaries with claims for multiple brands or forms within a year, I assign the brand and form representing the largest share of their annual utilization. Table 4 summarizes these choices.

Beneficiaries exhibit substantial persistence in both plan and insulin choices. Table 5 summarizes year-to-year choices of beneficiaries. Among beneficiaries taking long-acting insulin in consecutive years, 89% use the same insulin brand, and 97% use the same delivery form; 86% remain enrolled in the same Part D plan.

[30] This restriction is important to separately identify drug switching costs from persistent preferences in Section 4.2.

[31] Among beneficiaries with claims for multiple brands or forms, the most-used brand and form respectively account for 73% and 75% of utilization, on average.

Table 3: Beneficiary Sample

	Enrolled In SAPD Plan (1)	Long-Acting Insulin User (2)	Initial Insulin Choices Observed (3)
Age	74.1	74.5	75.8
Female	0.579	0.483	0.480
White	0.918	0.902	0.907
Black	0.036	0.052	0.051
Asian	0.008	0.008	0.007
Hispanic	0.004	0.005	0.005
Number of Chronic Conditions	3.622	5.801	5.928
Average Total Drug Expenditure	$2,276	$8,532	$8,235
Average Total Drug OOP	$579	$1,820	$1,789
Average Total Insulin Expenditure	—	$2,798	$2,738
Average Total Insulin OOP	—	$573	$575
Multiple Insulin Brands	—	0.058	0.062
Multiple Insulin Forms	—	0.040	0.041
Number of Beneficiary-Year Obs	13,672,370	437,735	269,381

Notes: Column (1) includes beneficiary-year observations enrolled in standalone Part D plans, excluding those who receive LIS subsidies, are dual enrolled in Medicaid, are enrolled in an employer-sponsored Part D plan, or switch during the contract year. Column (2) restricts the sample to beneficiary-year observations with claims for long-acting insulin. Column (3) further restricts the sample to beneficiary-year observations for beneficiaries who initiate insulin treatment after enrolling in Medicare Part D. Total Drug Expenditure and Total Insulin Expenditure represent the annual gross Part D spending for each beneficiary on prescription drugs and long-acting insulins, respectively; Total Drug OOP and Total Insulin OOP represent the OOP costs paid by the beneficiary. Multiple Insulin Brands and Multiple Insulin Forms represent the shares of beneficiaries with claims for more than one long-acting insulin brand or delivery form, respectively, within a year.

Table 4: Long-Acting Insulin Choices

	Insulin Sample	Final Sample
	(1)	(2)
Prefilled Pen	0.771	0.818
Insulin Brands		
Lantus	0.602	0.575
Levemir	0.248	0.265
Toujeo	0.055	0.059
Tresiba	0.057	0.060
Basaglar	0.035	0.036
Number of Beneficiary-Year Obs	437,735	269,381

Notes: Column (1) includes beneficiary-year observations with claims for long-acting insulin (2014–2019). Column (2) restricts the sample to beneficiary-year observations for beneficiaries who initiate insulin treatment after enrolling in Medicare Part D. Brand shares are based on each beneficiary's most-used brand. Prefilled Pen is the share of beneficiaries whose most-used delivery form is a prefilled insulin pen rather than insulin vials.

Table 5: Persistence in Plan and Insulin Choices

Same Long-Acting Insulin Molecule	0.929
Same Long-Acting Insulin Brand	0.893
Same Long-Acting Insulin Delivery Form	0.967
Same Part D plan	0.864
Number of Beneficiary-Year Obs	186,298

Notes: Includes consecutive beneficiary-year observations from the beneficiary sample (2014–2019).

3.3 Plan Sample

For estimation, I exclude Part D market-years with fewer than 1,000 beneficiaries. I further exclude plans representing less than 3% of enrollment within each market. For game simulations, I restrict attention to plans offered in 2019 by three insurers—Humana, SilverScript, and UnitedHealthcare. Table 6 summarizes all SAPD plans and the subset of plans included in estimation and game simulations.

3.4 OOP Calculator

Although beneficiary OOP costs are observed in the claims data, these realized costs depend on plan and drug choices and how these choices affect movement through the nonlinear

Table 6: Plan Sample

	All SAPD Plans (1)	Plan Estimation (2)	Game Simulations (3)
Panel A: 2014–2019			
Number of Plans	5,485	1,297	—
Number of Insurers	59	25	—
Number of Region-Years	204	154	—
Average Size of Choice Set	26.9	8.4	—
Aggregate Share of Enrollment	100%	57.9%	—
Panel B: 2019 Only			
Number of Plans	901	246	175
Number of Insurers	30	20	3
Number of Regions	34	27	27
Average Size of Choice Set	26.5	9.1	6.5
Aggregate Share of Enrollment	100%	61.8%	43.4%

Notes: Column (1) includes all standalone Part D plans. Column (2) excludes markets with fewer than 1,000 beneficiaries from the beneficiary sample and excludes plans accounting for less than 3% of enrollment of the beneficiary sample within a market. Column (3) further restricts the sample to plans offered by Humana, SilverScript, and UnitedHealthcare. Average Size of Choice Set summarizes the average number of SAPD plans available by region and year. Aggregate Share of Enrollment is the percentage of enrollment of the beneficiary sample accounted for by each subset.

phases of Part D benefit design. In order to estimate demand for Part D plans and long-acting insulins, I need a measure of OOP costs for counterfactual plans and insulins in the choice set. Furthermore, to solve the contracting game, I need OOP costs for counterfactual plans and insulins *and* for counterfactual formulary placement of each insulin (preferred, non-preferred, and excluded). Following the literature on Part D plan choice, I build an OOP calculator to map drug utilization into OOP costs, accounting for movement through Part D benefit phases and OOP costs in each phase.[32]

Risk Types. My implementation calculates expected OOP costs under each plan for each risk type, using simulated draws of total non-insulin drug expenditures for each type.[33] I first classify beneficiaries into 24 discrete risk types based on age groups (65–71, 72–77, and 78+), sex, and the number of diagnosed chronic conditions (0–3, 4–5, 6–7, and 8+). From the observed claims data, I calculate annual total drug expenditures for each beneficiary, excluding long-acting insulins. I assume drug expenditures follow a lognormal distribution (conditional on positive spending). For each risk type h and year y, I estimate the probability of positive expenditures on non-insulin drugs and the parameters (μ_{hy}, σ_{hy}) of a lognormal

[32] See for example: Abaluck and Gruber (2016), Olssen and Demirer (2024), and Coughlin (2026).

[33] Other papers typically use each beneficiary's full sequence of observed drug claims and construct OOP costs for every plan in their choice set. Since I need OOP costs for each plan×insulin×tier, I simplify the calculation and construct OOP costs using draws of total non-insulin expenditures ($) for each risk type.

distribution.

Beneficiaries of different risk types use different combinations of prescription drugs with differential cost-sharing, so they face different average coinsurance rates, even within the same plan. I calculate the weighted average coinsurance rate for each risk type, plan, and benefit phase using drug utilization weights. Let h index risk types, k index plan-years, b index benefit phases, and $d \in \mathcal{D}$ index non-insulin drugs. I calculate type-specific utilization weights:

$$w_{hd} = \frac{1}{N_h} \sum_{i:h(i)=h} \frac{q_{id}}{Q_i} \tag{1}$$

where N_h is the number of beneficiaries of risk type h. q_{id} represents beneficiary i's utilization of drug d in terms of days supplied; Q_i represents i's total utilization across all drugs, excluding long-acting insulins. The weights in (1) capture differences in the composition of prescription drug utilization across risk types. Then, I calculate the utilization-weighted average coinsurance for each type h, plan-year k, and phase b:

$$\overline{\theta}_{hkb} = \sum_{d \in \mathcal{D}} w_{hd} \theta_{dkb} \tag{2}$$

where θ_{dkb} is the observed coinsurance rate for drug d under plan k in benefit phase b.[34] The composition of prescription drugs varies over time, so I calculate type-specific utilization weights in (1) and coinsurance rates in (2) for each year in the sample (2014–2019).

Non-Insulin OOP Costs. I calculate non-insulin OOP costs for each risk type and plan by simulating draws of total non-insulin drug expenditures, calculating the implied OOP cost under each plan, and averaging across draws. Let $C_{hk}(T)$ denote the OOP calculator mapping total expenditures excluding insulin (T) to cumulative OOP costs in plan k, using the average coinsurance rates for risk type h from (2). For each risk type, I simulate $R = 1{,}000$ draws from the distribution of total non-insulin spending, with each realization denoted $T_h^{(r)}$. For each draw, I use the OOP calculator to get cumulative OOP costs $C_{hk}(T_h^{(r)})$ under each plan. I calculate expected non-insulin OOP costs by averaging across draws:

$$OOP_{hk} = \frac{1}{R} \sum_{r=1}^{R} C_{hk}\left(T_h^{(r)}\right) \tag{3}$$

Expected non-insulin OOP costs vary by risk type and plan and enter the plan demand model in Section 4.2. The estimation sample includes multiple years (2014–2019), so the

[34] If a plan specifies fixed copays for a drug, I calculate the implied coinsurance rate using the median pharmacy price of the drug.

OOP calculation differs by year; in particular, cost sharing in the coverage gap phase decreased during the sample period. Appendix Section **??** provides additional details on the construction of the OOP calculator.

Insulin OOP Costs. I calculate the additional OOP costs for each long-acting insulin under every plan in the beneficiary's choice set. Let $\Delta I_{j|k}(C_k, q)$ represent a function that calculates the OOP insulin cost for quantity q of insulin j under plan k, starting from cumulative OOP spending C_k. This function uses plan k's observed cost sharing amount for each benefit phase (cs_{jkb}) and accounts for progression through the nonlinear benefit design starting from C_k.[35]

For each risk type h and draw r of total non-insulin expenditures, I get the cumulative OOP costs $C_{hk}(T_h^{(r)})$ under each plan. For beneficiary i of risk type h, I calculate the incremental OOP cost of insulin j under plan k for each spending draw:

$$price_{ij|k}^{(r)} = \Delta I_{j|k}\left(C_{hk}(T_h^{(r)}), \overline{q}_i\right) \tag{4}$$

where $\overline{q}_i$ denotes beneficiary i's observed days supplied of long-acting insulin for their observed insulin choice, which I hold fixed across alternative insulin products.[36] I calculate the expected price of insulin by averaging across draws:

$$price_{ij|k} = \frac{1}{R}\sum_{r=1}^{R} price_{ij|k}^{(r)} \tag{5}$$

Expected insulin costs vary by beneficiary (depending on risk type and utilization), insulin product, plan, and year. Table 7 compares beneficiaries' realized insulin costs with the expected costs calculated for their observed insulin and plan choices. Expected insulin costs enter the drug demand model in Section 4.2.

3.5 Formulary Tiers

To simulate consumer demand and firm payoffs under alternative formulary arrangements, I construct counterfactual cost-sharing for each insulin product under three formulary placements: preferred (P), non-preferred (N), and excluded (E).

[35] Cost-sharing amounts are fixed copays ($) per day of treatment, or coinsurance rates (%) applied to the median pharmacy cost per day, depending on the formulary-tier rules.

[36] Days supplied is a standard measure of utilization that is comparable across products with different concentrations and delivery forms, unlike physical quantities such as milliliters, pens, or vials.

Table 7: Realized and Expected Insulin Costs

	Realized Insulin Cost (1)	Expected Insulin Cost (2)
2014	\$556	\$650
2015	\$617	\$646
2016	\$604	\$650
2017	\$580	\$577
2018	\$582	\$505
2019	\$520	\$472
2014–2019	\$575	\$573
Number of Beneficiary-Years	269,277	

Notes: Column (1) reports the average of actual OOP costs for long-acting insulins by year (2014–2019). Column (2) reports the average of expected insulin costs for beneficiaries' observed insulin and plan choices, constructed using the OOP calculator.

Part D plans can have up to six formulary tiers (numbered 1–6), with different cost-sharing rules. In the data, plans use up to two formulary tiers for long-acting insulins, as well as formulary exclusion. I define the average coinsurance rate for each plan k, tier τ, and benefit phase b as

$$\overline{\theta}_{\tau kb} = \frac{1}{|\mathcal{J}_{\tau k}|} \sum_{j \in \mathcal{J}_{\tau k}} \theta_{jkb} \tag{6}$$

where $\mathcal{J}_{\tau k}$ denotes the observed set of insulin products in tier τ in plan k. θ_{jkb} represents the observed coinsurance rate for product j, or the implied rate calculated as the observed copay for product j divided by its median pharmacy cost.

For plans with two insulin tiers, I designate the lower-numbered tier as preferred and the higher-numbered tier as non-preferred, using the respective coinsurance rates from (6). For plans with a single insulin tier, I designate the observed tier as preferred; I designate the next tier above it as non-preferred, and I calculate the average coinsurance rate for all branded drugs on the tier. For excluded insulins, I assume beneficiaries pay the full pharmacy price (100% coinsurance).

Table 8 summarizes the resulting coinsurance rates for preferred and non-preferred tiers during the initial coverage phase. Average preferred-tier coinsurance rates are similar across plans with one or two observed insulin tiers. However, the constructed non-preferred coinsurance rate for plans with a single insulin tier is substantially higher than the observed rate for plans with two insulin tiers (39% versus 25%). This difference may reflect variation in drug composition across tiers or differences in overall plan generosity. It also highlights uncertainty about insulin cost sharing under unobserved formulary tiers.[37]

This procedure constructs counterfactual coinsurance rates for each insulin tier (P, N, E)

[37] Future analyses will assess alternative assumptions about insulin cost sharing on non-preferred tiers.

Table 8: Observed Coinsurance Rates by Formulary Tier

	Two Insulin Tiers		One Insulin Tier	
	Lower Tier	Higher Tier	Insulin Tier	Next Brand Tier
	(1)	(2)	(3)	(4)
Average Coinsurance	0.095	0.248	0.098	0.394
Number of Plans	181		720	

Notes: Columns (1) and (2) report the average coinsurance rate in the initial coverage phase for insulin products in the lower-numbered and higher-numbered tiers, respectively, across SAPD plans with two observed insulin tiers in 2019. Column (3) reports the average coinsurance for insulin products across SAPD plans with a single insulin tier in 2019; Column (4) reports the average coinsurance for brand drugs in the next-highest tier.

for each plan and benefit phase. I use these rates to calculate beneficiary OOP costs and simulate plan and drug choice probabilities under alternative formularies, which determine insurer and manufacturer payoffs in the contracting game. Table 9 summarizes beneficiary insulin costs in 2019 by counterfactual formulary tiers.

Table 9: Insulin OOP Costs by Counterfactual Tiers

	Preferred	Non-Preferred	Excluded
	(1)	(2)	(3)
Lantus	$421	$886	$3,006
Levemir	$600	$1,105	$3,975
Toujeo	$550	$1,045	$3,704
Tresiba	$448	$922	$3,158
Basaglar	$338	$764	$2,521
Number of Beneficiary-Plan Obs	2,158,685		

Notes: Column (1) summarizes the counterfactual OOP cost of each insulin brand if assigned to the preferred tier, across beneficiaries and all SAPD plans in their choice set in 2019; I use costs for the prefilled pen of each brand. Columns (2) and (3) summarize the counterfactual OOP costs of each insulin brand if assigned to the non-preferred tier or excluded from coverage, respectively.

4 Model

The model consists of four stages. First, competing manufacturers of long-acting insulins offer rebate contracts to competing Part D insurers, with payments contingent on insurers' formulary choices. Second, insurers observe these contracts and choose their insulin formularies. Third, consumers choose among Part D insurance plans, and fourth, consumers choose among insulin products. Formulary choices affect consumers' plan and insulin choices, which determine insurer and manufacturer profits. I first describe the contracting game and its equilibrium and then specify consumer demand for Part D plans and long-acting insulins.

4.1 Contracting Game

4.1.1 Multi-Principal, Multi-Agent Contracting Game

I model negotiations between Part D insurers and insulin manufacturers as a multi-principal, multi-agent contracting game with outcome-contingent contracts and complete information. Drug manufacturers act as principals, offering incentive contracts to influence formulary actions taken by insurers, the agents in this game. Each manufacturer produces a single insulin brand, so let $j \in \mathcal{J}$ index manufacturers/insulins with $|\mathcal{J}| = J > 1$. Let $f \in \mathcal{F}$ index insurers with $|\mathcal{F}| = F > 1$, each offering a set of plans $\mathcal{K}_f$. Formularies have three tiers: preferred (P), non-preferred (N), and excluded (E). Let $a_{jf} \in \{P, N, E\}$ denote the tier of insulin j in insurer f's formulary. Each insurer f selects a single formulary design $a_f = (a_{jf})_{j\in\mathcal{J}} \in \mathcal{A}_f$, which applies to all plans; $\mathcal{A}_f$ denotes the set of feasible formulary actions for insurer f. The joint formulary action profile $a = (a_1, \ldots, a_F)$ collects the formulary choices of all insurers, where $a \in \mathcal{A} = \prod_f \mathcal{A}_f$. When there are no restrictions on insurers' formularies, $|\mathcal{A}_f| = 3^J$ and $|\mathcal{A}| = (3^J)^F$.[38]

Consumers' plan and insulin choices depend on the formulary choices for all plans in their choice set, and insurer and manufacturer profits depend on plan enrollment and drug utilization for each joint formulary profile. Let $\phi_j(a)$ and $\pi_f(a)$ denote the gross profits of manufacturer j and insurer f, respectively, before rebates. Rebate contracts for each manufacturer–insurer pair (j, f) are schedules $t_{jf} : \mathcal{A} \to \mathbb{R}_+$ that specify nonnegative lump-sum transfers paid by manufacturer j to insurer f for each joint formulary profile $a \in \mathcal{A}$. These transfers determine players' net profits:

$$\begin{aligned} \Phi_j(a; t) &= \phi_j(a) - \sum_{f\in\mathcal{F}} t_{jf}(a) \quad \text{for all manufacturers } j \in \mathcal{J} \\ \Pi_f(a; t) &= \pi_f(a) + \sum_{j\in\mathcal{J}} t_{jf}(a) \quad \text{for all insurers } f \in \mathcal{F} \end{aligned} \tag{7}$$

The timing of the game is as follows: First, each manufacturer simultaneously offers a rebate contract to each insurer. Second, insurers observe all contracts and simultaneously select their formularies. Finally, consumers observe drug formularies and other plan characteristics, and they choose a plan and then choose an insulin. Firms have complete information, so they know how joint formulary choices in the contracting game affect consumers' choices and, consequently, their payoffs.

[38] I describe feasible formularies under status quo rules in Section 6. Counterfactual policies in Section 7 restrict the set of actions available to insurers, for example, by removing formularies with excluded drugs (E).

4.1.2 Characterizing Equilibrium

An equilibrium of the contracting game consists of: (1) bilateral rebate contracts $\hat{t}_{jf}(a)$ for each manufacturer–insurer pair (j, f), (2) insurers' joint formulary choices $\hat{a}$, and (3) insurers' continuation strategies for any set of rebate offers. The equilibrium contracts specify transfers for *every* joint formulary profile $a \in \mathcal{A}$, not only the equilibrium $\hat{a}$. Consumer demand is known to firms and well-defined for every joint formulary profile $a \in \mathcal{A}$. Contract offers are public, so the solution concept is subgame-perfect equilibrium (SPE). Following Prat and Rustichini (2003) and Weber and Xiong (2006), I focus on a refinement referred to as Weakly Truthful Equilibrium (WTE) in which all manufacturers' incentives are aligned around the same formulary choices in Stage 2, given their offered contracts from Stage 1.

Three conditions must be true of any WTE (Prat and Rustichini 2003; Weber and Xiong 2006). First, **agent maximization (AM)** requires that each insurer's equilibrium formulary choice maximizes its net profits, given the rebate contracts it receives in equilibrium. This ensures that no insurer has an incentive to deviate from its equilibrium action.

$$\pi_f(\hat{a}) + \sum_j \hat{t}_{jf}(\hat{a}) \geq \pi_f(a_f, \hat{a}_{-f}) + \sum_j \hat{t}_{jf}(a_f, \hat{a}_{-f}) \qquad \forall f \in \mathcal{F}, \forall a_f \in \mathcal{A}_f \tag{AM}$$

Second, **weak truthfulness (WT)** requires that each manufacturer's net profit is maximized at the equilibrium formulary profile, holding equilibrium rebate contracts fixed. Thus, all manufacturers weakly prefer the equilibrium formulary profile over all alternatives, and their incentives are aligned at the equilibrium outcome from the insurer subgame.

$$\phi_j(\hat{a}) - \sum_f \hat{t}_{jf}(\hat{a}) \geq \phi_j(a) - \sum_f \hat{t}_{jf}(a) \qquad \forall j \in \mathcal{J}, \forall a \in \mathcal{A} \tag{WT}$$

Third, **principal cost minimization (PM)** stipulates that manufacturers' equilibrium rebates compensate insurers just enough to implement the equilibrium formulary profile.

$$\pi_f(\hat{a}) + \sum_{j' \in \mathcal{J}} \hat{t}_{j'f}(\hat{a}) = \max_{a_f \in \mathcal{A}_f} \left\{ \pi_f(a_f, \hat{a}_{-f}) + \sum_{j' \in \mathcal{J} \setminus j} \hat{t}_{j'f}(a_f, \hat{a}_{-f}) \right\} \quad \forall (j, f) \in \mathcal{J} \times \mathcal{F} \tag{PM}$$

As a consequence of **PM**, insurer f can recover its full equilibrium profit from its best formulary choice if manufacturer j withdraws its offers but contracts with other manufacturers $j' \neq j$ and other insurers' formulary choices are held fixed.

In my setting, insurers' profits depend on the formulary choices of competing insurers. If a manufacturer changes its rebate offers, it may induce one insurer to change its formulary choice; this affects competing insurers' profits and potentially their formulary choices as

well. Thus, a manufacturer may profitably deviate from its contracts and induce a different joint formulary outcome in the insurers' continuation game, even if the manufacturer weakly prefers the original formulary outcome under its existing contracts (WT). For this reason, rebate contracts satisfying **AM**, **WT**, and **PM** conditions do not necessarily constitute an equilibrium when insurers' profits depend on competing insurers' formulary choices.[39]

Weber and Xiong (2026) establish an additional condition for equilibrium in this setting: **Principal deviation (PD)** requires that no manufacturer can profitably change its rebate offers in the first stage, holding other manufacturers' offers fixed and accounting for insurers' equilibrium responses in the second stage. A deviating manufacturer can change its contracts with all insurers, but it must offer sufficient compensation to induce alternative formulary choices, given the rebates insurers continue to receive from competing manufacturers. A manufacturer has an incentive to deviate if it can induce alternative formulary choices at a cost that allows it to earn higher net profits than under its original contracts. Weber and Xiong (2026) characterize manufacturer deviations under different equilibrium selection rules in the continuation game. They distinguish a necessary condition ruling out profitable manufacturer deviations that induce pure-strategy Nash equilibria in the insurers' continuation game (PD-pure) from a stronger condition ruling out profitable deviations under any pure- or mixed-strategy continuation equilibrium (PD-mix). Appendix C.2 formally defines these conditions and describes how to evaluate profitable manufacturer deviations.

4.1.3 Solving for Equilibrium

The contracting game can have multiple equilibria with different formulary profiles and rebate contracts, including multiple contracts supporting the same formulary profile. I focus on constructing contracts that support a pure-strategy "efficient" outcome, where total surplus to insurers and manufacturers is maximized.[40] Even at the contract-efficient outcome, the equilibrium conditions do not uniquely determine bilateral rebate schedules.

Following Weber and Xiong (2006, 2026), I parameterize the allocation of surplus across firms and bilateral relationships to construct candidate rebate schedules satisfying AM, WT, and PM. I then evaluate whether manufacturers have profitable deviations from these schedules using the PD conditions. Section 6 describes the construction of rebate schedules and presents the simulated contracting outcomes.

[39] In Prat and Rustichini (2003), agents' payoffs depend only on their own actions, and AM, WT, and PM are necessary and sufficient conditions to characterize a WTE.

[40] Following Weber and Xiong (2006) and the literature on bilateral contracting games, I use "efficient" to indicate outcomes that maximize total surplus of the vertical structure.

4.2 Consumer Demand

Insurer and manufacturer payoffs depend on consumer choices. I model consumer demand for Medicare Part D plans and long-acting insulins as sequential choices, linked through the value consumers derive from plans' insulin formularies. Consumers first select a Part D plan based on the coverage and OOP costs of long-acting insulins, expected OOP costs for other drugs, annual premiums, and other plan characteristics. Next, they select a long-acting insulin based on their preferences for different insulins and the OOP cost of each insulin determined by the plan's formulary. Consumers incur switching costs if they switch from their previous Part D plan or previous insulin choice.

This specification links strategic formulary choices in the contracting game to both drug and plan demand. Formulary changes affect insulin OOP costs and drug choices conditional on enrollment, as well as the utility beneficiaries derive from a plan and their likelihood of enrolling in it. Beneficiaries may respond by switching insulins within their existing plan or switching plans, although switching costs limit their ability to substitute across both dimensions. Insurers and manufacturers anticipate these demand responses when contracting over formularies and rebates.

Below, I describe demand for long-acting insulins conditional on enrollment in a particular plan, followed by demand for Part D plans. Let $k \in \mathcal{K}_i$ index Part D plans in beneficiary i's choice set, which varies by region and year. Let $j \in \mathcal{J}_i$ index the set of available long-acting insulin products, defined by brand and form of delivery (pen or vial); availability varies by year as new brands enter the market (see Table 2).

4.2.1 Insulin Demand

Beneficiary i derives utility from long-acting insulin j conditional on enrollment in plan k:

$$\begin{aligned} v_{ij|k} = \alpha_i^p price_{ij|k} + \gamma_{ij}^{brand} + \gamma_{ij}^{form} \\ + \psi^{brand} \mathbb{1}\{brand_j = brand_i^{prev}\} \\ + \psi^{form} \mathbb{1}\{form_j = form_i^{prev}\} \\ + \psi^{mol} \mathbb{1}\{mol_j = mol_i^{prev}\} + \varepsilon_{ij|k} \end{aligned} \tag{8}$$

where $price_{ij|k}$ is beneficiary i's annual OOP cost of insulin j in plan k. I construct insulin OOP costs from each plan's formulary and cost-sharing rules, accounting for the nonlinear benefit design in Part D and expected spending on other prescription drugs. Details are provided in Section 3.

I allow for unobserved heterogeneity in sensitivity to insulin prices, brands, and de-

livery forms through random coefficients: $\alpha_i^p \sim N(\alpha, \sigma_\alpha)$, $\gamma_{ij}^{brand} \sim N(\gamma_j^{brand}, \sigma_j^{brand})$, and $\gamma_{ij}^{form} \sim N(\gamma_j^{form}, \sigma_j^{form})$. The brand-specific preferences capture persistent differences in beneficiaries' valuations of insulin brands beyond their OOP costs and observed characteristics. The form-specific preferences similarly capture persistent preferences for vial or pen insulin.

Beneficiaries also face three additive switching costs. The parameters ψ^{form}, ψ^{brand}, and ψ^{mol} measure the utility from remaining with the same insulin form, brand, or molecule, respectively, as their previous choice. Beneficiaries choosing insulin for the first time do not face switching costs, so the form, brand, and molecule indicators in (8) are all equal to zero. The idiosyncratic preference shocks $\varepsilon_{ij|k}$ are assumed to be independently and identically distributed according to a Type I Extreme Value distribution.

4.2.2 Plan Demand

Beneficiary i derives utility from insurance plan k:

$$\begin{aligned} u_{ik} = \beta_i^p prem_k + \beta^x x_k + \beta^o OOP_{h(i)k} + \beta^v v_{ik}^*(\nu_i) \\ + \psi^{plan} \mathbb{1}\{k = k_i^{prev}\} + \xi_k + \epsilon_{ik} \end{aligned} \tag{9}$$

where $prem_k$ is the annual premium or cost of insurance plan k, and x_k indicates whether the plan has zero deductible. Heterogeneity in premium sensitivity is specified through random coefficients: $\beta_i^p \sim N(\beta, \sigma_\beta)$. $OOP_{h(i)k}$ represents i's expected annual OOP spending on non-insulin prescription drugs, which depends on their risk type h. I simulate draws of total drug spending—excluding insulins—for discrete risk types and calculate the implied OOP cost under each plan; I take the average OOP costs across simulation draws for each risk type. Details are provided in Section 3.

The coefficient ψ^{plan} captures the utility gain from remaining in the same plan as the previous year. The term ξ_k represents unobserved plan quality (to the econometrician). Idiosyncratic plan preference shocks ϵ_{ik} are i.i.d. Type I Extreme Value and are observed by beneficiaries when choosing a plan. Beneficiaries also know their preferences for insulin brands, previous insulin choices, the OOP costs of each insulin under each plan, and their price sensitivity to insulin OOP costs. However, I assume that beneficiaries do not observe the idiosyncratic insulin preference shocks $\varepsilon_{ij|k}$ until after choosing a plan. When evaluating plans, beneficiaries anticipate the expected utility from their insulin choice in each plan.

Let $\overline{v}_{ij|k}(\nu_i)$ denote beneficiary i's utility from insulin j under plan k before realization

of the idiosyncratic insulin preference shock:

$$\begin{aligned}\overline{v}_{ij|k}(\nu_i) = (\alpha + \sigma_\alpha \nu_i^p)\ price_{ij|k} + (\gamma_j^{brand} + \sigma_j^{brand}\nu_{ij}^b) + (\gamma_j^{form} + \sigma_j^{form}\nu_{ij}^f) \\ + \psi^{form}\mathbb{1}\{form_j = form_i^{prev}\} \\ + \psi^{brand}\mathbb{1}\{brand_j = brand_i^{prev}\} \\ + \psi^{mol}\mathbb{1}\{mol_j = mol_i^{prev}\}\end{aligned} \tag{10}$$

Here, ν_i denotes the vector of realized drug taste shocks determining beneficiary i's random coefficients. The term v_{ik}^* represents beneficiary i's expected maximum utility from the insulin choice set under plan k:

$$v_{ik}^*(\nu_i) = \log\left(\sum_{j \in \mathcal{J}_i} \exp\left(\overline{v}_{ij|k}(\nu_i)\right)\right) \tag{11}$$

The logit inclusive value in (11) summarizes the expected utility beneficiaries derive from each plan's insulin formulary, accounting for their realized drug tastes (ν_i) and OOP costs while integrating over the uncertainty in their idiosyncratic shocks ($\varepsilon_{ij|k}$). This term enters plan utility in (9), linking consumers' plan and insulin choices.

5 Demand Estimation

I estimate consumer demand in two stages. First, I estimate the mixed logit model of insulin demand. The sample includes both continuing insulin users and beneficiaries initiating insulin treatment during the contract year. I use the estimated parameters to calculate the logit inclusive value for each beneficiary and every plan in their choice set. Second, I estimate plan demand for the subset of beneficiaries who took a long-acting insulin in the previous year. Below, I provide identification arguments for key demand parameters, and then outline the estimation procedure.

5.1 Identification

Inertia in plan and drug choices could be explained by multiple sources in the model—including switching costs and unobserved heterogeneity in tastes or price sensitivity. Below, I describe my approach to separately identify these sources of inertia.

Insulin Switching Costs. Inertia in drug choices could be attributed to high switching costs or persistent tastes for products or product characteristics—for instance, a benefi-

ciary might choose Lantus every year because they have a strong preference for Lantus or the cost of switching to another insulin is too high. To separately identify switching costs $(\psi^{brand}, \psi^{form})$ from unobserved preference heterogeneity, I construct instruments for beneficiaries' previous insulin choices using cohort-level initial choice probabilities, similar to the instrumental variables approach used by Feng and Maini (2024).

I define cohorts based on the year and pharmacy benefit manager (PBM) associated with the beneficiary's plan at the time of diagnosis.[41] Product availability varies across years due to entry, so beneficiaries who initiate treatment in the same year have access to the same set of insulins. Plans using the same PBM for rebate negotiations have similar drug formularies, so beneficiaries enrolled in those plans face similar OOP costs for different insulins. Appendix Figure A.3 shows the variation in initial insulin choices across years and PBMs.

For each beneficiary i, I construct leave-one-out cohort probabilities of initial form, brand, and molecule choices. Let $\mathcal{C}_i$ denote beneficiary i's insulin cohort, and let j_n^0 denote the initial choice of beneficiary $n \in \mathcal{C}_i$. I construct the following choice probabilities for each beneficiary:

$$\begin{aligned} Z_{ij}^{form} &= \frac{1}{|\mathcal{C}_i| - 1} \sum_{n \in \mathcal{C}_i \setminus \{i\}} \mathbb{1}\{form_j = form_{j_n^0}\} \\ Z_{ij}^{brand} &= \frac{1}{|\mathcal{C}_i| - 1} \sum_{n \in \mathcal{C}_i \setminus \{i\}} \mathbb{1}\{brand_j = brand_{j_n^0}\} \\ Z_{ij}^{mol} &= \frac{1}{|\mathcal{C}_i| - 1} \sum_{n \in \mathcal{C}_i \setminus \{i\}} \mathbb{1}\{mol_j = mol_{j_n^0}\} \end{aligned} \tag{12}$$

I use these cohort probabilities as instruments for the lagged choice indicators in Equation (8). First-stage estimates are reported in Appendix Table B.1. Beneficiaries' initial choices are correlated with the cohort initial choice probabilities since they face the same initial conditions. This correlation persists in subsequent years because beneficiaries face switching costs. The exclusion restriction requires that cohort choices are unrelated to beneficiary i's persistent tastes. The identifying assumption for exclusion is that beneficiaries do not select particular plans (and associated PBMs) based on their insulin coverage before initiating treatment.

Insulin Taste Heterogeneity. A second identification challenge is distinguishing heterogeneity in brand preferences from heterogeneity in price sensitivity. If the OOP cost of Lantus increases, beneficiaries with relatively high price sensitivity or relatively low tastes for Lantus are more likely to switch to alternative insulins. Identification relies on *differences* in pairwise substitution patterns in response to changes in the relative prices of competing

[41] I define diagnosis as the beneficiary's first long-acting insulin claim.

products.

Consider two plans in which the relative formulary placements of Lantus and Levemir reverse. In one plan, Lantus moves from preferred to excluded, and Levemir moves from excluded to preferred. In another plan, the opposite change occurs. Suppose mean brand preferences are pinned down by other identifying variation, such as aggregate market shares. Beneficiaries with high price sensitivity are more likely to switch toward the newly preferred insulin in either case, regardless of which brand becomes cheaper. However, the magnitude of substitution also depends on the degree of heterogeneity in brand preferences. For example, if consumer preferences for Lantus are highly dispersed, some beneficiaries with particularly strong preferences for Lantus will continue using it despite a substantial increase in its OOP cost. If preferences for Levemir are less dispersed, there will be relatively more substitution away from Levemir, for a similar increase in price, all else equal. Comparing substitution patterns following similar changes in relative prices, but in opposite directions, therefore helps distinguish heterogeneity in price sensitivity from heterogeneity in relative brand preferences.

In my setting, formulary tier changes provide substantial variation in relative insulin prices, including sharp changes in price rankings. Table 10 documents year-to-year transitions in which insulin brand is the lowest cost in a plan. This variation helps identify the random coefficients on OOP costs (σ_α) and insulin brand fixed effects (σ_j^{brand}).

Table 10: Year-to-Year Transitions in the Lowest Cost Insulin Within Plan (2014–2019)

		t			
		Basaglar	Lantus	Levemir	Toujeo
t-1	Basaglar	–	59	0	0
	Lantus	542	–	348	0
	Levemir	0	167	–	30

Notes: Counts indicate number of plan-year observations in which the lowest cost insulin brand in the plan changed from year $t-1$ to year t. I define the lowest cost insulin brands using the OOP cost for a 30-day supply of the prefilled pen in the initial coverage phase.

Plan Switching Costs. Inertia in plan choices may reflect switching costs or persistent preferences for particular plans or plan characteristics. I employ a similar identification strategy and construct cohort instruments for beneficiaries' previous plan choices to identify plan switching costs (ψ^{plan}).

For plan demand, I define cohorts of beneficiaries who turned 65 in the same calendar year and resided in the same Part D region. These beneficiaries had the same initial choice set of SAPD plans when they were first eligible for Medicare. Let $\mathcal{P}_i$ denote beneficiary i's plan cohort, and let k_m^0 denote the initial plan selected by beneficiary $m \in \mathcal{P}_i$. I construct

the leave-one-out initial choice probability for each beneficiary i and each plan k:

$$Z_{ik}^{plan} = \frac{1}{|\mathcal{P}_i| - 1} \sum_{m \in \mathcal{P}_i \setminus \{i\}} \mathbb{1}\{k_m^0 = k\} \tag{13}$$

I use these probabilities as instruments for the lagged plan choice indicators in Equation (9). First-stage estimates are reported in Appendix Table B.2.

Unobserved Plan Quality. Finally, plan premiums may be endogenous to plan characteristics that are valued by consumers but unobserved by the econometrician—such as pharmacy networks, customer service, or prior authorization practices. If plans with higher unobserved quality charge higher premiums, the estimated price coefficients will be smaller in magnitude or potentially positive. I address premium endogeneity by estimating unobserved plan quality ξ_k using market shares and plan-level instruments. I construct a Hausman instrument for each plan k equal to the average premium of similar plans offered by the same insurer in other regions.[42] These instruments capture variation in premiums associated with insurers' pricing decisions across markets. Their validity requires that premiums of comparable plans in other regions predict a plan's premium but are uncorrelated with its market-specific unobserved quality, conditional on controls. First-stage estimates are reported in Appendix Table B.3.

5.2 Estimation

Insulin Demand. I estimate the mixed logit model in Equation (8) using beneficiaries' observed insulin choices conditional on enrollment in their selected plan. The sample includes beneficiaries who used insulin in the previous year and new insulin users who initiate treatment during the contract year. I pool beneficiary-year observations across 2014–2019 and estimate the model cross-sectionally, treating each beneficiary-year independently.

Let k_i denote beneficiary i's observed insurance plan and j_i their observed insulin choice. Let θ^D denote the vector of drug demand parameters. Conditional on draws ν_i, the probability of choosing insulin j is given by:

$$s_{ij|k_i}(\nu_i; \theta^D) = \frac{\exp\left(\overline{v}_{ij|k_i}(\nu_i; \theta^D)\right)}{\sum_{j' \in \mathcal{J}_i} \exp\left(\overline{v}_{ij'|k_i}(\nu_i; \theta^D)\right)} \tag{14}$$

[42] I classify plans with zero deductible and positive deductible, and consider "similar" plans with the same deductible type. If the insurer does not offer comparable plans in other regions, I use the average premium of plans in other regions with the same deductible type.

I approximate the conditional drug-choice probability by integrating over the population distribution of the random coefficients using $R = 100$ simulated draws.[43]

$$\begin{aligned} s_{ij|k_i}(\theta^D) &= \int s_{ij|k_i}(\nu_i; \theta^D)\, dF(\nu_i) \\ &\approx \hat{s}_{ij|k_i}(\theta^D) \\ &= \frac{1}{R} \sum_{r=1}^{R} s_{ij|k_i}(\nu_i^{(r)}; \theta^D) \end{aligned} \tag{15}$$

I estimate the drug demand parameters by maximizing the log-likelihood:

$$\hat{\theta}^D = \arg\max_{\theta^D} \sum_i \log \hat{s}_{ij_i|k_i}(\theta^D) \tag{16}$$

I use two approaches to incorporate drug cohort instruments from Equation (12) to address endogeneity in the lagged choice indicators in drug demand. My primary specification follows the two-stage residual inclusion (2SRI) approach from Terza et al. (2008) and adds the first-stage residuals directly into the utility function as a control variable. I also implement the two-stage predictor substitution (2SPS) approach from Newey (1987) which replaces the endogenous variable in the second stage with the fitted values from the first stage. I report the 2SRI estimates in Table 11 and the 2SPS estimates in Appendix Table B.4.

I find significant heterogeneity in sensitivity to insulin OOP costs, which implies variation in their responses to formulary changes. However, estimated standard deviations of the brand random coefficients are not statistically significant. The estimated parameters also indicate substantial drug switching costs. On average, beneficiaries are willing to pay $82 per month to avoid switching insulin brands and $106 to avoid switching forms; they are willing to pay $2 per month to avoid switching molecules. These switching costs are additive, so beneficiaries moving from the Lantus pen to the Levemir pen face both brand and molecule switching costs, implying total switching costs of $84 per month. Beneficiaries switching from the Lantus pen to the Basaglar pen face switching costs of $82 per month, since both brands contain the same insulin analog (Insulin Glargine U-100). These estimates suggest substantial costs to switch between Lantus and the follow-on product Basaglar, on par with switching costs between brands of different molecules.

[43] Beneficiaries may select plans based on their unobserved drug preferences, leading to differences between the distribution of drug tastes conditional on plan enrollment and the population distribution ($F(\nu_i \mid k_i) \neq F(\nu_i)$). The current specification integrates over the population distribution and does not account for this selection on unobserved heterogeneity in drug tastes. Future analyses jointly estimate drug and plan demand parameters to account for this source of selection.

Table 11: Insulin Demand Parameters: 2SRI Specification

	Mean	SD
Insulin OOP cost ($100s)	-0.262	0.121
	(0.002)	(0.002)
Form = Prefilled Pen	1.340	0.013
	(0.012)	(0.007)
Brand = Basaglar	-0.223	0.083
	(0.026)	(0.215)
Brand = Lantus	0.760	0.003
	(0.013)	(0.018)
Brand = Levemir	-0.041	0.018
	(0.012)	(0.016)
Brand = Toujeo	-0.538	
	(0.014)	
Same Form as Previous Year	3.327	
	(0.037)	
Same Brand as Previous Year	2.598	
	(0.066)	
Same Molecule as Previous Year	0.069	
	(0.064)	

Notes: This table reports the estimated parameters from the insulin demand model described in Section 4.2.1. The first column reports the mean coefficients, and the second column reports the standard deviations for random coefficients. The indicators for same insulin form, brand, and molecule as the previous year are treated as endogenous, and I use cohort initial choice probabilities as instruments for these lagged choice indicators. The 2SRI specification includes these indicators and their first-stage residuals as additional controls; residual coefficients are omitted. First-stage estimates are reported in Appendix Table B.1. The reported standard errors do not account for uncertainty in the first-stage estimates.

Logit Inclusive Value. I use the estimated insulin demand parameters from Table 11 to calculate the logit inclusive value for each beneficiary and every plan in their choice set. Because beneficiaries know their realized random coefficients for insulin products when choosing a plan, I incorporate the distribution of drug tastes directly into the plan demand estimation.

I simulate $R = 50$ joint draws of the random coefficients for insulin and plan characteristics for each beneficiary. Let $\nu_i^{(r)}$ denote the insulin taste shocks and $\omega_i^{(r)}$ the plan premium taste shock for beneficiary i in draw r. I assume drug and plan taste shocks are independent, and I use the same set of simulated draws for all plans in the beneficiary's choice set. I

calculate the logit inclusive value for each beneficiary i, draw r, and plan k:

$$v_{ik}^{*}(\nu_i^{(r)};\hat{\theta}^D) = \log\left(\sum_{j\in\mathcal{J}_i}\exp\left(\overline{v}_{ij|k}\left(\nu_i^{(r)};\hat{\theta}^D\right)\right)\right) \tag{17}$$

Plan Demand. Two features of plan demand are important for estimation: (1) choices depend on individual-level characteristics, including prior plan choices and OOP costs, and (2) premiums may be endogenous to unobserved plan quality. For example, time-varying plan characteristics such as pharmacy networks and claims denials may be correlated with premiums but are not observed in my data.

I estimate plan demand using the Conformant Likelihood Estimator with Exogeneity Restrictions (CLEER) developed by Grieco et al. (2025). CLEER combines individual-level choice information and aggregate plan market shares with exogeneity restrictions on unobserved plan quality. Specifically, I minimize an objective function consisting of individual and aggregate log-likelihood components and a penalty for violations of the exogeneity restrictions:

$$(\hat{\boldsymbol{\theta}}^P, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) = \arg\min_{\theta^P,\delta,\beta}\left\{-\log L^{micro}(\boldsymbol{\theta}^P,\boldsymbol{\delta}) - \log L^{macro}(\boldsymbol{\theta}^P,\boldsymbol{\delta}) + \hat{\chi}(\boldsymbol{\beta},\boldsymbol{\delta})\right\} \tag{18}$$

where $\boldsymbol{\delta}$ is a vector of mean plan utilities, $\boldsymbol{\beta}$ contains the mean coefficients for premiums and zero-deductible plans, and $\boldsymbol{\theta}^P$ contains the remaining plan demand parameters governing individual-specific utility, including premium preference heterogeneity, expected OOP costs for non-insulins, logit inclusive value of insulins, and plan switching costs. L^{micro} and L^{macro} denote the individual and aggregate likelihood components, respectively, and $\hat{\chi}$ is a penalty function for violations of product exogeneity restrictions:

$$\hat{\chi}(\beta,\delta) = \frac{1}{2}\hat{m}(\beta,\delta)^\top \hat{W}\hat{m}(\beta,\delta) \qquad \text{where } \hat{m}(\beta,\delta) = \sum_{k\in\mathcal{K}}\mathbf{z}_k\left(\delta_k - \beta^\top\mathbf{x}_k\right) \tag{19}$$

where $\mathbf{x}_k$ is a vector of plan characteristics including annual premium and zero-deductible coverage. $\mathbf{z}_k$ contains Hausman instruments for plan premiums.

I construct observed market shares using observed choices for everyone in the beneficiary sample. Therefore, the individual and aggregate likelihood components jointly identify consumer preferences and plan mean utilities from observed choices and market shares.[44] The exogeneity restrictions require instruments to be uncorrelated with unobserved plan quality,

[44] In practice, the CLEER estimator may combine market share data with observed choices for a subset of consumers. The individual likelihood identifies coefficients on individual-level variables and preference heterogeneity, and the aggregate likelihood identifies mean plan utilities from market shares.

allowing identification of the effects of premiums and other observed plan characteristics. I use Hausman instruments described in Section 5.1 and implement CLEER using the Julia package Grumps developed by Grieco et al. (2025).[45] Additional details on the likelihood components and exogeneity restrictions are provided in Appendix Section D.

To address endogeneity in prior plan choices, I use cohort initial choice probabilities from Equation (13) as instruments for lagged plan choices. My primary specification uses two-stage residual inclusion (2SRI), including both the endogenous indicator and its first-stage residual in plan utility. Table 12 reports the 2SRI plan demand estimates, using logit inclusive values constructed from the corresponding 2SRI insulin demand estimates.[46] I also estimate a 2SPS specification that replaces the endogenous indicator with its first-stage fitted value (see Appendix Table B.5).

Table 12: Plan Demand Parameters: 2SRI Specification

	Mean	SD
Annual Premium ($100s)	-0.134	0.018
	(0.010)	(0.002)
Logit Inclusive Value of Insulins	0.020	
	(0.002)	
Non-Insulin OOP Cost ($100s)	-0.005	
	(0.001)	
Same Plan as Previous Year	3.326	
	(0.017)	
Zero-deductible plan	1.178	
	(0.068)	

Notes: This table reports the estimated parameters from the plan demand model described in Section 4.2.2. The first column reports the mean coefficients, and the second column reports the standard deviation for random coefficients. The logit inclusive value of insulins from Equation (11) is calculated using the 2SRI insulin demand estimates in Table 11. The indicator for same plan as the previous year is treated as endogenous, and I use cohort initial choice probabilities as an instrument. The 2SRI specification of plan demand includes the same plan indicator as well as its first-stage residual an as additional control; the coefficient for the residual is omitted. First-stage estimates are reported in Appendix Table B.2. The reported standard errors do not account for uncertainty in estimated parameters from the first stage, or estimated drug demand parameters used to construct the logit inclusive value.

Beneficiaries exhibit significant heterogeneity in sensitivity to plan premiums. The coefficient on the logit inclusive value term is positive, which indicates that beneficiaries are more likely to choose plans that provide greater value in terms of insulin coverage. The coefficient on non-insulin OOP costs is negative, so beneficiaries are less likely to choose

[45] https://nittanylion.github.io/Grumps.jl/dev/

[46] Future analyses will use bootstrapping to account for uncertainty across estimation stages.

plans if they expect to incur higher drug costs, all else equal. In addition, beneficiaries have strong preferences for plans with zero-deductible coverage.

On average, beneficiaries are willing to pay $208 per month to avoid switching prescription drug plans. This estimate is somewhat higher than previous papers that estimate Part D plan switching costs for a representative sample of SAPD beneficiaries.[47] However, Heiss et al. (2021) estimate plan switching costs ranging from $20 per month to $230 per month, reflecting differences in attention and acuity across beneficiaries. My estimate falls within this range, although my model of plan choice does not account for inattention.

Beneficiaries using long-acting insulin have substantial, ongoing prescription drug needs, and their plan switching costs may differ from those of the broader Part D population; understanding the sources of these differences is left for future research. Importantly, choice frictions for the subset of beneficiaries taking insulin shape firm incentives and the equilibrium coverage decisions for insulins.

6 Solving the Contracting Game

I simulate the contracting game between Part D insurers and insulin manufacturers using the estimated consumer demand parameters. First, I construct insurer and manufacturer payoffs under alternative joint formulary profiles. I then identify the contract-efficient formulary profile and construct candidate rebate schedules that support that profile. Finally, I evaluate whether these schedules are robust to manufacturer deviations and report the simulated contracting outcomes.

6.1 Firm Payoffs

Insurer Payoffs. Insurers collect premium revenue from enrolled beneficiaries and premium subsidies from the federal government, and they incur costs for enrollees' prescription drug utilization.[48] Different formulary arrangements affect who enrolls in each plan, which insulins they use, and the costs paid by the insurer. I calculate expected insurer profits under each joint formulary profile a using simulated plan and insulin choice probabilities.

Let $s_{ik}(a)$ denote the probability that beneficiary i enrolls in plan k under formulary profile a. Since beneficiaries' realized insulin preferences affect both plan and drug choices,

[47] Polyakova (2016) finds average switching costs around $100 per month. Yeo and Miller (2018) find switching costs of $140 per month.

[48] Insurers also receive risk adjustment payments depending on the risk profile of enrolled beneficiaries and reinsurance payments for very high catastrophic spending. The current version does not account for these sources of revenue in insurers' profit functions.

I calculate the joint probability of choosing plan k and insulin product j for each draw r:

$$s_{ijk}(a) = \frac{1}{R}\sum_{r=1}^{R} s_{ik}^{(r)}(a) \cdot s_{ij|k}^{(r)}(a) \tag{20}$$

Let $\pi_f(a)$ denote insurer f's intrinsic payoff before negotiated insulin rebates, summed across all offered plans $k \in \mathcal{K}_f$:

$$\begin{aligned}\pi_f(a) = &\sum_{k\in\mathcal{K}_f}\sum_{i} s_{ik}(a)\left[prem_k + sub_k - cost_{ik}^{N}\right] \\ &- \sum_{k\in\mathcal{K}_f}\sum_{i}\sum_{j\in\mathcal{J}} s_{ijk}(a)\left[cost_{ijk}^{I}(a) \times \overline{q}_i\right]\end{aligned} \tag{21}$$

The first term is expected premium revenue and subsidies net of spending on non-insulin drugs. For each enrollee, plan k receives a beneficiary premium $prem_k$ and a federal premium subsidy sub_k and incurs expected non-insulin costs $cost_{ik}^{N}$.[49] The second term is expected insulin spending. If beneficiary i enrolls in plan k and uses insulin j, the plan pays $cost_{ijk}^{I}(a)$ per day for $\overline{q}_i$ days of insulin. I use beneficiary i's observed insulin utilization (measured in days supplied).

Manufacturer Payoffs. Manufacturers earn revenue from insulin sales and incur production costs. For each formulary profile, I calculate expected insulin utilization across all plans using simulated plan and insulin choice probabilities. Let $\phi_j(a)$ denote manufacturer j's intrinsic payoff before negotiated insurer rebates:

$$\phi_j(a) = \sum_{k}\sum_{i} s_{ijk}(a) \times \overline{q}_i\left(\overline{p}_j - c_j\right) \tag{22}$$

where $\overline{p}_j$ is the median pharmacy price of insulin j per day and c_j is its production cost per day. I calibrate marginal production costs for each molecule-form using the midpoint of the range of estimated costs reported by Barber et al. (2024).[50]

Each insurer's formulary choices affect both substitution across competing insulins within plans and enrollment across insurers. Thus, insurer and manufacturer intrinsic payoffs depend on the joint formulary profile a. Rebate contracts $t_{jf}(a)$ specify transfers paid by manufacturer j to insurer f under each profile. Insurer and manufacturer *net profits* depend on the rebate contracts and formulary profile; these are both equilibrium objects from the

[49] The sample excludes beneficiaries who receive additional premium or cost-sharing subsidies, so beneficiary premiums and direct premium subsidies do not vary across beneficiaries i.

[50] The midpoint production costs per 30-day supply are: \$6.86 for Lantus vials, \$19.67 for Lantus and Basaglar pens, \$25.14 for Levemir vials, and \$37.27 for Levemir pens. I use the marginal production cost per day in Equation (22)

contracting game:

$$\begin{aligned} \Pi_f(a) &= \pi_f(a) + \sum_{j \in \mathcal{J}} t_{jf}(a) \\ \Phi_j(a) &= \phi_j(a) - \sum_{f \in \mathcal{F}} t_{jf}(a) \end{aligned} \tag{23}$$

6.2 Game Simulation

I simulate the contracting game in 2019 for three insurers and three manufacturers. The insurers are Humana, SilverScript, and UnitedHealth, which collectively account for approximately 70% of enrollment in the beneficiary sample. I include 175 plans offered by these insurers in 2019 across 27 Part D regions.

I restrict the strategic manufacturers to three insulin brands: Lantus, Levemir, and Basaglar, which account for 77% of long-acting insulin utilization in 2019. These brands comprise five insulin products: Lantus pen and vial, Levemir pen and vial, and Basaglar pen.

Each insurer chooses a single formulary arrangement specifying the tier placement of each insulin brand. The selected formulary applies to all plans offered by that insurer, and both pen and vial forms of a brand are assigned the same tier. Insurers have three tiers to which they can assign insulins: preferred (P), non-preferred (N), or excluded (E). I impose an additional feasibility restriction requiring each insurer to place at least one insulin on the preferred tier. This restriction is motivated by Part D requirements that non-standard benefit designs provide sufficient drug coverage.[51]

Under these rules, each insurer has 19 feasible formulary choices, and the joint action space with three insurers contains $19^3 = 6{,}859$ formulary profiles.[52] Each manufacturer offers a bilateral rebate schedule to each insurer, specifying a nonnegative transfer for every joint formulary profile. In the 3×3 game, there are nine manufacturer–insurer pairs, so the contracting space comprises nine bilateral contracts, each specifying 6,859 transfer values.

For every feasible formulary profile, I use the OOP calculator to get the implied insulin costs for each beneficiary, plan, and drug combination. I simulate plan and drug choice probabilities using $R = 30$ draws of each beneficiary's random coefficients. These choice probabilities determine insurer and manufacturer payoffs under each formulary profile.

[51] Standard benefit plans have a flat coinsurance rate for each tier that applies to all brand drugs—25% after the deductible in 2019. Plans with tiered formularies should provide similar generosity of coverage on average, but the coinsurance rate for a specific drug may be more or less than 25%.

[52] With three brands and three possible tiers, each insurer has $3^3 = 27$ potential formulary choices, but $2^3 = 8$ options do not have any preferred insulin, leaving 19 feasible formularies.

6.3 Candidate Equilibrium Construction

Following Weber and Xiong (2006, 2026), I construct candidate rebate schedules that satisfy the agent maximization (AM), weak truthfulness (WT), and principal cost minimization (PM) conditions at the contract-efficient formulary profile. These conditions do not uniquely determine bilateral rebate schedules. I therefore parameterize the allocation of surplus across firms and bilateral relationships to construct one set of candidate contracts.

First, I identify the joint formulary profile a^* that maximizes contract surplus:

$$a^* \in \arg\max_{a \in \mathcal{A}} W(a), \quad \text{where } W(a) = \sum_{f \in \mathcal{F}} \pi_f(a) + \sum_{j \in \mathcal{J}} \phi_j(a). \tag{24}$$

The efficient profile generates a surplus gain relative to every alternative joint formulary profile:

$$S(a) = W(a^*) - W(a) \geq 0, \qquad \forall a \in \mathcal{A}. \tag{25}$$

Second, I initially allocate the surplus difference $S(a)$ across manufacturers and insurers to satisfy WT and AM conditions. To support the contract-efficient equilibrium, WT requires that manufacturers weakly prefer the efficient formulary profile given their existing contracts; AM rules out profitable insurer deviations. These conditions thus compare firms' net profits across alternative formulary profiles and impose constraints on the allocation of excess surplus needed to support the efficient outcome.

Under the initial allocation, manufacturers and insurers split the surplus gain 50/50. Within each group, the allocation to each manufacturer j or insurer f is proportional to its drug utilization share or enrollment market share, evaluated at the efficient profile:

$$\rho_j = \frac{1}{2}\frac{q_j}{Q}, \qquad \rho_f = \frac{1}{2}\frac{e_f}{E}, \qquad \text{where } \sum_j \rho_j + \sum_f \rho_f = 1 \tag{26}$$

where q_j is utilization of drug j, e_f is enrollment of insurer f, and Q and E are total insulin utilization and enrollment, respectively, all evaluated at a^*. This sharing rule allocates the *surplus difference* as slack in firms' AM and WT incentive constraints, rather than splitting the total surplus generated for a given profile.

Let $\Delta_{jf}(a)$ denote the *difference* in bilateral rebates paid by manufacturer j to insurer f at profile a, relative to the efficient profile:

$$\Delta_{jf}(a) = t_{jf}(a) - t_{jf}(a^*). \tag{27}$$

These bilateral incentive contracts specify how rebates between each manufacturer–insurer

pair must vary across formulary profiles to incentivize player strategies that support the efficient outcome. The initial sharing rule in (26) restricts each firm's total change in incentive payments across its bilateral relationships, relative to the efficient profile:

$$\begin{aligned} \text{Mftr } j: &\quad \sum_f \Delta_{jf}(a) = \phi_j(a) - \phi_j(a^*) + \rho_j S(a) &\quad \forall a \in \mathcal{A} \\ \text{Insurer } f: &\quad \sum_j \Delta_{jf}(a) = \pi_f(a^*) - \pi_f(a) - \rho_f S(a) &\quad \forall a \in \mathcal{A} \end{aligned} \tag{28}$$

Each firm receives a share of the additional surplus generated at the efficient formulary profile. Therefore, manufacturers prefer the efficient profile because they pay less in rebates relative to their gross profits (WT), and insurers prefer it because they receive more in rebates relative to their gross profits (AM).[53]

Third, I distribute the firm-level incentives in (28) across bilateral manufacturer–insurer relationships. Since the firm-level constraints do not uniquely determine bilateral incentives, I select an allocation that minimizes deviations from utilization-based benchmarks. I calculate each manufacturer's share of insulin utilization within an insurer at the efficient profile and use these fixed shares to construct the bilateral benchmarks.

Last, I re-allocate any remaining slack in insurers' incentive constraint (AM) across bilateral relationships so insurers are indifferent across all formulary profiles. To pin down rebate *levels* I shift each bilateral schedule, so the manufacturer's minimum rebate over unilateral insurer deviations is zero (PM). This normalization preserves insurers' incentives across formularies and ensures that manufacturers pay just enough to induce the insurer's equilibrium formulary choice. Appendix Section C.1 details the sharing rule based on utilization, the allocation of remaining insurer incentive slack, and the normalization of rebate levels.

6.4 Manufacturer Offer Deviations

Following Weber and Xiong (2026), I check whether any manufacturer could profit by revising its offers to one or more insurers, holding competing manufacturers' offers fixed. For each feasible formulary profile, I compare the manufacturer's target net profit with its profit at that profile after paying each insurer the minimum rebate needed to make its formulary a best response.

The PD-pure condition requires that no manufacturer gain from a deviation followed by a pure-strategy equilibrium.[54] Where the initial contracts fail this test, I consider ad-

[53] This satisfies a stronger condition than AM because insurers prefer the efficient outcome to *any* alternative joint formulary profile, not only their own unilateral deviations.

[54] A stronger test, PD-mix, also covers mixed-strategy continuation equilibria.

justments to off-equilibrium rebates that raise the cost of a deviation while preserving AM, WT, and PM. If the adjusted schedules satisfy PD, they support the same contract-efficient formulary profile a^* and same bilateral transfers under a^* as the initial contracts, since these adjustments only occur to off-equilibrium profiles.

I illustrate the verification and adjustment procedures in a 2×2 game with two insurers and two insulins. I simulate choices in a market that includes only Humana and SilverScript plans and Lantus and Levemir products. In this game, there are 25 joint formulary profiles under status quo rules and nine profiles when exclusion is prohibited. The initial candidate rebates following the construction above do not satisfy the PD-pure condition for one manufacturer. Adjusting off-equilibrium rebates can raise the cost of a deviation enough to eliminate gains under pure-strategy continuations. The resulting contracts constitute a WTE equilibrium with no profitable manufacturer deviations to a pure-strategy NE in the insurer subgame. Bilateral rebates paid under the contract-efficient profile are identical under the initial and adjusted schedules, since adjustments occur only to *off*-equilibrium profiles. Appendix Section C.2 provides detailed verification and adjustment procedures, including results from the 2×2 game.

6.5 Simulation Results

Table 13 reports the contract-efficient formularies and the rebates paid under the candidate schedules in the 3×3 game. Under status quo rules, all three insurers place Levemir on the preferred tier and exclude Basaglar. Humana and UnitedHealth place Lantus on the non-preferred tier, and SilverScript excludes it. At these formularies, rebates under the candidate contracts average 29% of gross insulin revenue.

When Lantus and Levemir are both included on an insurer's formulary, both drug manufacturers pay positive rebates of approximately 28%. When an insurer excludes Lantus, its rival Levemir pays slightly higher rebates (32%), and Lantus pays no rebates. Basaglar pays a positive rebate only when Lantus is also excluded.

The initial candidate schedules satisfy AM, WT, and PM but fail PD-pure—At least one manufacturer can earn higher net profits by revising its rebate offers when insurers respond by choosing a different pure-strategy formulary profile. In future analyses, I will search for adjusted rebate schedules that eliminate these profitable deviations by increasing rebates at off-equilibrium profiles, making manufacturer deviations more costly. If such schedules exist, they support the same formularies reported in Panel A and same equilibrium rebates reported in Panel B.

Table 13: Simulated Rebate Contracts (3×3 Game)

	Humana	SilverScript	UnitedHealth	All Insurers
Panel A: Formulary Tiering Choices				
Lantus	N	E	N	–
Levemir	P	P	P	–
Basaglar	E	E	E	–
Panel B: Drug Manufacturer Rebates (% of Gross Revenue)				
Lantus	28%	0%	27%	27%
Levemir	29%	32%	28%	31%
Basaglar	0%	20%	0%	16%
All insulins	28%	31%	28%	29%

Notes: Panel A reports the contract-efficient formulary placement of each insulin by insurer: preferred (P), non-preferred (N), or excluded (E). Panel B reports manufacturer rebates paid to each insurer under candidate rebate contracts for formulary tier choices described in Panel A. Rebates are reported as percentage of manufacturer's gross revenue within each insurer. The final row reports average rebates received by each insurer, weighted by brand utilization. The final column reports average rebates paid by each manufacturer, weighted by insurer utilization.

7 Counterfactual Policies

I evaluate policies that expand access to long-acting insulins by restricting insurers' formulary choices or limiting beneficiary OOP costs. For each counterfactual, I calculate consumer demand and firm payoffs under the feasible formulary profiles and construct candidate rebate schedules using the same surplus-sharing rule as in the status quo. I consider three policies: (1) prohibiting drug exclusions, (2) requiring coverage of follow-on insulins, and (3) capping beneficiary OOP costs at $35 per month.

7.1 Prohibiting Drug Exclusions

I extend Medicare's protected drug class policy to long-acting insulins, requiring Part D plans to cover every insulin in the therapeutic class. Under this policy, insurers can assign drugs to preferred (P) or non-preferred (N) tiers but cannot exclude any insulin. This restriction reduces the number of feasible formularies per insurer from 19 to 7, leaving 343 joint formulary profiles in the 3×3 game.

7.2 Required Coverage of Follow-on Insulins

The second policy requires insurers to cover follow-on insulins, including biosimilars and interchangeable biosimilars. Basaglar, a follow-on version of Lantus approved through the FDA's 505(b)(2) pathway, entered the market in December 2016. I simulate this policy using 2019 formularies by requiring all insurers to cover Basaglar, while allowing them to assign it to either the preferred or non-preferred tier. Insurers retain the ability to exclude Lantus or Levemir. This restriction leaves 14 feasible formularies per insurer and 2,744 joint formulary profiles.

7.3 Cap on Beneficiary OOP Costs

The third counterfactual implements the $35 cap on monthly OOP costs for insulin established by the Inflation Reduction Act (IRA), which took effect in Medicare Part D in January 2023. The cap applies to all types of insulin, including long-acting and short-acting insulins, but only to products covered by the plan. Insurers retain discretion over which insulin brands to include on their formularies.

To implement the cap, I modify the OOP calculator so that beneficiary cost sharing for a 30-day supply of any covered insulin j in plan k is

$$\widetilde{cs}_{jkb} = \min\{35, cs_{jkb}\}, \tag{29}$$

where cs_{jkb} denotes the original cost-sharing amount in benefit phase b. Results for this counterfactual will be included in future versions. I will compare simulated formulary choices and insulin market shares under the cap with actual outcomes observed in 2023-2024.

7.4 Counterfactual Results

Table 14 compares the contract-efficient formulary profiles and candidate rebate outcomes under status quo rules, a prohibition on exclusions, and required coverage of Basaglar. I report percentage changes in contract surplus, insurer profits, manufacturer profits, and consumer surplus relative to the status quo.

Under the status quo, all three insurers place Levemir in the preferred tier and exclude Basaglar. Humana and UnitedHealth put Lantus in the non-preferred tier, and Silverscript excludes Lantus. Under both counterfactual policies insurers place Levemir on the preferred tier and both Lantus and Basaglar on the non-preferred tier.

Relative to the status quo, both policies increase consumer surplus by 2% and reduce aggregate insurer and manufacturer surplus by 2%. Thus, broader formulary coverage gener-

Table 14: Counterfactual Policy Results (3×3 Game)

	Observed Formularies	Status Quo Simulation	No Exclusion Counterfactual	Biosimilar Incl Counterfactual
Panel A: Insulin Formularies (LAN, LEV, BAS)				
Humana	(P, P, E)	(N, P, E)	(N, P, N)	(N, P, N)
SilverScript	(E, P, P)	(E, P, E)	(N, P, N)	(N, P, N)
UnitedHealth	(P, P, E)	(N, P, E)	(N, P, N)	(N, P, N)
Panel B: Rebates and Surplus				
Average rebate (%)	—	29%	6%	25%
Change in total surplus (%Δ)	—	0%	-2%	-2%
Change in insurer profit (%Δ)	—	0%	-65%	-6%
Change in mftr profit (%Δ)	—	0%	+30%	+1%
Change in consumer surplus (%Δ)	—	0%	+2%	+2%

Notes: Panel A reports observed and contract-efficient formulary placements for Lantus (LAN), Levemir (LEV), and Basaglar (BAS), where P denotes preferred, N non-preferred, and E excluded. Panel B reports average manufacturer rebates as a percentage of gross insulin spending and percentage changes relative to the status quo. Contract surplus is the sum of insurer and manufacturer profits before rebates, while firm profits are net of rebates. The reported rebate schedules satisfy AM, WT, and PM but have not been verified to satisfy PD.

ates modest consumer gains while reducing the total surplus available to firms in the vertical contracting relationship.

Despite generating identical formularies and consumer outcomes, the two counterfactual policies have substantially different effects on manufacturer rebates. Prohibiting exclusion reduces average rebates from 29% to 6% of gross insulin spending. In contrast, requiring coverage of Basaglar reduces rebates only to 25%. These differences translate into substantial changes in the distribution of profits between insurers and manufacturers. When all exclusions are prohibited, insurer net profits fall by 65%, while manufacturer net profits increase by 30%. Requiring only Basaglar coverage has much smaller effects: insurer profits fall by 6%, and manufacturer profits increase by 1%.

The difference arises because the policies restrict insurers' off-equilibrium formulary choices to different degrees. Under both policies, the contract-efficient formulary profile is identical, so consumer demand and firm payoffs before rebates are also identical. However, prohibiting all exclusions removes more of insurers' credible threat points than requiring coverage of Basaglar alone. These restrictions reduce insurers' leverage in contracting so they are not able to extract high rebates from manufacturers.

Consequently, policies that generate the same formulary coverage and consumer outcomes can have very different effects on rebate payments. In the contracting model, rebates depend not only on the selected formulary profile but also on firms' payoffs under alternative profiles and the resulting incentives embedded in rebate schedules.

These results are based on candidate rebate schedules satisfying AM, WT, and PM condi-

tions of equilibrium, but are not robust to manufacturer contract deviations in the first stage of the game. Therefore, the reported rebate and profit outcomes should be interpreted as outcomes of the parameterized candidate construction rather than robust WTE predictions.

8 Conclusion

This paper studies how competition between insurers and drug manufacturers shapes formulary coverage and manufacturer rebates in Medicare Part D. I model formulary design as a multi-principal, multi-agent contracting game in which manufacturers offer formulary-contingent rebate schedules to competing insurers, and insurers choose formularies that determine consumers' drug and insurance plan choices. Using Medicare claims data, I estimate demand for Part D plans and long-acting insulins and construct firm payoffs under alternative formulary profiles.

I use the contracting model to evaluate real-world policies to expand formulary coverage of competing insulins. Counterfacual simulations that prohibit drug exclusions or require coverage of follow-on insulins generate the same contract-efficient formularies, increasing consumer surplus by 2% and reducing aggregate firm surplus by 2%. However, the two policies have substantially different effects on manufacturer rebates. Average rebates fall from 29% under the status quo to 6% when all exclusions are prohibited, compared with 25% when only Basaglar coverage is required. These reductions in rebates translate into lower insurer net profits and higher manufacturer net profits.

These findings highlight the importance of off-equilibrium formulary choices in determining negotiated rebates. Even when policies generate identical formularies and consumer outcomes, they can have different effects on contracting leverage because they restrict insurers' ability to threaten exclusion of competing drugs. Restrictions that expand formulary coverage can therefore reduce manufacturer rebates without generating commensurate increases in consumer surplus. The results illustrate a tradeoff between patient access and insurers' ability to negotiate lower net drug costs.

Several limitations remain. First, I construct a single set of candidate contracts from a much larger class of contracts that support the contract-efficient outcome. My starting point is an equal split between manufacturers and insurers of the additional surplus they generate at the efficient outcome. Second, I will examine whether adjusting only the rebates promised at off-equilibrium profiles can deter manufacturer contract deviations. Future analyses will examine which division of surplus are consistent with equilibrium that is robust to such deviations and how equilibrium rebates vary as manufacturers receive a larger or smaller share of those gains. Finally, I will estimate plan demand for both insulin users and non-

users to quantify how changes in net drug costs consumer surplus from plan premiums and government spending on Part D subsidies. These extensions will allow a more complete assessment of the distributional and welfare effects of formulary regulations.

References

Abaluck, Jason and Jonathan Gruber (2016). "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance". *American Economic Review* 106(8): pp. 2145–2184.

Adams, Christopher, Meagan Madden, and Alexander Olssen (2025). "Formulary Restrictions and Drug Rebates: Evidence from Medicare's Protected Drug Classes". Working paper.

Barber, Melissa J., Dzintars Gotham, Helen Bygrave, and Chris Cepuch (2024). "Estimated Sustainable Cost-Based Prices for Diabetes Medicines". *JAMA Network* 7(3).

Bernheim, B. Douglas and Michael D. Whinston (1986). "Menu Auctions, Resource Allocation, and Economic Influence". *QJE* 101(1): pp. 1–31.

Conti, Rena, Brigham R. Frandsen, Michael Powell, and James B. Rebitzer (2026). "Common Agent or Double Agent? Pharmacy Benefit Managers in the Prescription Drug Market". *RESTAT*. Forthcoming.

Coughlin, Maura (2026). "Insurance Choice with Non-Monetary Plan Attributes: Limited Consideration in Medicare Part D". *The Review of Economics and Statistics*. Forthcoming.

Feng, Josh and Luca Maini (2024). "Demand Inertia and the Hidden Impact of Pharmacy Benefit Managers". *Management Science* 70(12): pp. 8940–8961.

Ghili, Soheil (2022). "Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance". *Marketing Science* 41(3): pp. 501–527.

Grieco, Paul L. E., Charles Murry, Joris Pinkse, and Stephan Sagl (2025). "Optimal Estimation of Discrete Choice Demand Models with Consumer and Product Data". Working paper.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou (2021). "Inattention and Switching Costs as Sources of Inertia in Medicare Part D". *American Economic Review* 111(9): pp. 2737–2781.

Ho, Kate and Robin S. Lee (2019). "Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets". *AER* 109(2): pp. 473–522.

Ho, Kate and Robin S. Lee (2026). "Contracting Over Pharmaceutical Formularies and Rebates". *Econometrica* 94(3): pp. 689–728.

Jackson, Matthew O. and Simon Wilkie (2005). "Endogenous Games and Mechanisms: Side Payments Among Players". *Review of Economic Studies* 72(2): pp. 543–566.

Kakani, Pragya, Michael Anne Kyle, Amitabh Chandra, and Luca Maini (2024). "Medicare Part D Protected-Class Policy Is Associated With Lower Drug Rebates". *Health Affairs* 43(10): pp. 1420–1427.

Kyle, Michael Anne, Pragya Kakani, Samantha L. Burn, and Luca Maini (2026). "Trends in Prescription Drug Coverage Restrictions in Medicare, Medicaid, and Commercial Insurance Plans, 2011–2019". *Health Affairs Scholar* 4(3): pp. 1–9.

Medicare Payment Advisory Commission (2023). "Assessing Postsale Rebates for Prescription Drugs in Medicare Part D". *June 2023 Report to the Congress: Medicare and the Health Care Delivery System*. Chapter 2.

Newey, Whitney K. (1987). "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables". *Journal of Econometrics* 36(3): pp. 231–250.

Olssen, Alexander and Mert Demirer (2024). "Drug Rebates and Formulary Design: Evidence from Statins on Medicare Part D". Conditionally accepted at *JPE Micro*.

Polyakova, Maria (2016). "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D". *American Economic Journal: Applied Economics* 8(3): pp. 165–195.

Prager, Elena (2015). "Tiered Hospital Networks, Health Care Demand, and Prices". Job market paper.

Prager, Elena and Nicholas Tilipman (2026). *Regulating Out-of-Network Hospital Payments: Disagreement Payoffs, Negotiated Prices, and Access*. R&R at *JPE*.

Prat, Andrea and Aldo Rustichini (2003). "Games Played Through Agents". *Econometrica* 71(4): pp. 989–1026.

Segal, Ilya (1999). "Contracting with Externalities". *QJE* 114(2): pp. 337–388.

Segal, Ilya and Michael D. Whinston (2003). "Robust Predictions for Bilateral Contracting with Externalities". *Econometrica* 71(3): pp. 757–791.

Starc, Amanda and Ashley Swanson (2021). "Preferred Pharmacy Networks and Drug Costs". *AEJ: Economic Policy* 13(3): pp. 406–446.

Strulovici, Bruno and Thomas A. Weber (2009). "Constructing Efficient Equilibria in Games Played Through Agents". Working paper.

Terza, Joseph V., Anirban Basu, and Paul J. Rathouz (2008). "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling". *Journal of Health Economics* 27(3): pp. 531–543.

U.S. Government Accountability Office (2023). *Medicare Part D: CMS Should Monitor Effects of Rebates on Drug Coverage and Spending*. GAO Report 23-107056.

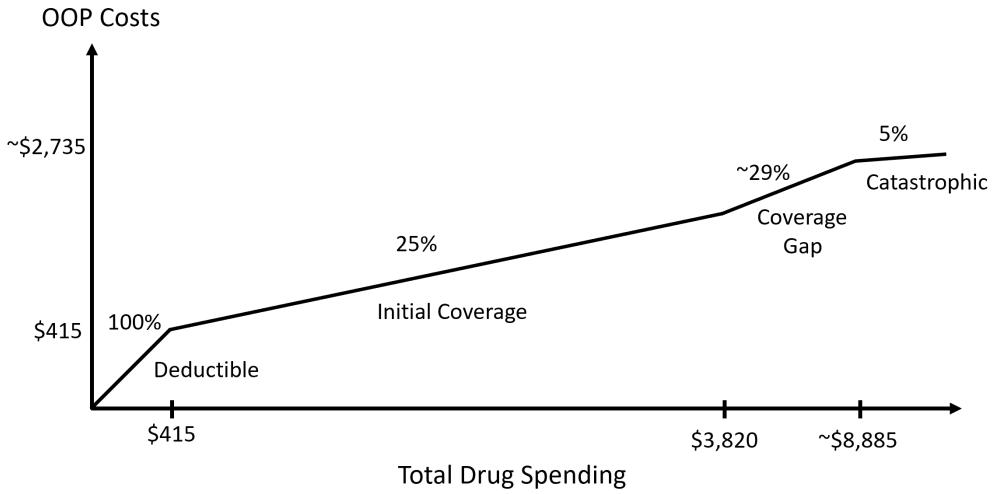
Weber, Thomas A. and Hongxia Xiong (2006). "Efficient Contract Design in Multi-Principal Multi-Agent Supply Chains". Working paper.

Weber, Thomas A. and Hongxia Xiong (2026). "Efficient Multi-Principal Multi-Agent Contract Design". Working paper.

Yeo, Jungwon and Daniel P. Miller (2018). "Estimating Switching Costs with Market Share Data: An Application to Medicare Part D". *International Journal of Industrial Organization* 61: pp. 459–501.

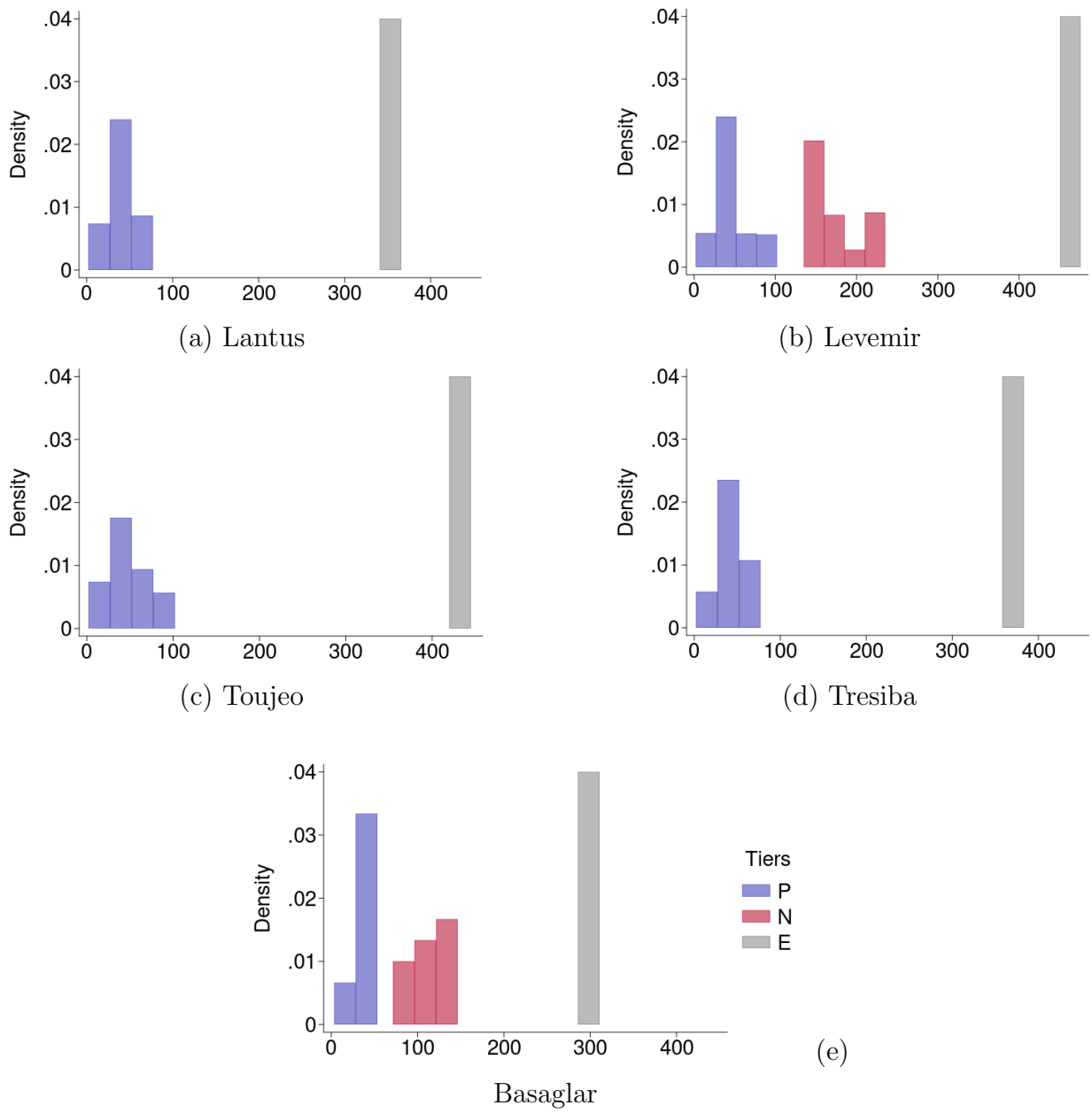
A Additional Figures

Figure A.1: Standard Benefit Design (2019)



Notes: Nonlinear benefit design for standard Part D plan in 2019. The slopes represent coinsurance for each benefit phase. The standard deductible was $415 and the initial coverage limit was $3,820 in total drug expenditures. The catastrophic phase starts at $5,100 in TrOOP costs, which depends on brand and generic utilization in the coverage gap phase; the beneficiary depicted here reaches the catastrophic phase after $2,735 in OOP costs or $8,885 in total drug expenditures.

Figure A.2: OOP Costs by Formulary Tier (2019)



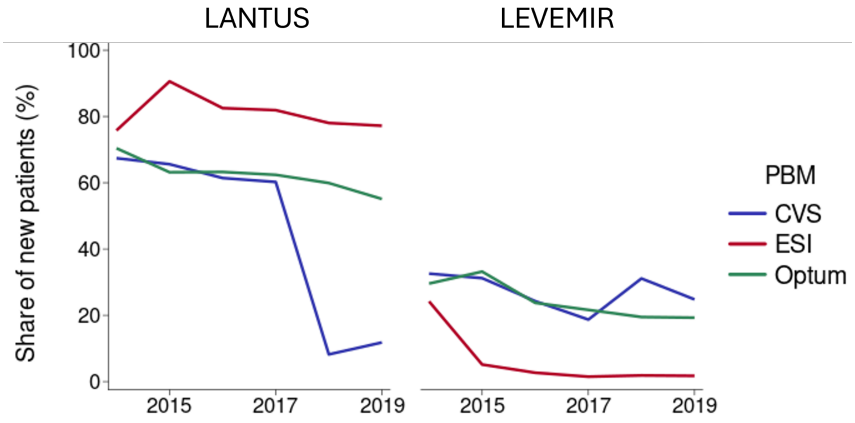
(a) Lantus (b) Levemir

(c) Toujeo (d) Tresiba

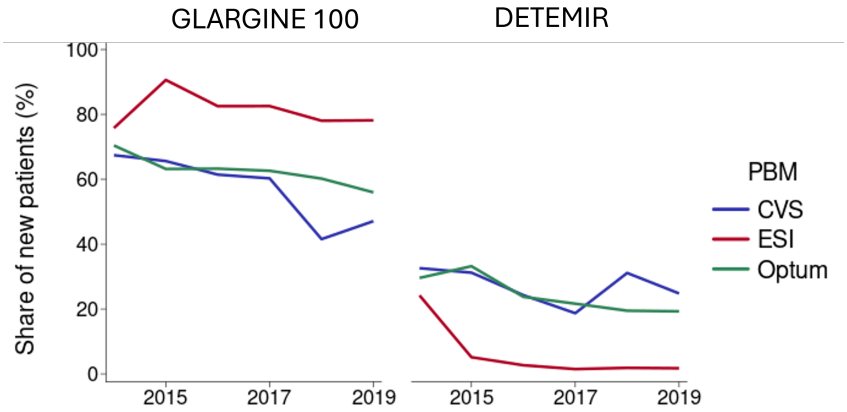
(e) Basaglar

Notes: Distribution of OOP costs ($) in the initial coverage phase for a 30-day supply of prefilled insulin pen for each brand. Bars represent the density of OOP costs across SAPD plans conditional on observed formulary tier.

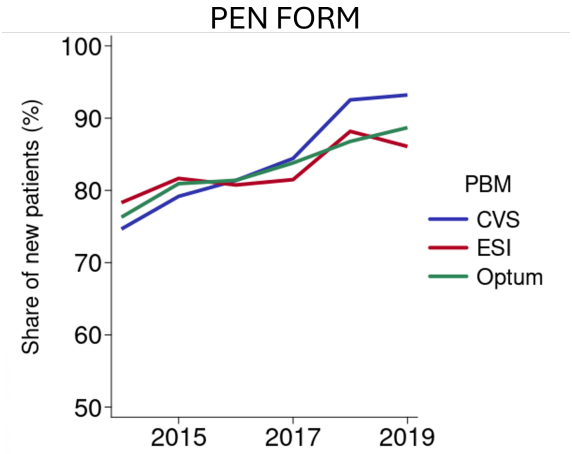
Figure A.3: Initial Insulin Choices by PBM and Diagnosis Year



(a) Brand Choice



(b) Molecule Choice



(c) Form Choice

B Additional Tables

Table B.1: First Stage Estimates for Lagged Drug Choices

	Beneficiary Previous Drug Choice		
	FORM	BRAND	MOLECULE
	(1)	(2)	(3)
Cohort Initial FORM Choice Probability	0.889	0.056	0.062
	(0.001)	(0.001)	(0.001)
Cohort Initial BRAND Choice Probability	-0.132	0.786	0.026
	(0.003)	(0.003)	(0.003)
Cohort Initial MOLECULE Choice Probability	0.162	0.006	0.769
	(0.003)	(0.003)	(0.003)
Insulin OOP cost	0.001	-0.002	-0.001
	(0.000)	(0.001)	(0.000)
Prefilled Pen	0.162	-0.017	-0.021
	(0.001)	(0.001)	(0.001)
Basaglar	-0.085	-0.000	0.011
	(0.002)	(0.002)	(0.002)
Lantus	-0.007	-0.016	-0.002
	(0.001)	(0.001)	(0.001)
Levemir	0.004	-0.007	-0.000
	(0.001)	(0.001)	(0.001)
Toujeo	0.006	-0.018	-0.012
	(0.001)	(0.001)	(0.001)
Intercept	-0.080	0.051	0.034
	(0.001)	(0.001)	(0.001)
Number of Beneficiary-Drug-Years	1.7 million	1.7 million	1.7 million
R-squared	0.475	0.363	0.357

Table B.2: First Stage Estimates for Lagged Plan Choices

	Beneficiary Previous Plan Choice
Cohort Initial Plan Choice Probability	0.882
	(0.002)
Annual Premium	-0.000
	(0.000)
Zero Deductible Plan	0.063
	(0.001)
Logit Inclusive Value	0.004
	(0.001)
Non-Insulin OOP Cost	-0.001
	(0.000)
Number of Beneficiary-Plan-Years	6.5 million
R-squared	0.101

Table B.3: First Stage Estimates for Endogenous Premiums

	Annual Premium
Hausman Instrument	1.000
	(0.015)
Zero Deductible Plan	-3.817
	(9.484)
Intercept	7.905
	(8.451)
Number of Plan-Years	3,993
R-squared	0.649

Table B.4: Insulin Demand Parameters: 2SPS Specification

	Mean	SD
Insulin OOP cost ($100s)	-0.221	0.096
	(0.002)	(0.001)
Form = Prefilled Pen	1.156	0.546
	(0.033)	(0.123)
Brand = Basaglar	0.018	0.005
	(0.022)	(0.015)
Brand = Lantus	0.941	-0.013
	(0.011)	(0.034)
Brand = Levemir	0.207	1.137
	(0.016)	(0.038)
Brand = Toujeo	-0.229	
	(0.013)	
Same Form as Previous Year	1.072	
	(0.028)	
Same Brand as Previous Year	2.004	
	(0.053)	
Same Molecule as Previous Year	-0.519	
	(0.051)	

Table B.5: Plan Demand Parameters: 2SPS Specification

	Mean	SD
Annual Premium ($100s)	-0.239	0.508
	(0.016)	(0.026)
Logit Inclusive Value of Insulins	0.015	
	(0.003)	
Non-Insulin OOP Cost ($100s)	-0.006	
	(0.001)	
Same Plan as Previous Year	0.554	
	(0.023)	
Zero-deductible plan	0.810	
	(0.072)	

C Weakly Truthful Equilibrium

C.1 Constructing Candidate Contracts

C.2 Principal Contract Deviations

Consider candidate rebate contracts $\hat{t}$ that satisfy Agent Maximization (AM), Weakly Truthful (WT), and Principal Cost Minimization (PM) conditions to support the contract-efficient formulary profile. These conditions do not rule out profitable contract deviations by manufacturers in the first stage. A manufacturer may change its contract offers to multiple insurers, potentially changing their formulary choices in the continuation game.

Principal Deviation (PD)

The insurer subgame may have multiple Nash equilibria after a manufacturer changes its offers. A selection rule specifies which equilibrium insurers play after each set of offers. Candidate contracts $\hat{t}$ support $\hat{a}$ as a WTE if and only if *some* selection rule chooses $\hat{a}$ and deters all manufacturer deviations (Weber and Xiong 2026, Theorem 1). In other words, after each deviation, at least one insurer equilibrium must leave the deviating manufacturer no better off than at the candidate contracts. A stronger requirement is that *every* selection rule that chooses $\hat{a}$ at $\hat{t}$ deters all deviations. This requires that no deviation be profitable under any pure or mixed continuation equilibrium, as characterized by (PD-mix) below.

Pure-strategy equilibria (PD-pure)

To evaluate deviations, let $h_f^{-j}(a)$ denote the payoff insurer f retains at formulary profile a when manufacturer j's rebates are removed:

$$h_f^{-j}(a) = \pi_f(a) + \sum_{m \neq j} \hat{t}_{mf}(a), \qquad \forall j \in \mathcal{J},\ \forall f \in \mathcal{F}. \tag{30}$$

With the other insurers' formularies fixed at a_{-f}, the minimum rebate manufacturer j must pay insurer f to make a_f a best response equals the retained payoff the insurer forgoes by choosing a_f:

$$c_{jf}(a) = \max_{a'_f \in \mathcal{A}_f} h_f^{-j}(a'_f, a_{-f}) - h_f^{-j}(a). \tag{31}$$

Since a_f is itself feasible, this cost is nonnegative. The pure-strategy test compares manufacturer j's profit at the candidate equilibrium contracts with its profit at alternative feasible profiles, net of the compensation to induce the alternative:

$$\phi_j(\hat{a}) - \sum_f \hat{t}_{jf}(\hat{a}) \geq \phi_j(a) - \sum_f c_{jf}(a), \qquad \forall j \in \mathcal{J},\ \forall a \in \mathcal{A}. \tag{PD-pure}$$

If this condition fails, then some manufacturer j has a profitable deviation that induces a pure-strategy equilibrium, and no selection rule can deter the deviation. For *some* continuation selection rule to support $\hat{a}$ at the candidate contracts $\hat{t}$, condition (PD-pure) must hold. The stronger condition below tests whether manufacturer deviations are unprofitable under *every* selection rule that selects $\hat{a}$, including rules that select a mixed-strategy equilibrium after a deviation.

Mixed-strategy equilibria (PD-mix)

Let $\zeta = (\zeta_f)_{f \in \mathcal{F}}$ denote a profile of independent mixed strategies. Each insurer chooses its formulary before observing its rivals' realized choices, so its best alternative maximizes its *expected* retained payoff given ζ_{-f}. The minimum expected rebate needed to make ζ_f a best response is

$$C_{jf}(\zeta) = \max_{a'_f \in \mathcal{A}_f} \mathbb{E}_{\zeta_{-f}}\left[h_f^{-j}(a'_f, a_{-f})\right] - \mathbb{E}_{\zeta}\left[h_f^{-j}(a)\right]. \tag{32}$$

This can be less than the average cost of supporting each realized pure profile separately. The condition for ruling out profitable deviations under every pure or mixed continuation equilibrium is

$$\phi_j(\hat{a}) - \sum_f \hat{t}_{jf}(\hat{a}) \geq \mathbb{E}_{\zeta}[\phi_j(a)] - \sum_f C_{jf}(\zeta), \qquad \forall j \in \mathcal{J},\ \forall \zeta. \tag{PD-mix}$$

C.2.1 Verifying PD Condition

C.2.2 Off-Equilibrium Adjustments

C.2.3 Example: 2×2 Game

D Conformant Likelihood Estimator with Exogeneity Restrictions

I estimate plan demand using the Conformant Likelihood Estimator with Exogeneity Restrictions (CLEER) developed by Grieco et al. (2025). The objective function includes individual and aggregate log-likelihood components and a penalty for violations of the exogeneity restrictions:

$$(\hat{\boldsymbol{\theta}}^P, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) = \arg\min_{\theta^P, \delta, \beta} \left\{ -\log L^{micro}(\boldsymbol{\theta}^P, \boldsymbol{\delta}) - \log L^{macro}(\boldsymbol{\theta}^P, \boldsymbol{\delta}) + \hat{\chi}(\boldsymbol{\beta}, \boldsymbol{\delta}) \right\} \tag{33}$$

where $\boldsymbol{\delta}$ is a vector of mean utilities for each plan k:

$$\delta_k = \boldsymbol{\beta}^\top \mathbf{x}_k + \xi_k, \tag{34}$$

$\mathbf{x}_k$ includes observed plan characteristics (premiums and zero-deductible status), $\boldsymbol{\beta}$ contains their mean utility coefficients, and ξ_k denotes unobserved plan quality. $\boldsymbol{\theta}^P$ contains the remaining plan demand parameters governing individual-specific utility, including premium preference heterogeneity, expected OOP costs for non-insulins, logit inclusive value of insulins, and plan switching costs. L^{micro} and L^{macro} denote the individual and aggregate likelihood components, respectively, and $\hat{\chi}$ is a penalty function for violations of product exogeneity restrictions.

Plan choice probabilities integrate over the joint distribution of beneficiaries' random tastes for insulin and plan characteristics. Let $\omega_i^{(r)}$ denote beneficiary i's random draw for premium sensitivity and $\nu_i^{(r)}$ denote the vector of random draws for insulin preferences. Conditional on draw r, the plan choice probability is:

$$s_{ik}(\nu_i^{(r)}, \omega_i^{(r)}; \boldsymbol{\theta}^P, \boldsymbol{\delta}) = \frac{\exp\left(\bar{u}_{ik}(\nu_i^{(r)}, \omega_i^{(r)}; \boldsymbol{\theta}^P, \boldsymbol{\delta})\right)}{\sum_{k' \in \mathcal{K}_i} \exp\left(\bar{u}_{ik'}(\nu_i^{(r)}, \omega_i^{(r)}; \boldsymbol{\theta}^P, \boldsymbol{\delta})\right)} \tag{35}$$

where $\bar{u}_{ik}$ denotes plan utility excluding the idiosyncratic Type I Extreme Value shock. In particular, it includes the logit inclusive value $v_{ik}^*(\nu_i^{(r)})$ for all plans in the choice set, calculated using the same draw of random preference shocks. I integrate over the joint distribution

of plan and insulin taste shocks to calculate beneficiary i's plan choice probability:

$$\begin{aligned} s_{ik}(\boldsymbol{\theta}^P, \boldsymbol{\delta}) &= \int s_{ik}(\nu, \omega; \boldsymbol{\theta}^P, \boldsymbol{\delta})\, dF(\nu, \omega) \\ &\approx \frac{1}{R} \sum_{r=1}^{R} s_{ik}(\nu_i^{(r)}, \omega_i^{(r)}; \boldsymbol{\theta}^P, \boldsymbol{\delta}) \end{aligned} \tag{36}$$

Let $s_k(\boldsymbol{\theta}^P, \boldsymbol{\delta})$ denote the predicted market share for plan k offered in market g with N_g beneficiaries:

$$s_k(\boldsymbol{\theta}^P, \boldsymbol{\delta}) \approx \frac{1}{N_g} \frac{1}{R} \sum_{i=1}^{N_g} \sum_{r=1}^{R} s_{ik}(\nu_i^{(r)}, \omega_i^{(r)}; \boldsymbol{\theta}^P, \boldsymbol{\delta}) \tag{37}$$

The micro and macro log-likelihood components are:

$$\begin{aligned} \log L^{micro}(\boldsymbol{\theta}^P, \boldsymbol{\delta}) &= \sum_i \sum_{k \in \mathcal{K}_i} \widetilde{y}_{ik} \log \left(\frac{s_{ik}(\boldsymbol{\theta}^P, \boldsymbol{\delta})}{s_k(\boldsymbol{\theta}^P, \boldsymbol{\delta})} \right) \\ \log L^{macro}(\boldsymbol{\theta}^P, \boldsymbol{\delta}) &= \sum_{k \in \mathcal{K}} \widetilde{s}_k \log s_k(\boldsymbol{\theta}^P, \boldsymbol{\delta}) \end{aligned} \tag{38}$$

where $\widetilde{y}_{ik}$ indicates beneficiary i's observed plan choice and $\widetilde{s}_k$ is plan k's observed market share. The penalty function contains moments enforcing the product-level exogeneity restrictions:

$$\hat{\chi}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{1}{2} \hat{\mathbf{m}}(\boldsymbol{\beta}, \boldsymbol{\delta})^\top \hat{\mathbf{W}} \hat{\mathbf{m}}(\boldsymbol{\beta}, \boldsymbol{\delta}), \qquad \hat{\mathbf{m}}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \sum_{k \in \mathcal{K}} \mathbf{z}_k \left(\delta_k - \boldsymbol{\beta}^\top \mathbf{x}_k \right), \tag{39}$$

$\mathbf{z}_k$ is a vector of instruments including exogenous plan characteristics (zero-deductible coverage) and Hausman instruments for endogenous premiums.