Utilization Thresholds in Risk Adjustment Systems*

Eran Politzer†

June 2022

Abstract

Risk adjustment systems, that reallocate funds among competing health insurers, often use risk adjustors that are based on utilization documented in medical claims. The level of utilization that triggers an adjustor, i.e. the utilization threshold, is frequently chosen implicitly and uniformly. I empirically study utilization thresholds in the setting of the U.S. Marketplaces. Simulating alternative levels of thresholds for adjustors, I find thresholds that improve the prediction fit, by up to 9.6% in some disease groups. Using newly-defined measures for the incentives to game the system, I show that for some thresholds a tradeoff between fit and gaming-incentives does not exist. To guide a choice of multiple utilization thresholds, I employ a regression tree algorithm that considers both fit and gaming incentives.

Keywords: Health insurance, Risk adjustment, Utilization threshold
JEL Classification: I11, I13

*I am grateful to Tom McGuire for his guidance. For their helpful comments and explanations, I thank Konstantin Beck, John Bertko, Itamar Caspi, Randall Ellis, Keith Ericson, Kobi Glazer, Lukas Kauer, Richard van Kleef, Tim Layton, Michael McWilliams, Joe Newhouse, Sonja Schillo, Erik Schokkaert, Wynand van de Ven, and participants of the Harvard Health Care Policy Department's Health Economics seminar, and the Risk-Adjustment Network meeting.

†The Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue Boston, MA 02115 (email: eran.politzer@mail.huji.ac.il)

1 Introduction

Risk adjustment schemes are a cornerstone of a functioning managed competition market for health insurance. They reallocate funds among competing health plans based on the risk of their enrollees, and by that decrease plans' incentives to select profitable (typically healthier) enrollees and deter unprofitable (sicker) ones (see Ellis, Martins, and Rose (2018) and Layton et al. (2018) for an in-depth review of risk adjustment). To improve their predictive accuracy, risk adjustment systems have long ago advanced from relying only on age and gender adjustors that are exogenous to plans' influence, and now often use adjustors established from medical claims.[1] All these adjustors depend on enrollees' utilization of services either directly, e.g. adjustors based on the utilization of prescription drugs, or indirectly, e.g. diagnoses-based adjustors that are established during provider-patient interactions (Geruso and McGuire (2016)). Any adjustor based on utilization requires a decision regarding the minimum level of utilization that will trigger the adjustor - the utilization threshold. This decision is often made implicitly and uniformly for all adjustors. For example, when setting a diagnoses-based adjustor, policy makers rarely debate what is the number of times the diagnosis has to appear in claims over the year, implicitly choosing a utilization threshold of a single appearance in one medical encounter. This is not the only possible choice - for example, the risk adjustment system in Germany's Social Health Insurance requires that out-patient diagnoses appear twice over the year, in two separate quarters (see Table A1 in the Appendix). Outside of the U.S., explicit thresholds are common for adjustors based on the use of prescription drugs. To limit the incentive for gaming these adjustors by manipulating the prescription behavior, most countries that use them require a minimum level of utilization before a prescription affects a patient's risk score - often 90 or 180 days of supply. These thresholds mostly apply to all drug-adjustors uniformly.

This paper studies utilization thresholds and examines how their level may affect the performance of the risk adjustment model. Explicit thresholds would be desirable when a certain level of utilization is more predictive of spending and less prone to gaming, comparing to the baseline. I show that finding such thresholds is an empirical question, and they may be unique for each adjustor. Intuition may suggest that a higher utilization threshold always harms the fit of the risk adjustment model, as

[1] Better predictive accuracy, i.e. higher fit of the actual costs, may decrease plans' incentive for cost saving - a tradeoff long acknowledged in the literature. Moreover, using adjustors that are endogenous to plans' influence may create opportunities for manipulation.

information about some utilizers seems to be ignored. However, this intuition is wrong as a rule. A simple counter-example, presented in section 2, demonstrates how a threshold for a drug-based adjustor may increase the fit, relative to a zero-threshold baseline, when patients' total costs increase with higher utilization of the drug. The same threshold would decrease the fit, when total costs are a decreasing function of the drug's utilization. Furthermore, a higher threshold does not necessarily reduce the incentives for gaming, as the opportunity for gaming depends on the existing utilization patterns - a higher threshold may increase the number of patients susceptible for gaming if a larger group is left just below it. For example, if most patients in a group are prescribed with 28 days' supply of a certain drug, then setting a 30-days threshold for it will create a strong gaming incentive. Higher thresholds may also increase the potential revenues from gaming if the payment per patient above the threshold rises as the threshold increases.

While several established measures exist for the fit of a risk adjustment system, there is no consensus on how to measure the incentives to game the system by increasing utilization. I introduce new measures, based on the potential net revenue plans can gain from gaming an adjustor. The measures differ, first, by the group considered susceptible to gaming - those that already have some utilization (e.g. filled a short prescription for Insulin, included in a drug-adjustor), those with only a potential for new utilization (e.g. diagnosed with Diabetes, but have no Insulin prescription), or both. Second, the measures differ by the scope of the gaming activity - unlimited or marginal (e.g. limited to no more than 30 additional days of supply). While this paper studies gaming-incentives, it does not examine the actual response of insurers to these incentives.

In addition to setting the level of the utilization threshold for an existing adjustor, I examine also the general case in which multiple thresholds may be set, possibly adding new adjustors to the system. The choice of the number of thresholds and their levels is again an empirical question. This choice could be guided by machine learning algorithms that tackle a parallel challenge of splitting data - dividing observations into groups with the most homogeneity (See Ellis, Martins, and Rose (2018) for a review of the use of such algorithms for risk adjustment). I employ a regression tree algorithm - CART - to choose the utilization thresholds. The algorithm's loss function favors splits that increase the system's fit, while minimizing the potential revenue to insurers from gaming the system, with some possible weights on fit vs. gaming-incentives.

I study utilization thresholds empirically in the setting of the ACA Marketplaces, mainly examining days' supply thresholds for prescription-drug adjustors. I also explore thresholds for diagnosis-related adjustors, based on the number of times a diagnosis appears in patients' claims. To simulate thresholds, I use the IBM Truven Marketscan database, that holds claims from employers and commercial health plans, and was used to develop the Marketplaces' risk adjustment model (Kautter et al. 2014). I use the data from 2015 and 2016 for calibration and learning, and apply the risk adjustment model to enrollees in 2017.

For ten drug-adjustors (RXCs), I simulate multiple days' supply thresholds between zero and 360 days, re-estimating the model's coefficients in each iteration. The results show a unique pattern of the fit[2] and the gaming incentives for each adjustor. For five adjustors, a non-zero utilization threshold would improve both the overall fit (by 0.07% to 0.17% of the baseline fit) and the fit for the disease group related to the drug (by 2.1% to 9.6% of the baseline fit). The fit-maximizing threshold is 60 days' supply for anti Hepatitis-C agents, 120 days for Multiple Sclerosis agents and for Immune Suppressants, and 180 days for Anti-HIV agents and for Cystic Fibrosis agents. For five other drug adjustors - Antiarrhythmics, Phosphate Binders, Inflammatory Bowel Disease, Insulin, and Other Anti-Diabetic Agents - fit is maximized with the implicit zero-days threshold. Non-zero thresholds reduce the potential net revenue from unlimited gaming for most adjustors. In an important finding, for four drug-adjustors, setting a utilization threshold may pose no trade-off between fit and gaming incentives as both are improved. Relative to the zero-threshold scenario, setting a 60-days threshold for anti Hepatitis-C agents (RXC2), improves the individual fit for the disease group by 3.1%, while reducing the potential net revenue from unlimited gaming by 54%. A 120-days threshold improves the fit by 9.6% for RXC8 (Multiple Sclerosis agents) and by 2.1% for RXC9 (Immune Suppressants and Immunomodulators), decreasing the incentive for unlimited gaming by about 40%. Lastly, setting a 180-days threshold for Cystic Fibrosis agents (RXC10) improves the fit by 8.6%, while the incentive for gaming decreases by 5%.

I show how non-zero thresholds improve the fit when patients' costs increase with the number of days' supply in their prescriptions - the fit-maximizing threshold reduces the payment to over-compensated patients with short prescription by turning off the drug-adjustor for them. It increases the payment to under-compensated pa-

[2]I calculate the "individual fit" - the R-square of a model that predicts individuals' costs using the adjustors included in the risk-adjustment formula. I also present the calculation of the Cumming's Prediction Measure in the Appendix.

tients with longer prescriptions, as the payment above the threshold is re-estimated.

Using regression trees, I find that multiple thresholds could be beneficial for some adjustors. For example, optimizing fit, the tree algorithm suggests five thresholds for RXC1: 0, 120, 180, 270, and 330 days' supply.[3] However, when the loss function assigns some weight also to the gaming incentives, the algorithm's results change. With a weight of 0.2 on gaming incentives (vs. 0.8 on fit), there are only three suggested thresholds: 150, 270, and 300 days. With a weight of a third on the incentives, only a single threshold of 150 days' supply is recommended. In some cases, multiple thresholds may improve fit, but violate common principles of risk adjustment systems, e.g. the regression tree algorithms recommends four thresholds for RXC3 (0, 60, 180, 270), but requires adjustors with negative coefficients for all but the first threshold, in contrast to the monotonicity principle stated in Pope et al. (2004).

The rest of the paper proceeds as follows. Section 2 demonstrates that finding a fit-maximizing threshold is an empirical question. Section 3 introduces measures for the incentives to game the system. Section 4 describes the risk adjustment model in the Marketplaces, and section 5 describes the data. Section 6 presents the simulations and the regression tree algorithm used to select thresholds. Section 7 examines the impact of thresholds for drug-adjustors on fit and incentives for gaming, and Section 8 explores these effects for morbidity-based adjustors. I discuss the results in section 9. Section 10 concludes.

2 The Choice of Utilization Thresholds

2.1 The fit-maximizing level of a single utilization threshold

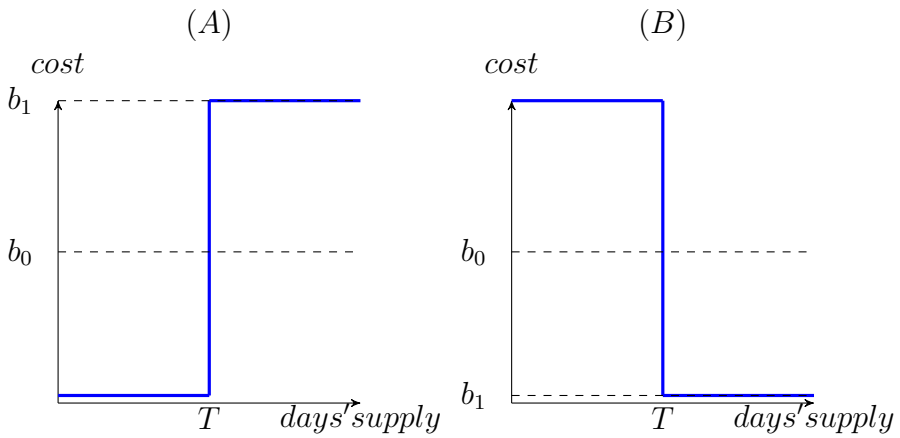
Intuition may suggest that a higher utilization threshold always harms the fit, as information about some utilizers seems to be ignored. However, this intuition is wrong as a rule. Consider a simple risk adjustment system with only one adjustor that indicates the use of drug X. For each patient using the drug, a plan receives a risk-adjustment payment that equals the average of the additional costs for all drug-X users.[4] With a non-zero utilization threshold, payment is the average additional cost

[3] I limit the algorithm to find regression trees with at most 3 levels, i.e. limit the number of thresholds to 7. I also restrict the thresholds to be a multiple of 30 between 0 and 360.

[4] This will be the payment if the coefficient for the drug-X adjustor comes from an OLS estimation of enrollees' costs on a constant and the single adjustor. This kind of estimation is the typical way to set coefficients for adjustors in risk adjustment formulas.

of patients with utilization above the threshold. Suppose that the cost of drug X is negligible and the number of users is small relative to the number of non-users. Figure 1 presents two possible distributions of the additional cost of drug users, ordered by the number of days' supply in their prescriptions.

Figure 1. Possible distributions of enrollees' costs, by days' supply of drug X



The figure presents two possible distributions of the additional costs for users of a certain drug ("drug X") by the days' supply of the filled prescriptions for each patient during the year. In panel A, high-cost patients have high utilization of the drug, while in panel B low-cost patients have higher utilization. b_0 is the average additional cost over all users of drug X. Hence, b_0 equals also the payment to the plan for each such patient. Setting a utilization threshold of T days' supply changes the payment due to patients below the threshold to zero. Patients above the T-days threshold have an average cost of b_1, and hence the plans receive a payment of b_1 for them. The threshold increases the individual fit in panel A as each type of patient receives the correct payment. However, the threshold decreases the fit in panel B, as all patients receive zero payment.

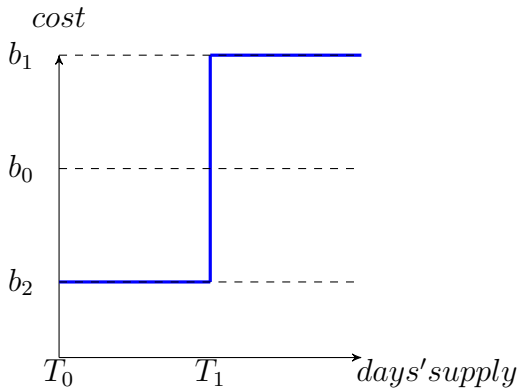
Costs may be higher for high utilizers of the drug (panel A) if a higher use indicates a severe chronic condition with additional co-morbidities. Alternatively, costs may be lower for high utilizers (panel B) if higher and continuous use indicates a patient with good drug adherence and a controlled disease. With a zero-days threshold, the average additional cost for all patients with the drug-adjustor turned on is b_0, and hence the payment to the plan is also b_0. Setting a utilization threshold that requires prescriptions of at least T days' supply has a very different impact on the individual fit in these two cases. In panel A, setting a threshold T would improve the fit, as the adjustor would pay zero for low-cost patients and the correct cost b_1 for high cost patients. In panel B, setting an identical threshold of T days' supply, would decrease the fit as both high-cost patients and low-cost patients would receive zero payment. This is equivalent to eliminating the adjustor altogether and fit must be better with the adjustor than without it.

2.2 Fit with multiple utilization thresholds (TBC)

The case in the simple example above restricts the drug-adjustor to have a single utilization threshold. This section relaxes the restriction and discusses the choice of multiple thresholds, practically allowing to add new adjustors to the risk adjustment scheme. An additional adjustor would weakly increase the fit of the scheme, as long as overfitting does not become an issue due to small sample sizes or inflation of adjustors. However, some additional adjustors may not be acceptable from policymakers' point of view. Returning to panel B of Figure 1, consider the choice of two thresholds - zero and T. Patients with some utilization of the drug, below T days would receive a high compensation, while patients with utilization higher than T days would receive b_1 (or zero). It is easy to see that fit would be better with two thresholds - plans receive exactly the additional cost of each patient. However, a lower compensation when utilization increases would violate monotonicity - one of the common principles guiding the development of risk adjustment systems (Pope et al. 2004). This principle states that insurers should not be penalized for additional recording of diagnoses, or in our case - recording of additional days of supply provided to a patient. Violation of this principle may incentivize insurers to game the system by skimping, or by under-reporting the number of days' supply.

Figure 2 presents another possible distribution of the additional costs of patients using a drug, by the number of days' supply in their filled prescriptions. Again, it is easy to see that adopting two thresholds - T_0, T_1, paying b_2, b_1 accordingly, would lead to a higher fit than using any single threshold.

Figure 2. A possible distribution of enrollees' costs, by days' supply (two thresholds)



The figure presents a possible distribution of the additional costs for users of a certain drug ("drug X") by the days' supply of the filled prescriptions for each patient during the year.

In Figure 2, a shift from a single threshold at T_1 to two thresholds at T_0 and T_1

has a mixed effect on the incentives for gaming: The incentives increase for patients with no utilization of the drug, as even a single additional day of supply now yields a gaming revenue of b_2. For the group of patients already prescribed with a positive number of days' supply, but less than T_1, the gaming incentives are lower now, as the potential revenue from gaming decreases from b_1 to $b_1 - b_2$. The total effect on gaming incentives is thus an empirical question that depends on the distribution of patients by the length of their prescriptions.

3 Measuring the Incentives for Gaming

A major concern when choosing adjustors for risk adjustment systems is that plans and providers should not be able to readily manipulate them to increase plan payments (Ellis, Martins, and Rose 2018). Unlike age or gender, adjustors based on utilization are susceptible to gaming. When adding drug adjustors to the Marketplace risk-adjustment model, CMS acknowledged that it "may provide an incentive to overprescribe medications" (CMS 2016b). This concern has led CMS to exclude some drug groups from the model because medical professionals judged that they are "particularly subject to intentional or unintentional discretionary prescribing variation or inappropriate prescribing by health plans or providers" (CMS 2017).

Gaming of drug-adjustors is a concern also in other risk adjustment systems, e.g. Lamers and van Vliet (2003) discuss the potential for gaming the drug adjustors in the Dutch risk-adjustment scheme.[5] However, there is no consensus about how to measure the incentives for gaming.[6] One indirect measure appears in Lamers, van Vliet, and van de Ven (1999) that examines the proposed Pharmacy-based Cost Group (PCG) model in the Netherlands. They calculate the ratio between the capitation payment for those assigned to one of the PCGs and their pharmacy costs. They find that revenues are on average about four times as high as the cost of drugs. While

[5]The authors propose several measures to decrease the opportunities for gaming - setting days' supply thresholds, preventing payments due to multiple chronic conditions by assigning only the most expensive drug-adjustor, reducing the compensation or restricting it, and avoiding drug adjustors that lead only to a small increase in predicted costs.

[6]The literature is wider on the incentives for cost saving, that are related to incentives for gaming as both may depend on the effect of current plan's spending on its future revenue. To quantify incentives for cost-saving, Geruso and McGuire (2016) use the "power" concept to measure the share of costs borne by the plan at the margin. However, the power measure seems insufficient to measure incentives for gaming. Some gaming activities may entail no further utilization or cost at the enrollee level. Even when gaming requires additional utilization, e.g. prescribing additional days' supply, the resulting increase in revenue could be much higher than the additional cost, leading to a negative "power" measure - below the usual 0 to 1 range of this statistic.

this ratio may suggest that there is some room for insurers to increase revenues by prescribing more drugs, it doesn't quantify the gaming incentives directly. The approach in this paper is similar to the way Behrend, Felder, and Busse (2007) examine the gaming opportunities of the then-proposed drug adjustors in the German risk adjustment system (IPHCC+RxGroups). The authors simulate three specific cases of plans' gaming behaviors and calculate their net monetary returns per insured person: substitution to an alternative drug for hypertension that leads to a higher risk score[7]; increasing the prevalence of antidepressants use among patients already diagnosed with depression[8]; and increasing the use of diabetes drugs by diagnosing previously unidentified diabetic patients and supplying them with very short prescriptions.[9]

Before turning to defining new measures, I note that I quantify the *incentives* to prescribe more and not the actual response of the plans.[10] The question of how plans respond to the incentives is not addressed in this paper. The reluctance of providers to prescribe more and the cost for patients to purchase more drugs may inhibit gaming. However, as plans can incentivize providers, change formularies and set non-linear price schedules for patients, strong incentives to prescribe more may still distort prescribing and purchasing behaviors in practice.

To define measures for the incentives for gaming one has to choose first the relevant population, for which gaming is examined. This choice creates a tradeoff – widening the population may enable a more comprehensive examination of incentives, but in most cases these incentives will be less and less actionable. For example, a plan may have a theoretical incentive to prescribe *everyone* with a drug if the resulting payment is higher than the drug's cost. However, it will be very hard for a plan to act on such an incentive - make providers prescribe unnecessary drugs to healthy individuals, and convince individuals to fill these prescriptions. In contrast to that, it will be most likely much easier for a plan to make providers and patients lengthen justified prescriptions for patients that already use a drug. A similar tradeoff exists when choosing the gaming activity that the measure examines - limiting the scope of the gaming activity will most likely make the incentives more actionable. For example,

[7]The simulation moves all the patients prescribed with ACE inhibitors to Angiotensin II receptor blockers - an alternative drug that leads to a higher risk score.

[8]To simulate the change, the authors randomly assign antidepressant drugs to patients diagnosed with depression, so the prevalence of the drug use increases by 30%.

[9]This simulation examines an increase of 4 percentage points in the prevalence of diabetes treatment among the relevant age groups - a 33% increase of the baseline prevalence.

[10]I avoid using the term "over-prescribing", which may hint that the baseline prescription behavior is optimal. Incentives to prescribe more may lead to optimal utilization if plans would have skimped on drugs without them.

an incentive for a gaming activity that includes prescribing additional 180 days of supply is likely less actionable than an incentive to game the system by prescribing one more day of supply (regardless of the cost of the drug).

With these tradeoffs in mind, the measures I define focus either on the population of patients with an existing adjustor-related utilization (e.g. prescription for a drug included in a drug-adjustor), or on the group of patients with at least a potential for such utilization (e.g. patients diagnosed with a disease related to the drug). Regarding the scope of the gaming activity, I examine the case of no limits to the additional utilization (e.g. allowing any number of additional days of supply when gaming a prescription). I also examine measures that limit the additional utilization in gaming (e.g. prescribing no more than 30 days of additional supply). To quantify the extent of the gaming incentives I use the potential net revenue to the plan from the gaming activity. Table A4 in the Appendix also presents the share of enrollees that are "gameable", i.e. the enrollees for whom the gaming activity yields a profit to the plan. The three measures I define are presented below for the case of additional days of supply in drug adjustors, but can be easily adopted to reflect other types of utilization (e.g. Section 8 examines incentives for gaming the number of times a diagnosis appears in claims by adding additional office visits).

3.1 Measure A: Unlimited gaming of the disease group

The first measure examines the potential revenue from gaming a disease group, i.e. the patients that are either prescribed with a drug included in a drug-adjustor or diagnosed with a related disease. To make *all* these patients cross a utilization threshold, plans may prescribe an unlimited number of additional days of supply. I measure the net revenue to the plan from this gaming activity, per person in the disease group. The definition is:

$$MeasureA = \frac{1}{N}\sum_i \Delta R_i - \min_j(\Delta sup_{ij} * cost_j) \\ s.t.\ \forall i\ \exists j\ s.t.\ sup_{ij} + \Delta sup_{ij} > T \tag{1}$$

where N is the number of patients in the disease group, ΔR_i is the additional revenue to the plan due to pushing patient i over the threshold, sup_{ij} is the annual number of days' supply in patient i's prescriptions of drug j (included in the drug-adjustor), Δsup_{ij} is the number of additional days' supply prescribed as part of the gaming, and $cost_j$ is the daily cost of drug j. The incentive measure is calculated by using the

cheapest way for patient i to cross the threshold of T days' supply.

3.2 The incentives for a limited gaming activity

The next two measures limit the scope of the gaming activity examined, allowing plans to prescribe no more than 30 additional days of supply to push patients across the threshold. As these measures focus on patients at the margin of passing the threshold, gaming of their prescriptions is arguably easier and the incentives to do so are more actionable.

Measure B: Gaming patients with an existing prescription The second measure examines only patients with an existing prescription, excluding those with a related diagnosis alone.[11] The examined gaming activity allows to extend the existing prescription by no more than 30 additional days of supply. The measure calculates the average net revenue from the gaming activity, per patient in the disease group. The measure's definition:

$$\begin{aligned} &MeasureB = \frac{1}{N}\sum_i \Delta R_i - \min_j(\Delta sup_{ij} * cost_j) \\ &s.t.\ sup_{ij} + \Delta sup_{ij} > T \ \ \& \ \ \Delta sup_{ij} <= 30 \ \forall i \in \{i|sup_{ij} > 0\} \end{aligned} \tag{2}$$

The incentive is calculated as in equation (1), limiting the number of additional days of supply to 30 and changing prescriptions for a narrower group of patients. N is the number of patients in the disease group, ΔR_i is the additional revenue to the plan due to pushing patient i over the threshold, sup_{ij} is the annual number of days' supply in patient i's prescriptions of drug j (included in the drug-adjustor), Δsup_{ij} is the number of additional days' supply prescribed as part of the gaming, and $cost_j$ is the daily cost of drug j. The incentive measure is calculated by using the cheapest way for patient i to cross the threshold of T days' supply.

Measure C: Gaming diagnosed patients with no prescription The third measure examines gaming of diagnosed patients in the disease group that have no prescription for one of the drugs in the drug-adjustor. It includes providing new prescriptions of up to 30 additional days of supply for these patients. The definition

[11] This measure is applicable only when there is a non-zero utilization threshold. When the threshold is zero, the drug-adjustor is turned on for all the patients with a prescription, making gaming for this population futile.

of this measure is similar to the definition of measure B (Equation 2), changing the group of patients susceptible to gaming to patients with no prescription related to the drug-adjustor (i.e. $sup_{ij} = 0$).

4 Drug-adjustors in the Marketplaces' Risk Adjustment

The risk adjustment scheme for the plans in the U.S. Marketplaces includes two components: the Department of Health and Human Services' (HHS) risk adjustment model and a transfer formula (Layton, Montz, and Shepard 2018). The basic model predicts this year's plan liability for enrollees based on their age, sex and the diagnoses drawn from their claims, producing a risk score for each person.[12] The transfer formula redistributes plans' premium revenues by the average risk score in each plan and other factors.

Beginning in the 2018 benefit year, The Centers for Medicare & Medicaid Services (CMS), the federal agency that administers this risk adjustment system, started using a "hybrid drug-diagnosis" risk adjustment model in the Marketplaces, adding adjustors indicating a filled prescription for the included drugs (CMS 2016a, 2016b). For example, a patient who filled a prescription for insulin will have a higher risk score, potentially increasing the risk adjustment transfer to her plan, whether she has a diabetes diagnosis in one of her claims or not. The drug adjustors are meant to indicate health risk when a diagnosis is missing. This can happen due to a mistake, to avoid stigma, or because the patient did not visit a physician. However, the drug adjustors appear independently in the risk adjustment model, and are not used only to turn on a related diagnosis-adjustor. The drug adjustors may also provide information on the severity of a diagnosed illness. To do this, the model includes interactions of drug-adjustors and their related diagnosis-adjustors. In the model, no minimum utilization is required for a prescription to increase a patient's risk score, e.g. a prescription of insulin for a single day will suffice to increase the score, and will have the same effect as a prescription for a year's supply. CMS considered setting days' supply thresholds for some potential drug adjustors, but eventually set no thresholds.

The baseline risk adjustment model in this paper is the CMS 2019 model (*HHS-HCC V0519*), that includes ten drug-adjustors (RXCs). Each RXC is a prescription drug category that may include several drugs, identified by their National Drug Code

[12]The prediction model produces 15 sets of risk adjustment coefficients: three age-specific models (adult, child and infant), and five models specific for each coverage level in the Marketplaces (platinum, gold, silver, bronze, catastrophic).

(NDC). CMS chose RXCs that are closely related to diagnoses that were already included in the model within Hierarchical Condition Categories adjustors (HCC), that group diagnoses. Each RXC appears in the model as both an independent adjustor and within an interaction with its paired HCCs. Table 1 describes the RXC-HCC disease groups in the 2019 model.

Table 1. Drug-Diagnosis Pairs in the 2019 Marketplaces Risk Adjustment Model

RXC	RXC Label	Related Diagnoses (HCCs)
1	Anti-HIV Agents	HIV/AIDS
2	Anti-Hepatitis C (HCV) Agents	Chronic Hepatitis C, Cirrhosis of Liver, End-Stage Liver Disease, and Liver Transplant
3	Antiarrhythmics	Specified Heart Arrhythmias
4	Phosphate Binders	End Stage Renal Disease, Kidney Transplant, Chronic Kidney Disease - Stage 5, Chronic Kidney Disease - Severe (Stage 4)
5	Inflammatory Bowel Disease Agents	Inflammatory Bowel Disease, Intestine Transplant
6	Insulin	Diabetes, Pancreas Transplant
7	Anti-Diabetic Agents, Except Insulin and Metformin Only	Diabetes, Pancreas Transplant
8	Multiple Sclerosis Agents	Multiple Sclerosis
9	Immune Suppressants and Immunomodulators	Rheumatoid Arthritis and Specified Autoimmune Disorders, Systemic Lupus Erythematosus and Other Autoimmune Disorders, Inflammatory Bowel Disease, Intestine Transplant
10	Cystic Fibrosis Agents	Cystic Fibrosis, Lung Transplant

The coefficients in the 2019 version of the model are based on an average of the coefficients separately estimated for the years 2014, 2015 and 2016 (CMS 2018). CMS adjusts the coefficients post-estimation for clinical reasonableness and to decrease gaming. Fearing inappropriate prescribing when an inexpensive drug treats a medically expensive condition, CMS restricted the payment for two of the drug-adjustors included in the model - RXC3 (Antiarrhythmics) and RXC4 (Phosphate Binders) - to less than the average cost of supplying the drugs.[13]

[13]Payment due to these RXC adjustors was a priori set to be equal to the average annual per capita cost of the drugs in the RXC (in the calibration dataset). In addition to that, the RXC-HCC interaction term was set to zero for both RXCs.

CMS considered in a 2016 White Paper (CMS 2016a) whether to require a utilization threshold to trigger a drug indication - either require multiple prescriptions for the same drug, or prescriptions totalling at least 30 or 60 days' supply. CMS' clinical consultants suggested that for some potential RXCs, a minimum days' supply utilization threshold would be useful to distinguish severely ill patients from those with milder conditions. However, CMS decided to not include a days' supply restriction in the model, requesting feedback from the public.

Prescription drugs serve as adjustors in risk adjustment models in other countries as well. In most cases, some minimum utilization threshold is required to trigger an indication. In Germany, 183 days' supply are required for drug adjustors to validate most chronic diseases, 42 days are required for diseases with medication to be taken as needed, and 10 days are required for acute diseases. Switzerland, the Netherlands and the Czech Republic demand prescriptions of at least 180 days' supply for most drug groups. The Netherlands has a 90 days threshold for some specific groups, and no threshold at all for extremely high-cost drugs. See Table A1 in the appendix for more details on the use of drug adjustors in these countries.

5 Data

This paper uses the IBM Truven MarketScan database of medical claims from the employer-sponsored insurance market to measure spending, record diagnoses, and examine the utilization of prescription drugs. Utilization of drugs is measured by the number of days' supply, i.e. - the number of days for which supply will last for the patient when using the maximum dose prescribed.[14] The Truven database was used to develop the original Marketplace payment system (Kautter et al. 2014), and until recently was used exclusively in updating it.[15] I estimate the risk adjustment coefficients using the 2015 and 2016 versions of the database, and use the 2017 version to simulate payments under different utilization thresholds. The analytic sample is composed of adults, between ages 21 and 65. It includes individuals who had coverage for both prescription drugs and mental health, were continuously enrolled for twelve months, and had fee-for-service claims data for the whole period (i.e. no

[14]This measure, appearing in U.S. pharmacy claims, is different than the number of Defined Daily Doses - a uniform standard dose defined for each drug by the World Health Organization.

[15]Starting in the model for 2019, CMS is gradually shifting to using claims data from the plans in the Marketplaces themselves (EDGE data), instead of the Truven database.

encounter data from managed care plans).[16] Table 2 reports summary statistics for the 12,227,124 individuals in the analytic sample. The plans' average annual spending on these enrollees is $5,741 (this sum does not include any out-of-pocket payments by the enrollees. It includes plans' spending on in-patient and out-patient services, as well as on prescription drugs). 5.9% of enrollees have a prescription for a drug included in one of the ten RXC drug-adjustors. The cost of treating these patients is 4.5 times higher than the cost of the average enrollee. Table 2 also presents the share of patients and the average cost for each RXC-HCC disease group, and for patients with a prescription for the RXC drugs.

Table 2. Descriptive statistics for the 2017 sample (N=12,227,124)

Variable	Share of Enrollees (%)	Mean Spending ($)		
All	100	5,741		
Share of:				
Females	52.6	6,354		
21-29	17.6	3,063		
30-39	20.3	4,142		
40-49	23.6	5,199		
50-65	38.5	8,139		
Patients w. Any RXC	5.9	25,585		
	Diagnosed or Prescribed		Prescribed Only	
	Share of Enrollees (%)	Mean Spending ($)	Share of Enrollees (%)	Mean Spending ($)
RXC-HCC 1 (HIV)	0.27	34,480	0.23	38,624
RXC-HCC 2 (Hepatitis-C)	0.27	47,056	0.03	92,208
RXC-HCC 3 (Antiarrhythmics)	1.30	30,387	0.36	39,373
RXC-HCC 4 (Phosphate Binders)	0.24	76,717	0.07	126,120
RXC-HCC 5 (IBD)	0.71	29,790	0.35	27,628
RXC-HCC 6 (Insulin)	7.33	15,770	1.59	27,012
RXC-HCC 7 (Other Diabetes)	8.31	16,628	3.20	16,387
RXC-HCC 8 (Multiple Sclerosis)	0.27	57,269	0.15	77,545
RXC-HCC 9 (Immunosuppressants)	2.28	28,260	0.92	39,281
RXC-HCC 10 (Cystic Fibrosis)	0.02	122,968	0.01	146,288

This table presents summary statistics of the analytic sample used in the paper. For each RXC-HCC disease group (named by the drug category), the share of enrollees and the plans' mean spending on them is shown for both the whole group, and the subgroup of patients prescribed with an RXC drug. Spending does not include any out-of-pocket payments by patients.

For each RXC drug group, Figure 3 presents the distribution of patients with a prescription by their annual number of days supplied. The share of the prescribed

[16] I allow for negative claims, but drop enrollees with a negative sum of their total spending for the year, as well as enrollees with a negative sum of spending for one of the categories: in-patient care, out-patient care, or drugs.

patients among the RXC-HCC group is noted at the upper-left corner of each graph.

The numbers of days' supply are mostly bunched in multiples of 30-days or 28-days.[17] Patients with more than a year-worth of supply are top coded in the figure and included in the 365 days' supply category.[18] Figure 3 also shows the average cost of patients for each days' supply category. Costs increase as a function of the number of days' supply for RXCs 1,2,8 and 10, decrease with days' supply for RXC 3, and are mostly stable for RXCs 4, 5, 6 and 7. For RXC 9, costs are stable for patients with 30-days multiples of prescriptions, and are higher and increasing for patients with 28-days multiples.

I use the 2019 HHS-HCC risk adjustment methodology, implemented in CMS' *HHS-HCC V0519* software,[19] to calculate the risk scores, and thus the risk adjustment payment for each person. The V0519 methodology is used to decide which adjustors should be turned on for each enrollee. To calculate the risk scores for each enrollee the adjustors' vector is multiplied by their corresponding coefficients, which I re-estimate in each step (see details in the next section). While CMS adjusts the risk adjustment coefficients post-estimation for clinical reasonableness and to decrease gaming, I apply no restrictions on the estimated coefficients.

6 Empirical Methods

6.1 Simulations of multiple levels for a single threshold

I use simulations to examine the effect of days' supply utilization thresholds on the model's fit and the incentives for gaming. I use the CMS software to turn on the risk adjustors for all the data years and then, for each drug group (RXC), simulate thirteen thresholds - all the 30-days multiples between 0 and 360 days. Each simulation of a single threshold includes four steps:

1. **Adjust the drug-adjustor** - Turn off the drug adjustor for all enrollees whose annual number of days' supply is lower than the simulated threshold. This step is done for all the data years.[20]

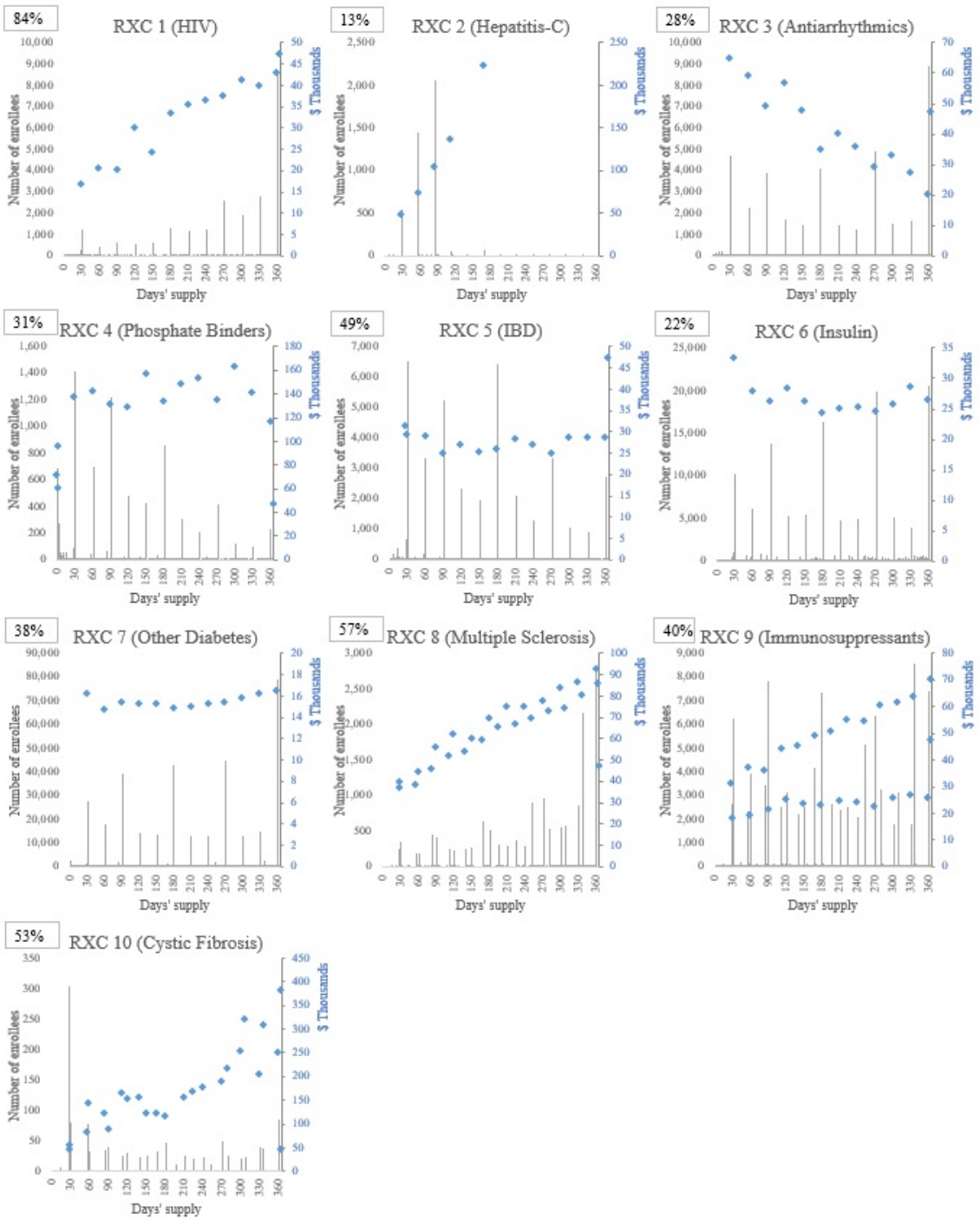
[17] This bunching presumably reflects the common packaging of the RXC drugs. It happens despite the growing popularity of 90-days prescriptions in both mail-orders and retail pharmacies.

[18] During a single year, some patients may fill prescriptions with more than a year's worth of supply if, for example, a long prescription is filled toward the end of the year.

[19] The 2019 HHS Risk adjustment software can be downloaded here: https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/

[20] The number of days' supply is summed up at the single-drug level (NDC-by-NDC) and not at

Figure 3. The distribution of prescribed patients and their average annual cost, by annual number of days' supply



For each RXC, the graph shows: 1. The share of patients in the disease group prescribed with an RXC drug (box at the upper left corner); 2. The number of patients by the number of annual days' supply in their prescriptions (bars, left axis). Patients with no prescription are excluded. All patients with 365 days' supply or more are top-coded to the 365 days category; 3. The average costs of patients to their plans by the number of annual days' supply in their prescriptions (dots, right axis). Costs are shown only for days' supply categories with at least 1% of the prescribed patients in the RXC.

2. **Reestimate model coefficients** - Reestimate the risk adjustment model using the enrollees' revised risk adjustors (from the previous step) for each of the years 2015 and 2016. The dependent variable in these estimations is the annual total cost of the enrollee.[21] Average the coefficients from the two estimations to get the payment coefficients for the modified 2017 risk adjustment model.

3. **Recalculate risk scores** - Calculate new risk scores for the 2017 enrollees, using the revised adjustors for this year (step 1), and the new coefficients (step 2). Calculate the payment (i.e. the predicted cost) for each enrollee.

4. **Calculate fit and gaming incentives** - Calculate the fit measures, comparing the annual 2017 costs with the updated payments (step 3). Calculate the incentives to prescribe more to cross the simulated threshold.

For each simulated threshold, two individual R^2 fit measures are calculated[22]: First, the individual fit for the entire population of enrollees. Second, the individual fit for the patients in each RXC-HCC disease group. This group includes the patients with any prescription for a drug in the RXC and the patients that are diagnosed with one of the related diseases (described in Table 2).

To assess the gaming incentives under each utilization threshold, I calculate for each patient the additional revenue the plan would receive if the patient crosses the threshold.[23] I also calculate the minimum cost of the additional days of supply required for the crossing.[24] [25] The direct cost used in the calculation is most likely just a part of the total cost of gaming, that may include overcoming providers' reservations

the RXC level (by drug group). e.g. two 90-days prescriptions for different drugs in the same RXC do not sum up to cross a 180-days threshold.

[21] I use the cost not covered by the Marketplace risk sharing scheme, that pays 60% of costs above $1 million. This cost is also used for fit calculations. "Payment system fit" is a more general fit statistic to calculate the fit including these omitted costs. See Layton et al. (2018) for details on this measure.

[22] Risk adjustment models are essentially prediction models, and a fit statistic measures the accuracy of their prediction. The most common fit statistic is the R^2 individual-level fit. This is the R^2 of a regression of enrollees' cost on the adjustors included in the risk model.

[23] This calculation uses the coefficients of the risk adjustors re-estimated in each simulation.

[24] I assume that an additional supply that allows the patient to *just* cross the threshold is possible. In practice, the cost of gaming could be higher if prescriptions must be rendered in multiples of certain days' supply due to availability of specific dosages. Costs could also be higher if adding days of supply requires an additional service from a provider (e.g. an office visit). The costs of gaming could be lower if plans shift patients from one drug, not included in the RXC, to another one that is included in the RXC. I ignore all such options in the calculations.

[25] For each patient with an existing RXC prescription, the cost of an additional day of supply is the average of the daily cost in her own prescriptions. For diagnosed patients with no prescription, the cost of an additional day is the average daily cost for all prescribed RXC patients.

due to professional ethics and intrinsic concern for their patients. Such costs may become prohibitively high, especially when gaming is extensive or is outside of any gray area around the proper prescription behaviour. In that sense, the incentive measures defined here may serve as an upper bound for the true incentives for gaming.

6.2 Regression trees for multiple thresholds

Regression trees may guide the choice of multiple thresholds for a drug-based risk adjustor, or any other adjustor based on utilization. This choice requires partitioning the range of patients' utilization by a splitting variable so more cost-homogeneous groups of patients are formed, improving the model's fit. This task is performed by a common algorithm to identify regression trees - the Classification And Regression Tree algorithm (CART). For drug-based adjustors, the splitting variable in this paper is the number of days' supply in patients' prescriptions. I use a version of CART to choose days' supply thresholds: At the root node, the algorithm splits the entire range of days of supply, finding the first threshold. It recursively searches for an additional threshold in each subset of this range (i.e. in each child node), below and above the chosen threshold. CART is a "greedy" algorithm - it finds a *local* optimum for a partition in each iteration and not a *global* optimum, but it provides an efficient way to guide the choice of thresholds. To identify the best threshold at each tree node, I simulate multiple (additional) thresholds as described in the previous section. The chosen threshold at each node is the one that minimizes a loss function that considers changes in both fit and the cumulative 30-days gaming incentives, relative to the baseline of a single zero-days threshold:

$$(1-\lambda)(1-\frac{Fit_{n,t}}{Fit_{1,0}})+\lambda\frac{GamingIncentives_{n,t}}{GamingIncentives_{1,0}} \tag{3}$$

where $Fit_{n,t}$ is the R-squared individual fit for the disease group after adding the utilization threshold t at node n, $Fit_{1,0}$ is the baseline fit with a single zero threshold. $GamingIncentives_{n,t}$ is the cumulative incentives for gaming by prescribing up to 30 more days of supply to patients in the disease group (the sum of measures B and C from Section 3), after adding threshold t at node n. $GamingIncentives_{1,0}$ is the baseline gaming incentive. λ is a parameter that determines the relative weight of the change in gaming incentives in the loss function versus the change in fit.

I limit the regression tree to three levels, i.e. no more than seven thresholds, restrict the possible thresholds to multiples of 30 days, between 0 and 360, and avoid

thresholds with a difference of 30 days or less from the previous threshold in a parent node. These restrictions are mainly employed to limit the computational resources required to create the regression trees, and may be relaxed in alternative versions. However, increasing the number of thresholds and reducing the distance between them may increase the risk of overfitting the data, and may require other stopping criteria for the recursion (e.g. a minimal number of patients within the partition).

7 Results

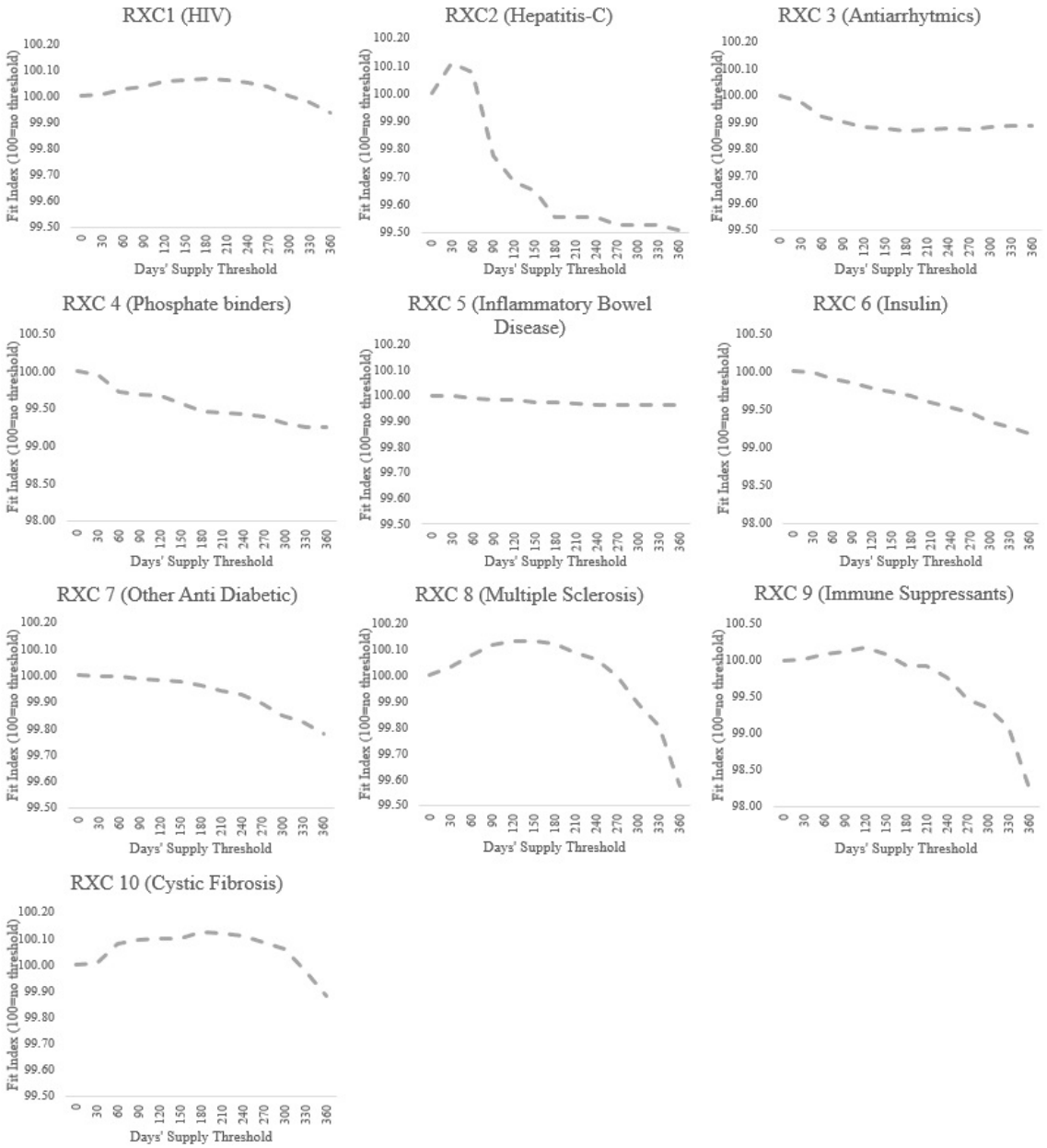
7.1 Simulation of days' supply thresholds for a single adjustor

For each threshold, two individual R^2 fit measures are presented: First, Figure 4 presents the individual fit for the entire population of enrollees. Second, Figure 5 presents the individual fit for the patients in each RXC-HCC disease group. This group includes the patients with any prescription for a drug in the RXC and the patients that are diagnosed with one of the related diseases (described in Table 2). Both figures present the results by days' supply thresholds of each RXC. They normalize the baseline zero-threshold fit to 100, and hence show the percent change in the fit, relative to the baseline.[26]

The results show a unique impact of thresholds on the fit for each RXC. While the overall fit in the population doesn't vary a lot since the number of patients in most groups is quite small (Figure 4), the fit for patients within each RXC-HCC disease group may be affected in a significant way by the choice of the threshold (Figure 5). For five drug groups - RXC3 (Antiarrhythmics), RXC4 (Phosphate Binders), RXC5 (Inflammatory Bowel Disease), RXC6 (Insulin), and RXC7 (Other Anti Diabetic) - both fit measures are maximized with the baseline zero-days utilization threshold. For five other RXCs, non-zero days' supply thresholds may improve both the overall fit and the fit for the disease group: For RXC1 (Anti HIV Agents), a utilization threshold of 180 days is fit-maximizing. It increases the overall R^2 fit by 0.07% relative to the baseline fit (of 34%) and increases the individual R^2 fit in the disease group by 4.4%; RXC2 (Anti Hepatitis-C Agents) benefits from a 60 days' supply threshold, that improves overall fit by 0.07%, and increases the fit in the disease group by 3.1%; For RXC8 (MS Agents), a 120 days' supply threshold maximizes fit: overall fit increases

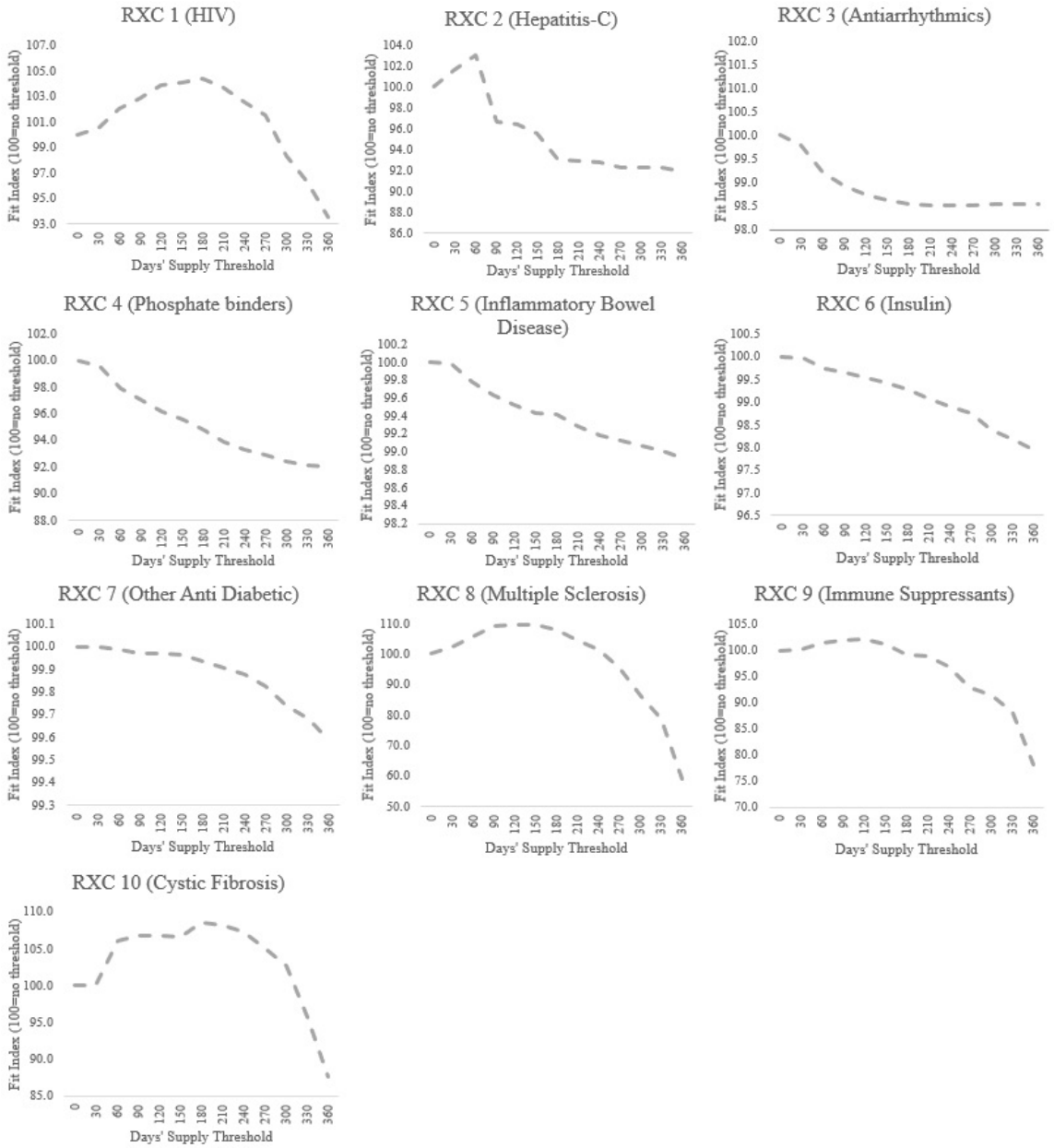
[26]The actual individual fit results for each RXC appear in Table A3 in the Appendix. The appendix also presents an alternative fit measure - the Cumming's Prediction Measure (CPM), that does not square the prediction-cost deviations as the R^2 measure. Table A2 in the appendix presents the model coefficients for the RXC and HCC adjustors and their interactions in each simulation.

Figure 4. Individual fit for all the enrollees, by days' supply threshold (0-days=100)



For each RXC drug-adjustor, the figure presents the individual R^2 fit statistic for all enrollees in the the simulated days' supply thresholds that are multiples of 30 days, between 0 (i.e. no-threshold) and 360.

Figure 5. Individual fit for enrollees in the RXC-HCC disease group, by days' supply threshold (0-days=100)



For each RXC drug-adjustor, the figure presents the individual R^2 fit statistic, for enrollees in the RXC-HCC disease group, in all the simulated days' supply thresholds that are multiples of 30 days, between 0 (i.e. no-threshold) and 360. The RXC-HCC group includes patients prescribed with a drug included in the RXC or diagnosed with a related diseases (HCC).

by 0.13% and the fit for the disease group increases by 9.6%; A 120-days threshold maximizes fit also for RXC9 (Immune Suppressants) - overall fit increases by 0.17% and the fit in the disease group rises by 2.1%; Lastly, for RXC10 (CF Agents), fit is maximized with a 180 days' supply threshold. With such a threshold, the overall fit increases by 0.12% and the fit for the disease group increases by 8.6%.

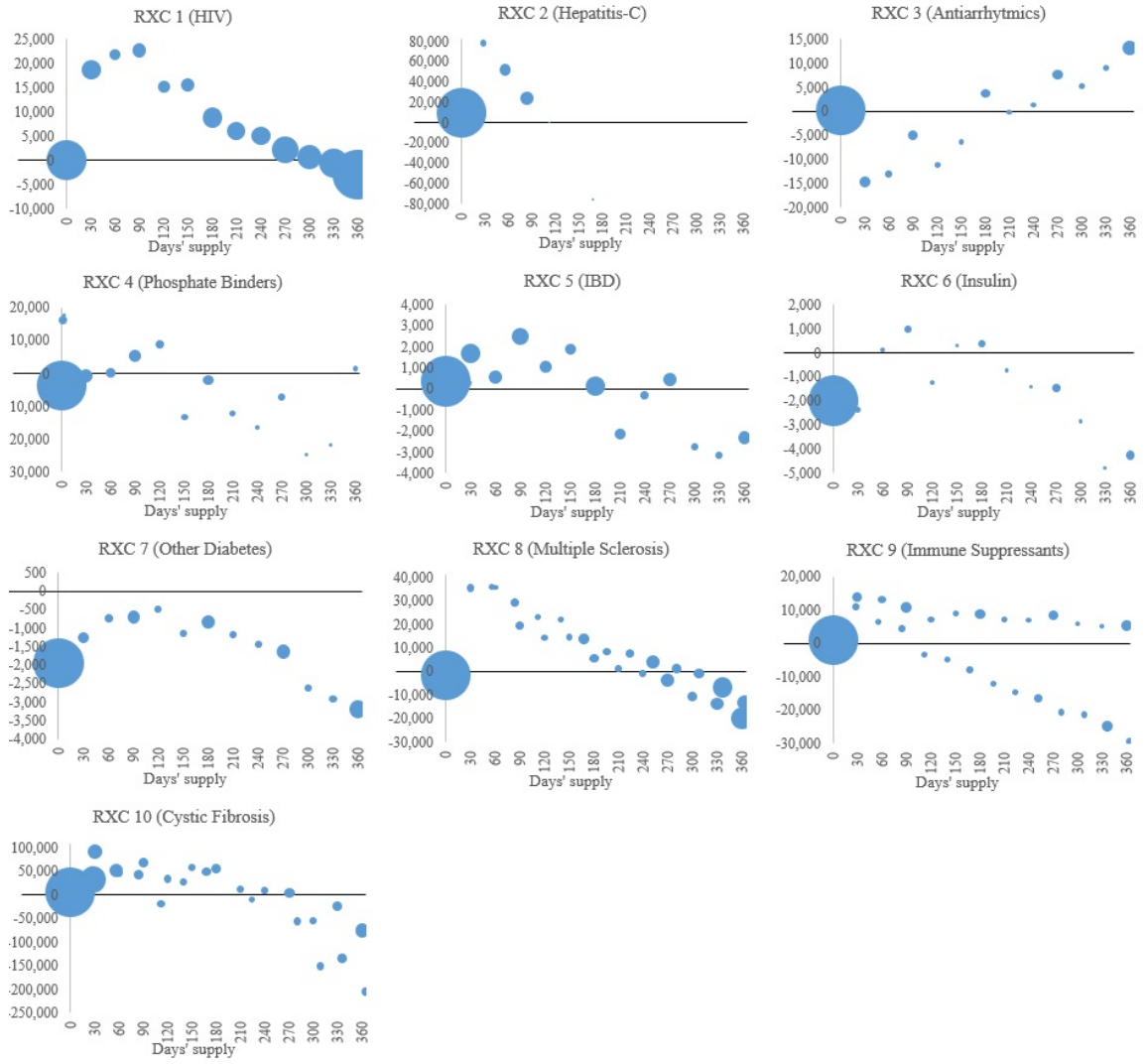
7.2 Determinants of the impact of thresholds on fit

The individual fit statistic is based on the difference between each enrollee's actual costs and the costs predicted by the risk adjustment model. To better understand the effect of non-zero thresholds on the fit, Figure 6 explicitly examine these prediction errors, i.e. the over- or under-compensation to patients, at the baseline with a zero-days threshold. The figure groups patients by the annual number of days' supply in their prescriptions for drugs in the RXC. The size of the bubbles in the graph indicates the number of people in each days' supply group (normalized separately for each disease group).

In the groups related to RXCs 1,2,8,9, and 10, patients with shorter prescriptions are mostly over compensated while patients with longer prescriptions are under-compensated (or less over-compensated). This may explain why a non-zero threshold improves the fit for these RXCs - it decreases payments for the over-compensated below the threshold, as the drug-adjustor is turned off for them, and increases payments to the under compensated above the threshold, as the adjustor's coefficient is re-estimated to better fit their higher costs. This also confirms the intuition from the simple example presented in Figure 1 - for these RXCs, cost is an increasing function of the number of days' supply (see Figure 3), essentially similar to the distribution in panel A of the example. In such cases, a non-zero threshold may improve the fit.

RXC3 presents an opposite example. The cost for patients in the RXC-HCC disease group decreases as a function of the number of days' supply, basically similar to the distribution in panel B of Figure 1. As a result of the decreasing cost, patients with shorter prescriptions are under-compensated, while patients with longer prescriptions are over compensated. In such cases, a non-zero threshold will make things worse for the under-compensated below the threshold, further lowering their predicted costs. For the patients above the threshold, the payment will be reduced to better match their actual costs. The total effect on the fit depends on the relative amounts of under and over compensation in each days' supply group, weighted by a quadratic loss function. In RXC3 any non-zero threshold leads to a worse fit.

Figure 6. Over/Under-compensation in the disease group with a zero-days threshold, by patients' days' supply



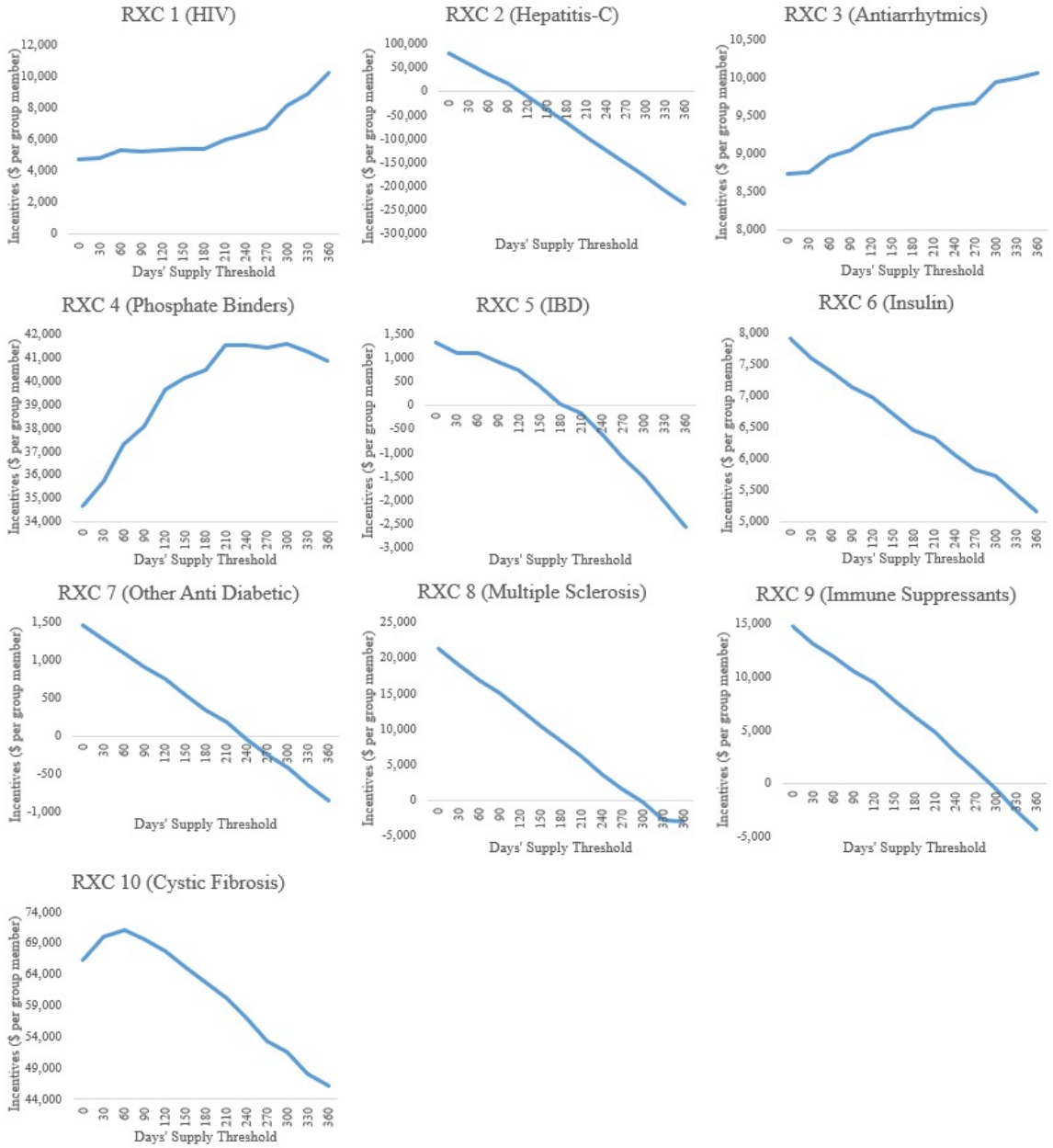
For each disease group, the figure presents the average over- or under-compensation for patients, by the annual number of days' supply in their prescriptions for the RXC's drugs. It shows the compensation gap in the baseline scenario of a zero-days threshold, examining patients with 0 days of supply (i.e. diagnosed only) up to 365 days of supply. The dollar gap in compensation is calculated by the difference between the average payment to the patients in each days' supply subgroup, and the average baseline costs to the plans (which are the actual costs minus the costs covered by the reinsurance program). The size of the bubble around every point indicates the number of patients in each subgroup (normalized separately for each graph).

7.3 Incentives for gaming

Figure 7 presents measure A of gaming incentives - the potential net revenue from prescribing an unlimited amount so all the patients in the disease group cross the threshold. It presents the measure for each RXC, by the simulated days' supply thresholds. At the baseline scenario of zero-days threshold, the measure essentially examines a gaming activity in which all the diagnosed patients in the disease group, without an existing RXC prescription, receive a new one-day prescription of a drug included in the RXC. This gaming behaviour is always profitable, and yields the highest net revenue in RXC2 (Hepatitis-C) - $80,740 per patient in the disease group. In seven disease groups - related to RXCs 2, 5, 6, 7, 8, 9, and 10 - a higher threshold mostly decreases the incentives for gaming. For most of these RXCs there is a threshold in the examined range that eliminates the incentive for gaming. For example, with a threshold just above 180 days' for RXC 5 (IBD), the cost of prescribing more exactly equals the revenue from this gaming activity. In contrast to that, in the disease groups related to RXCs 1,3, and 4 the incentive for gaming mostly *increases* with a higher threshold. This implies that the revenue from turning on the RXC adjustor to a growing number of patients increases faster than the additional cost of prescribing more. This may happen if the drug is inexpensive relative to the high cost of treating patients with long prescriptions. The result may justify CMS's choice to restrict the coefficients of RXCs 3 and 4. The full results of this measure appear in Table A4 in the Appendix.

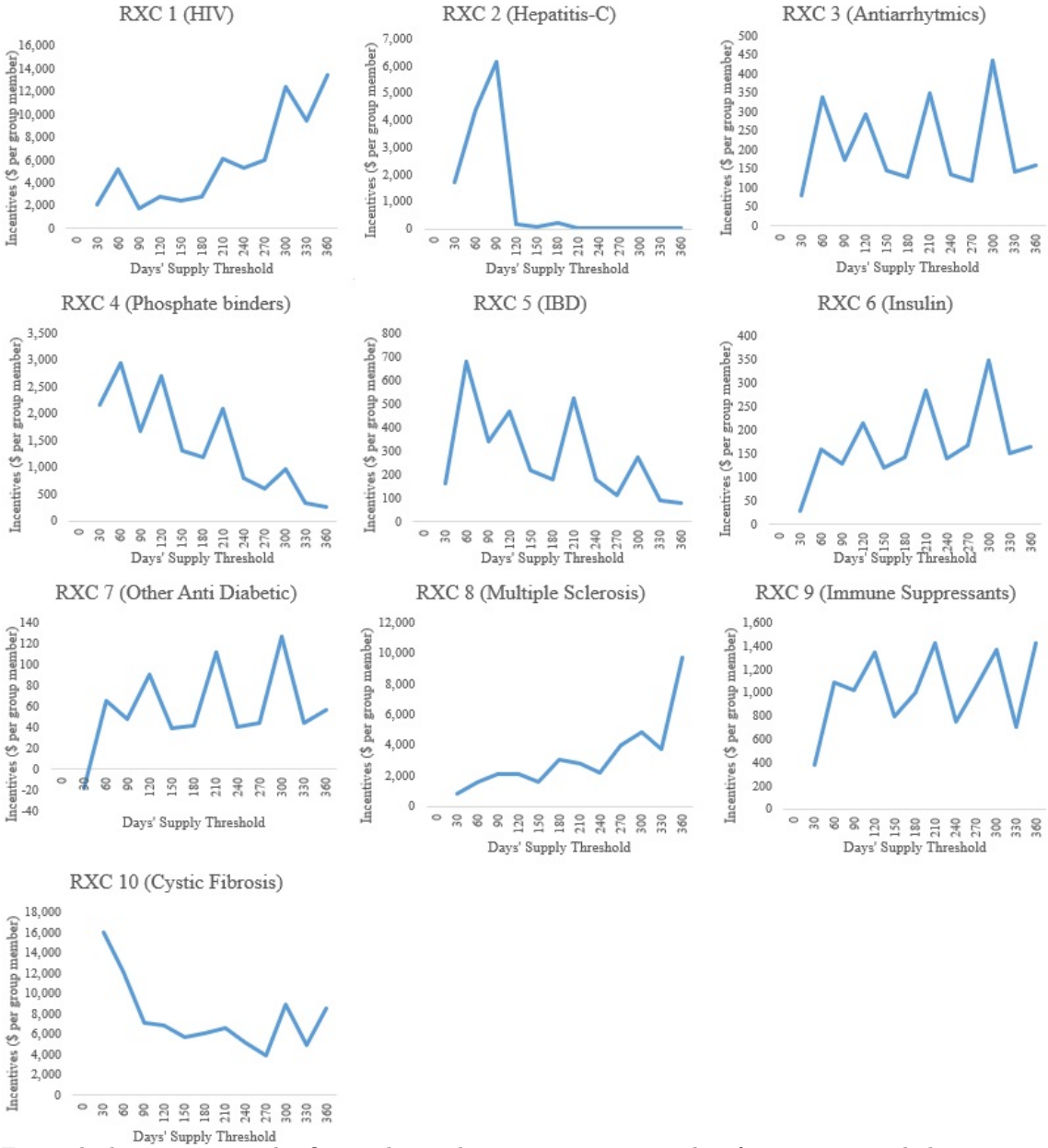
Figure 8 presents measure B of gaming incentives - the net revenue from prescribing up to 30 more days of supply so patients with an existing prescription cross the threshold. It presents the incentives for each RXC, by the simulated days' supply thresholds. The magnitude of this measure is naturally smaller than measure A, as it focuses on gaming for a narrower group of patients. The measure is strongly affected by the distribution of patients by days' supply (shown in figure 3). For example, a large share of the patients treated for HIV (RXC1) fill prescriptions with 270 days' supply or more, making the gaming incentive higher when the threshold is above 270 days. For RXC1, a threshold of 90 days' supply creates the lowest incentive for this marginal gaming. With such a threshold, the gaming yields a net revenue of $1,773 per patient in the HIV disease group. For none of the RXCs the incentives are monotonous, emphasizing the need for an empirical analysis to identify the effect of a threshold on these incentives. The full results for measure B are presented in table A4 in the Appendix.

Figure 7. Net revenue from gaming prescriptions so everyone in the disease group crosses the threshold, by days' supply threshold of each RXC



For each disease group, the figure shows the net revenue to a plan from a gaming behaviour that prescribes more of the RXCs drugs, so all the patients in the group cross a days' supply threshold. The net revenue, per member of the disease group, is shown for all thresholds that are multiples of 30-days, between 0 and 360. The net revenue of each patient is calculated by subtracting the minimal cost of the additional drugs, required to cross the threshold, from the additional revenue accrued to the plan from having the patient cross the threshold and have a higher risk score.

Figure 8. Net revenue from prescribing up to 30 days more for patients with an existing prescription, by days' supply threshold



For each disease group, the figure shows the net revenue to a plan from a gaming behaviour that prescribes up to 30 additional days' supply of the RXC's drugs to push patients with an existing prescription across a days' supply threshold. The net revenue, per member of the disease group, is shown for all thresholds that are multiples of 30-days, between 0 and 360. The net revenue of each patient is calculated by subtracting the minimal cost of the additional drugs, required to cross the threshold, from the additional revenue accrued to the plan from having the patient cross the threshold and have a higher risk score.

Lastly, Table 3 presents measure C, that examines the net revenue from prescribing up to 30 more days of supply so patients without a prescription cross the threshold. The table shows the potential net revenue (average per member of the disease group) with the baseline zero-days threshold and with a 30-days threshold. In all RXCs, a threshold of 30 days' supply lowers the gaming incentive relative to the baseline, by 0.5% to 30%.

Table 3. Net revenue from prescribing up to 30 days more for diagnosed patients with no RXC prescription, dollars per patient in the disease group

	RXC									
	1	2	3	4	5	6	7	8	9	10
# patients in disease group	33,623	33,386	158,635	114,753	87,017	896,307	916,779	32,879	278,230	2,332
Share w.o. prescription (%)	16	87	72	69	51	78	62	43	60	47
Net revenue ($):										
No threshold	29,048	92,531	12,085	50,338	2,578	10,089	2,385	48,888	24,646	142,163
30-days threshold	27,359	65,164	12,029	49,697	1,961	9,670	2,094	42,848	21,697	134,197
Change in net revenue with a 30-days threshold (%)	-5.8	-29.6	-0.5	-1.3	-23.9	-4.2	-12.2	-12.4	-12.0	-5.6

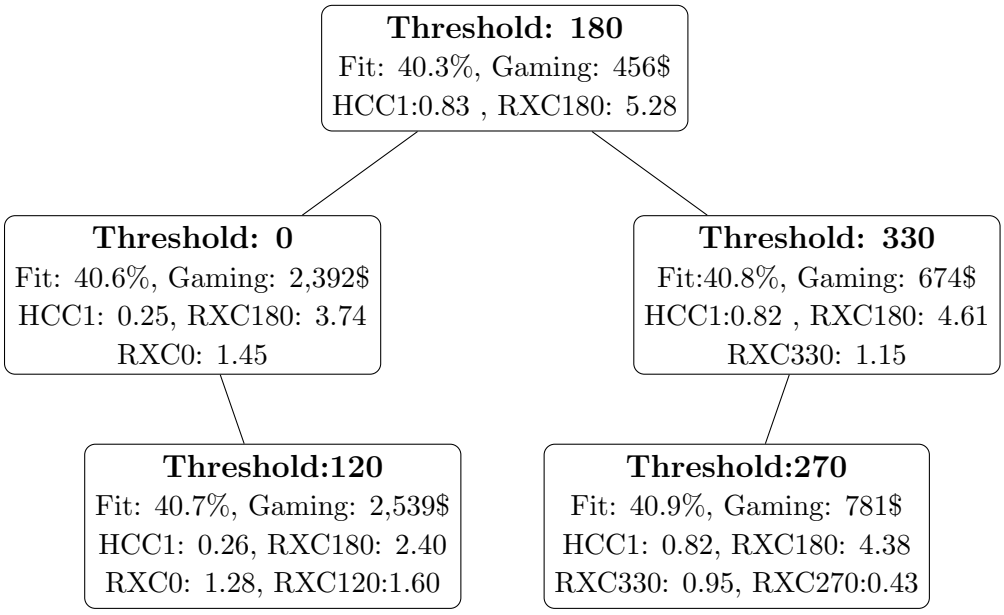
The table presents plans' potential net revenue from giving new RXC prescriptions to patients that are diagnosed with a disease related to the RXC but have no RXC prescription. The net revenue is the revenue to the plan from turning on patients' RXC adjustor, minus the cost of the additional drugs supplied. The measure presented is the average net revenue per member of the disease group (that includes both diagnosed and prescribed patients).

7.4 Fit and incentives for gaming with multiple thresholds

The empirical analysis so far was restricted to a single threshold, exploring only different levels that trigger the existing drug-adjustor. This section eliminates this restriction and tries to find a better partitioning of the days' supply space, potentially using more than a single threshold. The regression tree found by the CART algorithm for RXC1 (Anti-HIV agents), when the loss function puts no weight on gaming incentives ($\lambda = 0$), is presented in Figure 9. The algorithm identifies a threshold of 180 days' supply as the fit-maximizing threshold in the root node (as identified in the simulations above). It then recursively searches for additional thresholds and finds: 0 and 120 days on the range below 180, and 330 and 270 days on the range above 180.

The results change as the loss function puts more weight on the gaming incentives. With a weight of $\lambda = \frac{1}{5}$ on the incentives, the regression tree includes only three thresholds: 150, 270, and 300 days, as further thresholds do not improve the fit enough to compensate for the higher incentives for gaming. When the weight of the gaming incentives increases to $\lambda = \frac{1}{3}$, only a single threshold of 150 days' supply is included in the regression tree.

Figure 9. Best thresholds for RXC1, as chosen by the regression tree algorithm, with a zero weight on gaming incentives

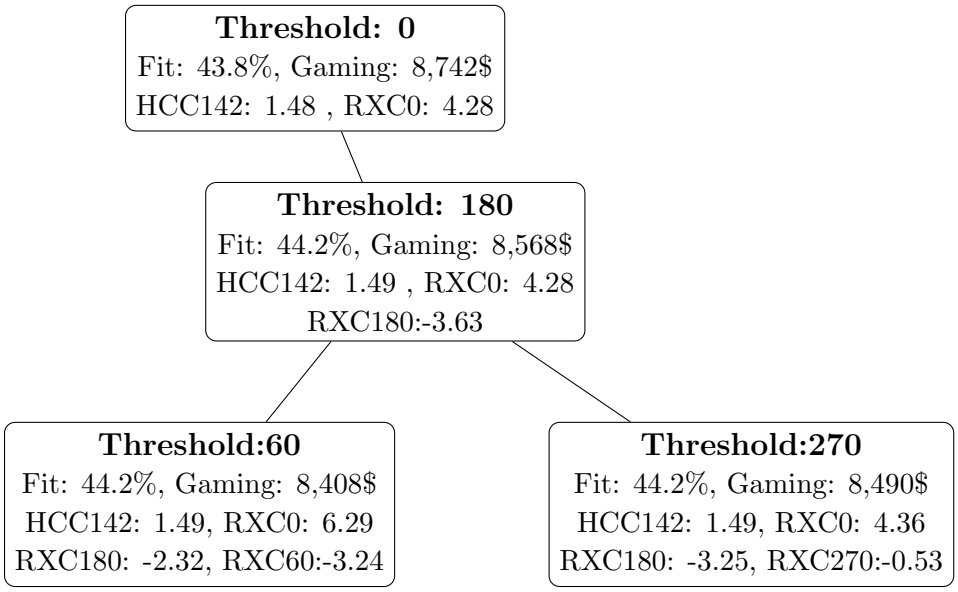


The figure presents the tree created by the CART algorithm that recursively searches for utilization thresholds to maximize the individual fit for the RXC1 disease group (i.e. patients either diagnosed with HIV or prescribed with Anti-HIV agents). Each node shows the fit-maximizing threshold, the individual r-squared fit for the group, and the cumulative incentive for gaming by prescribing up to 30 more days of supply of the RXC drugs. Each node also presents the coefficients of the relevant adjustors: HCC1 is the diagnosis-based adjustor for HIV; RXC180 is the adjustor for Anti-HIV agents, with a utilization threshold of 180, and similarly RXC0, RXC330, RXC120, and RXC270 signify RXC adjustors with different utilization thresholds.

The regression tree for RXC3 (Antiarrhythmics), when all the weight is on fit, is presented in Figure 10 and serves as another example. The fit-maximizing threshold in the root node is 0-days (as in the simulations above). The algorithm continues to search for thresholds above zero that maximize the fit, and finds 180-days in the second level of the tree, and 60 and 270 at the third level.

However, the coefficients estimated for the new adjustors of the non-zero thresholds are negative. At the second level of the tree, adding one day of supply to a patient with 179 days' supply prescription, will *decrease* her risk score by 3.63, leading to a payment lower by $20,839. Thus, an additional adjustor for a 180-days threshold would violate the monotonicity principle, set in Pope et al. (2004) and widely adopted by policy makers in developing risk adjustment systems. This principle states that insurers should not be penalized for additional recording of diagnoses, and in the current context - for additional recording of days of supply in patients' prescriptions. Violation of monotonicity may create an incentive for gaming the system by either

Figure 10. Best thresholds for RXC3, as chosen by the regression tree algorithm, with a zero weight on gaming incentives



The figure presents the tree created by the CART algorithm that recursively searches for utilization thresholds to maximize the individual fit for the RXC3 disease group (i.e. patients either diagnosed with Specified Heart Arrhythmias or prescribed with Antiarrhythmics). Each node shows the fit-maximizing threshold, the individual r-squared fit for the group, and the cumulative incentive for gaming by prescribing up to 30 more days of supply of the RXC drugs. Each node also presents the coefficients of the relevant adjustors: HCC142 is the diagnosis-based adjustor for HIV; RXC0 is the adjustor for Antiarrhythmics, with a utilization threshold of 0-days, and similarly RXC180, RXC60, and RXC270 signify RXC adjustors with different utilization thresholds.

skimping on the provision of the drug, under-reporting its use, or providing it in ways that are not recorded in claims (e.g. sending a coupon to cover out-of-pocket purchases). This kind of gaming is not examined in this paper, that measures incentives to game the risk adjustment system only by prescribing more or providing additional services.

8 Utilization Thresholds for Non-drug Adjustors

Utilization thresholds may apply not only to adjustors based directly on utilization, like the consumption of prescription drugs, but also to any adjustor that is related to utilization indirectly. An example for the latter are morbidity-based adjustors. They are based on data that appears in claims, and thus depend on the utilization of services (Geruso and McGuire 2016). In this section I examine two examples of utilization thresholds for diagnosis-based adjustors to demonstrate how the impact

of these thresholds is again an empirical question. The examined thresholds are applied to the number of times a diagnosis appears in a patient's claims, regardless of the setting in which the diagnosis is recorded (in-patient or out-patient).[27] When counting the number of appearances, a diagnosis is counted at most once per hospital admission. Figure 11 examines the effect of simulated thresholds for two adjustors: To the left, CC19 that indicates "Diabetes with acute complications", and on the right, CC21, that indicates "Diabetes without complications".

The bars in Panel A in the figure present the distribution of patients for which the CC adjustor is turned on, by the number of times the diagnosis appear in their claims[28]. The dots in panel A present the average annual cost of patients by the number of appearances. Most patients with a CC19 diagnosis have only one claim that denotes it, and for 45% of them it is an in-patient claim from an hospital admission. Average costs are higher for patients with one appearance (40,828$) than for patients with two (35,891$) or three appearances (34,092$). In contrast to that, most patients with a CC21 diagnosis (a much larger group) have more than one appearance of this diagnosis, and almost all of these appearances are recorded in an out-patient setting. The annual cost of patients with one appearance is 11,740$ and cost monotonously increases as patients have a higher number of appearances.

Panel B presents the individual fit for the group of patients with a diagnosis included in the CC group, by appearances thresholds. It present the R^2 measure (solid line) and the Cumming's Prediction Measure (dashed line), with the fit at the baseline scenario - a single appearance threshold - indexed to 100. The baseline scenario provides the best fit for CC19. For CC21 a 2-appearances threshold improves the R^2 fit (and a 4-appearances threshold maximize the CPM).

Panel C presents the net revenue from a gaming activity that adds at most a single additional appearance of a CC diagnosis to patients that are already diagnosed (a measure equivalent to Measure B for drug-adjustors). The net revenue for each gamed patients is the revenue due to crossing the threshold and turning the CC on,[29] minus the cost of gaming, defined here as the cost of an additional out-patient visit in which

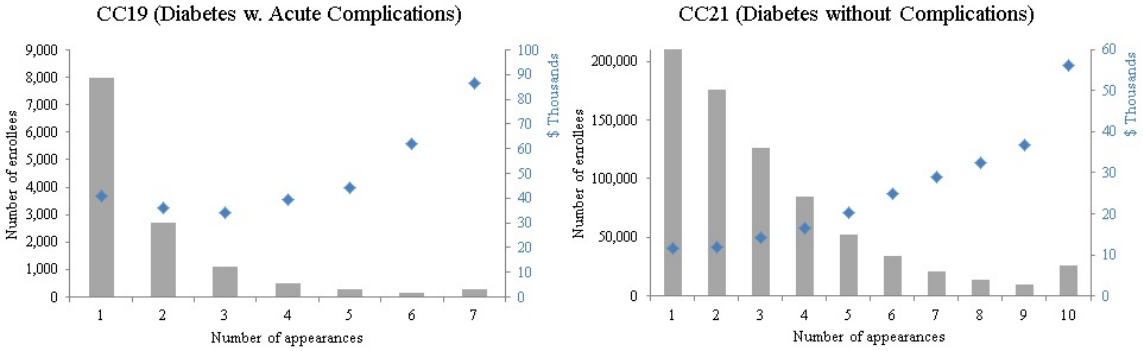
[27] Alternative thresholds may be applied only to out-patient diagnoses, examine whether the diagnoses were recorded in separate quarters of the years, require a minimum cost of a claim to allow the diagnosis included in it to be used for calculating risk scores, etc.

[28] If a patient has several diagnoses that are included in the same CC, the diagnosis with the highest number of appearances is used.

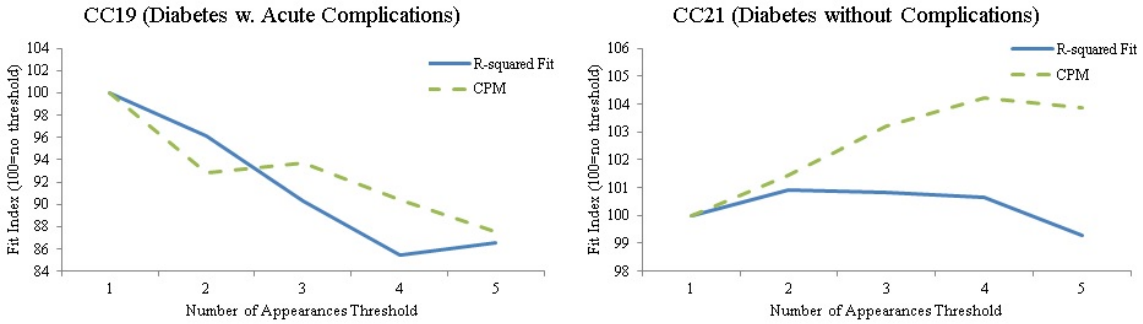
[29] Turning the CC on may yield no additional revenue if a CC higher in the hierarchy of Condition Categories is already turned on.

Figure 11. Appearances thresholds for diagnoses-group adjustors: Distribution of enrollees, fit, and incentives for gaming

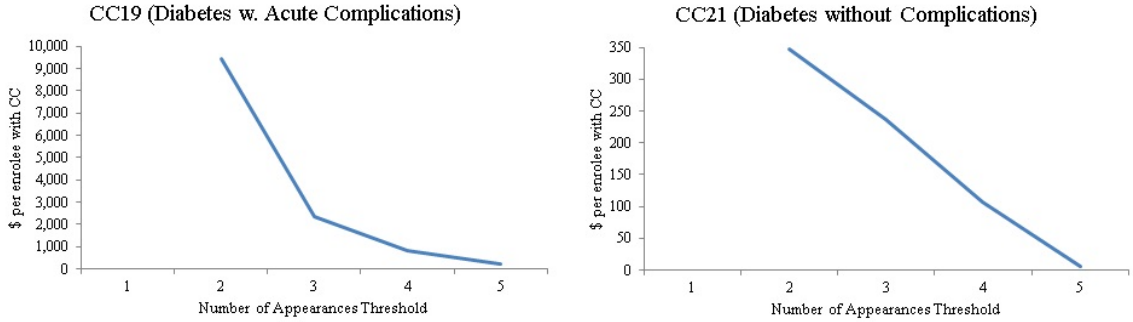
(a) The distribution of patients with a CC diagnosis and their average annual cost, by the number of annual appearances of the diagnosis



(b) Individual fit for patients with CC diagnosis, by appearances threshold



(c) Net revenue from gaming CC adjustors by adding at most a single appearance of a diagnosis to patients with an existing CC diagnosis, by appearances threshold



Panel A presents the distribution and costs of patients, by the number of times that a diagnosis included in the Condition Category (CC) appears in their annual claims. Bars show the number of patients in each number-of-appearances category (left axis). The dots show the average cost of patients in each category (right axis). Panel B show the calculated individual fit for patients with a diagnosis included in the CC, by the number-of-appearances threshold. It presents both the R^2 fit statistic (solid line) and the Cumming's Prediction Measure (dashed line). Panel C presents the potential net revenue to plans from gaming the number of times a diagnosis appears in a patient's claims. The target population for gaming includes patients with an existing CC diagnosis, and the gaming activity adds at most a single appearance, by inducing an additional out-patient claim that will record the diagnosis. Revenue is measured as dollars per enrollee with an existing CC diagnosis.

the diagnosis is coded.[30] This kind of incentive for gaming decreases monotonously with higher thresholds for both CCs.

To conclude, a utilization threshold that requires more than one appearance of a diagnosis to turn on a CC-adjustor may increase fit for some adjustors, and decrease it for others. Higher thresholds may be more beneficial for CCs in which a large number of patients have more than one appearance of the diagnosis (e.g. chronic conditions rather than acute episodes), and when patients' costs increase in the number of appearances (then patients with a higher number of appearances tend to be under-compensated). Here again, simulations can serve as a tool to estimate the effect of utilization thresholds on the fit of the model and the incentives for gaming it.

9 Discussion

9.1 The lack of a tradeoff between fit and the incentives for gaming

When implementing the new drugs-diagnoses risk adjustment system, CMS declared that it is seeking to "strike a reasonable balance between increasing predictive accuracy and reducing incentives for overprescription" (CMS 2016b). This reflects a common belief in the existence of a tradeoff between fit and the incentives for gaming. However, this paper shows that such a tradeoff doesn't always exist. For four out of the ten RXCs, non-zero thresholds can both improve the fit and reduce the incentives for gaming the prescription behavior: A 60-days threshold for RXC2 (Anti Hepatitis-C agents) improves fit in the disease group by 3.1%. It also reduces by 54% the net revenue to the plan from gaming the prescription behavior so all patients cross the threshold (measure A) - from $80,740 to $37,015 per patient in the disease group; For RXC8 (MS Agents), a 120 days' supply threshold increases the disease group's individual fit by 9.6%. The net revenue from gaming under such threshold is $12,808 per member of the group, 40% lower than the incentive with no threshold; In RXC9 (Immune suppressants), a 120-days threshold increases the individual fit in the disease group by 2.1%, and decreases the gaming incentive by 36%; Lastly, a 180-days threshold for RXC10 (Cystic Fibrosis) allows to improve the fit in the disease group (by 8.6%), while decreasing the incentive for gaming by 5%. The lack of tradeoff between fit and incentives for gaming is apparent also in cases where both measures

[30] If the patient has out-patient visits in which a CC-diagnosis appears, than their average cost is used. If the patient has only in-patient claims, then the average cost of an out-patient visit for the whole group is used.

become worse with a threshold. For example, a single 180-days threshold for RXC3 (Antiarrhythmics) decreases the overall fit by 0.13% and hurts the fit in the disease group by 1.4%, while increasing the incentive for gaming by 15%.

9.2 Dynamic vs. Static Incentives for Gaming

A caveat to the incentive measures defined in this paper is that the measures are static in nature, i.e. they measure the incentives for gaming within a single year (in the concurrent payment system used at the Marketplaces). The expected incentives may be lower for later years, especially when the risk adjustment model is re-estimated using claims data from the Marketplaces themselves (data that CMS began using in 2019), and when more plans, with a larger share of relevant patients, game the system. In such a case, even if the absolute return to gaming decreases, a plan that avoids gaming alone may suffer financially, as the predicted costs of gamed RXCs decreases. Behrend, Felder, and Busse (2007) show that in a prospective payment system, gaming may be lucrative to health plans as long as the share of plans gaming and the share of patients gamed are not too high.

Another dynamic aspect is the potential effect of gaming on patients' selection into plans. Gaming activity that easily provides longer prescriptions to drugs that treat a certain disease, may attract patients with the disease to the plan. Such adverse selection may change the incentives in the following years. In this paper, I abstract away from such issues.

10 Conclusions

I examine the impact of utilization thresholds on the performance of risk adjustment systems, specifically the model fit and the incentives for gaming. The sign and size of this impact is an empirical question that, inter alia, depends on the cost distribution of enrollees by the threshold categories. I study this question in the setting of the U.S. Marketplaces, that added prescription-drug adjustors to their risk-adjustment model in 2018, but set no thresholds on the number of days' supply in these prescriptions. I show that for some drug adjustors, there is no tradeoff between fit and incentives for gaming - a non-zero days' supply threshold can both improve fit, and decrease the potential net revenue from gaming. The paper demonstrates how regression trees may guide the choice of multiple utilization thresholds and allow policy makers to explicitly balance their desire for a better fit with their concerns from gaming.

References

Behrend, C., S. Felder, and R. Busse. 2007. "Susceptibility to Strategy of the Drug Component of the IPHCC+RxGroups Classification System in a Risk-Adjusted Morbidity Compensation Scheme - A Conceptional and Data-Supported Analysis (German)". *Das Gesundheitswesen* 69 (1): 1–10.

Bryndová, Lucie, Pavel Hroboň, and Henrieta Tulejová. 2019. "The 2018 Risk-Adjustment Reform in the Czech Republic: Introducing Pharmacy-Based Cost Groups and Strengthening Reinsurance". *Health Policy* 123 (8): 700–705.

CMS. 2017. *Creation of the 2018 Benefit Year HHS-Operated Risk Adjustment Adult Models Draft Prescription Drug (RXCUIs) to HHS Drug Classes (RXCs) Crosswalk.*

— . 2016a. *March 31, 2016, HHS-Operated Risk Adjustment Methodology Meeting, Discussion Paper.*

— . 2016b. *Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2018.*

— . 2018. *Updated 2019 Benefit Year Final HHS Risk Adjustment Model Coefficients.*

Ellis, Randall P., Bruno Martins, and Sherri Rose. 2018. "Risk Adjustment for Health Plan Payment". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef, 55–104. London: Academic Press (Elsevier).

Geruso, Michael, and Thomas G. McGuire. 2016. "Tradeoffs in the Design of Health Plan Payment Systems: Fit, Power and Balance". *Journal of health economics* 47:1–19.

Gilmer, Todd. 2013. *Risk Adjustment Using CDPS.*

Israel Ministry of Health, General Manager Regulations. 2009. *National Health Insurance Regulations (Deducting the Cost of Severe Diseases) - Measures and Reporting Procedure (Hebrew).*

Kautter, John, et al. 2014. "The HHS-HCC Risk Adjustment Model for Individual and Small Group Markets under the Affordable Care Act". *Medicare & Medicaid research review* 4 (3).

Lamers, Leida M., and Rene C. J. A. van Vliet. 2003. "Health-Based Risk Adjustment: Improving the Pharmacy-Based Cost Group Model to Reduce Gaming Possibilities". *The European Journal of Health Economics* 4 (2): 107–114.

Lamers, Leida M., Rene C. J. A. van Vliet, and Winand PMM van de Ven. 1999. "Farmacie Kosten Groepen: Een Verdeelkenmerk Voor Normuitkeringen Gebaseerd Op Medicijngebruik in Het Verleden (Pharmacy-Based Cost Groups: A Risk Adjuster for Capitation Payments Based on the Utilization of Prescribed Drugs)".

Layton, Timothy J., E. Montz, and Mark Shepard. 2018. "Health Plan Payment in US Marketplaces: Regulated Competition With A Weak Mandate". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef. London: Academic Press (Elsevier).

Layton, Timothy J., et al. 2018. "Evaluating the Performance of Health Plan Payment Systems". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef, 133–180. London: Academic Press (Elsevier).

Pope, Gregory C., et al. 2004. "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model". *Health Care Financing Review* 25 (4): 119–141.

Schmid, Christian P. R., Konstantin Beck, and Lukas Kauer. 2018. "Health Plan Payment in Switzerland". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef, 453–490. London: Academic Press (Elsevier).

Schokkaert, Erik, Joeri Guillaume, and Carine van de Voorde. 2018. "Risk Adjustment in Belgium: Why and How to Introduce Socioeconomic Variables in Health Plan Payment". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef, 209–234. London: Academic Press (Elsevier).

van Kleef, Richard C., et al. 2018. "Health Plan Payment in the Netherlands". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef, 397–430. London: Academic Press (Elsevier).

Wasem, Jurgen, et al. 2018. "Health Plan Payment in Germany". In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, ed. by Thomas G. McGuire and Richard C. van Kleef, 295–330. London: Academic Press (Elsevier).

A Appendix

A.1 Appendix tables

Appendix Table A1. The use of information on drug prescriptions in risk adjustment systems

Country	Insurance Market (year added)	Purpose	# of groups (year)	Utilization-based thresholds
U.S.A[31]	ACA Marketplaces	Independent adjustors; Severity indicators for related diagnoses	10 drug classes (RXC) (2019)	None
U.S.A[32]	Medicaid managed care (*Medicaid Rx*[33]: CA, DC; *CDPS+Rx*: DE, NJ, MO, PA, OH, FL)	Independent adjustors	Medicaid Rx: 45 therapeutic categories; CDPS+Rx: 15 drug categories (MRX)	None

[31] CMS (2016b)

[32] Gilmer (2013)

[33] Medicaid Rx and Chronic Illness & Disability Payment System (CDPS) are two risk adjustment models that U.S. states use to pay Medicaid managed care plans.

Country	Insurance Market (year added)	Purpose	# of groups (year)	Utilization-based thresholds
Germany[34]	Social Health Insurance (2009 onwards)	Validation of some outpatient and minor inpatient diagnoses; Severity interaction with some diagnoses; Independent Insulin use	21 (2019)	**10 days** for validation of acute-recurrent disease; **42 days** for medication to be taken as needed; **183 days** for validation of chronic diseases; For some diseases, different thresholds for children under age 12
Switzerland[35]	Social Health Insurance (2020)	Independent adjustors	35 Pharmaceutical Cost Groups (PCG) (2020)	**180 Days** (DDD); **3 days** for drugs for cancer; **15 days** for drugs for "complex cancer"

[34]Wasem et al. (2018), German Federal Office for Social Security (Bundesamt für soziale Sicherung) website (https://www.bundesamtsozialesicherung.de/de/themen/risikostrukturausgleich/festlegungen/)

[35]Schmid, Beck, and Kauer (2018), Swiss Federal Office of Public Health Website (https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-versicherer-aufsicht/risikoausgleich.html)

Country	Insurance Market (year added)	Purpose	# of groups (year)	Utilization-based thresholds
Netherlands[36]	Social Health Insurance (2002: Somatic Care, 2008: Mental Care)	Independent adjustors	Somatic care model: 33 PCGs; Mental care model: 7 PCGs	**181 days** (DDD) for most groups; **91 days** for 15 specific groups[37], and for four more groups for under 18 years old only[38]; **3 prescriptions** for Cancer and Immunoglobulins; **None** for extremely high-cost drugs
Belgium[39]	(2017)	Independent adjustors; Severity indicators for age categories	16	**90 days** (DDD)
Israel[40]	National Health Insurance (1994)	Validation only for the "severe illnesses" adjustors	3 (2019)	Consumed in both halves of the year

[36] van Kleef et al. (2018), Lamers and van Vliet (2003)

[37] 91 days threshold applies to the groups: Glaucoma, Psychosis and addiction, Diabetes type I, Crohn's disease / ulcerative colitis, HIV / AIDS, Transplants, Parkinson's disease, cystic fibrosis / pancreatic enzymes, brain / spinal cord disorders: MS, hormone-sensitive tumors, pulmonary (arterial) hypertension, growth disorders, acromegaly, chronic anticoagulation, and hypertension.

[38] 91 days threshold for under 18 years old applies for the groups: Thyroid disorders, depression, asthma, and epilepsy.

[39] Schokkaert, Guillaume, and van de Voorde (2018)

[40] Israel Ministry of Health (2009)

Country	Insurance Market (year added)	Purpose	# of groups (year)	Utilization-based thresholds
Czech Republic[41]	(2018)	Independent adjustors	25 PCGs	**181 days** (DDD) for all PCG; MoH can set per-PCG thresholds between 121 and 365
Slovakia[42]	(2012)	Independent	24 PCGs (2012)	

[41] Bryndová, Hroboň, and Tulejová (2019)

[42] Health Policy Institute Website: http://www.hpi.sk/2012/08/poistovne-s-chorlavejsim-kmenom-budu-mat-viac-penazi/

Appendix Table A2. Risk score coefficients of RXC adjustors and their related HCC adjustors, by days' supply threshold

	Days' Supply Thresholds												
Coefficient	0	30	60	90	120	150	180	210	240	270	300	330	360
RXC1	4.27	4.50	4.91	4.98	5.13	5.19	5.28	5.41	5.44	5.50	5.64	5.76	5.88
HCC1	0.25	0.26	0.28	0.35	0.47	0.63	0.82	1.27	1.62	1.90	2.54	2.81	3.12
RXC1 X HCC	0.80	0.57	0.22	0.13	-0.03	-0.19	-0.42	-0.87	-1.19	-1.46	-2.20	-2.47	-2.79
RXC2	15.29	16.84	18.94	31.77	34.92	36.35	54.73	58.62	61.58	64.42	67.97	78.24	87.15
HCC34	9.11	9.13	9.22	9.65	9.78	10.00	10.42	10.44	10.43	10.43	10.45	10.45	10.45
HCC35	4.06	4.10	4.17	4.44	4.55	4.66	5.12	5.14	5.15	5.15	5.15	5.16	5.16
HCC36	1.67	1.75	1.83	2.25	2.39	2.50	3.07	3.09	3.11	3.11	3.11	3.12	3.13
HCC37_1	0.27	0.53	1.62	4.36	4.50	4.61	5.03	5.06	5.07	5.08	5.10	5.12	5.12
RXC2 X HCC	0.99	0.54	-0.27	-3.08	-2.25	-1.80	-6.07	-0.77	1.58	-0.59	-1.93	-5.32	-21.40
RXC3	1.98	1.75	1.27	1.09	0.91	0.78	0.71	0.63	0.57	0.55	0.47	0.46	0.42
HCC142	1.48	1.51	1.67	1.77	1.87	1.94	1.99	2.04	2.07	2.09	2.09	2.10	2.10
RXC3 X HCC	0.13	0.30	0.30	0.16	0.00	-0.17	-0.37	-0.61	-0.73	-0.91	-1.02	-1.20	-1.32
RXC4	5.06	4.79	3.28	2.98	3.02	2.95	2.37	2.18	2.07	2.03	1.63	1.51	1.68
HCC183	3.03	3.14	3.34	3.50	3.69	3.81	3.89	3.98	4.01	4.04	4.10	4.11	4.12
HCC184	11.63	11.64	12.46	13.11	13.91	14.48	14.97	15.43	15.70	15.88	16.26	16.39	16.46
RXC4 X HCC	3.71	4.17	5.43	5.29	4.92	4.46	4.48	4.66	4.55	4.66	4.13	4.14	4.43
RXC5	1.67	1.69	1.71	1.69	1.75	1.74	1.76	1.84	1.86	1.90	2.00	2.06	1.96
HCC41	31.20	31.20	31.20	31.20	31.21	31.21	31.21	31.21	31.21	31.21	31.21	31.21	31.21
HCC48	2.04	2.04	2.09	2.09	2.09	2.10	2.11	2.11	2.12	2.13	2.15	2.17	2.19
RXC5 X HCC	-1.21	-1.23	-1.30	-1.26	-1.26	-1.24	-1.25	-1.24	-1.24	-1.29	-1.34	-1.39	-1.32
RXC 6	1.51	1.50	1.40	1.41	1.46	1.48	1.52	1.61	1.62	1.65	1.75	1.81	1.89
HCC18	4.83	4.84	4.87	4.88	4.92	4.94	4.96	4.99	5.00	5.01	5.00	5.00	5.01
HCC19	1.86	1.88	2.00	2.07	2.15	2.21	2.26	2.34	2.39	2.44	2.51	2.55	2.57
HCC20	0.61	0.62	0.65	0.67	0.71	0.74	0.77	0.82	0.85	0.88	0.94	0.97	0.99
HCC21	0.37	0.37	0.38	0.39	0.41	0.42	0.43	0.45	0.46	0.47	0.48	0.49	0.50
RXC6 X HCC	0.25	0.25	0.36	0.35	0.31	0.29	0.26	0.19	0.18	0.15	0.15	0.12	0.07
RXC 7	0.53	0.52	0.52	0.52	0.55	0.56	0.57	0.61	0.61	0.61	0.67	0.68	0.66
HCC18	4.83	4.83	4.84	4.84	4.84	4.84	4.84	4.85	4.85	4.85	4.86	4.86	4.87
HCC19	1.86	1.86	1.87	1.87	1.88	1.89	1.89	1.91	1.92	1.93	1.95	1.96	1.97
HCC20	0.61	0.61	0.62	0.62	0.63	0.64	0.65	0.66	0.67	0.68	0.71	0.72	0.73
HCC21	0.37	0.37	0.38	0.38	0.38	0.39	0.40	0.41	0.41	0.42	0.44	0.45	0.46
RXC7 X HCC	-0.11	-0.10	-0.09	-0.09	-0.09	-0.09	-0.09	-0.10	-0.10	-0.11	-0.13	-0.13	-0.12
RXC8	10.02	10.19	10.47	10.74	11.00	11.17	11.48	11.74	11.91	12.18	12.47	12.72	13.40
HCC118	2.83	2.86	2.94	3.09	3.29	3.46	3.87	4.31	4.65	5.15	5.61	5.89	6.61
RXC8 X HCC	-1.47	-1.57	-1.68	-1.85	-2.07	-2.26	-2.73	-3.18	-3.58	-4.11	-4.61	-4.95	-5.70
RXC9	4.31	4.38	4.76	5.01	5.42	5.59	5.73	6.26	6.44	6.54	7.28	7.41	6.95
HCC41	31.20	31.20	31.19	31.18	31.16	31.16	31.14	31.13	31.12	31.11	31.10	31.10	31.09
HCC48	2.04	2.07	2.11	2.16	2.19	2.23	2.29	2.34	2.38	2.47	2.53	2.61	2.81
HCC56	1.67	1.70	1.72	1.79	1.87	1.94	2.07	2.16	2.25	2.44	2.55	2.67	2.96
HCC57	0.77	0.78	0.79	0.80	0.83	0.84	0.86	0.89	0.91	0.92	0.95	0.97	0.99
RXC9 X HCC41_48	2.29	2.38	2.19	2.03	1.75	1.65	1.54	1.03	0.93	0.84	0.22	0.14	0.82
RXC9 X HCC56	-0.91	-0.96	-1.16	-1.31	-1.55	-1.67	-1.81	-2.03	-2.18	-2.41	-2.73	-2.92	-3.21
RXC9 X HCC57	-3.66	-3.73	-4.08	-4.31	-4.68	-4.84	-4.97	-5.44	-5.61	-5.74	-6.28	-6.53	-6.38
RXC9 X (41/48)&(56/57)	-0.79	-1.03	-1.04	-0.94	-0.77	-0.70	-0.58	0.02	0.14	0.29	0.70	0.71	0.00
RXC10	8.22	11.01	13.63	14.97	15.33	16.65	18.57	19.38	20.46	20.79	22.41	21.65	20.42
HCC158	19.39	19.83	20.37	20.74	20.87	20.98	21.06	21.19	21.20	21.29	21.43	21.43	21.49
HCC159	1.98	2.51	3.11	3.76	4.67	5.35	6.41	7.59	8.33	9.40	10.72	11.79	13.17
RXC10 X HCC	16.59	13.61	11.75	10.88	11.59	10.62	9.23	8.66	8.34	7.63	6.63	6.91	7.42

Appendix Table A3. Individual fit for all enrollees and for each RXC-HCC disease group, by days' supply threshold

RXC	Days' Supply Thresholds												
	0	30	60	90	120	150	180	210	240	270	300	330	360
	R^2 Fit For All Enrollees (%)												
1	34.033	34.035	34.042	34.046	34.052	34.054	34.056	34.054	34.050	34.046	34.033	34.024	34.012
2	34.032	34.070	34.058	33.956	33.924	33.912	33.880	33.881	33.880	33.871	33.871	33.871	33.866
3	34.033	34.023	34.006	33.999	33.992	33.990	33.988	33.989	33.990	33.990	33.992	33.994	33.994
4	34.034	34.016	33.937	33.928	33.924	33.884	33.849	33.842	33.838	33.829	33.795	33.780	33.780
5	34.033	34.032	34.028	34.027	34.026	34.024	34.024	34.022	34.020	34.020	34.021	34.020	34.020
6	34.033	34.027	33.998	33.981	33.961	33.941	33.922	33.892	33.868	33.848	33.808	33.780	33.751
7	34.033	34.032	34.031	34.028	34.027	34.025	34.019	34.013	34.007	33.997	33.982	33.973	33.958
8	34.033	34.043	34.058	34.073	34.077	34.077	34.074	34.063	34.053	34.032	33.996	33.966	33.887
9	34.033	34.035	34.064	34.077	34.091	34.064	34.007	34.005	33.954	33.847	33.810	33.720	33.435
10	34.033	34.034	34.060	34.065	34.066	34.066	34.074	34.073	34.069	34.061	34.052	34.025	33.991
	R^2 Fit For the RXC-HCC Disease Group (%)												
1	38.89	39.08	39.68	39.99	40.42	40.50	40.59	40.33	39.90	39.47	38.25	37.46	36.38
2	38.35	38.99	39.54	37.07	36.99	36.66	35.69	35.61	35.60	35.39	35.38	35.37	35.26
3	38.15	38.07	37.84	37.74	37.66	37.62	37.59	37.58	37.58	37.58	37.58	37.59	37.59
4	32.10	31.97	31.42	31.14	30.86	30.68	30.41	30.13	29.96	29.83	29.67	29.58	29.53
5	39.22	39.21	39.14	39.08	39.03	39.00	38.99	38.94	38.90	38.88	38.85	38.83	38.80
6	39.16	39.14	39.06	39.02	38.98	38.93	38.88	38.80	38.73	38.67	38.53	38.44	38.35
7	38.99	38.99	38.99	38.98	38.98	38.98	38.97	38.96	38.94	38.92	38.89	38.87	38.83
8	26.59	27.24	28.13	28.98	29.15	29.10	28.75	27.76	26.93	25.36	22.97	20.96	15.70
9	29.72	29.79	30.12	30.25	30.36	30.08	29.47	29.37	28.81	27.64	27.18	26.21	23.24
10	34.60	34.61	36.71	36.93	36.93	36.88	37.59	37.44	37.11	36.32	35.54	33.14	30.30
	Cummings's Prediction Measure For the RXC-HCC Disease Group (%)												
1	21.65	22.66	25.16	25.73	26.34	26.41	26.01	24.12	22.20	20.43	17.10	14.82	11.58
2	20.65	21.32	22.11	18.42	17.88	17.28	14.87	14.76	14.71	14.65	14.63	14.60	14.53
3	26.38	26.44	26.22	26.06	25.90	25.82	25.76	25.74	25.74	25.72	25.78	25.78	25.80
4	28.03	28.77	28.67	28.46	28.20	27.93	27.63	27.45	27.30	27.16	26.92	26.82	26.78
5	28.66	28.74	29.01	29.15	29.34	29.42	29.45	29.64	29.65	29.62	29.60	29.54	29.46
6	30.92	30.96	31.10	31.20	31.35	31.39	31.40	31.34	31.27	31.16	30.84	30.68	30.46
7	30.76	30.77	30.86	30.92	31.04	31.09	31.12	31.21	31.22	31.21	31.20	31.16	31.05
8	28.38	29.04	30.11	31.06	31.24	31.11	30.14	28.46	27.00	24.49	21.35	18.72	11.39
9	22.55	22.67	23.39	23.57	23.92	23.62	22.67	22.71	21.93	19.95	19.44	17.97	13.71
10	27.13	27.56	29.51	29.43	29.17	28.84	28.76	28.69	27.63	26.15	24.43	21.93	18.70

Appendix Table A4. Incentives for gaming, by days' supply threshold

RXC	Days' Supply Thresholds												
	0	30	60	90	120	150	180	210	240	270	300	330	360
	Measure A: Net revenue from unlimited gaming so all patients cross the threshold ($ per patient in the disease group)												
1	4,695	4,754	5,300	5,245	5,333	5,341	5,382	5,941	6,319	6,732	8,140	8,874	10,157
2	80,740	58,339	37,015	15,988	-12,030	-40,157	-68,183	-96,458	-124,758	-153,043	-181,350	-209,655	-237,963
3	8,743	8,760	8,966	9,055	9,233	9,300	9,353	9,573	9,629	9,669	9,943	9,998	10,065
4	34,650	35,693	37,291	38,080	39,621	40,115	40,484	41,519	41,549	41,427	41,588	41,243	40,835
5	1,321	1,087	1,107	903	734	400	26	-183	-616	-1,102	-1,513	-2,043	-2,584
6	7,898	7,592	7,377	7,135	6,960	6,702	6,457	6,322	6,060	5,816	5,711	5,436	5,167
7	1,468	1,277	1,102	913	747	543	338	176	-39	-254	-416	-641	-860
8	21,251	18,994	16,907	14,926	12,808	10,358	8,418	6,193	3,549	1,560	-271	-2,870	-3,059
9	14,691	13,163	11,923	10,587	9,374	7,753	6,185	4,787	2,917	1,173	-480	-2,593	-4,320
10	66,204	69,966	71,075	69,497	67,643	65,229	62,580	60,216	57,059	53,212	51,565	48,032	46,182
	Share of the disease group with potential profitable unlimited gaming (%))												
1	16.2	17.5	21.0	22.2	23.9	25.5	27.3	31.2	34.7	38.5	46.3	52.0	60.1
2	87.3	88.9	93.2	99.4	11.6	8.3	4.6	2.1	1.3	0.6	0.3	0.2	0.2
3	72.4	72.8	74.9	76.0	77.9	78.8	79.7	81.9	82.8	83.5	86.2	87.1	88.1
4	68.8	73.0	77.5	80.2	84.5	86.6	88.5	91.9	93.1	94.1	95.7	96.2	96.6
5	51.3	52.7	59.9	63.5	67.9	18.1	18.9	24.5	25.2	23.9	25.0	24.0	22.1
6	78.3	78.5	79.8	80.9	82.7	83.7	84.9	87.2	88.3	89.7	92.5	93.7	95.0
7	61.5	61.8	63.5	64.8	67.4	68.3	69.3	72.1	11.2	11.8	14.7	15.2	16.3
8	43.5	44.2	45.8	47.7	49.7	51.2	54.1	13.3	15.2	18.1	21.7	23.7	31.7
9	59.6	60.6	63.3	65.7	68.9	70.7	73.0	75.8	76.9	18.9	21.3	21.9	24.5
10	46.6	59.8	66.6	69.6	72.6	75.0	77.4	79.7	81.2	76.1	76.8	76.9	78.9
	Measure B: Net revenue from prescribing up to 30 days more to the prescribed ($ per patient in the disease group)												
1		2,059	5,214	1,773	2,770	2,424	2,816	6,074	5,297	5,954	12,405	9,411	13,450
2		1,693	4,375	6,146	160	86	226	22	3	21	3	5	5
3		79	341	174	294	146	127	350	135	118	434	141	158
4		2,155	2,940	1,681	2,712	1,317	1,193	2,089	797	595	974	342	269
5		159	678	341	466	219	178	526	178	111	273	87	78
6		29	159	127	216	120	142	284	139	168	350	150	166
7		-19	65	48	91	38	42	112	40	44	127	44	57
8		849	1,559	2,132	2,084	1,630	3,025	2,761	2,162	3,953	4,839	3,716	9,738
9		385	1,083	1,016	1,350	792	993	1,422	747	1,043	1,370	710	1,430
10		16,044	12,113	7,096	6,853	5,693	6,096	6,544	5,228	3,923	8,903	4,857	8,496
	Share of the disease group with potential profitable 30-days gaming to the prescribed (%)												
1		1.4	3.5	1.1	1.7	1.5	1.7	3.7	3.3	3.6	7.6	5.7	8.2
2		1.6	4.3	6.2	0.2	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0
3		0.5	2.1	1.1	1.8	0.9	0.8	2.1	0.8	0.7	2.6	0.9	1.0
4		4.2	4.5	2.4	3.9	1.9	1.7	3.0	1.2	0.9	1.4	0.5	0.4
5		1.4	7.0	3.7	5.6	2.8	2.4	7.4	2.6	1.6	3.9	1.3	1.2
6		0.2	1.3	1.0	1.7	1.0	1.1	2.3	1.1	1.4	2.8	1.2	1.3
7		0.2	1.8	1.3	2.5	1.1	1.2	3.1	1.1	1.2	3.5	1.2	1.6
8		0.8	1.6	2.0	2.0	1.5	2.8	2.7	2.1	3.8	4.7	3.6	9.5
9		1.0	2.7	2.5	3.4	2.1	2.6	3.6	1.9	2.8	3.5	1.9	4.0
10		13.2	6.7	3.0	2.8	2.3	2.6	2.4	2.0	1.5	3.3	1.8	3.2