Investigating the effect of Gleevec list price on adherence and outcomes in Medicare patients with chronic myeloid leukemia.

Samantha Clark

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Anirban Basu, Chair

Zachary Marcum

Jerry Radich

Program Authorized to Offer Degree:

School of Pharmacy

University of Washington

Abstract

Investigating the effect of Gleevec list price on adherence and outcomes in Medicare patients with chronic myeloid leukemia.

Samantha Clark

Chair of the Supervisory Committee:
Anirban Basu, Stergachis Family Endowed Director & Professor of Health Economics
School of Pharmacy

Gleevec is a specialty drug that revolutionized the treatment of chronic myeloid leukemia (CML). Its relatively early introduction and high list price have made it emblematic of drug pricing concerns in the US. In Medicare Part D, out-of-pocket payments (OOP) in the catastrophic coverage phase are co-insurance based and a direct function of list price. The financial impact of this benefit structure on CML patients taking Gleevec may have significant downstream consequences in the form of reduced adherence and suboptimal health outcomes.

This dissertation comprises two aims focused on better understanding the complex relationship between list-price-based OOP, adherence, and outcomes in newly-diagnosed CML patients on Gleevec who are enrolled in Medicare Part D. In the first aim, a fuzzy donut regression discontinuity design was used to estimate the causal effect of Medicare Part D

enrollment at age 65 on adherence and OOP. Results were then combined to calculate the corresponding price elasticity of demand (PED) and determine the degree of price sensitivity in these patients with respect to Gleevec OOP. In the second aim, we applied the alternating conditional estimation (ACE) algorithm alongside a backward elimination procedure to identify time dependence and nonlinearity in the relationship between adherence and time-to-remission. An extended Cox model allowing for the flexible modeling of each effect via unpenalized B-splines was subsequently fit and assessed.

Results from aim 1 indicate that, while there is a large and significant increase in initial OOP of $232 for patients on Part D at diagnosis, the effect on adherence is a smaller and non-significant decrease of 6 percentage points. The corresponding PED of -0.02 indicates that the demand for Gleevec with respect to OOP is highly inelastic for these patients. Aim 2 provides insight into the shape of the dose-response curve (DRC) for Gleevec and demonstrates how the strength of this curve varies over the first year following treatment initiation. Results suggest that the DRC is non-linearly monotonically increasing and that the strength of this effect on time-to-remission is strongest in the first three to four months of treatment.

This study provides evidence regarding potential drivers of the substantial non-adherence observed in CML patients, as their minimal responsiveness to OOP indicates that other factors may be playing a more important role. The finding that cumulative adherence is most important early on can potentially inform monitoring and intervention strategies for meeting treatment milestones. Results from both aims can additionally be used to help guide the design of future studies regarding important analytic decisions and assumptions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Getting to this point was a team effort, and I would like to acknowledge all of the amazing people who have supported me throughout the process. First, my dissertation committee for taking the time to join me on this long (sometimes strange) trip. In particular, I would like to thank Dr. Zach Marcum, whom I only half-jokingly refer to as "a literal angel." I could write an entire other dissertation describing all the ways in which you have gone above and beyond for me during my time in CHOICE. I would not have made it this far without your mentorship. I would also like to thank my dissertation chair, Dr. Anirban Basu, for his support, guidance, and seemingly endless patience with my loose interpretation of internal deadlines.

An acknowledgements section would not be complete without my fellow students in CHOICE. It has truly been the privilege of my academic career getting to know and work with all of you. I would like to especially thank Lauren, Lizzy, Tricia, Enrique, Eddy, Justin, Tommy, Felipe, and Rachel for your support and, more importantly, laughing at my jokes. You're all great researchers and even better friends.

Finally, I would like to thank my best friends from back home (Allie, Annie, Mer, Emily, and Alice), who have all known me for at least 15 years, borne witness to my absolute lack of maturation during that time, and stuck by me regardless.

DEDICATION

To my mom, whom I still miss every day. What is best in me, I owe to you.

My dad, whose combination of brilliance, humility, and deep kindness continues to inspire me.

And my sister, who is my best friend and eternal partner in crime. I love you dearly.

Chapter 1. INTRODUCTION

The use of specialty medications has grown dramatically in the United States over the last 20 years (Anderson-Cook et al., 2014). While many brand-name specialty drugs confer survival and quality of life improvements, their high prices have drawn increasing scrutiny from media, policymakers, and patients. This is especially true within the Medicare Part D program, where they accounted for only 1 percent of all prescriptions dispensed but 30 percent of total net spending on prescription drugs in 2015 (Frakt & Miller, 2018). Due to the benefit structure in Part D, where cost-sharing is based on list price, patients prescribed specialty drugs often enter the catastrophic coverage phase ($7,050 TrOOP threshold in 2022) after their first prescription fill (Dusetzina & Keating, 2016). With 5% coinsurance and no spending cap in this phase, many beneficiaries face high annual medication costs (Dusetzina, 2021). The financial burden imposed by these treatments can have important behavioral consequences, including failure to initiate, medication rationing, and discontinuation (Doshi et al., 2015; Goldman et al., 2007). While it is often assumed that medication non-adherence will negatively impact patients' health, little is known about the downstream effect of high prices on clinical outcomes.

Since the RAND Health Insurance Experiment (RAND HIE) was conducted in the late 1970s and 1980s, conventional wisdom has held that individuals are inelastic with respect to the cost of healthcare (i.e., less responsive to price changes) (Aron-Dine et al., 2013). Results from this study continue to inform policy today, including the design of the Medicare Part D program initiated in 2006. While spending thresholds have changed and the coverage gap has closed, the overall structure of the standard benefit plan remains intact. This includes a deductible followed by three coverage phases with varying levels of co-insurance. Cost-sharing levels were less

consequential at the time that Part D was drafted since common prescriptions were relatively inexpensive compared to contemporary drug pricing (Kirchhoff, 2015). The advent of high-priced specialty drugs in the 2010s has changed that formulation, with co-insurance that's a function of list price and no spending cap leading to high monthly OOPs for many seniors (Anderson-Cook et al., 2014).

As the increasing strain of high OOPs in Medicare beneficiaries has become impossible to ignore, there have been numerous calls for a redesign of the Part D benefit structure (Dusetzina, 2021). The recently implemented Inflation Reduction Act (IRA) represents an important step towards addressing many of these concerns; most notably, by eliminating 5% co-insurance in catastrophic coverage and capping total OOPs at $2,000 in 2024 and 2025, respectively (KFF, 2022). Rebuffing the types of legal challenges and threats to repeal that the Affordable Care Act has and continues to face will require rigorous evidence on the potential impact of reverting to a coinsurance-based payment structure with no spending cap. Additionally, as private insurers move towards benefit plan designs with higher deductibles and coinsurance as a method of cost containment, it is important that the full implications of shifting costs onto patients are known (Agarwal et al., 2017; Dusetzina et al., 2016, 2022). Key to future policy discussions will be a better understanding of how high OOP affects medication adherence in patients and what, if any, implications this might have regarding patient outcomes.

The use of Gleevec (imatinib), a tyrosine kinase inhibitor (TKI), in the treatment of CML provides an ideal case study for assessing the relationship between list-price-based OOP, adherence, and clinical outcomes.

CML is a rare cancer of the bone marrow with a historically poor life expectancy of only two to three years following diagnosis (Bower et al., 2016; Deininger et al., 2020). The introduction of Gleevec in 2001 revolutionized treatment and prognosis (Jabbour & Kantarjian, 2020; Marin et al., 2010; Noens et al., 2009). Hailed as a "magic bullet," it, along with second- and third-generation TKIs, has allowed patients to achieve an almost normal lifespan with proper disease management (Gambacorti-Passerini et al., 2021; Iqbal & Iqbal, 2014; Noens et al., 2009). Importantly, Gleevec is not generally curative, and patients must remain continuously adherent for the duration of their lifetime to realize and maintain full treatment benefit (Deininger et al., 2020; Jabbour & Kantarjian, 2020).

In addition to being one of the most successful oral chemotherapy agents ever developed, Gleevec is also among the costliest (Abboud et al., 2013; Dusetzina et al., 2014). Although OOP has been identified as a potential driver of non-adherence in the approximately 30% of CML patients who exhibit this behavior, a definitive relationship between the two has not been established (Dusetina 2014; Dusetzina 2022; Marin et al., 2010; Noens et al., 2009; Phuar et al., 2019; Winn et al., 2016). Further, the differential impact of varying adherence levels on the ability and time to achieve remission in this population is not well understood (Darkow 2007; Jabbour & Kantarjian, 2020; Noens 2009).

The application of causal inference and flexible survival modeling techniques in this area can generate novel insight by addressing important methodological limitations and providing a comprehensive view of the process by which high list-price-based OOPs affect patients. This dissertation does this through two separate, but related aims. First, the causal effect of Medicare Part D enrollment at age 65 on OOP and adherence in newly-diagnosed CML patients initiating Gleevec is determined through the implementation of a fuzzy donut regression discontinuity

design. The corresponding price elasticity of demand PED is then estimated from these results to identify the sensitivity of patients to the increase in out-of-pocket payments associated with the Part D benefit structure

Second, we apply the ACE algorithm alongside a backward elimination procedure to detect the presence of time dependence and nonlinearity in the relationship between cumulative Gleevec adherence and time-to-remission. From this analysis, an extended Cox model allowing for the flexible modeling of these effects is fit via unpenalized B-splines. Results are then used to model the full dose-response curve for Gleevec and assess the extent to which the timing and duration of exposure affect this relationship.

Chapter 2. THE PRICE ELASTICITY OF GLEEVEC IN PATIENTS WITH CHRONIC MYELOID LEUKEMIA ENROLLED IN MEDICARE PART D: EVIDENCE FROM A REGRESSION DISCONTINUITY DESIGN

2.1 INTRODUCTION

The use of Gleevec (imatinib), a tyrosine kinase inhibitor (TKI), in the treatment of chronic myeloid leukemia (CML) provides an ideal case study for assessing the relationship between list price, OOP, and adherence in Medicare Part D since almost half of patients are diagnosed over the age of 65 (Jabbour & Kantarjian, 2020). In addition to being one of the most successful oral chemotherapy agents ever developed, Gleevec is also among the costliest (Abboud et al., 2013; Dusetzina et al., 2014). Although OOP has been identified as a potential driver of non-adherence in the approximately 30% of CML patients who exhibit this behavior, a definitive relationship between the two has not been established (Marin et al., 2010; Noens et al., 2009; Phuar et al., 2019; Winn et al., 2016).

Analyses since the RAND HIE study have supported initial findings regarding the inelasticity of demand for prescription drugs; however, their focus was largely on non-specialty drugs for which the corresponding financial burden is low (Chandra et al., 2010, 2014; Gatwood et al., 2014).

Evidence from the literature assessing specialty cancer therapies is more mixed, with some analyses suggesting a higher elasticity for these drugs than for their low-cost counterparts (Doshi et al., 2016). In a 2016 review on cost sharing and specialty drug use, Doshi et al found that high OOP was more strongly associated with non-initiation and abandonment than adherence, which appeared to be insensitive to price (Doshi et al., 2016). These findings are

supported by similar studies indicating that patient demand for specialty drugs is fairly inelastic, with estimates ranging from -0.04 to -0.26 (Goldman 2007; Gatwood 2014; Doshi 2016).

Recent work, however, demonstrated that patients could be more responsive to price than initially thought. Using an instrumental variable design, Jung et al identified the overall elasticity for specialty drugs to be between -0.72 and -0.75, with leukemia-specific estimates of -0.96 to -0.99 (Jung et al., 2017). Importantly, both incident and prevalent cases were included and a significant effect was not identified once treatment had been initiated or in newly diagnosed patients.

Gleevec's high initial list price, early introduction relative to other costly specialty drugs, and the importance of adherence in realizing treatment outcomes have made it an area of considerable interest regarding the impact of drug pricing. The objective of this work is to contribute novel evidence to the growing literature on the price elasticity of specialty drugs and our understanding of the drivers of non-adherence in CML patients by using a regression discontinuity design to retrieve the causal effects of the Medicare Part D benefit structure on initial OOP, adherence, and elasticity of demand in patients with CML diagnosed right before and right after age 65. Our analytic approach combined with the information available in our data allows us to account for the endogeneity associated with plan choice (patients with higher drug costs tend to opt for more generous plans) and the distortionary effect of coupon use (Doshi et al., 2015; Jung et al., 2017; Phuar et al., 2019, 2020; Seymour et al., 2021).

2.2 METHODS

2.2.1 *Data*

We use data from the TriNetX Diamond Network (TDN) database for the period from first availability in 2011 through the end of patent exclusivity following the introduction of

generic Gleevec (imatinib) in early 2016. TDN comprises a convenience sample of claims and encounter data from 1.8 million providers representing 99 percent of U.S. health plans, with longitudinal data covering over 200 million patients (TriNetX, 2022). Data are pulled from open claims clearinghouses and include de-identified medical and pharmacy claims for Commercial, Disability/Workers Compensation, Dual Coverage, Medicare, Medicaid, and VA/Other plans.

The decision to use TDN for the primary analysis rather than other available claims databases (SEER-Medicare, MarketScan) is based on the following considerations: ability to capture hypothesized discontinuities across comparable pre- and post-Medicare populations and inclusion of coupon use as a form of payment assistance.

2.2.2 *Study design*

The study period consists of the time from the first Gleevec fill, defined as the index date, to the end of three-months of follow-up, with the six months preceding the index date serving as a washout period for confirming incident diagnoses (no previous claims with CML diagnosis or treatment) and the baseline period for covariate measurement. A new user design is adopted to account for survivor bias (the most elastic patients will have the worst outcomes), healthy user bias (non-adherers are more likely to switch treatment due to resistance or lack of response), and to minimize the impact of anticipatory plan switching.

Individuals are included if they had no previous indication of CML diagnosis or treatment during the washout period, two or more diagnoses consistent with International Classification of Diseases 9th and 10th revision (ICD-9 and ICD-10) diagnosis codes for "Chronic myeloid leukemia, BCR/ABL-positive, not having achieved remission" (205.10 and C92.10, respectively), and at least one Gleevec order within 12 months of confirmatory diagnosis. Patients meeting the following criteria are excluded: treatment switching within the

follow-up period, non-commercial insurance in the pre-Medicare period (Medicaid, VA), subsidized Medicare coverage (dual coverage), and inpatient or outpatient claims listing a diagnosis code consistent with any additional indications for imatinib (Trivedi et al., 2014).

We have two key exposures of interest, a "running variable" and a binary "cutoff," or "threshold," variable representing the actual treatment (or what we are interested in estimating the causal effect of) that is a function of the running variable. In this analysis, the running variable is age and the threshold variable is age 65. Elasticity is estimated using the adherence and OOP measures described in Table 1.

Table 1. Exposure and outcome measurement

Variable	Estimation method
Exposure	
Age	Because TDN does not include age at claim information, year of birth is used to get an approximate estimate of age at baseline (year of index date – birth year).
Outcomes (standardized to 90-day follow-up period)	
Initial adherence	Based on proportion of days covered (PDC), defined as the number of days supplied of Gleevec over the follow-up period, with maximum adherence capped at 100%.
Initial out-of-pocket payment (OOP)	OOP includes total patient responsibility for the prescription (deductible + co-pay/co-insurance) minus coupon value (where applicable).

2.2.3 *Empirical strategy*

The estimation of PED for Gleevec using observational data is complicated by the presence of multiple sources of bias - primarily, endogeneity associated with plan choice due to individuals most sensitive to price choosing more generous plans either at eligibility or in response to their CML diagnosis. Omitted variable bias is also a concern due to the lack of information on several important confounders in TriNetX data (e.g., socioeconomic status (particularly disposable income) and plan structure). Additionally, selection bias regarding which

patients remain on some form of employer-sponsored health plan and which enroll in Part D post eligibility is also present.

To address these concerns, we use a regression discontinuity design (RD) to retrieve the causal effect of Medicare Part D enrollment at age 65 on initial OOP and adherence. If identifying assumptions are met, the exogenous shock of universal Part D eligibility at age 65 should lead to treatment assignment that is "as-if random" (Hernán & Robins, 2020). These designs also offer the advantage of demonstrating high internal validity in practice and requiring what many consider to be less stringent and more easily testable identifying assumptions than other causal inference methods (Chaplin et al., 2018; Imbens & Lemieux, 2008).

The core identifying assumption of RD is continuity, which requires that the threshold (age 65 in this case) is itself not endogenous with respect to another "shock" or treatment occurring simultaneously (Lee & Lemieux, 2010). In other words, there is a "smoothness" assumption that, in the absence of universal eligibility for Medicare enrollment at age 65, adherence and OOP would have remained continuous across this time period. Importantly, continuity allows for there to be differences in both observed and unobserved confounders around the cutoff as long as they vary "smoothly" with respect to age (Hahn et al., 2001). We evaluate this assumption using the following variables measured during the baseline period: sex, comorbidities (based on Charlson Comorbidity Index), mean 30-day OOP in the 90 days prior to diagnosis, diagnosis year, and diagnosis month.

Aside from the need for a discontinuous jump in the probability of treatment at the threshold (Medicare Part D eligibility at age 65 in this case), the other key assumption of the RD is that individuals must not be able to precisely manipulate the threshold for treatment (Lee &

Lemieux, 2010). This assumption should hold, as individuals cannot manipulate their age or health status to qualify at any other time.

In a sharp regression discontinuity design (SRD) *all* individuals in the sample above the threshold are eligible for and comply with treatment (compliance is 100%). Because individuals may qualify for Part D early at age 62 (through Social Security disability insurance (SSDI) end-stage renal disease, or ALS) and individuals post-65 may choose to not to enroll, this analysis represents a fuzzy regression discontinuity design (FRD). For this design, while there's still a jump in treatment (from 0% to <100%), receipt post-threshold is now probabilistic (Gelman & Imbens, 2019).

We implement an FRD using the Wald estimator, which effectively estimates the intention to treat effect (numerator) scaled by the proportion of compliers (denominator) at the threshold. The numerator is estimated using a reduced form RD model that is equivalent to that used for a sharp design. The scaling factor is derived based on the same design, but with the probability of treatment (enrollment in Part D at age 65) as the outcome (Cunningham, 2021; Hahn et al., 2001; Lee & Lemieux, 2010). The general equation used to estimate each model is provided below.

$$y_i = \gamma * 1(age > 65) + f(age) * 1(age \leq 65) + g(age) * 1(age > 65) + \varepsilon_i(age)$$

where y_i represents a study outcome (OOP, adherence, or Part D enrollment) for patient i, $f(age)$ and $g(age)$ are smooth functions of age for individuals below and above the age 65 threshold, respectively, ε_i represents the error term, and γ is the effect of Medicare eligibility at age 65.

The canonical RD relies on the use of a continuous "running" variable, where the presence of a cutoff or threshold splits the sample by treatment status (Cattaneo et al., 2019). The TDN data, however, do not include age, but rather birth year from which we can derive age using the index date. Because we are estimating adherence over a three-month period, measurement error at age 65 would run the risk of "contamination" in our price elasticity estimates if OOP and adherence were attributed to the wrong treatment.

When the running variable is discrete, as in this case, traditional RD methods become less reliable (Barreca et al., 2011). To address this issue, a "donut RD" (DRD) combined with an "honest confidence" interval approach based on Goldsmith-Pinkham et al and Kolesar et al is adopted (Goldsmith-Pinkham et al., 2021; Kolesár & Rothe, 2018). The DRD design gets around the lack of precision in the running variable by omitting values at the threshold (individuals with an index date sometime in the year they turned 65 in this case) (Barreca et al., 2011).

Bias-adjusted 95% confidence intervals are estimated via "honest" methods, which have guaranteed coverage properties and account for the bias caused by the additional extrapolation required with a discrete running variable and lack of observations at the threshold (Armstrong & Kolesár, 2018; Kolesár & Rothe, 2018). The impact of different scaling parameter values used to estimate the tuning parameter from this method are explored in sensitivity analyses.

Price elasticity of demand is calculated using results from each FRD to represent the change in quantity demanded $(Q_{post} - Q_{pre})$ and price $(P_{post} - P_{pre})$ at age 65 among those who enroll in Part D due to expanded eligibility and average adherence (Q_{pre}) and OOP (P_{pre}) immediately prior to age 65 to estimate the ratio of the % percent change in quantity (90-day adherence) divided by the percent change in price (initial OOP) from the pre- to post-period (see below).

$$\frac{(Q_{post} - Q_{pre})/Q_{pre}}{(P_{post} - P_{pre})/Pre} = \frac{\%\Delta\ adherence}{\%\Delta\ price}$$

We do not consider substitution effects to be necessary since the study period occurs prior to generic Gleevec (imatinib) introduction when only costlier, imperfect substitutes (second-generation TKIs) were also available as first-line treatment.

Bootstrap resampling was used to construct standard errors and confidence intervals (Choquet et al., 1999).

2.3 RESULTS

2.3.1 *Change in Part D enrollment, OOP, and adherence at age 65*

In Figure 1, we estimate the effect of age 65 on Part D enrollment (scaling factor), OOP (reduced form RD), and adherence (reduced form RD). These plots represent sharp donut regression discontinuity designs where the exposure is age 65 (i.e., when universal Medicare Part D eligibility takes effect). The results presented are from local linear or polynomial regressions using a 10-year bandwidth on either side of the threshold.

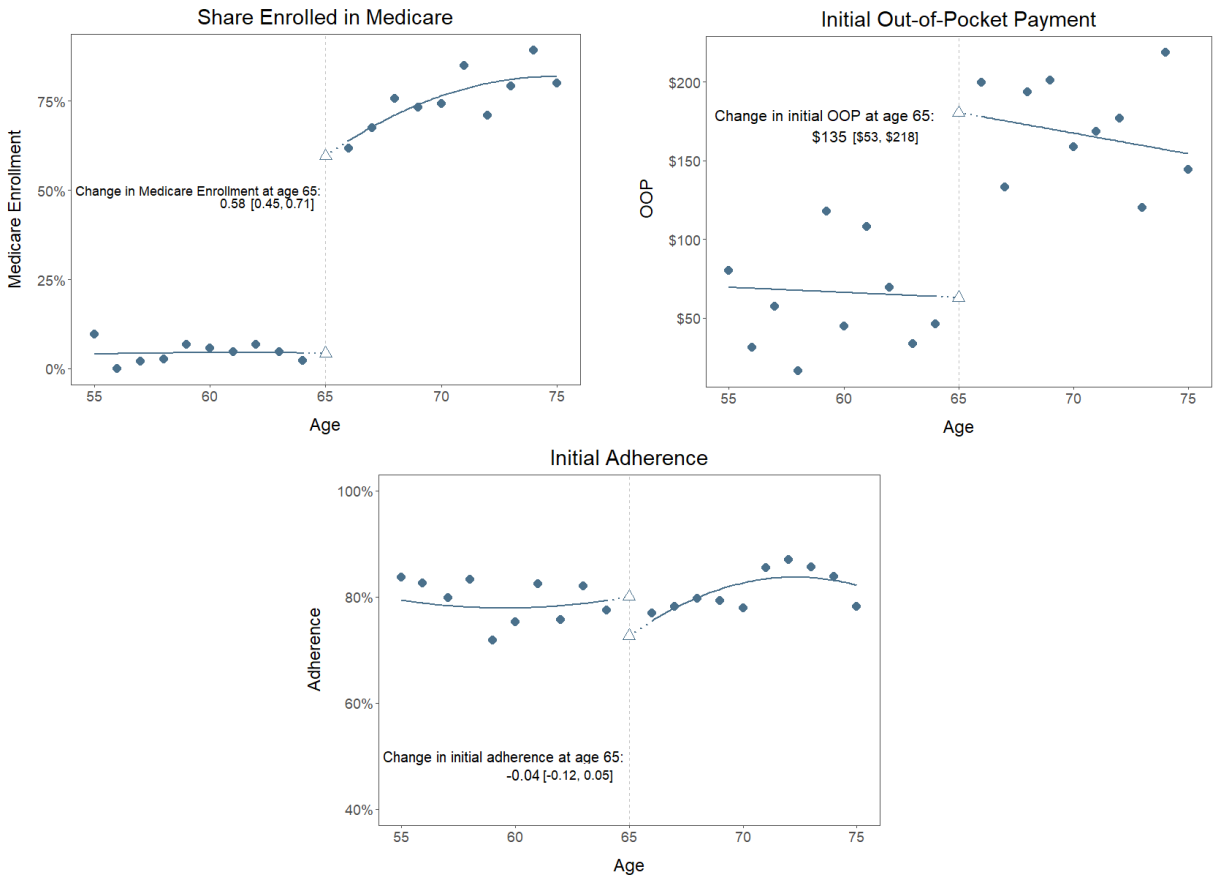

Figure 1. Effect of age 65 on Part D enrollment, OOP, and adherence

Each circle represents the average outcome by age at index date from the data. The hollow triangles at age 65 are estimates of values at age 65 from local regressions as the threshold is approached from above and below. The lines are fitted values from a local linear regression for OOP and local polynomial regressions (quadratic) for Part D coverage and adherence with a triangular kernel and a bandwidth of 10. The dotted lines represent the additional extrapolation required for a donut RD due to dropping observations at the threshold (age 65). Adherence is based on proportion of days covered. Each plot reports the sharp RD effect with bias-adjusted 95% confidence intervals.

We find that there is a marked and significant increase in Part D enrollment at age 65 of 58 percentage points (95% CI: 0.45 to 0.71). Initial OOP at this age also jumps significantly by $135 (95% CI: $53 - $218). While initial adherence does change as well, the reduction of 4 percentage points (95% CI: -0.12 to 0.05) is small and non-significant.

2.3.2 *Impact of Part D enrollment on OOP and adherence at age 65*

To better understand the effect of Medicare Part D enrollment on patients, the previously discussed sharp RD results were used to form the numerator (change in adherence and OOP) and denominator (change in Part D enrollment) of the Wald estimator. Results from this fuzzy regression discontinuity design looking only at individuals that enrolled in Part D due to expanded eligibility at age 65 found a significant increase in average OOP of $232 (95% CI: $102 to $362). Similar to the sharp RD results, the corresponding decrease in average adherence was small, at only 6 percentage points (95% CI: -0.21 to 0.08), and non-significant.

2.3.3 *Price elasticity of demand*

To get an idea of how responsive newly diagnosed CML patients are to the cost for an initial Gleevec prescription, we combined results from our separate RDs and pre-65 adherence and OOP to estimate the price elasticity of demand (percent change in adherence per 1% change in OOP). We are using OOP here rather than list price because our research question of interest is centered on how patients respond to the price that they actually face when filling their first prescription (i.e., OOP) and how responsiveness varies across different benefit structures. The list price for Gleevec may or may not be changing, but this shouldn't affect our results given that index year is balanced around the threshold.

For our main results, PED in those individuals that enrolled in Medicare Part D due to expanded eligibility at age 65 were found to be extremely inelastic with respect to price, with a PED of -0.02. These results were non-significant and fairly imprecise (95% CI: -0.056, 0.015), most likely due to our smaller sample size and the uncertainty in our adherence estimate (Table 2).

Table 2. Primary results and sensitivity to model specifications

Model	Main Analysis (n = 1,416)	Global/parametric			Local		Honest CI Scaling Factor	
		Linear	Quadratic	Cubic	Quadratic	Cubic	2	6
First Stage	0.582*** (0.067)	0.650*** (0.042)	0.555*** (0.066)	0.523*** (0.097)	0.582*** (0.067)	0.505** (0.199)	0.582*** (0.054)	0.582*** (0.078)
Fuzzy – OOP	$232*** ($66)	$141.53*** ($48.17)	$288.152*** ($76.999)	$247.370** ($105.945)	$315.785** ($123.493)	$307.010 ($257.940)	$232*** ($48)	$232** ($91)
Fuzzy – Adherence	-0.063 (0.075)	-0.030 (0.042)	-0.045 (0.074)	-0.068 (0.110)	-0.063 (0.075)	-0.015 (0.254)	-0.063 (0.059)	-0.063 (0.094)
PED	-0.020 (-0.056, 0.015)	-0.016 (-0.038, 0.021)	-0.012 (-0.029, 0.011)	-0.021 (-0.049, 0.007)	-0.009 (-0.031, 0.063)	-0.004 (-0.083, 0.264)	-0.020 (-0.042, 0.017)	-0.020 (-0.077, 0.008)

*** = significant at the 1% level; ** = significant at the 5% level. Standard errors in parentheses (bootstrap confidence intervals for PED). All estimates rounded to three decimal places.

PED estimates in our ITT population (all individuals eligible for Medicare Part D at age 65) from our sharp RD were the same as our main results (-0.02). This makes sense intuitively, as the values used in the numerator and denominator for our main results are all scaled by the same factor in the FRD.

2.3.4 *Robustness checks*

In Table 2, we also examine the sensitivity of our results to model specification by estimating discontinuities and PED using different functional forms for parametric and non-parametric models. We show that, while magnitudes are affected, overall model results are robust to polynomial degree and global vs local fit, with estimated PED ranging from -0.004 to -0.021. Only the estimate for OOP in the local cubic model moved from highly significant to non-significant. This is not surprising given polynomial fits over the 2nd degree have been demonstrated to perform poorly at the boundaries, tend to overfit the data, and are not recommended in practice (Gelman & Imbens, 2019).

The main assumption that we make when calculating Honest CIs relates to the value that we assign to our "tuning parameter," which limits the extent to which the function that we're using to extrapolate from our boundary point at age 64 and 66 to age 65 can change (Armstrong

& Kolesár, 2018; Kolesár & Rothe, 2018). We follow the approach described by Kolesár et al and implemented by Goldsmith-Pinkham et al in estimating this parameter where, because we want to account for how wrong we could potentially get our extrapolation, we use the coefficient from a quadratic global regression model and multiply it by a pre-determined scaling factor (four in this case) (Goldsmith-Pinkham et al., 2021; Kolesár & Rothe, 2018). The results of varying this scaling factor for both the adherence and OOP analyses are presented in Table 1 and indicate that our estimates are robust with respect to choice of scaling factor; although, the change in OOP was no longer significant at the 0.01 level (but still significant at the 0.05 level).

In Figure 2, we test the core identifying assumption of the regression discontinuity design – that there is "smoothness" in other observed characteristics at age 65. If this assumption is violated, it indicates that the discontinuities present in the data may be due to other factors that jump discontinuously at age 65 other than enrollment in Medicare Part D. Based on figure 2, there do not appear to be any discontinuities in other covariates at the threshold and observed effects are non-significant.

While can empirically test observed characteristics, we have to assume that there are no unobserved "shocks" or treatments that would cause either OOP or adherence to jump at the age 65 threshold. An important potential shock to consider is the impact of retirement, as this traditionally occurs at age 65 in the US. If retirement is the primary factor leading to decreased income and subsequent changes in other lifestyle factors, this could also drive any observed discontinuities in OOP and subsequently adherence. Previous work and current Bureau of Labor Statistics data, however, indicate that retirement age in the US has become more of a smooth function over time (Card, 2009; Goldsmith-Pinkham, 2021).

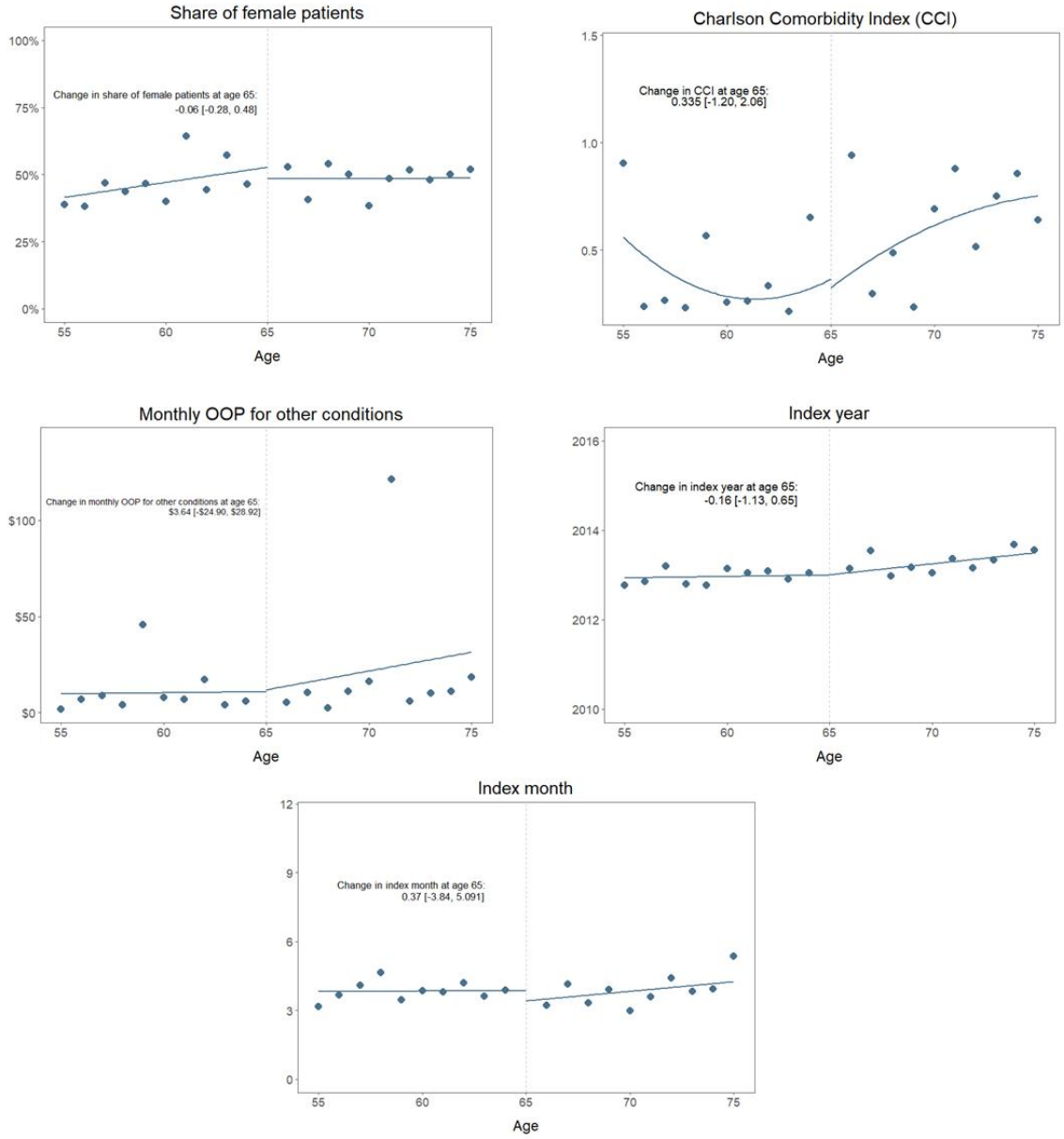

Figure 2. Regression discontinuity plots and estimates for pre-treatment variables at age 65
Each circle represents the average outcome by age at index date from the data. The lines are fitted values from local linear or polynomial regressions with triangular kernel and bandwidth of 10. Each plot reports the sharp RD effect with bias-adjusted 95% confidence intervals.

We further test the robustness of analysis results by varying the bandwidth used in the local linear regressions for OOP and adherence. Figure A1 demonstrates that OOP results were insensitive to bandwidths ranging from 5 to 15, with significance and directionality consistent across all values. Adherence results are robust with respect to significance (none of the

specifications resulted in a significant difference). While the directionality of the effect did change from slightly negative to slightly positive for bandwidths of 5 and 15, this is likely a reflection of the smaller sample available for the former leading to increased variance/extremely wide 95% CIs (increased variance) and increased bias due to less comparable treatment and comparison groups in the latter.

Placebo tests are additionally performed to assess whether the discontinuity in outcomes observed at age 65 could potentially be due to random noise in the data. Figure A2 indicates that results are robust to these placebo thresholds, with the main finding of a significant discontinuity in average OOP holding only at age 65. For adherence, the null hypothesis of no effect size could not be rejected for any of the placebo thresholds. More broadly, the largest effect sizes for both outcomes were seen at the age 65 threshold.

Figure A3 provides average outcomes and enrollment by age from age 50 to age 80 to give a better sense of overall trends in the data using global polynomial (quadratic) regressions.

2.4 CONCLUSION

For policy makers and payers to strike the optimal balance between cost sharing and financial risk protection, it's essential that robust estimates of the downstream impact of high drug prices on patients are available. This is particularly true in the Medicare Part D program, which has seen an explosion in spending in recent years due to the increase in high-priced specialty drugs entering the market (Frakt & Miller, 2018). The coinsurance-based benefit design in Part D means that patients bear a significant cost for these medications as well due a lack of any spending cap in the catastrophic coverage phase (Dusetzina, 2021; Dusetzina & Keating, 2016). In addition to the financial burden placed on patients, these high OOPs can have serious consequences in terms of non-initiation and sub-optimal adherence. This has important

implications for oral cancer therapies like Gleevec, where adherence as high 90% is required to achieve and maintain remission.

In this analysis, we implement a fuzzy donut regression discontinuity design to assess the impact of moving onto Medicare Part D at age 65 on average initial OOP and adherence for Gleevec in newly diagnosed CML patients. Combining results from each, we estimate how much OOP increases affect patient adherence. Our finding that there is a large and significant increase in both Part D enrollment (58 percentage points) and average OOP ($232) is in line to what has been found in similar studies (Dusetzina et al., 2014, 2022; Parasrampuria et al., 2020; Shen et al., 2017). Interestingly, although associational studies have identified a link between OOP and adherence, we find only a small and non-significant reduction of 6 percentage points at age 65 (Dusetzina et al., 2014; Phuar et al., 2019, 2020; Seymour et al., 2021; Winn et al., 2016). This difference is potentially due to our ability to account for the endogeneity associated with plan choice, the distortionary effect of coupon use, and our new user design. Results from both analyses were robust to placebo cutoffs and different model specifications.

The corresponding PED of only -0.02 suggests that CML patients are extremely inelastic with respect to their initial Gleevec prescription cost. This is largely consistent with the broader literature on the price elasticity of demand for specialty drugs (Doshi et al., 2015; Goldman et al., 2007, 2010; Karaca-Mandic et al., 2010). This PED value is much lower than the estimates from Jung et al for leukemia patients (-0.96 to -0.99), but this is likely due to differences in study scope. Given that Jung et al is looking at Part D patients at all ages, for all types of leukemia, and includes both prevalent cases and non-initiators, it is not unexpected that results would be different. This would be particularly true if the conditional expectation function for each

outcome given age is highly heterogenous since Jung et all include all patients over 65 (Jung et al., 2017).

While there are significant advantages to implementing an RD, external validity is limited by its inherently local nature. This means that results from this analysis are relevant only to patients who enroll in Medicare Part D at age 65. Because we are estimating OOP and PDC longitudinally, there is a chance we may capture some plan switching in response to CML diagnoses/Gleevec. The short length of follow-up (90 days) and restricted window for plan switching (October 15 to December 7 each year) should limit the impact of any anticipatory behavioral changes. It's also important to note that the substantial improvements in outcomes associated with consistent Gleevec use make it unique among specialty drugs, many of which demonstrate only incremental treatment benefits. This may limit the generalizability of this study, as it's likely that patient demand is more elastic when the effects of non-adherence are less severe.

Additionally, we need to use part A and B claims and the absence of commercial claims as a proxy for Part D enrollment because we do not have this information directly available in TriNetX. This could also potentially bias results if it leads to significant misclassification of patients by enrollment status. We found that our estimates of Part D enrollment are consistent with government sources like ASPE-HH, Kaiser Family Foundation, and a similar analysis by Card et al (KFF, 2016.; KFF 2022; Card et al., 2009; Tarazi et al., 2022). Given that our fuzzy donut RD results are scaled versions of our sharp donut RD estimates, the latter could be considered to be a lower bound on the impact of enrollment since it includes individuals that retain commercial insurance as well as those that enroll in Part D (under the assumption that patients that retain commercial coverage after age 65 do so because it's more generous).

Our results suggesting that high OOP is not a determinant of nonadherence in the approximately 30% of patients that were observed to exhibit this behavior indicate that there may be other factors driving noncompliance. The non-responsiveness of CML patients to initial Gleevec OOP suggests that high-cost sharing plans, particularly those implementing price-based coinsurance, may be putting patients at unnecessary financial risk. Results highlight the importance of future work in focusing on other potential causes of nonadherence in the older CML population on Medicare Part D and assessing the impact of high-cost-sharing benefit designs on patient costs and adherence for other specialty drugs.

Chapter 3. TEMPORAL EFFECT OF GLEEVEC ADHERENCE ON TIME TO REMISSION IN CHRONIC MYELOID LEUKEMIA PATIENTS

3.1 INTRODUCTION

Our understanding of the impact of adherence on outcomes in patients with CML is largely based on evidence from observational studies. Seminal work in this area indicates that suboptimal Gleevec adherence significantly affects a patient's ability to reach important treatment milestones. In a long-term study, Gleevec adherence of less than 90% was strongly associated with a failure to achieve a molecular response by the end of the six-year follow-up period (Marin et al., 2010). Results from a follow-up analysis suggest that adherence rates less than 85% were the primary driver in the loss of complete cytogenic response (CCyR) (Ibrahim et al., 2011). The ADAGIO study found that achievement of CCyR within 12 months of initiation was significantly associated with higher adherence (Noens et al., 2009). A recent analysis by Ganesan et al indicates that nonadherent patients were significantly less likely to achieve CCyR at any point in the three-year follow-up period, reinforcing previous findings (Ganesan et al., 2011). Due to the rarity of CML, the small size of these studies prohibited the analysis of these data in detail and at more than a descriptive level in most cases. This means that our understanding of the dose-response curve for Gleevec with respect to the timing and ability to achieve remission, is limited.

One key characteristic of these previous studies is that they incorporated medication use as a single, static measure rather than allowing its level and effect to vary over time. These factors are important to consider, as evidence indicates that Gleevec adherence trajectories fluctuate during the course CML treatment (Clark et al., 2021; Yeager et al., 2020). Relying

instead on a single measure of adherence can obscure important changes in usage that affect a patient's prognosis. For example, deceleration, acceleration, and intermittent adherence may all impact outcomes differently but result in the same static adherence value. Additionally, it's unlikely that the effect of adherence at any given time remains constant, as Gleevec is not curative and continual use is needed to maintain treatment benefit (Ganesan et al., 2011; Ibrahim et al., 2011; Jabbour & Kantarjian, 2020).

There is little evidence regarding the shape of the full dose-response curve (DRC) for Gleevec, as analyses typically use a pre-determined threshold of 80-90% to dichotomize adherence (Ward et al., 2015; Wu et al., 2010; Yood et al., 2012). This use of what is essentially a step function is unlikely to provide a clinically accurate approximation of the true relationship, and the limitations to this approach are well documented. Drawbacks include a significant loss of information, reduced statistical power in normally distributed predictors, and increased likelihood of type I (false positive) or type II errors (false negative) (Altman et al., 1994; Altman & Royston, 2006; Tueller et al., 2016).

Although it is considered the standard for analyzing time-to-event survival analyses, the conventional Cox proportional hazards model (Cox PH) is limited in its ability to properly model the relationship between cumulative Gleevec adherence and time-to-remission due to its strict assumptions regarding the proportionality of hazards, time invariance of the exposure, and linearity in the relationship between an exposure and the log-hazard ratio of the outcome (Breslow, 1978; Kalbfleisch & Prentice, 2011). Fortunately, many elegant methods relaxing some or all of these assumptions have been developed. These include multivariable fractional polynomials, restricted cubic splines, weighted cumulative dose, the product model, and the iterative ACE algorithm adapted to the context of survival analysis by Wynant et al and later

extended by Wang et al (Abrahamowicz & MacKenzie, 2007; Royston, 2001; Sauerbrei et al., 2006; Sylvestre & Abrahamowicz, 2009; Wang et al., 2020; Wynant & Abrahamowicz, 2016). Of these approaches, only the ACE algorithm provides a straightforward method for modeling the time-dependent (TD) and nonlinear (NL) effects of a time-varying continuous exposure while allowing for the flexible modeling of covariates as well.

Given the complex nature of adherence and clinical characteristics of CML, it's unclear whether the proportional hazards, linearity, and time-invariance restrictions imposed by the conventional Cox PH model hold in this setting. The objective of this analysis is to simultaneously and flexibly relax these restrictions through a novel application of the ACE algorithm. The use of a backward elimination procedure combined with visual assessment allows us to identify which model covariates require flexible modeling, ensuring their proper incorporation into the final analysis. Results provide insight into the shape of the DRC for Gleevec and illustrate how the strength of this effect varies over time. The improvements to model fit and substantial divergence from conventional Cox PH assumptions observed for Gleevec adherence and other covariates in this study highlight the importance of properly accounting for these effects.

3.2 METHODS

3.2.1 *Data*

We use data from the TriNetX Diamond Network (TDN) database for the period from first availability in 2011 through the end of 2020 (TriNetX, 2022). TDN comprises a convenience sample of claims and encounter data from 1.8 million providers representing 99 percent of U.S. health plans, with longitudinal data covering over 200 million patients. Data are pulled from open claims clearinghouses and include de-identified medical and pharmacy claims

for Commercial, Disability/Workers Compensation, Dual Coverage, Medicare, Medicaid, and VA/Other plans.

3.2.2 *Study design*

The study period consists of the time from the initial Gleevec prescription fill, defined as the index date, to the first of the following terminal events: remission, a censoring event (death, dropout/loss to follow-up, treatment switching, or treatment discontinuation (≥ 90 days without a claim for a prescription fill)), or the end of 12 months of follow-up (Figure 3). We used the six months preceding the index date as a washout period for confirming incident diagnoses and the baseline period for covariate measurement. The length of follow-up is based on the definition of suboptimal response set by European LeukemiaNet (Hochhaus et al., 2020).

A new user design is adopted to account for survivor bias (non-adherent patients are more likely to die or progress to a more advanced phase of CML) and healthy user bias (non-adherence typically leads to treatment switching due to resistance or lack of response).

Individuals over age 65 are included if they had no previous indication of CML diagnosis or treatment during the washout period, two or more primary inpatient or outpatients diagnoses consistent with International Classification of Diseases 9th and 10th revision (ICD-9 and ICD-10) diagnosis codes for "Chronic myeloid leukemia, BCR/ABL-positive, not having achieved remission" (205.10 and C92.10, respectively), and at least one Gleevec, or its generic equivalent - imatinib, order within 12 months of confirmatory diagnosis. Patients with inpatient or outpatient claims listing a diagnosis code consistent with any additional indications for Gleevec (e.g., GIST) or evidence of initiating another CML treatment at baseline are excluded (Trivedi et al., 2014). The final sample following the application of inclusion/exclusion criteria contained 613 individuals.

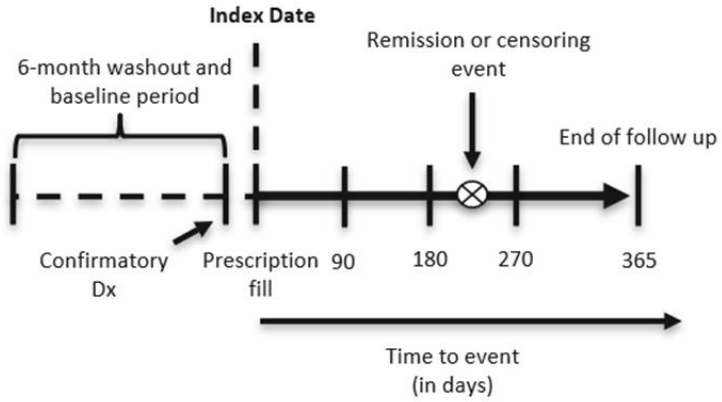

Figure 3. Study design

3.2.3 *Exposure and outcome*

The exposure in this analysis is adherence to Gleevec during the follow-up period. Because patient adherence is subject to change over time with cumulative effects, it is incorporated into the analysis as a time-varying covariate in the form of current cumulative proportion of days covered (current cumulative PDC) (Abrahamowicz et al., 2012; Sylvestre & Abrahamowicz, 2009). We define current cumulative PDC as the number of days supplied of Gleevec over the total days of follow-up elapsed, with the maximum value capped at 100% (Loucks et al., 2022). Current cumulative PDC is updated daily to allow for the most accurate measure of adherence for all patients remaining in the risk set immediately prior to a given event time.

Our outcome is the achievement of remission. Although patient response is typically assessed via laboratory testing, in the absence of this data, we relied on the first instance of an ICD-9 or ICD-10 code for "Chronic myeloid leukemia, in remission" (205.11 and C92.11, respectively). Our use of diagnostic codes presents a challenge in that there are multiple clinical benchmarks used to assess remission in CML patients (complete hematologic response, complete cytogenic response, and major molecular response). Additionally, there is likely intra-physician

heterogeneity regarding which definition is used when assigning patient status in claims data and the timing and accuracy of these codes (Horsky et al., 2018). Due to our inability to more precisely define which level of response is associated with a specific claim, remission here represents a composite measure of complete hematologic, cytogenic, or major molecular response.

3.2.4 *Covariates*

The complexity of our model and potential for overfitting require the careful selection of only those covariates with a strong clinical and empirical basis for acting as confounders. CML is unique in that, aside from adherence, patient prognosis at baseline is affected by a limited number of factors (Deininger et al., 2020). Based on National Comprehensive Cancer Care Network (NCCN) guidelines and clinical evidence, the following covariates will be measured at baseline or in the pre-index period: age, sex, diagnosis year, starting Gleevec dose, Adjusted Deyo-Charlson comorbidity index score (excluding CML), and Darkow cancer complexity index (DCCI). CCI represents a composite score of overall health status and is calculated using the presence of ICD codes for up to 17 comorbid conditions, excluding CML (Charlson et al., 1987; Quan et al., 2011). DCCI is a diagnosis-code-based score indicating the clinical complexity associated with managing a given CML patient's condition that is frequently used as a proxy measure for CML phase/staging information (Darkow et al., 2007). We defined starting Gleevec dose as "Normal" (≤400mg) or "High" (≥400mg) based on commonly used clinical thresholds and NCCN guidelines (Deininger et al., 2020). All covariates are incorporated into the model as time-invariant, meaning that their value at baseline remains fixed for the duration of their follow-up.

3.2.5 *Statistical Analysis*

In assessing the relationship between cumulative adherence and time to remission, there are three important and interrelated factors to account for: the shape of the dose-response curve, the time-varying effect of current cumulative PDC, and the fact that current cumulative PDC itself will vary over the duration of the follow-up period. The traditional Cox PH is limited in its ability to accommodate these factors due to restrictive assumptions regarding the time invariance of covariate values, their effects, and, in the case of continuous variables, the strict linearity in its relationship with the log hazard ratio (Kalbfleisch & Prentice, 2011).

We relax these assumptions by adopting the iterative ACE algorithm proposed by Wynant et al and extended in subsequent work by Wang et al (Wang et al., 2020; Wynant & Abrahamowicz, 2014). This framework uses the previously developed product model from Abrahamowicz et al to allow for the simultaneous modeling of time-dependent and non-linear effects in multiple time-varying and/or continuous covariates (Abrahamowicz & MacKenzie, 2007).

3.2.6 *Product model*

Figure 4 illustrates how the conventional Cox PH model below can be extended to allow for the joint estimation of nonlinear (NL) and time-dependent (TD) effects of a given exposure using the product model.

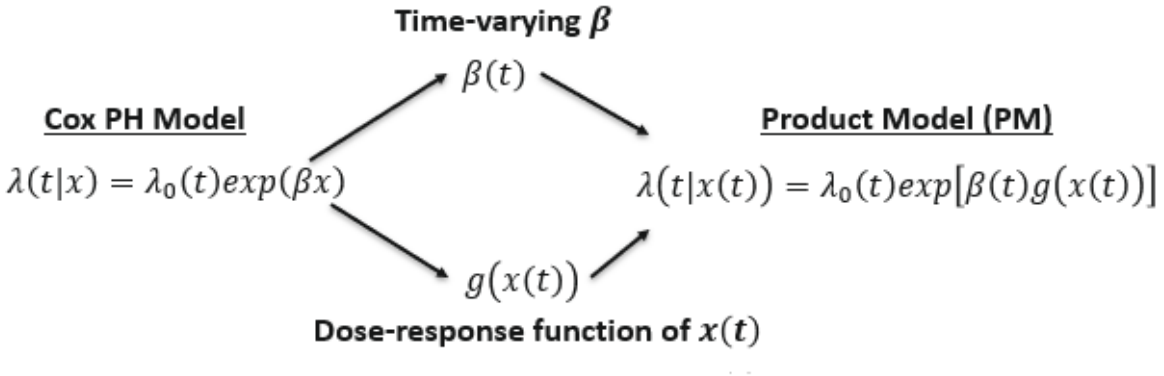

Figure 4. Product model

This approach introduces additional flexibility by modeling the hazard ratio for a continuous exposure x at time t as the product of a β that is a function of time t and g(x(t)) that is a non-linear function of the exposure. Here, $\lambda(\mathrm{t}|x(t))$ represents the hazard at time t for a given covariate x at that time, λ_0 is the baseline hazard function at time t, g(x(t)) represents the shape of the DRC for the values of x at time t, and $\beta(t)$ can be thought of as the change in the magnitude of this effect over the duration of the follow-up period.

We use unpenalized polynomial regression B-splines, which allow for smooth effects and estimation via standard maximum likelihood methods, to model both functions (Abrahamowicz & MacKenzie, 2007). This approach offers the additional advantage of enabling comparisons of fit across nested models, which include the conventional Cox PH and marginal models for TD or NL effects alone. The general equation for a given covariate x with TD and NL effects is shown below.

$$\lambda(\mathrm{t}|X(t)) = \lambda_0(t)\exp\left\{\sum_{k=1}^{K} a_k\, A_k(t) * \sum_{l=1}^{L} b_l\, B_l(X(t))\right\}$$

where *Ak* and *Bl* are the basis B-splines and *ak* and *bl* are the coefficients for the associated splines (27).

Given the low number of events (n=260), small sample size, and potentially large number of parameters, we use quadratic splines with one internal knot in an effort to ensure model stability and minimize degrees of freedom (Abrahamowicz & MacKenzie, 2007; Sauerbrei et al., 2007). Previous work indicates that this provides a good balance between the bias associated with allowing too much model flexibility and variance associated with adopting an approach that is too parsimonious (Abrahamowicz et al., 1996; Wynant & Abrahamowicz, 2014). Knot selection for both flexible functions was based on the median value of either uncensored event times (TD) or the sample distribution of cumulative adherence (NL) (Wang et al., 2020).

3.2.7 *Assessment of NL and TD effects for model covariates*

Prior to the implementation of the ACE algorithm, one must decide which variables to flexibly incorporate into the analysis. Unnecessarily including TD or NL effects for a given covariate could lead to overfitting, an important concern given the small sample size in this analysis. Alternatively, leaving out an influential covariate effect could result in biased estimates for both the covariate in question and any interrelated model covariates (Benedetti & Abrahamowicz, 2004; Wynant & Abrahamowicz, 2014). While this would ideally be informed by subject area knowledge, little evidence exists in the CML space upon which an a priori assumption can be made regarding the presence of NL and/or TD effects. The data-driven backward elimination procedure implemented alongside the ACE algorithm by Wang et al can help guide this decision in the absence of such data. This approach offers an advantage over common assessment methods that check for these effects in isolation because the presence or absence of TD and/or NL effects in multivariable analyses is likely a function of how other covariates are modeled (Abrahamowicz & MacKenzie, 2007; Binquet et al., 2008; Wynant & Abrahamowicz, 2014). Use of backward elimination has additionally been demonstrated to

perform well in this context and in simulation studies (Mantel, 1970; Sauerbrei et al., 2007; Wynant & Abrahamowicz, 2014).

The backward elimination procedure starts from a model incorporating all covariates as TD and NL (for those that are continuous) and iteratively compares the impact of removing a single covariate effect using the difference in partial log likelihood from these separately run models. Subsequent retention or removal of a given covariate effect was based on the extent to which its inclusion improved model fit to the data using relative partial deviance and Akaike's Information Criterion (AIC) value. Because all model covariates besides current cumulative PDC are fixed at baseline, we also visually assessed the extent of any deviations from linearity or PH using Martingale or Schoenfeld residual, respectively, plots.

3.2.8 *ACE algorithm*

A major challenge associated with implementation of the product model is that, because the TD and NL functions are directly multiplied by each other, we cannot simultaneously estimate each effect. Adoption of the ACE algorithm allows us to avoid this identification issue by breaking the product model estimation process down into two iteratively repeated steps:

1. NL effect estimated conditional on time-invariant effects (PH) in the first iteration and subsequently estimated based on the TD effects from the second stage of the previous iteration
2. TD effects estimated conditional on NL effects from the first stage acting as fixed constants

Algorithm steps continue until convergence is reached based on a threshold of a 0.0001 or less reduction in the partial log-likelihood from one iteration to the next (Wang et al., 2020). This approach has been validated in multiple contexts, including psychometric applications and

the flexible modeling of survival data (Abrahamowicz & Ramsay, 1992; Wang et al., 2020; Wynant & Abrahamowicz, 2016).

To properly account for the covariance of the spline coefficients for both effects and assess model stability, we drew 300 bootstrapped samples and ran the ACE algorithm independently on each (Wang et al., 2020). The standard deviation of the log-hazard ratio at specific follow-up times or current cumulative PDC levels for these samples was used as a proxy for standard errors to construct 95% pointwise confidence intervals based on normal-theory methods (Austin et al., 2022).

3.3 RESULTS

3.3.1 *Patient Characteristics*

Descriptive statistics for the 613 patients in the final sample criteria are presented in Table 2.

Table 3. Patient demographics and baseline clinical characteristics

Characteristic	**Overall Sample (n = 613)**
Current cumulative PDC at terminal event, mean (SD)	0.64 (0.33)
Age, mean (SD)	73.0 (5.6)
Female patients, n(%)	301 (49.1%)
Darkow score, n(%)	
Usual	466 (76.1%)
Moderate/High	147 (23.9%)
Starting dose, n(%)	
Normal (≤400mg)	558 (90.7%)
High (≥400mg)	55 (9.3%)
Charlson comorbidity index, mean (SD)	2.13 (2.8)
Year of diagnosis, mean (SD)	2014 (1.8)

The mean age at TKI index date was 73 (standard deviation (SD): 5.6), with women accounting for approximately half (49.1%) of the sample. On average, patients had a CCI of 2.13 (SD: 2.8) and the majority (90.7%) started on the standard Gleevec dose of 400mg per day. Mean diagnosis year was 2014 (SD: 1.8) and 76.1% of patients had usual cancer complexity. Mean cumulative adherence (PDC) at patients' terminal event was 64.2% (SD: 32.9%). Terminal events for patients broke down as follows, 12.1% (n=74) discontinued, 2.2% (n=14) had a death recorded, 9.3% (n=57) switched to another TKI, 17.1% (n=105) reached the end of follow up, 42.7% (n=262) achieved remission, and 16.5% (n=101) were lost to follow up.

The Sankey plot in Figure 1 depicts temporal changes in patient adherence from the second month of follow up through their terminal event time. By breaking the sample into groups based on cumulative PDC, we can see that broader patterns of adherence were fairly stable over time. The majority of patients exhibited "High" adherence, based on the standard definition of 80% PDC or greater, and remained in that category until their terminal event (de Almeida 2014). Only 19% of these patients moved from the "High" category to the "Moderate" adherence category and none transitioned to "Low" adherence. For those patients that were initially moderately adherent, 13% became highly adherent by their terminal time and 3% moved to the "Low" adherence category. No one that started in the "Low" adherence category became highly adherent, and only 14% increased their adherence to moderate levels. It's important to note that, because we're only looking at two points in time using broad adherence groups, there is likely a greater degree of variation within and between adherence categories than is captured by this diagram.

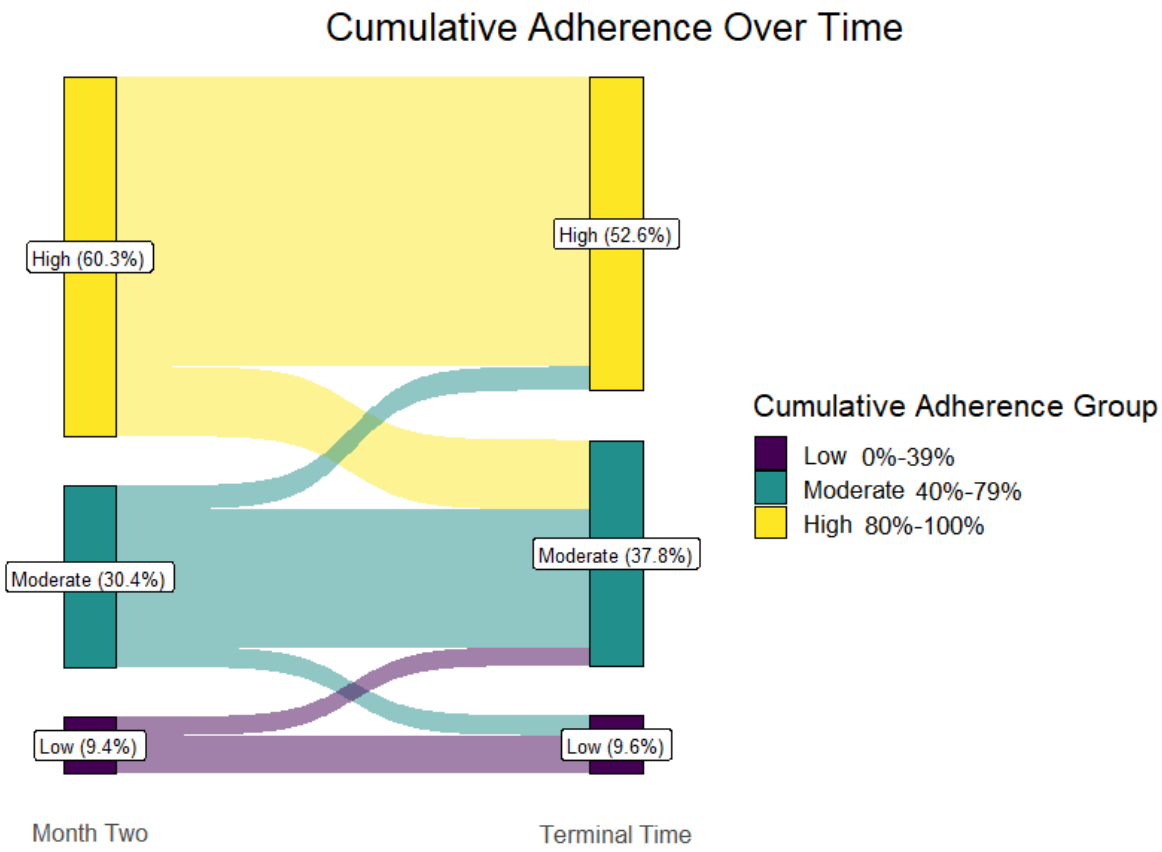

Figure 5. Sankey diagram of changes in current cumulative adherence over time

The best fitting model selected by the backward elimination procedure included NL effects of cumulative adherence and TD effects for current cumulative PDC, starting dose, and CCI. All other covariates (including those for age and DCCI) were modeled as linear with respect to the log-hazard ratio for remission and constant over time. Incorporating this model into the ACE algorithm, convergence was reached after fifteen iterations. For those covariates not flexibly modeled, the appropriateness of linearity and PH assumptions were additionally checked via Shoenfeld and Martingale residual plots.

Comparing selected model fits, the conventional Cox PH demonstrated the highest AIC (2,733), while the previously described final model provided the best fit to the data with an AIC of 2,660 (Table 3). A fully flexible model relaxing linearity and PH assumptions for all relevant

covariates possessed a lower partial deviance than our final model; however, the overall AIC was much higher due to the large number of additional parameters required and minimal improvement in model fit.

Table 4. Measures of model fit for selection

Model	Degrees of Freedom	Partial Deviance*	AIC**
Cox PH	5	2,723	2,733
NL effects only	11	2,700	2,722
TD effects only	20	2,648	2,688
Final flexible model***	17	2,626	2,660
Full flexible model	26	2,621	2,673

*Partial Deviance = -2 * partial log-likelihood
**AIC = 2k + partial deviance; where k = number of model parameters
***Fitted using backward elimination procedure

3.3.2 *Final Model*

Use of the product model means that the log-hazard ratio will always be a function of both the TD and NL effects (i.e., the individual contribution of each cannot be separately estimated). Therefore, results for TD and NL effects are presented at specific values of current cumulative PDC or time, respectively, holding all other covariate values constant.

Figure 6 illustrates how the effect of updated current cumulative PDC values (or the DRC presented in Figure 7) changes over the course of the follow-up period using a reference value of time 0. Uncertainty in point estimates is represented via 95% pointwise bootstrap-based confidence intervals at 90 days, 240 days, and 360 days of follow up.

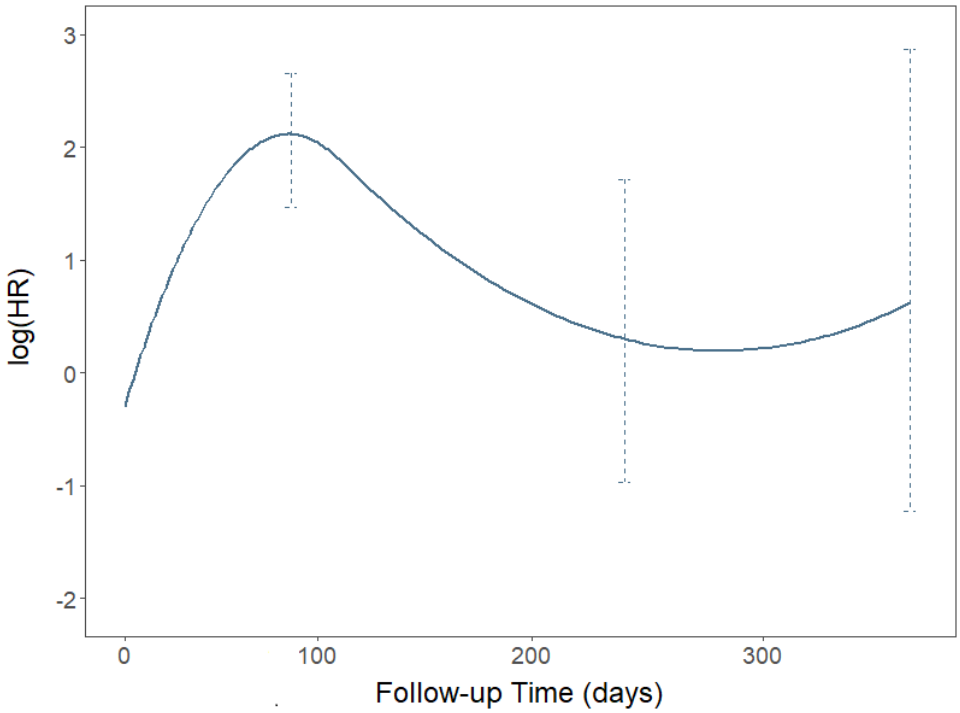

Figure 6. Time-varying strength of the effect of current cumulative PDC on time-to-remission
95% pointwise bootstrap-based confidence intervals at select times (90 days, 240 days, and 360 days of follow-up, respectively) are represented by dashed vertical bars.

The marked increase in the log-hazard ratio from the beginning of follow-up through months three to four suggests a lagged and/or cumulative dose effect, where Gleevec use most strongly affects remission status once a given level of adherence is maintained for a certain duration of time. The overall shape of the curve is nonmonotone, reaching a peak around 90 days of follow-up (log-hazard ratio: 2.12, 95% CI: 1.47 to 2.66). The curve then decreases from its initial peak and subsequently plateaus after approximately 240 days (log-hazard ratio: 0.30, 95% CI: -0.97 to 1.61). This is likely driven by two factors. First, the majority of remissions (69%) occurred early in the follow-up period and in individuals with relatively high adherence (93% exhibited cumulative adherence over 80%). This is consistent with the clinical literature regarding the effectiveness of Gleevec and studies indicating that hematological response usually occurs within a few weeks and complete cytogenic response within a few months in adherent patients (de Almeida 2014; Deininger 2020; Jabbor 2020). Second, our cumulative adherence measure does

not adequately account for recency and may not fully reflect important changes in adherence at later points in follow-up. Essentially, because most of the remissions occurred relatively early and in highly adherent patients, this effect compared to non-adherent patients peaked at around three to four months. Remissions that occurred later in the follow-up period are likely are result of increased adherence that won't be captured as well if current cumulative PDC is effectively "diluted" by poor initial adherence. This may be the case, as only 29% of individuals with remission after 240 days had current cumulative PDC over 80% at the time of remission and only 18% had current cumulative PDC over 80% at month four. The small increase at the end of follow-up is likely an artifact of the functional form chosen for our splines (quadratic) and poor boundary performance for these polynomial splines due to the small sample remaining (Austin et al., 2022). High variability in estimates at increasing follow-up times is also a function of the small risk set remaining (n=135 from day 240 onward).

Figure 7 shows how the dose-response curve (constructed using a reference value of 0% current cumulative PDC) for Gleevec changes over the course of the follow-up period. Unsurprisingly, the relationship is monotonically increasing, with a slight attenuation in the rate of increase in the log-hazard ratio at higher levels of cumulative adherence. Consistent with the shape observed in Figure Six, the effect is strongest at three months, with a log-hazard ratio for individuals exhibiting 100% current cumulative PDC of 4.16 (95% CI 3.81 to 4.51) and wanes substantially thereafter. In fact, there appears to be only a small effect at 240 days, with the increase at 360 days (0.98 (95% CI -0.33 to 1.31)) likely a function of other, previously mentioned factors. Like Figure 4, the wide confidence intervals at 240 days and beyond means that we're less able to reliably assess this relationship.

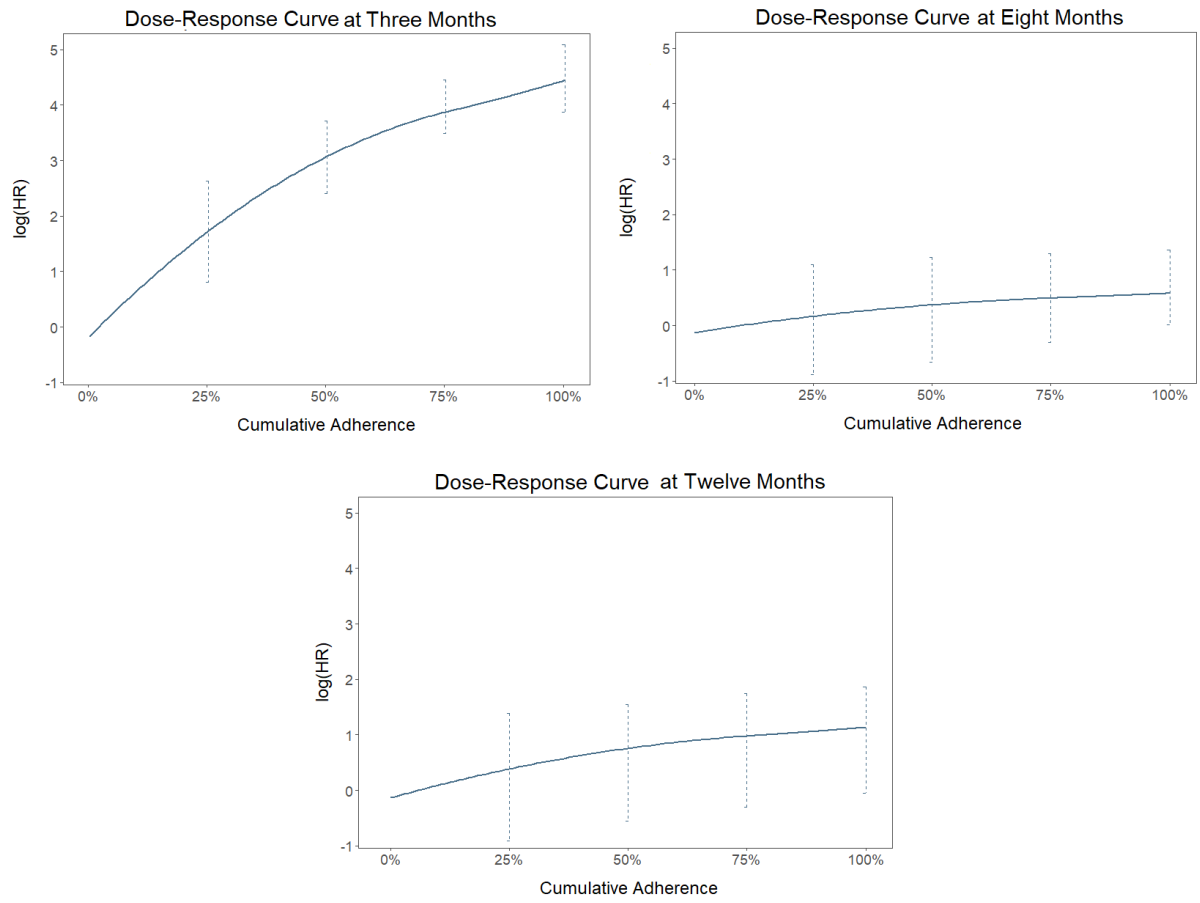

Figure 7. Estimated dose-response curve for current cumulative PDC at selected times
Log-hazard ratios are relative to a reference level of 0%. 95% pointwise bootstrap-based confidence intervals at selected cumulative adherence levels (25%, 50%, 75%, 100%) are represented by dashed vertical bars.

3.3.3 *Discussion*

In this study, we present a novel application of the previously validated ACE algorithm in the context of newly diagnosed CML patients initiating Gleevec. By combining this algorithm with a backward elimination procedure, we identified time dependence and nonlinearity in relevant covariates and fit an extended Cox model allowing for the flexible modeling of each effect via unpenalized B-splines.

Results highlight the importance of properly accounting for the complex, interrelated nature of these effects. The traditional Cox PH generates a single, constant hazard ratio over the

duration of follow-up (Breslow, 1978). If the true effect is time-varying, the resulting hazard ratio will be non-representative and potentially biased (O'Quigley & Pessione, 1991). Figure 6 suggests that there is temporal variability in this measure, with the strength of the effect peaking at three months (90 days) post-initiation and steadily decreasing thereafter. The fact that most patients that achieved remission did so by month four (120 days) with consistently high current cumulative PDC suggests that this could be a critical time and duration for realizing treatment benefit. Unfortunately, the small remaining risk set beyond month eight (240 days) makes drawing concrete conclusions about this relationship later in the follow-up period difficult. This underscores both the challenge of working with a rare condition like CML and the potential for future analyses with access to larger and more robust datasets to build upon these results.

Figure 7 indicates a subtler departure from the Cox PH linearity assumption, with the increase in the log-hazard ratio attenuating at higher levels of current cumulative PDC. The importance of time is apparent when comparing the DRCs at different points in the follow-up period. For example, a low current cumulative PDC of 25% at month three is associated with a large log-hazard ratio (1.67) relative to 0% current cumulative PDC. By month eight, even the effect of 100% current cumulative PDC is barely differentiable from a log-hazard ratio of 0.

The DRC at month three additionally demonstrates the importance of modeling adherence as a continuous measure. Studies in this space typically treat adherence as a dichotomous, piecewise function, keeping the log-hazard ratio for remission uniform above and below a pre-defined threshold (80% - 90% for CML depending on the outcome). The loss of information that this approach entails by assigning the same log-hazard ratio to all adherence levels below 80%, for example, is highlighted by the substantial difference in log-hazard ratio at 1% and 79% current cumulative PDC observed in our study (0.23 vs 3.89, respectively).

There are several limitations to the approach adopted in this study. The dynamic nature of adherence means that its relationship to patient outcomes is often confounded by time-varying factors that may additionally lie in the causal pathway. Conventional survival analysis methods alone, including the extended Cox model that we use, are not able to account for the bias caused by these factors (Robins, 1999). We considered this to be less of a concern in the context of CML due to the limited number of potential time-varying confounders (dose escalation and treatment response) that would be expected to significantly affect our relationship of interest. For the former, minimal dose escalation (n=4) was observed in the study. For the latter, because the majority of CML present in the largely asymptomatic chronic phase (90%), it's unlikely that there is reverse causality between treatment response/how a patient feels and adherence.

Similarly, we are unable to account for the impact of unobserved confounders, which could lead to biased results if they are driving the relationship between adherence and outcomes in our study. While our data did not contain information on all prognostic factors for CML, the limited number of these factors and fact that Gleevec adherence is unlikely to be affected by most of them indicates that any resulting omitted variable bias may be small (Deininger et al., 2020). Additionally, the use of an internal, or endogenous, time-varying exposure like adherence that represents a measure taken directly on a patient precludes estimation of the survival function if its observation is affected by the failure mechanism under study (Kalbfleisch & Prentice, 2011). While an issue in the study of survival, this was deemed less impactful in the present analysis since our outcome is remission and Gleevec is not curative (i.e., treatment does not stop after remission). It's important to note that the ability of some patients to achieve a "deep molecular response" and stop treatment is less well understood and only occurs after many years of successful treatment, making it irrelevant during our study period. Thus, the presence or

absence of Gleevec claims at any given time does not provide perfect information about the actual time to or achievement of remission.

On the data side, the primary limitation was the lack of clinical laboratory results available for assessing patient response. This necessitated the use of a broader, composite measure for assigning remission status based on ICD codes, which may combine different levels of treatment response (hematological, cytogenic, and molecular). There is consequently likely measurement error in our outcome due to both inaccurate timing and, potentially, outcome status if these codes are not accurately assigned in patient claims. The consistency between our static, average current cumulative PDC at terminal event and what has been observed in other CML studies (~70%) does indicate that measurement error may be less of an issue on the exposure side (an important assumption in Cox modeling) (Ibrahim et al., 2011; Jabbour & Kantarjian, 2020; Marin et al., 2010; Noens et al., 2014). Our analysis was also likely not sufficiently powered to detect a meaningful difference in log-hazard ratios further out in the follow-up period (eight months and beyond) due to the small proportion of the sample (22%) remaining in the risk set at this time. Another important assumption in Cox modeling is that of non-informative censoring (i.e., censored patients do not systematically differ in outcomes than uncensored patients) (Kalbfleisch & Prentice, 2011). Although this assumption is not testable, the fact that dropout or discontinuation accounted for the majority of censoring (excluding right censoring due to reaching the end of the follow-up period) observed in the study (71%) suggests that systematic differences may be less likely. This is due to the fact that we're using open claims data, so the reason for dropout or no longer seeing pharmaceutical claims cannot be attributed to a single specific cause (e.g., loss of insurance) and likely reflects patients moving or switching to another doctor/pharmacy that did not contract with a TriNetX-covered clearinghouse.

This study presents the unique application of flexible survival analysis methods to determine the impact of current cumulative Gleevec adherence on time-to-remission in newly diagnosed CML patients over age 65. Implementation of the backward elimination procedure alongside the ACE algorithm led to substantial improvements in model fit over the conventional Cox PH model. Our findings suggest that the a priori assumptions regarding the proportional hazards, linearity, and time-invariance of our continuous exposure imposed by this model do not hold in this setting. Results provide insight into the shape of the DRC for Gleevec and demonstrate how the strength of this curve varies over the first year following treatment initiation. From a research perspective, this can help guide the design of future studies regarding important analytic decisions and assumptions. From a clinical perspective, these results can inform monitoring and intervention strategies for meeting treatment milestones. Future work can build upon this analysis using larger, more robust datasets with access to laboratory results for assigning remission status.

Chapter 4. SUMMARY

Gleevec's high initial list price, early introduction relative to other specialty drugs, and the importance of adherence in realizing treatment outcomes have made it an area of considerable interest regarding the downstream impact of drug pricing in co-insurance-based systems. This dissertation attempts to add to the evidence base in this area through the novel application of advanced statistical methods able to overcome previously unaddressed analytic challenges.

Our finding that newly-diagnosed CML patients on Medicare Part D are highly unresponsive to their initial Gleevec prescription costs suggests that high-cost sharing plans may be putting vulnerable patients at unnecessary financial risk. This concern is particularly consequential for Medicare-age individuals given the significant levels of financial insecurity that this population faces (Dusetzina 2022). It also underscores the importance of legislation like the IRA aimed at reducing the financial burden of prescription drugs in the elderly.

While the inelasticity of CML patients to Gleevec makes sense intuitively given the seriousness of the condition, the effectiveness of Gleevec, and its relatively tolerable side effects, it runs counter to what has been observed in associational studies (Dusetzina et al., 2014; Phuar et al., 2019, 2020; Seymour et al., 2021; Winn et al., 2016). This reinforces the importance of adopting a robust identification strategy capable of accounting for the endogeneity associated with plan choice and distortionary effect of coupon use (Doshi et al., 2015; Jung et al., 2017; Phuar et al., 2019, 2020; Seymour et al., 2021). The non-responsiveness of patients to OOP also prompts the question of what, if not high OOP, is driving the substantial levels of non-adherence observed in this population?

Our use of flexible survival modeling methods provides new insight into the shape and strength of the DRC for Gleevec in the first year following treatment initiation. The observed peak at three months (90 days) post-initiation, steady decrease, and eventual plateau indicates that early adherence may critical to the achievement of remission. Additionally, the non-linearity and time-dependence identified in the relationship between cumulative adherence and time-to-remission highlights the importance of flexibly modeling these previously unaccounted for effects in empirical work. From a clinical perspective, these results can inform monitoring and intervention strategies for meeting treatment milestones.

This dissertation provides robust and novel evidence regarding the relationship between list-price-based OOP, adherence, and patient outcomes in Medicare Part D patients that can inform decision making at the clinical and policy level. The methodological approach that we adopt may be adapted to other settings and can help guide the design of similar studies regarding important analytic decisions and assumptions. There is ample room for future work to build off this analysis, particularly in the area of determinants of non-adherence in CML patients and the use of larger, more robust datasets with access to laboratory results for identifying remission status.

Chapter 5. BIBLIOGRAPHY

Abboud, C., Berman, E., Cohen, A., Cortes, J., DeAngelo, D., Deininger, M., Devine, S., Druker, B., Fathi, A., Jabbour, E., Jagasia, M., Kantarjian, H., Khoury, J., Laneuville, P., Larson, R., Lipton, J., Moore, J. O., Mughal, T., O'Brien, S., … Salem, Z. (2013). The price of drugs for chronic myeloid leukemia (CML) is a reflection of the unsustainable prices of cancer drugs: From the perspective of a large group of CML experts. *Blood*, *121*(22), 4439–4442. https://doi.org/10.1182/blood-2013-03-490003

Abrahamowicz, M., Beauchamp, M.-E., & Sylvestre, M.-P. (2012). Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine*, *31*(11–12), 1014–1030. https://doi.org/10.1002/sim.4343

Abrahamowicz, M., & MacKenzie, T. A. (2007). Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in Medicine*, *26*(2), 392–408. https://doi.org/10.1002/sim.2519

Abrahamowicz, M., Mackenzie, T., & Esdaile, J. M. (1996). Time-Dependent Hazard Ratio: Modeling and Hypothesis Testing with Application in Lupus Nephritis. *Journal of the American Statistical Association*, *91*(436), 1432–1439. https://doi.org/10.1080/01621459.1996.10476711

Abrahamowicz, M., & Ramsay, J. O. (1992). Multicategorical spline model for item response theory. *Psychometrika*, *57*(1), 5–27. https://doi.org/10.1007/BF02294656

Agarwal, R., Mazurenko, O., & Menachemi, N. (2017). High-Deductible Health Plans Reduce Health Care Cost And Utilization, Including Use Of Needed Preventive Services. *Health Affairs*, *36*(10), 1762–1768. https://doi.org/10.1377/hlthaff.2017.0610

Altman, D. G., Lausen, B., Sauerbrei, W., & Schumacher, M. (1994). Dangers of Using "Optimal" Cutpoints in the Evaluation of Prognostic Factors. *JNCI Journal of the National Cancer Institute*, *86*(11), 829–835. https://doi.org/10.1093/jnci/86.11.829

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ : British Medical Journal*, *332*(7549), 1080.

Anderson-Cook, A., Maeda, J., & Nelson, L. (2014.). *Prices for and Spending on Specialty Drugs in Medicare Part D and Medicaid: An In-Depth Analysis*. 62.

Armstrong, T. B., & Kolesár, M. (2018). *Simple and Honest Confidence Intervals in Nonparametric Regression*. https://doi.org/10.2139/ssrn.3243943

Aron-Dine, A., Einav, L., & Finkelstein, A. (2013). The RAND Health Insurance Experiment, Three Decades Later. *Journal of Economic Perspectives*, *27*(1), 197–222. https://doi.org/10.1257/jep.27.1.197

Austin, P. C., Fang, J., & Lee, D. S. (2022). Using fractional polynomials and restricted cubic splines to model non-proportional hazards or time-varying covariate effects in the Cox regression model. *Statistics in Medicine*, *41*(3), 612–624. https://doi.org/10.1002/sim.9259

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving Babies? Revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, *126*(4), 2117–2123. https://doi.org/10.1093/qje/qjr042

Benedetti, A., & Abrahamowicz, M. (2004). Using generalized additive models to reduce residual confounding. *Statistics in Medicine*, *23*(24), 3781–3801. https://doi.org/10.1002/sim.2073

Binquet, C., Abrahamowicz, M., Mahboubi, A., Jooste, V., Faivre, J., Bonithon-Kopp, C., & Quantin, C. (2008). Empirical study of the dependence of the results of multivariable flexible survival analyses on model selection strategy. *Statistics in Medicine*, *27*(30), 6470–6488. https://doi.org/10.1002/sim.3447

Bower, H., Björkholm, M., Dickman, P. W., Höglund, M., Lambert, P. C., & Andersson, T. M.-L. (2016). Life Expectancy of Patients With Chronic Myeloid Leukemia Approaches the Life Expectancy of the General Population. *Journal of Clinical Oncology*, *34*(24), 2851–2857. https://doi.org/10.1200/JCO.2015.66.2866

Breslow, N. (1978). The proportional hazards model: Applications in epidemiology. *Communications in Statistics - Theory and Methods*, *7*(4), 315–332. https://doi.org/10.1080/03610927808827626

Card, D., Dobkin, C., & Maestas, N. (2009). DOES MEDICARE SAVE LIVES? *The Quarterly Journal of Economics*, *124*(2), 597–636. https://doi.org/10.1162/qjec.2009.124.2.597

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *ArXiv:1911.09511 [Econ, Stat]*. https://doi.org/10.1017/9781108684606

Chambers, J. D., Thorat, T., Pyo, J., Chenoweth, M., & Neumann, P. J. (2014). Despite High Costs, Specialty Drugs May Offer Value For Money Comparable To That Of Traditional Drugs. *Health Affairs*, *33*(10), 1751–1760. https://doi.org/10.1377/hlthaff.2014.0574

Chandra, A., Gruber, J., & McKnight, R. (2010). Patient Cost-Sharing and Hospitalization Offsets in the Elderly. *The American Economic Review*, *100*(1), 193–213. https://doi.org/10.1257/aer.100.1.193

Chandra, A., Gruber, J., & McKnight, R. (2014). The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts. *Journal of Health Economics*, *33*, 57–66. https://doi.org/10.1016/j.jhealeco.2013.10.008

Chaplin, D. D., Cook, T. D., Zurovac, J., Coopersmith, J. S., Finucane, M. M., Vollmer, L. N., & Morris, R. E. (2018). The Internal and External Validity of the Regression Discontinuity Design: A Meta-Analysis of 15 Within-Study Comparisons. *Journal of Policy Analysis and Management*, *37*(2), 403–429. https://doi.org/10.1002/pam.22051

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383. https://doi.org/10.1016/0021-9681(87)90171-8

Choquet, D., L'Ecuyer, P., & Léger, C. (1999). Bootstrap confidence intervals for ratios of expectations. *ACM Transactions on Modeling and Computer Simulation*, *9*(4), 326–348. https://doi.org/10.1145/352222.352224

Clark, S. E., Marcum, Z. A., Radich, J. P., & Bansal, A. (2021). Predictors of tyrosine kinase inhibitor adherence trajectories in patients with newly diagnosed chronic myeloid leukemia. *Journal of Oncology Pharmacy Practice*, *27*(8), 1842–1852. https://doi.org/10.1177/1078155220970616

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Darkow, T., Henk, H. J., Thomas, S. K., Feng, W., Baladi, J.-F., Goldberg, G. A., Hatfield, A., & Cortes, J. (2007). Treatment interruptions and non-adherence with imatinib and associated healthcare costs: A retrospective analysis among managed care patients with chronic myelogenous leukaemia. *PharmacoEconomics*, *25*(6), 481–496. https://doi.org/10.2165/00019053-200725060-00004

Darrow, J. J., & Kesselheim, A. S. (n.d.). Nearly One-Third Of New Drugs Are No Better Than Older Drugs, And Some Are Worse. *Health Affairs Forefront*. Retrieved November 29, 2022, from https://www.healthaffairs.org/do/10.1377/forefront.20171021.268271/full/

Deininger, M. W., Shah, N. P., Altman, J. K., Berman, E., Bhatia, R., Bhatnagar, B., DeAngelo, D. J., Gotlib, J., Hobbs, G., Maness, L., Mead, M., Metheny, L., Mohan, S., Moore, J. O., Naqvi, K., Oehler, V., Pallera, A. M., Patnaik, M., Pratz, K., … Sundar, H. (2020). Chronic Myeloid Leukemia, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network: JNCCN*, *18*(10), 1385–1415. https://doi.org/10.6004/jnccn.2020.0047

Doshi J.A., Li P., Huo H., Pettit A.R., Kumar R., Weiss B.M., Huntington S.F. (2016). High cost sharing and specialty drug initiation under Medicare Part D: a case study in patients with newly diagnosed chronic myeloid leukemia. Am J Manag Care, 22(4 Suppl):s78-86.

Doshi, J. A., Li, P., Ladage, V. P., & Taylor, E. A. (2016). Impact of cost sharing on specialty Drug Utilization and outcomes: A review of the evidence and Future directions. *Am J Manag Care. 2016 Mar;22(3):188-97.*

Dusetzina, S. B. (2021). Relief in Sight—Estimated Savings under Medicare Part D Redesign. *New England Journal of Medicine*, *385*(26), e93. https://doi.org/10.1056/NEJMp2116586

Dusetzina, S. B., Huskamp, H. A., Rothman, R. L., Pinheiro, L. C., Roberts, A. W., Shah, N. D., Walunas, T. L., Wood, W. A., Zuckerman, A. D., Zullig, L. L., & Keating, N. L. (2022). Many Medicare Beneficiaries Do Not Fill High-Price Specialty Drug Prescriptions. *Health Affairs (Millwood, Va.)*, *41*(4), 487–15. https://doi.org/10.1377/hlthaff.2021.01742

Dusetzina, S. B., & Keating, N. L. (2016). Mind the Gap: Why Closing the Doughnut Hole Is Insufficient for Increasing Medicare Beneficiary Access to Oral Chemotherapy. *Journal of Clinical Oncology*, *34*(4), 375–380. https://doi.org/10.1200/JCO.2015.63.7736

Dusetzina, S. B., Winn, A. N., Abel, G. A., Huskamp, H. A., & Keating, N. L. (2014). Cost Sharing and Adherence to Tyrosine Kinase Inhibitors for Patients With Chronic Myeloid Leukemia. *Journal of Clinical Oncology*, *32*(4), 306–311. https://doi.org/10.1200/JCO.2013.52.9123

Frakt, A., & Miller, M. (2018). The Case for Restructuring the Medicare Prescription Drug Benefit. *Health Services Research*, *53*(6), 4132–4137. https://doi.org/10.1111/1475-6773.13019

Gambacorti-Passerini, C., Chen, C., Davis, C., Sen, G. P., Guyan, C., Hehlmann, R., Michallet, M., Paquette, R., & Goldberg, S. L. (2021). Treatment patterns and clinical outcomes of tyrosine kinase inhibitors in chronic-phase CML in clinical practice: 3-year European SIMPLICITY data. *European Journal of Haematology*, *106*(1), 82–89. https://doi.org/10.1111/ejh.13524

Ganesan, P., Sagar, T. G., Dubashi, B., Rajendranath, R., Kannan, K., Cyriac, S., & Nandennavar, M. (2011). Nonadherence to Imatinib adversely affects event free survival in chronic phase chronic myeloid leukemia. *American Journal of Hematology*, *86*(6), 471–474. https://doi.org/10.1002/ajh.22019

Gatwood, J., Gibson, T. B., Chernew, M. E., Farr, A. M., Vogtmann, E., & Fendrick, A. M. (2014). Price Elasticity and Medication Use: Cost Sharing Across Multiple Clinical Conditions. *Journal of Managed Care Pharmacy*, *20*(11), 1102–1107. https://doi.org/10.18553/jmcp.2014.20.11.1102

Gelman, A., & Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

Goldman, D. P., Jena, A. B., Lakdawalla, D. N., Malin, J. L., Malkin, J. D., & Sun, E. (2010). The Value of Specialty Oncology Drugs. *Health Services Research*, *45*(1), 115–132. https://doi.org/10.1111/j.1475-6773.2009.01059.x

Goldman, D. P., Joyce, G. F., & Zheng, Y. (2007). Prescription drug cost sharing: Associations with medication and medical utilization and spending and health. *JAMA*, *298*(1), 61–69. https://doi.org/10.1001/jama.298.1.61

Goldsmith-Pinkham, P., Pinkovskiy, M., & Wallace, J. (2021). The Great Equalizer: Medicare and the Geography of Consumer Financial Strain. *ArXiv:2102.02142 [Econ, q-Fin, Stat]*. http://arxiv.org/abs/2102.02142

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, *69*(1), 201–209. https://doi.org/10.1111/1468-0262.00183

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.

Hochhaus, A., Baccarani, M., Silver, R. T., Schiffer, C., Apperley, J. F., Cervantes, F., Clark, R. E., Cortes, J. E., Deininger, M. W., Guilhot, F., Hjorth-Hansen, H., Hughes, T. P., Janssen, J. J. W. M., Kantarjian, H. M., Kim, D. W., Larson, R. A., Lipton, J. H., Mahon, F. X., Mayer, J., … Hehlmann, R. (2020). European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia*, *34*(4), Article 4. https://doi.org/10.1038/s41375-020-0776-2

Horsky, J., Drucker, E. A., & Ramelson, H. Z. (2018). Accuracy and Completeness of Clinical Coding Using ICD-10 for Ambulatory Visits. *AMIA Annual Symposium Proceedings*, *2017*, 912–920.

Ibrahim, A. R., Eliasson, L., Apperley, J. F., Milojkovic, D., Bua, M., Szydlo, R., Mahon, F.-X., Kozlowski, K., Paliompeis, C., Foroni, L., Khorashad, J. S., Bazeos, A., Molimard, M., Reid, A., Rezvani, K., Gerrard, G., Goldman, J., & Marin, D. (2011). Poor adherence is the main reason for loss of CCyR and imatinib failure for chronic myeloid leukemia patients on long-term therapy. *Blood*, *117*(14), 3733–3736. https://doi.org/10.1182/blood-2010-10-309807

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Iqbal, N., & Iqbal, N. (2014). Imatinib: A Breakthrough of Targeted Therapy in Cancer. *Chemotherapy Research and Practice*, *2014*, 357027. https://doi.org/10.1155/2014/357027

Jabbour, E., & Kantarjian, H. (2020). Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *American Journal of Hematology*, *95*(6), 691–709. https://doi.org/10.1002/ajh.25792

Jung, J. K., Feldman, R., & McBean, A. M. (2017). The Price Elasticity of Specialty Drug Use: Evidence from Cancer Patients in Medicare Part D. *Forum for Health Economics & Policy*, *20*(2), 20160007.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data* (2. Aufl.). Wiley-Interscience.

Karaca-Mandic, P., Joyce, G. F., Goldman, D. P., & Laouri, M. (2010). Cost Sharing, Family Health Care Burden, and the Use of Specialty Drugs for Rheumatoid Arthritis. *Health Services Research*, *45*(5p1), 1227–1250. https://doi.org/10.1111/j.1475-6773.2010.01117.x

KFF (2016). *70% of Medicare beneficiaries enrolled in Part D: KFF | Fierce Healthcare*. Retrieved October 10, 2022, from https://www.fiercehealthcare.com/payer/70-medicare-beneficiaries-enrolled-part-d

KFF (2022). How Will the Prescription Drug Provisions in the Inflation Reduction Act Affect Medicare Beneficiaries? *KFF*. http://www.kff.org/medicare/issue-brief/how-will-the-prescription-drug-provisions-in-the-inflation-reduction-act-affect-medicare-beneficiaries/

KFF. (2022). Key Facts About Medicare Part D Enrollment and Costs in 2022. *KFF*. http://www.kff.org/medicare/issue-brief/key-facts-about-medicare-part-d-enrollment-and-costs-in-2022/

Kirchhoff, S. M. (2015). *Specialty Drugs: Background and Policy Concerns*. 29.

Kolesár, M., & Rothe, C. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review*, *108*(8), 2277–2304. https://doi.org/10.1257/aer.20160945

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Loucks, J., Zuckerman, A. D., Berni, A., Saulles, A., Thomas, G., & Alonzo, A. (2022). Proportion of days covered as a measure of medication adherence. *American Journal of Health-System Pharmacy: AJHP: Official Journal of the American Society of Health-System Pharmacists*, *79*(6), 492–496. https://doi.org/10.1093/ajhp/zxab392

Mantel, N. (1970). Why Stepdown Procedures in Variable Selection. *Technometrics*, *12*(3), 621–625. https://doi.org/10.1080/00401706.1970.10488701

Marin, D., Bazeos, A., Mahon, F.-X., Eliasson, L., Milojkovic, D., Bua, M., Apperley, J. F., Szydlo, R., Desai, R., Kozlowski, K., Paliompeis, C., Latham, V., Foroni, L., Molimard, M., Reid, A., Rezvani, K., de Lavallade, H., Guallar, C., Goldman, J., & Khorashad, J. S. (2010). Adherence Is the Critical Factor for Achieving Molecular Responses in Patients With Chronic Myeloid Leukemia Who Achieve Complete Cytogenetic Responses on Imatinib. *Journal of Clinical Oncology*, *28*(14), 2381–2388. https://doi.org/10.1200/JCO.2009.26.3087

Noens, L., Hensen, M., Kucmin-Bemelmans, I., Lofgren, C., Gilloteau, I., & Vrijens, B. (2014). Measurement of adherence to BCR-ABL inhibitor therapy in chronic myeloid leukemia: Current situation and future challenges. *Haematologica*, *99*(3), 437–447. https://doi.org/10.3324/haematol.2012.082511

Noens, L., van Lierde, M.-A., De Bock, R., Verhoef, G., Zachée, P., Berneman, Z., Martiat, P., Mineur, P., Van Eygen, K., MacDonald, K., De Geest, S., Albrecht, T., & Abraham, I. (2009). Prevalence, determinants, and outcomes of nonadherence to imatinib therapy in patients with chronic myeloid leukemia: The ADAGIO study. *Blood*, *113*(22), 5401–5411. https://doi.org/10.1182/blood-2008-12-196543

O'Quigley, J., & Pessione, F. (1991). The Problem of a Covariate-Time Qualitative Interaction in a Survival Study. *Biometrics*, *47*(1), 101–115. https://doi.org/10.2307/2532499

Parasrampuria, S., Aditi P. Sen, P., & Anderson, G. (2020). Comparing Patient OOP Spending for Specialty Drugs in Medicare Part D and Employer-Sponsored Insurance. *The American Journal of Managed Care*, *26*(09). http://www.ajmc.com/view/comparing-patient-oop-spending-for-specialty-drugs-in-medicare-part-d-and-employer-sponsored

Phuar, H. L., Begley, C. E., Chan, W., & Krause, T. M. (2019). Tyrosine Kinase Inhibitors Initiation, Cost Sharing, and Health Care Utilization in Patients with Newly Diagnosed Chronic Myeloid Leukemia: A Retrospective Claims-Based Study. *Journal of Managed Care & Specialty Pharmacy*, *25*(10), 1140–1150. https://doi.org/10.18553/jmcp.2019.25.10.1140

Phuar, H. L., Begley, C. E., Chan, W., & Krause, T. M. (2020). Tyrosine Kinase Inhibitors and the Relationship With Adherence, Costs, and Health Care Utilization in Commercially Insured Patients With Newly Diagnosed Chronic Myeloid Leukemia: A Retrospective Claims-Based Study. *American Journal of Clinical Oncology*, *43*(7), 517–525. https://doi.org/10.1097/COC.0000000000000700

Quan, H., Li, B., Couris, C. M., Fushimi, K., Graham, P., Hider, P., Januel, J.-M., & Sundararajan, V. (2011). Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *American Journal of Epidemiology*, *173*(6), 676–682. https://doi.org/10.1093/aje/kwq433

Robins, J. M. (1999). Association, Causation, and Marginal Structural Models. *Synthese*, *121*(1), 151–179.

Royston, P. (2001). Flexible Parametric Alternatives to the Cox Model, and more. *The Stata Journal: Promoting Communications on Statistics and Stata*, *1*(1), 1–28. https://doi.org/10.1177/1536867X0100100101

Sauerbrei, W., Meier-Hirmer, C., Benner, A., & Royston, P. (2006). Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Computational Statistics & Data Analysis*, *50*(12), 3464–3485. https://doi.org/10.1016/j.csda.2005.07.015

Sauerbrei, W., Royston, P., & Binder, H. (2007). Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, *26*(30), 5512–5528. https://doi.org/10.1002/sim.3148

Seymour, E. K., Ruterbusch, J. J., Winn, A. N., George, J. A., Beebe-Dimmer, J. L., & Schiffer, C. A. (2021). The costs of treating and not treating patients with chronic myeloid leukemia with tyrosine kinase inhibitors among Medicare patients in the United States. *Cancer*, *127*(1), 93–102. https://doi.org/10.1002/cncr.33267

Shen, C., Zhao, B., Liu, L., & Shih, Y.-C. T. (2017). Financial Burden for Patients With Chronic Myeloid Leukemia Enrolled in Medicare Part D Taking Targeted Oral Anticancer Medications. *Journal of Oncology Practice*, *13*(2), e152–e162. https://doi.org/10.1200/JOP.2016.014639

Sylvestre, M.-P., & Abrahamowicz, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine*, *28*(27), 3437–3453. https://doi.org/10.1002/sim.3701

Tarazi, W., Welch, W. P., Nguyen, N., Bosworth, A., Sheingold, S., Lew, N. D., & Sommers, B. D. (2022). *Medicare served nearly 63 million beneficiaries in 2019. 62 percent were enrolled in Part A or Part B, and the rest (37 percent) were in Medicare Advantage (Part C). 74 percent were enrolled in Part D drug coverage, 13 percent had private drug coverage, and nearly 9 percent had no drug coverage. Demographic characteristics and health status varied across these groups.* 13.

Taylor, E. A. (2013). *The Effect of Specialty Tier Placement on Enrollment and Utilization in Medicare Part D*. 159.

TriNetX. (2022). TriNetX. Retrieved October 10, 2022, from https://trinetx.com/healthcare-organizations/

Trivedi, D., Landsman-Blumberg, P., Darkow, T., Smith, D., McMorrow, D., & Mullins, C. D. (2014). Adherence and Persistence Among Chronic Myeloid Leukemia Patients During Second-Line Tyrosine Kinase Inhibitor Treatment. *Journal of Managed Care Pharmacy*, *20*(10), 1006–1015. https://doi.org/10.18553/jmcp.2014.20.10.1006

Tueller, S. J., Deboeck, P. R., & Van Dorn, R. A. (2016). Getting less of what you want: Reductions in statistical power and increased bias when categorizing medication adherence data. *Journal of Behavioral Medicine*, *39*(6), 969–980. https://doi.org/10.1007/s10865-016-9727-9

Wang, Y., Beauchamp, M.-E., & Abrahamowicz, M. (2020). Nonlinear and time-dependent effects of sparsely measured continuous time-varying covariates in time-to-event analysis. *Biometrical Journal. Biometrische Zeitschrift*, *62*(2), 492–515. https://doi.org/10.1002/bimj.201900042

Ward, M. A., Fang, G., Richards, K. L., Walko, C. M., Earnshaw, S. R., Happe, L. E., & Blalock, S. J. (2015). Comparative evaluation of patients newly initiating first-generation versus second-generation tyrosine kinase inhibitors for chronic myeloid leukemia and medication adherence, health services utilization, and healthcare costs. *Current Medical Research and Opinion*, *31*(2), 289–297. https://doi.org/10.1185/03007995.2014.991440

Winn, A. N., Keating, N. L., & Dusetzina, S. B. (2016). Factors Associated With Tyrosine Kinase Inhibitor Initiation and Adherence Among Medicare Beneficiaries With Chronic Myeloid Leukemia. *Journal of Clinical Oncology*, *34*(36), 4323–4328. https://doi.org/10.1200/JCO.2016.67.4184

Wu, E. Q., Johnson, S., Beaulieu, N., Arana, M., Bollu, V., Guo, A., Coombs, J., Feng, W., & Cortes, J. (2010). Healthcare resource utilization and costs associated with non-adherence to imatinib treatment in chronic myeloid leukemia patients. *Current Medical Research and Opinion*, *26*(1), 61–69. https://doi.org/10.1185/03007990903396469

Wynant, W., & Abrahamowicz, M. (2014). Impact of the model-building strategy on inference about nonlinear and time-dependent covariate effects in survival analysis. *Statistics in Medicine*, *33*(19), 3318–3337. https://doi.org/10.1002/sim.6178

Wynant, W., & Abrahamowicz, M. (2016). Validation of the alternating conditional estimation algorithm for estimation of flexible extensions of Cox's proportional hazards model with nonlinear constraints on the parameters. *Biometrical Journal*, *58*(6), 1445–1464. https://doi.org/10.1002/bimj.201500035

Yeager, K. A., Waldrop-Valverde, D., Paul, S., Bruner, D. W., Klisovic, R., Burns, E., Mason, T. A., Patel, N., & Jennings, B. M. (2020). Adherence trajectories in oral therapy for chronic myeloid leukemia: Overview of a research protocol. *Research in Nursing & Health*, *43*(5), 443–452. https://doi.org/10.1002/nur.22069

Yood, M. U., Oliveria, S. A., Cziraky, M., Hirji, I., Hamdan, M., & Davis, C. (2012). Adherence to treatment with second-line therapies, dasatinib and nilotinib, in patients with chronic myeloid leukemia. *Current Medical Research and Opinion*, *28*(2), 213–219. https://doi.org/10.1185/03007995.2011.649849

Chapter 6. APPENDIX

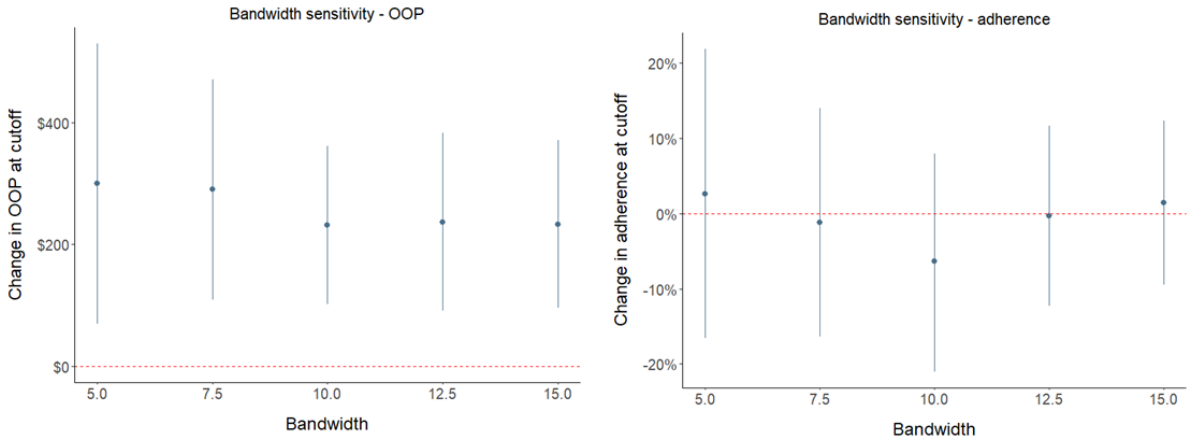

Figure A1. Sensitivity of results to bandwidth selection

These plots represent the impact of various bandwidths (selected based on Kolesár et al (Kolesár & Rothe, 2018)), with circles representing point estimates of average effects, vertical lines representing the corresponding 95% confidence intervals, and dashed red line representing a null finding of no effect.

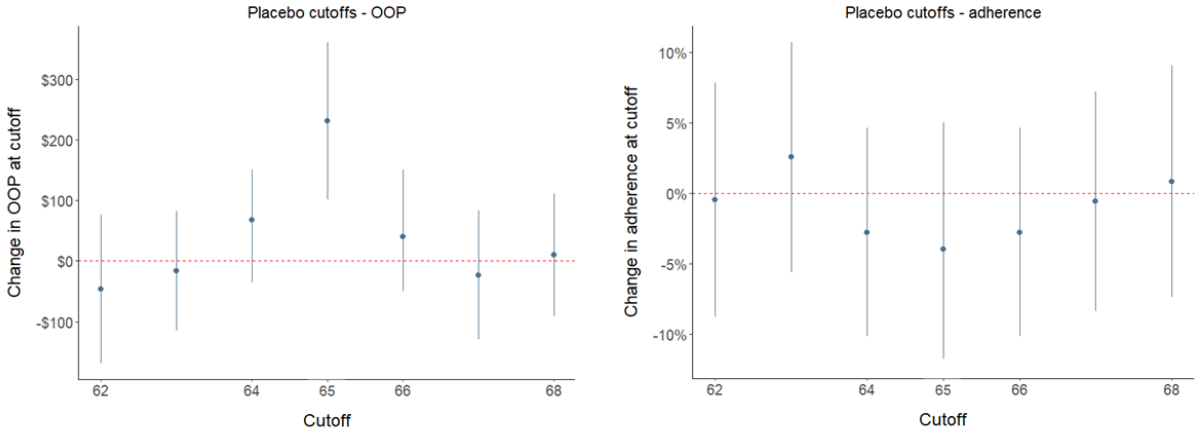

Figure A2. Robustness of results to placebo thresholds

These plots represent the impact of using an RD threshold other than age 65 on results, with circles representing point estimates of average effects, vertical lines representing the corresponding 95% confidence intervals, and dashed red line representing a null finding of no effect.

.

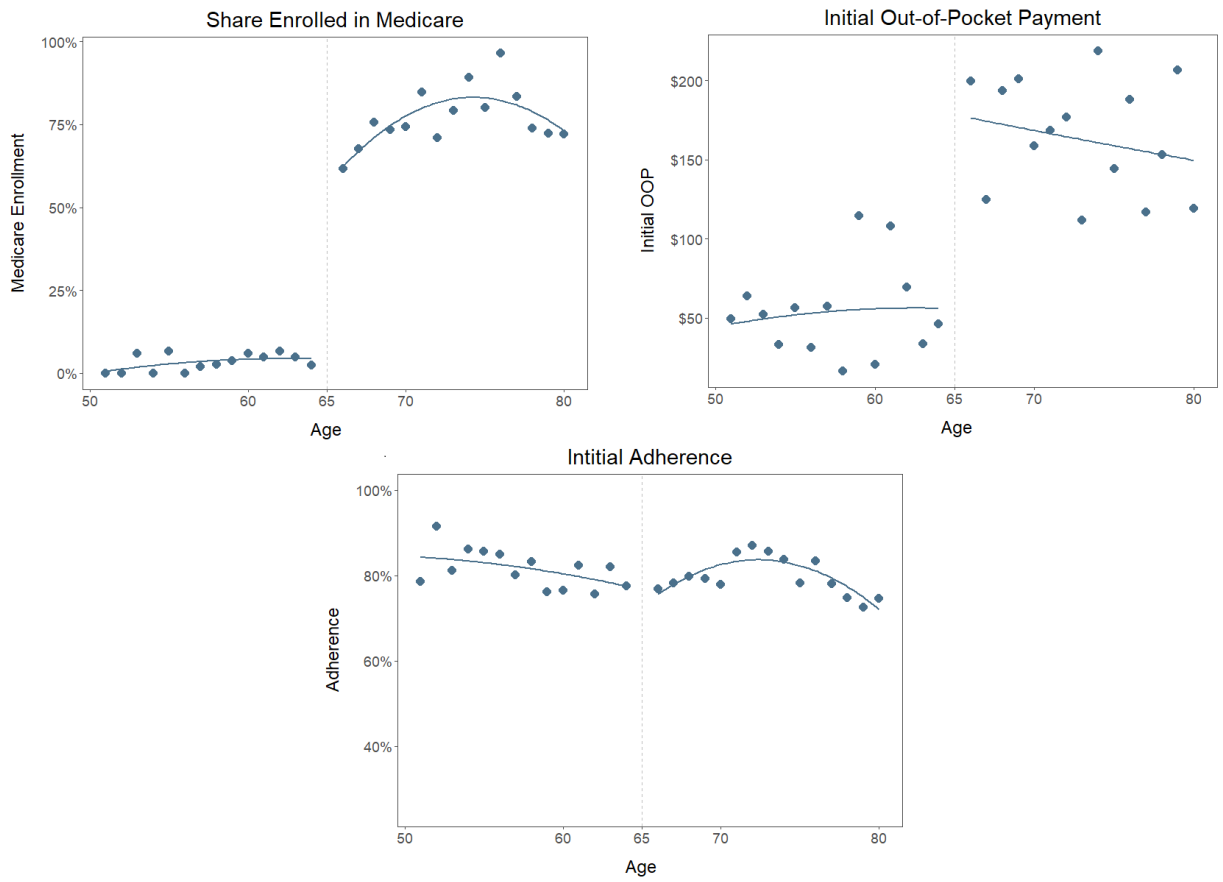

Figure A3. Global trends in Part D coverage, OOP, and adherence

Each circle represents the average outcome by age at index date from the data. The lines are fitted values from a global polynomial (quadratic) regression.