Healthcare Costs of Treating Privately Insured Patients with Acute Myeloid Leukemia in the United States 2004–2014: A Generalized Additive Modeling Approach

Lih-Wen Mau, PhD, MPH[1,2], **Jaime M. Preussler, MS**[1,2,†], **Linda J. Burns, MD**[1,2], **Susan Leppke, MPH**[1], **Navneet S. Majhail, MD, MS**[3], **Christa L. Meyer, MS**[1,2], **Tatenda Mupfudze, PhD**[1,2], **Wael Saber, MD, MS**[4], **Patricia Steinert, PhD**[4], **David J. Vanness, PhD**[5]

[1]National Marrow Donor Program/Be The Match, Minneapolis, MN

[2]Center for International Blood and Marrow Transplant Research, Minneapolis, MN

[3]Blood & Marrow Transplant Program, Cleveland Clinic, Cleveland, OH

[4]Center for International Blood and Marrow Transplant Research, Milwaukee, WI

[5]Apriori Bayesian Consulting, LLC, State College, PA

Address for Correspondence: David J. Vanness, PhD, Apriori Bayesian Consulting, LLC, 2643 Sleepy Hollow Drive, State College, PA 16803, Phone: 608-345-0521, dvanness@aprioribayesian.com.
[†]Co-first author
[*]Corporate Members

Data Availability Statement

Optum's de-identified Clinformatics® Data Mart Database is commercially available. According to the terms of the data purchase agreement, the dataset used in this study cannot be shared. Analysis code is available at: https://github.com/djvanness/pharmacoecon_GAM_AML

Abstract

Objectives—To predict healthcare cost trajectories for patients with newly diagnosed acute myeloid leukemia (AML) receiving allogeneic hematopoietic cell transplantation (alloHCT), as a function of days since chemotherapy initiation, days relative to alloHCT and days before death or last date of insurance eligibility (LDE). An exploratory objective examined patients with AML receiving chemotherapy only.

Methods—We used Optum's de-identified Clinformatics® Data Mart Database to construct cumulative cost trajectories from chemotherapy initiation to death or LDE (through December 31, 2014) for United States patients aged 20–74 years diagnosed between March 1, 2004 and December 31, 2013 (n=187 alloHCT; n=253 chemotherapy only). We used generalized additive modeling (GAM) to predict expected trajectories and bootstrapped confidence intervals (CI) at user-specified intervals conditional on dates of alloHCT and death or LDE relative to chemotherapy initiation.

Results—Expected costs for a hypothetical patient receiving alloHCT 60 days after chemotherapy initiation and followed for five years were $572,000 (95% CI:$517,000-$633,000); $119,000 (95% CI:$51,000-$192,000); $102,000 (95% CI:$0-$285,000); $79,000 (95% CI:$0-$233,000), for years 1–4 respectively, and either $494,000 (95% CI:$212,000-$799,000) or $108,000 (95% CI:$0-$230,000) in year 5, whether the patient died or was lost to follow-up on day 1,825, respectively.

Conclusions—Rates of cost accrual varied over time since chemotherapy initiation with accelerations around the time of alloHCT and death. GAM is a potentially useful approach for imputing longitudinal costs relative to treatment initiation and one or more intercurrent or terminal events in randomized controlled trials or registries with unrecorded costs or for dynamic decision-analytic models.

Keywords

Hematopoietic cell transplantation; chemotherapy; administrative claims; generalized additive modeling; real world evidence; healthcare cost

Introduction

Acute myeloid leukemia (AML) is a cancer of the bone marrow, characterized by proliferation of cancerous cells that prevent the marrow from producing healthy blood cells [1,2]. Following initial chemotherapy, consolidation chemotherapy is recommended for patients with favorable risk disease whereas those with intermediate or poor risk disease are considered candidates for allogeneic hematopoietic cell transplantation (alloHCT) [3].

AlloHCT is resource-intensive [4–7]. Several studies have examined costs and utilization associated with alloHCT or chemotherapy but not alloHCT (chemotherapy only) for patients with AML [8–11]. A recent study estimated total costs in the first year after AML diagnosis for patients age 50–64 years who received chemotherapy only to be $280,799 (95% confidence interval [CI]:$258,473-$305,030) or who received alloHCT to be $544,178 (95% CI:$489,994-$604,353) [12]. However, this study was limited by its short time horizon and

inability to account for the relationship between mortality and cost. All else equal, patients who survive longer are expected to incur higher total costs simply because they receive additional healthcare while alive. From the perspective of third-party payers, the healthcare system or society, costs associated with increased longevity are real costs that must be considered [13]. However, intensive resource use around the time of death may incur substantial costs [14–16] with little therapeutic benefit [17,18].

Our primary objective was to estimate healthcare cost trajectories as a function of days since chemotherapy initiation, days relative to alloHCT and days before death or loss to follow-up for patients with newly-diagnosed AML receiving alloHCT. An exploratory objective examined trajectories for patients with AML receiving chemotherapy only. Better understanding of the dynamics of healthcare costs may shed light on the potential value of alloHCT and may further help identify future research opportunities for higher quality end-of-life care for patients with AML.

To best inform decision-making, cost-effectiveness and budget impact analyses should account for variation in healthcare costs over the course of treatment. Treatment intensity, and therefore the rate at which costs accumulate, is likely to vary over time for complex conditions with heterogeneous clinical courses such as AML. Common approaches to estimating costs for patients with cancer [19] stratify total costs by phase of care (initial, continuing and terminal) [20–22] or over specific windows of time since diagnosis such as the first year after diagnosis [12,23]. Analysis by fixed time windows since diagnosis combines the heterogeneous experiences of patients who die within the period (potentially contributing substantial end-of-life costs, but lower costs of continuing care) with those who survive. Analysis by phase of care also typically uses fixed windows of time since diagnosis but adds a fixed window of time before death and defines the continuing phase of care as all time in-between. For example, Brown et al (2002) define the initial phase of colorectal cancer care to be the month of diagnosis and five months afterwards and defines the terminal phase as the last twelve months of life [20]. The treatment phase approach blends potentially heterogeneous patient experiences within phases, varying by timing of initial and potentially subsequent treatments. Cost analysis by phase also requires patients in the initial treatment phase to have survived into the continuing phase; patients dying within the specified initial phase window are assigned to the terminal phase despite having initiated treatment in that same timeframe.

Longitudinal approaches estimating trajectories of costs conditional on survival time offer potentially more useful information for decision-makers relative to analyses of the total cost of care over wide time windows or by phase of care [24–27]. Such approaches are useful for imputing costs for dynamic decision analytic models -- particularly discrete event simulation [28] or partitioned survival models [29] that simulate individual times to events. We introduce an approach to analyzing trajectories of cumulative healthcare costs from claims data using natural cubic splines in a generalized additive modeling (GAM) framework assuming gamma-distributed costs and a log link function. We then apply our approach to predict expected cost trajectories of patients with AML who were treated with either alloHCT or chemotherapy only, conditional on when alloHCT was received as well as when death or loss to follow-up occurred.

Methods

Data Source

Optum's de-identified Clinformatics® Data Mart Database (Optum) included commercial and Medicare Advantage inpatient, outpatient, professional, ancillary services, and pharmacy claims from January 1, 2004 through December 31, 2014, for members from all 50 states in the United States (U.S.). Optum linked claims to the Social Security Administration Death Master File (DMF), allowing joint investigation of cost and survival. While specificity for deaths reported in the DMF is nearly 100% [30], due to lack of a deterministic link to the death master file some deaths among patients are unreported [31]. Because Optum reported month and year of death only, we imputed death dates following an algorithm using claims information at the last date of service (see Appendix Table A1).

Our study was considered exempt by the National Marrow Donor Program's Institutional Review Board.

Study Population and Patient Selection

Our primary study population comprised U.S. commercially insured and Medicare Advantage adults aged 20–74 years with de novo AML diagnosed between March 1, 2004 and December 31, 2013 who received alloHCT; an exploratory aim included patients of the same age range and diagnosis who received chemotherapy only. Cohort identification followed procedures from a previous claims-based study of patients with newly-diagnosed AML treated with alloHCT or chemotherapy only [12]. Patients were identified by International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes (see Appendix Table A2). Inclusion and exclusion criteria (see Appendix Table A3) were designed to balance sensitivity (including as many patients with newly-diagnosed AML who underwent treatment with curative intent as possible) with specificity (excluding patients who had spurious AML-related codes, AML likely to be secondary to other diseases, or recurrent previously-diagnosed AML) [32].

Cumulative Healthcare Cost Trajectories

We constructed cumulative healthcare cost trajectories as daily running totals for each patient. Trajectories for each individual i began at their date of chemotherapy initiation ($t_{CI[i]} \equiv 0$), and ended at the earlier of their date of death $t_{D[i]}$ or last date of eligibility (LDE) $t_{LDE[i]}$ (capped at the end of the study period, December 31, 2014). Trajectories for patients with LDE preceding death or who had no recorded date of death ended at $t_{LDE[i]}$. Dates of alloHCT ($t_{HCT[i]}$), death, and eligibility were counted relative to chemotherapy initiation. Individual cost trajectories therefore ranged from $t \in [0, \min\{t_{D[i]}, t_{LDE[i]}\}]$, and overall expected cost trajectories ranged from $t \in [0, \max_i\{\min\{t_{D[i]}, t_{LDE[i]}\}\}]$.

Let $j = 1, \ldots J_i$ index the set of all J_i claims of patient i sorted by service date. Then $\ell = 1, \ldots \mathscr{L}_{i,j,t}$ indexes the line items for claim j of patient i having date of service t relative to the date of chemotherapy initiation. Line-item service dates for outpatient facility, professional and ancillary services, and pharmacy services were considered reliable, since such services are delivered on a single date. Line-item inpatient facility costs observed

during an inpatient stay were summed and divided by the length of stay in days, and then allocated equally to each day of the inpatient stay:

$$c_{i,j,t} = \frac{1}{(t1 - t0)} \sum_{t=t0}^{t1} \left(\sum_{\ell=1}^{\mathscr{L}_{i,j,t}} c_\ell \right) \tag{1}$$

for $t = t0,\ldots,t1$, where $t0$ and $t1$ are the admission and discharge dates, respectively. Costs were then summed across claims by date:

$$\{c_{i,t}\} = \sum_{j=1}^{J_i} c_{i,j,t} \tag{2}$$

for $t = 0,\ldots \min \{t_{D[i]}, t_{LDE[i]}\}$ to obtain a daily sequence of costs $\{c_{i,t}\}$. By convention, $c_{i,j,t} = 0$ for all dates t in which no service was provided under claim j. Given each patient's cost sequence, we defined the cumulative cost trajectory as the daily running total of all healthcare costs:

$$\{C_{i,t}\} = \sum_{T=1}^{t} c_{i,T} \tag{3}$$

Note that if each element of $\{c_{i,t}\}$ ⩾ 0, then the cumulative cost trajectory $\{C_{i,t}\}$ is non-decreasing.

Optum applied administrative cost adjustments as negative values on either the same or a separate claim. We used an algorithm matching negative line-items with their corresponding positive line-items, if possible, and adjusted the date of each negative claim accordingly to match its associated positive claim. If no match was observed, negative claims were applied on the date indicated. Optum provided inflation-adjusted to the year 2017 standardized line-item costs, which account for variations in payment rates, contract terms, and geographic areas and therefore can be considered representative of the U.S. third-party payer perspective.

Statistical Analysis

For an introduction to GAM, see the Methodological Appendix. Our specification used four time-varying covariates, each of which can be thought of as a running stopwatch with different reference points measuring time relative to chemotherapy initiation and the dates of alloHCT, and death or loss to follow-up. Specifically, we modeled:

$$C_{i,t} \sim \text{Gamma}\left(\frac{\mu_{i,t}}{\gamma}, \gamma \right) \tag{4a}$$

$$\ln(\mu_{i,t}) = \sum_{k \in K_t} \beta_{1k} N_k(t_i) + 1[t_{HCT[i]} \neq NA] \sum_{k \in K_{HCT}} \beta_{2k} N_k(t_i - t_{HCT[i]})$$
$$+ 1[t_{D[i]} \leq t_{LDE[i]}] \cdot \sum_{k \in K_D} \beta_{3k} N_k(t_i - t_{D[i]}) \qquad (4b)$$
$$+ 1[t_{D[i]} = NA \text{ or } t_{LDE[i]} < t_{D[i]}] \cdot \sum_{k \in K_D} \beta_{4k} N_k(t_i - t_{LDE[i]})$$

where γ is an estimated shape parameter, $N()$ is the natural cubic spline basis function, β_{1k}, β_{2k}, β_{3k}, and β_{4k} are sets of estimated natural spline coefficients for each time covariate, and $K_t = \{0,30,60,90,120,150,180,365,550,730,1095,\max_i t_{D[i]}\}$, $K_{HCT} = \{-30,-14,0,14,30,100,180,365\}$, and $K_D = \{-30,-14,-7,0\}$ are sets of knots representing time since chemotherapy initiation t_i, time relative to alloHCT ($t_i - t_{HCT[i]}$), and time before death ($t_i - t_{D[i]}$) or loss-to-follow-up ($t_i - t_{LDE[i]}$), respectively, whichever comes first. The indicator function $1[t_{HCT[i]} \neq NA]$ turns on the additive component for time relative to alloHCT only for patients who received alloHCT (i.e., had a date of transplant that was not missing). Similarly, indicator functions $1[t_{D[i]} \leq t_{LDE[i]}]$ and $1[t_{D[i]} = NA \text{ or } t_{LDE[i]} < t_{D[i]}]$ turn on the appropriate additive component for time relative to whichever terminal event occurred (death and loss-to-follow-up, respectively). Knots were chosen by our transplant hematologist-oncologist co-authors to represent clinically relevant time points. Models were estimated separately by treatment received using the 'gam' package [33] in R (3.6.0) [34].

We used the 'predict.gam' method to predict expected cumulative cost through any date of interest. Given estimated $\hat{\beta}$ and $\hat{\gamma}$ and an input vector of known dates of alloHCT and death or loss to follow-up for a hypothetical representative individual j, the tool predicted expected cumulative costs $E[C_{j,t*}|t_{HCT[j]}, t_{D[j]}, t_{LDE[j]}; \hat{\beta}, \hat{\gamma}]$ through any specified date $t_* \in [0,\min\{t_{D[j]}, t_{LDE[j]}\}]$. To predict expected costs incurred in a time period, we took first-differences of a sequence of expected cumulative costs evaluated at specified time points t_n, for $n = 1,2,$ …N: $E[c_{j,t_n}| \cdot] = E[C_{j,t_n}| \cdot] - E[C_{j,t_{n-1}}| \cdot]$, where $E[C_{j,t_{n-1}}| \cdot] = 0$ if $t_{n-1} = 0$ by definition. For illustration and comparison with other published alloHCT cost analyses, we generated predictions on an annual basis, although any regular or irregular sequence of dates could be chosen.

To quantify uncertainty, we conducted 1,000 unweighted bootstrap replications by resampling trajectories with replacement, separately by treatment group. We constructed 95% confidence intervals (CI) for spline smooths using the percentile method. 95% CIs for predicted trajectories using the lookup tool imposed a monotonicity constraint by recursively choosing $E[C_{j,t_n}| \cdot] = \max\{E[C_{j,t_n}| \cdot], E[C_{j,t_{n-1}}| \cdot]\}$, for $n = 2,…N$, thus requiring each bootstrapped sequence to be non-decreasing.

RESULTS

Four hundred forty patients met inclusion criteria (n=187 alloHCT; n=253 chemotherapy only), (see Appendix Figure A1 and Appendix Table A3). Plots of observed cumulative cost trajectories for alloHCT and chemotherapy only recipients are presented in Figure 2. Due to

data sparsity beyond 5 years, we truncated all model plots and predictions at 1,825 days post-chemotherapy initiation.

Generalized Additive Model Results

GAM results are shown in Figure 2 for alloHCT recipients and Figure 3 for patients receiving chemotherapy only. Each panel represents the weighted sum of basis functions for each running time covariate and therefore corresponds to one of the Σ terms in (4b). Confidence intervals widen markedly approximately 1,000 days after chemotherapy initiation and 800 days after receipt of alloHCT due to small numbers of observed trajectories in which individuals have survived or were not lost to follow-up at that time. Base expected costs (i.e., not attributed by the GAM model to timing of alloHCT, death or loss to follow-up) rapidly accrue within the first three months after chemotherapy initiation and accrue less rapidly beginning around six months (Figures 2a, 3a). For patients treated with chemotherapy only surviving beyond approximately 500 days, the base rate of cost accrual reduces appreciably, while for patients treated with alloHCT, the base rate of cost accrual stabilizes, with cumulative costs increasing at a higher rate than for chemotherapy only.

For patients receiving alloHCT, the additive component for time relative to the date of alloHCT (Figure 2b) indicates rapidly increasing costs beginning 30 days prior to transplant and continuing at a high rate for the first 100 days. Cost accrual relative to the date of alloHCT then abates significantly and remains stable for approximately the first year, after which it begins to decline.

Both for patients receiving alloHCT and those receiving chemotherapy only, rates of cost accrual during the 30 days prior to death are significantly elevated (Figures 2c, 3c). While the magnitude of results may look small relative to the base rate of cost accrual, these plots are on the log-dollar scale, so the effect on the dollar scale is multiplicative. For patients receiving alloHCT, there appears to be little change in the rate of cost accumulation around the time of loss to follow-up (Figure 2d). However, for patients receiving chemotherapy only, the pattern of cost accrual around loss to follow-up (Figure 3d) is qualitatively similar to the pattern around time of death.

Expected Cost Trajectories

Allowing time for initial therapy and evaluation for alloHCT candidacy, we evaluated estimated cost trajectory equations at annual intervals for a hypothetical representative patient receiving alloHCT 60 days after chemotherapy initiation who either died (Figure 4a) or was lost to follow-up (Figure 4b) on day 1,825 after chemotherapy initiation. Yearly (non-cumulative) expected costs (rounded to the nearest $1,000) for patients receiving alloHCT on day 60 and surviving up to day 1,825 were $572,000 in year one (95% CI: $517,000-$633,000); $119,000 in year two (95% CI: $51,000-$192,000); $102,000 in year three (95% CI: $0-$285,000); and $79,000 in year four (95% CI: $0-$233,000). Expected costs in year five were substantially larger if death occurred on day 1,825: $494,000 (95% CI: $212,000-$799,000) than if the patient was lost to follow-up on day 1,825: $108,000 (95% CI:$0-

$230,000), indicating costs of approximately $386,000 attributable to end-of-life for alloHCT recipients.

Figures 4c and 4d assume the same scenario as 4a and 4b, except alloHCT occurs on day 425. While year one costs were approximately $221,000 lower than if alloHCT occurred on day 60, cumulative costs in years two through five are approximately $60,000 to $90,000 higher, suggesting that later transplantation is associated with slightly higher long-run cumulative costs.

Figures 4e and 4f represent scenarios where patients received alloHCT at day 60 but died or were lost to follow-up on day 960. Comparing 4f to 4b shows slightly lower cumulative costs at three years, reflecting loss to follow-up occurring 135 days before the full third year. Markov models with fixed cycle length should use appropriate methods to account for death or loss to follow-up occurring prior to the end of the full cycle.

Figure 5 represents scenarios where a hypothetical representative patient receiving chemotherapy only died (Figure 5a) or was lost to follow-up (Figure 5b) on day 1,825; or died (Figure 5c) or was lost to follow-up (Figure 5d) on day 960. Yearly costs for patients receiving chemotherapy only and surviving up to day 1,825 were $304,000 in year one (95% CI: $271,000-$339,000); $21,000 in year two (95% CI: $0-$61,000); $16,000 in year three (95% CI: $0-$69,000); and $43,000 in year four (95% CI: $0-$120,000). For patients receiving chemotherapy only, expected costs in year five were predicted to be higher than in year four, regardless of whether the patient died on day 1,825: $194,000 (95% CI: $0-$376,000) or was lost to follow-up: $131,000 (95% CI: $0-$260,000). Differences suggest expected end-of-life costs of approximately $63,000 for patients receiving chemotherapy only.

Discussion

Our analysis provides the longest follow-up to date of costs for patients with newly-diagnosed AML receiving alloHCT or chemotherapy only. Notably, our estimates for a hypothetical patient receiving alloHCT on day 60 after chemotherapy initiation are close to published one-year post-diagnosis costs using data from Truven [12], which included patients surviving less than one year: $572,000 vs. $544,178 for patients receiving alloHCT and $304,000 vs. $280,788 for patients receiving chemotherapy only.

Costs associated with end-of-life appear lower among patients treated with chemotherapy only than patients treated with alloHCT. However, end-of-life costs could be underestimated for patients receiving chemotherapy only if such patients are more likely than those receiving alloHCT to have unreported deaths. We note recent development of an algorithmic approach to identifying under-reported deaths in claims data [35]. While we can think of no plausible mechanism why patients receiving chemotherapy only who died would be less likely to have deaths reported in the DMF, we note that Optum uses clinical information including hospital discharge disposition to supplement the DMF. If patients receiving alloHCT are more likely to use healthcare, and particularly to die while hospitalized, then we might expect more accurate death status for alloHCT patients. We find suggestive

evidence in our results. For patients receiving alloHCT, healthcare cost trajectories in the 30 days before death (Figure 2c) and loss to follow-up (Figure 2d) were markedly different. However, for patients receiving chemotherapy only, the trajectories before death (Figure 3c) and loss to follow-up (Figure 3d) were qualitatively similar. The pattern we observe for alloHCT is what we would expect if loss to follow-up were unrelated to the intensity of healthcare use around the time of the last date of insurance eligibility. Overall, year five costs for chemotherapy only patients who are lost to follow-up on day 1,825 were predicted to be about $88,000 higher than year four costs, while the difference between cost in years four and five was only $29,000 for patients receiving alloHCT.

In our case study, confidence intervals become wider over time uncertainty due to decreasing numbers of trajectories lasting through later dates. Indeed, uncertainty in later years is substantial enough that in the absence of a monotonicity constraint, some of the 95% CIs for annual (non-cumulative) costs would have included negative numbers. We recommend when estimates are incorporated into decision-analytic models in practice that bootstrapped trajectories be used directly in second-order Monte Carlo probabilistic sensitivity analysis rather than treating expenditures at each time point as independent draws matching the point estimates and confidence intervals. This will preserve correlation among the parameters and ensure monotonicity of cumulative expenditures.

Another substantial limitation of our analysis was its inability to infer causality from directly observed differences in cost trajectories between non-randomized treatments. Patients receiving alloHCT and surviving to day 1,825 are unlikely to have the same clinical characteristics as patients who survive to day 1,825 receiving chemotherapy only, and therefore direct comparison of the two trajectories is not appropriate. However, we note that it is possible in other applications with larger datasets to account for observed potential confounders by adding them as additional covariates in the GAM specification. As GAM is simply a generalization of GLM, other covariates, whether continuous, ordinal or factor, can be incorporated if there are enough observations with variation to support their estimation. We also note that propensity scores could be used as inverse probability of treatment weights in GAM to achieve balance on observable factors. The two approaches could be combined in doubly-robust estimation [36]. However, we caution that administrative claims may not provide enough clinical information (such as cytogenetic risk in the case of AML) to balance patients on all important confounders between treatments.

GAM is a flexible semi-parametric approach with many applications in medical research [37], ecology [38], epidemiology [39] and geospatial analysis [40]. However, its use in health economics and outcomes research has so far been limited. We were unable to find any published applications of GAM to modeling of cumulative healthcare costs. One study by Pullenayegum et al (2012) demonstrated the potential usefulness of GAM for modeling the association of EQ-5D utility as a quasi-continuous measure with diabetes-related complications in a cross-sectional study of adults with diabetes [41]. GAM and restricted splines have been proposed as an alternative to fractional polynomials [42] for the modeling of time-varying non-proportional hazard curves in survival analysis [43,44], including network meta-analysis [45].

Our analysis shows GAM to be a potentially useful approach for imputing cost trajectories where variation in the intensity of health care utilization occurs around known events (such as receipt of subsequent treatment and death or loss to follow-up). Target applications include imputing costs in studies where RCTs or registries record relevant clinical events but do not track costs – a common situation faced by applied health economics and outcomes researchers. Imputation of cost trajectories with GAM is also potentially useful for dynamic decision analytic modeling, where simulated events occur with probability and timing dictated by parameters obtained through synthesis of secondary data sources. As demonstrated in Figures 4 and 5, we can evaluate expected cumulative cost trajectories at one or more user-specified time points, conditional on timing of intercurrent and terminal events. Such conditional cost curves can be used directly in decision-analytic models based on discrete event simulation or partitioned survival analysis, where simulated event times are drawn from empirical or parametric distributions. Alternatively, the trajectories can be evaluated at regular intervals and first-differenced for inclusion in Markov-like models. Of course, analysts would still need access to administrative data with sufficient sample sizes representing relevant treatments and populations. Fortunately, such data sources are widely available from several commercial vendors and government databases.

Our approach can be compared with other longitudinal cost models [24–27]. Wilson et al. (2006) used mixed effects recurrent event regression to estimate a "mean cumulative function," which accumulated average costs per patient per six-month period, excluding patients who were censored in each period [27]. Liu et al. (2007) jointly modeled survival and monthly cost using shared random effects to account both for censoring and for higher costs near death [26]. Liu (2009) expanded upon their previous work to incorporate two-part models accounting for periods with zero costs [25]. Recent work by Li et al. (2018) also jointly modeled survival and costs, in this case using Kaplan-Meier estimates to obtain the marginal distribution of survival followed by a bivariate penalized spline to model monthly cost conditional on estimated survival [24].

In comparison, our approach accounted for censoring by incorporating additive terms in GAM allowing the shape of cost trajectories to vary smoothly by time before death or loss to follow-up. While our approach cannot jointly predict survival and cost as in Liu et al. (2007), Liu (2009) or Li et al. (2018), our target applications (imputation of cost trajectories when event times are known) only require prediction of costs conditional on survival. The joint survival-cost models discussed above are fairly highly parameterized and require relatively sophisticated estimation techniques including the EM algorithm [24,26] and gaussian quadrature [25]. Furthermore, the additive nature of GAM allows us to predict cost trajectories conditional on more than one event – i.e., in addition to survival, we can also condition on loss to follow-up among patients who were not observed to have died as well as timing of alloHCT. In other applications, this approach could account for timing of multiple events, including treatment escalations, development of complications, disease recurrence or progression. GAM also easily accommodated non-normality of cost data, in our case using a gamma distribution with a log link function [46]. By using cumulative trajectories, we were also able to avoid the need to "bin" costs into monthly (or other fixed length) periods and/or use two-part models to account for periods with zero expenditures.

Conclusions

Our study demonstrated a novel approach to analysis of healthcare cost trajectories using generalized additive models. We found the rates at which costs accrue for patients diagnosed with AML vary substantially over time since chemotherapy initiation and around the time of alloHCT. We also found evidence of substantially increased costs in the 30 days prior to death, particularly for patients treated with alloHCT. Because treatment is self-selected in our claims-based dataset, causal interpretation of observational cost-differences between treatments is not warranted. However, our approach may have general application for imputing cost trajectories conditional on known intercurrent and terminal events, such as in randomized controlled trials where cost data were not collected or for decision dynamic decision-analytic models where such events are simulated based on synthesized evidence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Schiffer C, Anastasi J. Classification of acute myeloid leukemia. UpTodate [Internet]. 2019 [cited 2019 Dec 23]. Available from: https://www.uptodate.com/contents/classification-of-acute-myeloid-leukemia

2. BeTheMatch.org. Acute myeloid leukemia (AML) [Internet]. 2019 [cited 2019 Dec 23]. Available from: https://bethematch.org/patients-and-families/about-transplant/diseases-treated-by-transplant/acute-myeloid-leukemia/

3. O'Donnell MR, Tallman MS, Abboud CN, Altman JK, Appelbaum FR, Arber DA, et al. Acute myeloid leukemia, version 3.2017, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2017;15:926–957. [PubMed: 28687581]

4. Preussler JM, Denzen EM, Majhail NS. Costs and cost-effectiveness of hematopoietic cell transplantation. Biol Blood Marrow Transplant. 2012;18:1620–8. [PubMed: 22484549]

5. Majhail NS, Mau L-W, Denzen EM, Arneson TJ. Costs of autologous and allogeneic hematopoietic cell transplantation in the United States: a study using a large national private claims database. Bone Marrow Transplant. 2013;48:294. [PubMed: 22773126]

6. Majhail NS, Mothukuri JM, Brunstein CG, Weisdorf DJ. Costs of hematopoietic cell transplantation: comparison of umbilical cord blood and matched related donor transplantation and the impact of posttransplant complications. Biol Blood Marrow Transplant. 2009;15:564–573. [PubMed: 19361748]

7. Khera N, Zeliadt SB, Lee SJ. Economics of hematopoietic cell transplantation. Blood. 2012;120:1545–1551. [PubMed: 22700725]

8. Ashfaq K, Yahaya I, Hyde C, Andronis L, Barton P, Bayliss S, et al. Clinical effectiveness and cost-effectiveness of stem cell transplantation in the management of acute leukaemia: a systematic review. Health Technol Assess. 2010;14:1–141.

9. Yang H, Huang S, Zhu C-Y, Gao L, Zhu H-Y, Lv N, et al. The superiority of allogeneic hematopoietic stem cell transplantation over chemotherapy alone in the treatment of acute myeloid leukemia patients with mixed lineage leukemia (MLL) rearrangements. Med Sci Monit. 2016;22:2315–23. [PubMed: 27373985]

10. Leunis A, Blommestein HM, Huijgens PC, Blijlevens NM, Jongen-Lavrencic M, Uyl-de Groot CA. The costs of initial treatment for patients with acute myeloid leukemia in the Netherlands. Leuk Res. 2013;37:245–250. [PubMed: 23069745]

11. Van de Velde AL, Beutels P, Smits EL, Van Tendeloo VF, Nijs G, Anguille S, et al. Medical costs of treatment and survival of patients with acute myeloid leukemia in Belgium. Leuk Res. 2016;46:26–29. [PubMed: 27111858]

12. Preussler JM, Meyer CL, Mau L-W, Majhail NS, Denzen EM, Edsall KC, et al. Healthcare costs and utilization for patients age 50 to 64 years with acute myeloid leukemia yreated with chemotherapy or with chemotherapy and allogeneic hematopoietic cell transplantation. Biol Blood Marrow Transplant. 2017;23:1021–8. [PubMed: 28263920]

13. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-Effectiveness in Health and Medicine. Oxford University Press; 2016.

14. Bramley T, Antao V, Lunacsek O, Hennenfent K, Masaquel A. The economic burden of end-of-life care in metastatic breast cancer. J Med Econ. 2016;19:1075–1080. [PubMed: 27248201]

15. Fletcher SA, Cronin AM, Zeidan AM, Odejide OO, Gore SD, Davidoff AJ, et al. Intensity of end-of-life care for patients with myelodysplastic syndromes: Findings from a large national database. Cancer. 2016;122:1209–1215. [PubMed: 26914833]

16. Chastek B, Harley C, Kallich J, Newcomer L, Paoli CJ, Teitelbaum AH. Health care costs for patients with cancer at the end of life. J Oncol Pract. 2012;8:75s–80s. [PubMed: 23598848]

17. Huynh TN, Kleerup EC, Wiley JF, Savitsky TD, Guse D, Garber BJ, et al. The frequency and cost of treatment perceived to be futile in critical care. JAMA Intern Med. 2013;173:1887–1894. [PubMed: 24018712]

18. Earle CC, Landrum MB, Souza JM, Neville BA, Weeks JC, Ayanian JZ. Aggressiveness of cancer care near the end of life: is it a quality-of-care issue? J Clin Oncol. 2008;26:3860–6.

19. Barlow WE. Overview of methods to estimate the medical costs of cancer. Med Care. 2009;47:S33–6. [PubMed: 19536013]

20. Brown ML, Riley GF, Schussler N, Etzioni R. Estimating health care costs related to cancer treatment from SEER-Medicare data: Med Care. 2002;40:IV104–17.

21. Yabroff KR, Lamont EB, Mariotto A, Warren JL, Topor M, Meekins A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 2008;100:630–641. [PubMed: 18445825]

22. Mariotto AB, Robin Yabroff K, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010–2020. JNCI J Natl Cancer Inst. 2011;103:117–28. [PubMed: 21228314]

23. Guy GP, Ekwueme DU, Yabroff KR, Dowling EC, Li C, Rodriguez JL, et al. Economic burden of cancer survivorship among adults in the United States. J Clin Oncol. 2013;31:3749–57. [PubMed: 24043731]

24. Li L, Wu C-H, Ning J, Huang X, Shih Y-CT, Shen Y. Semiparametric Estimation of Longitudinal Medical Cost Trajectory. J Am Stat Assoc. 2018;113:582–92. [PubMed: 30853736]

25. Liu L Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. Stat Med. 2009;28:972–86. [PubMed: 19040210]

26. Liu L, Wolfe RA, Kalbfleisch JD. A shared random effects model for censored medical costs and mortality. Stat Med. 2007;26:139–55. [PubMed: 16526006]

27. Wilson LS, Tesoro R, Elkin EP, Sadetsky N, Broering JM, Latini DM, et al. Cumulative cost pattern comparison of prostate cancer treatments. Cancer Interdiscip Int J Am Cancer Soc. 2007;109:518–527.

28. Karnon J Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. Health Econ. 2003;12:837–48. [PubMed: 14508868]

29. Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of survival probabilities for use in cost-effectiveness analyses: A comparison of a multi-state modeling survival analysis approach with partitioned survival and Markov decision-analytic modeling. Med Decis Making. 2017;37:427–39. [PubMed: 27698003]

30. Hauser TH, Ho KK. Accuracy of on-line databases in determining vital status. J Clin Epidemiol. 2001;54:1267–1270. [PubMed: 11750196]

31. da Graca B, Filardo G, Nicewander D. Consequences for healthcare quality and research of the exclusion of records from the Death Master File. Circ Cardiovasc Qual Outcomes. 2013;6:124–128. [PubMed: 23322808]

32. Preussler JM, Mau L-W, Majhail NS, Meyer CL, Denzen EM, Edsall KC, et al. Administrative claims data for economic analyses in hematopoietic cell transplantation: challenges and opportunities. Biol Blood Marrow Transplant. 2016;22:1738–46. [PubMed: 27184624]

33. Hastie T gam: Generalized Additive Models [Internet]. 2019 [cited 2019 Nov 7]. Available from: https://CRAN.R-project.org/package=gam

34. R Core Team. R: A language and environment for statistical computing. [Internet]. R Found. Stat. Comput. Vienna Austria. 2019 [cited 2019 Nov 7]. Available from: https://www.r-project.org/

35. Reps JM, Rijnbeek PR, Ryan PB. Identifying the DEAD: development and validation of a patient-level model to predict death status in population-level claims data. Drug Saf. 2019;42:1377–86. [PubMed: 31054141]

36. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol. 2011;173:761–7. [PubMed: 21385832]

37. Hastie T, Tibshirani R. Generalized additive models for medical research. Stat Methods Med Res. 1995;4:187–196. [PubMed: 8548102]

38. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol. 2002;156:193–203. [PubMed: 12142253]

39. Setodji CM, Scheuner M, Pankow JS, Blumenthal RS, Chen H, Keeler E. A graphical method for assessing risk factor threshold values using the generalized additive model: the multi-ethnic study of atherosclerosis. Health Serv Outcomes Res Methodol. 2012;12:62–79. [PubMed: 22593642]

40. Webster T, Vieira V, Weinberg J, Aschengrau A. Method for mapping population-based case-control studies: an application using generalized additive models. Int J Health Geogr. 2006;5:26. [PubMed: 16764727]

41. Pullenayegum EM, Wong HS, Childs A. Generalized additive models for the analysis of EQ-5D utility data. Med Decis Making. 2013;33:244–251. [PubMed: 23132902]

42. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC Med Res Methodol. 2011;11:61. [PubMed: 21548941]

43. Tsujitani M, Tanaka Y. Analysis of heart transplant survival data using generalized additive models. Comput Math Methods Med. 2013;Article ID 609857.

44. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21:2175–2197. [PubMed: 12210632]

45. Freeman SC, Carpenter JR. Bayesian one-step IPD network meta-analysis of time-to-event data using Royston-Parmar models. Res Synth Methods. 2017;8:451–464. [PubMed: 28742955]

46. Blough DK, Ramsey SD. Using generalized linear models to assess medical care costs. Health Serv Outcomes Res Methodol. 2000;1:185–202.

47. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. J Health Serv Res Policy. 2004;9:197–204. [PubMed: 15509405]

48. Hastie T, Tibshirani R. Generalized Additive Models. Stat Sci. 1986;1:297–310.

49. Durrleman S, Simon R. Flexible regression models with cubic splines. Stat Med. 1989;8:551–561. [PubMed: 2657958]

50. De Boor C On calculating with B-splines. J Approx Theory. 1972;6:50–62.

- For patients with newly diagnosed acute myeloid leukemia (AML), receipt of allogeneic hematopoietic cell transplantation (alloHCT) is associated with increased intensity of healthcare cost accumulation from 30 days before transplant through about 100 days after, remaining elevated for about the first year.

- For patients with AML receiving alloHCT, the rate at which healthcare costs are incurred accelerates substantially in the last 30 days of life.

- Generalized additive models may be useful for imputing trajectories of healthcare costs that vary with respect to the timing of treatment initiation and receipt of subsequent treatment as well as death or other events.

a. AlloHCT (n=187)
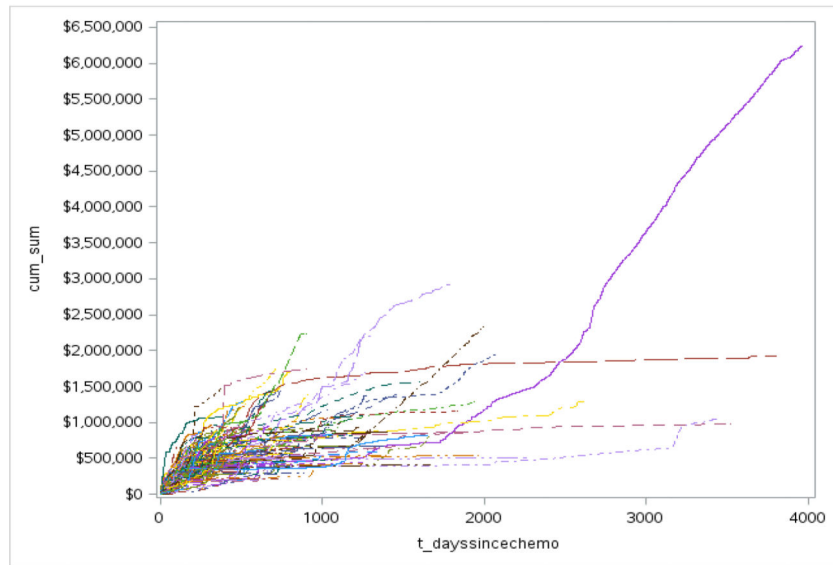
b. Chemotherapy only (n=253)
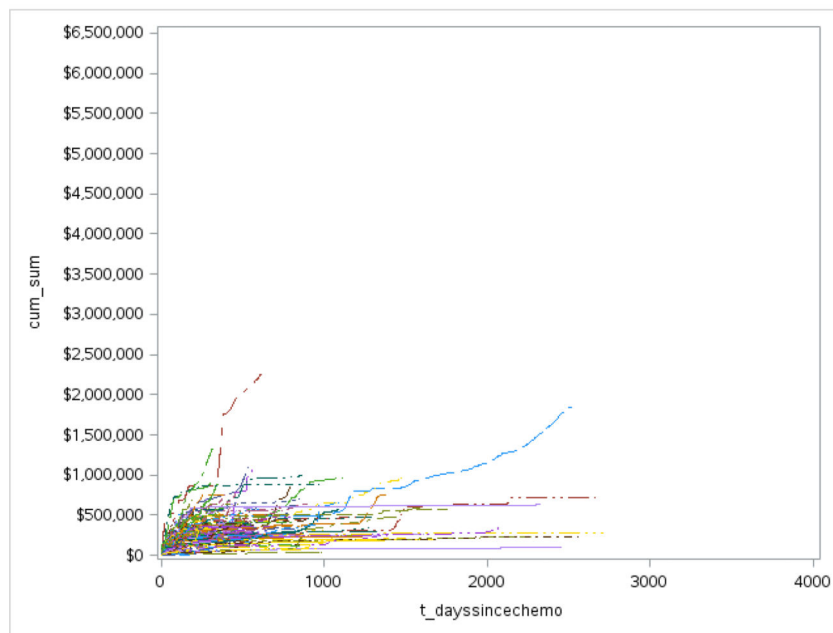
Figure 1. Cost trajectories from chemotherapy initiation until death or loss to follow-up for patients diagnosed with AML between March 2004 and December 2013
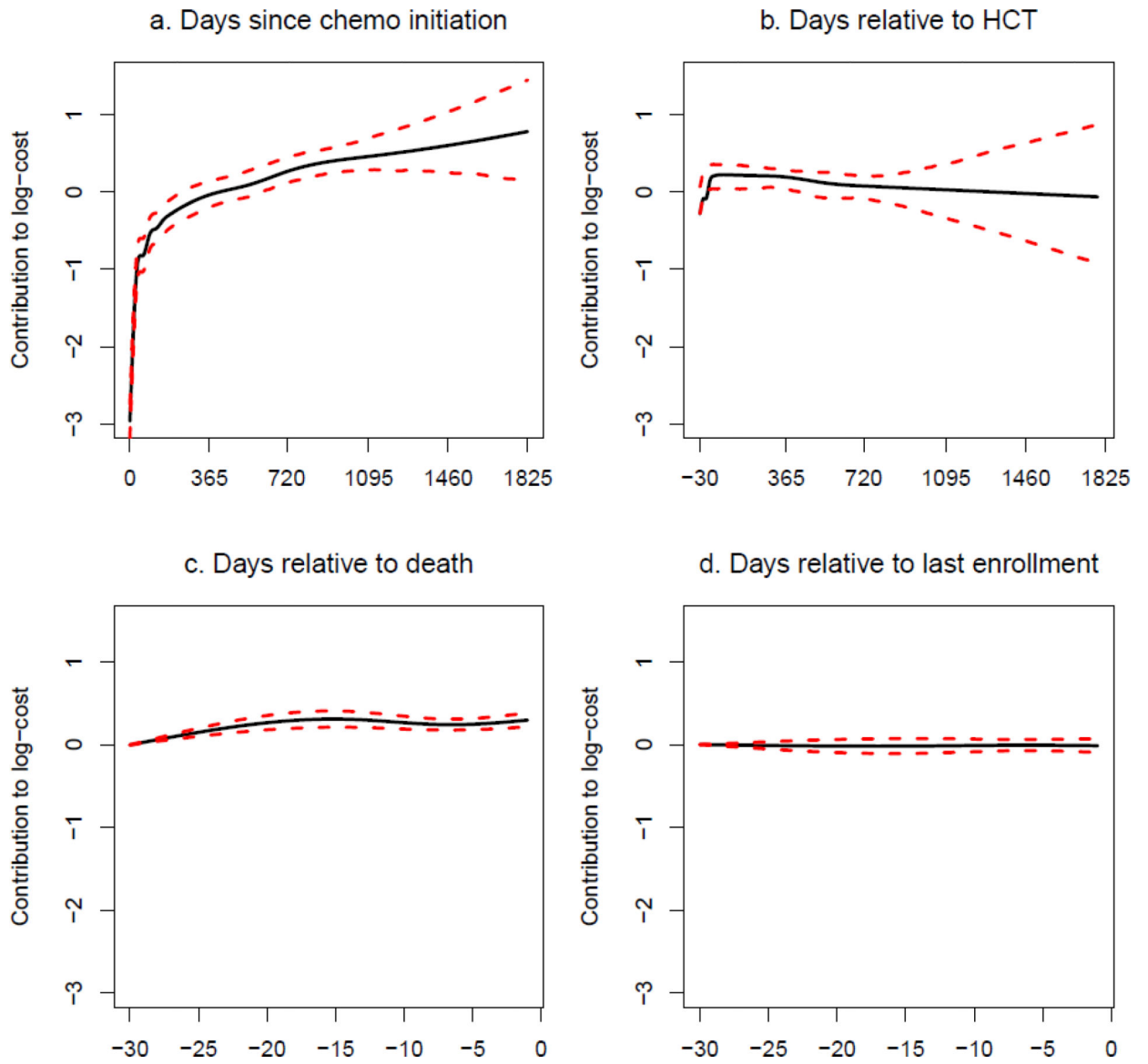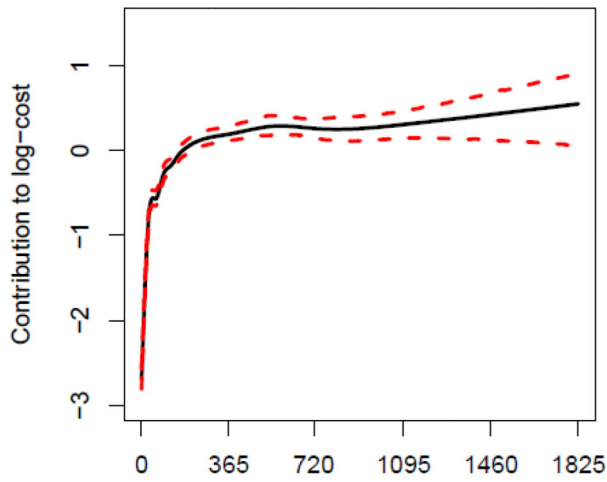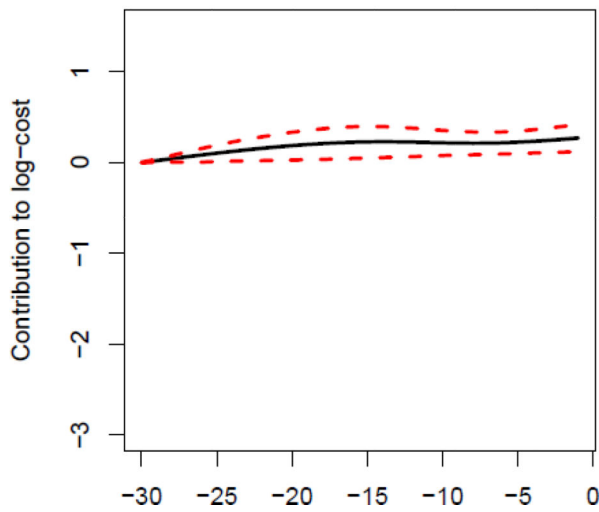
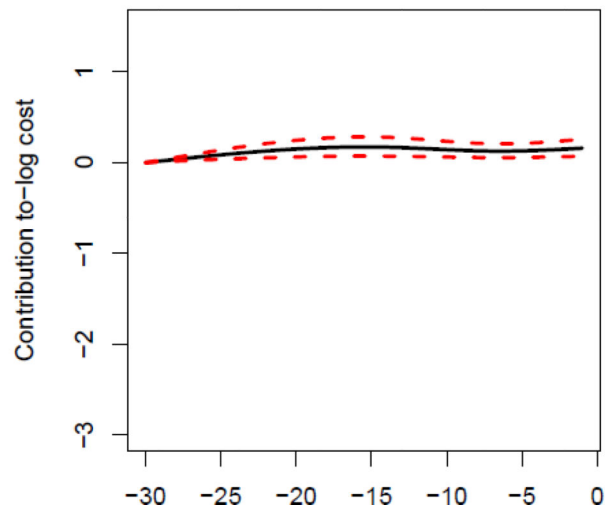**Figure 2. GAM modeling: AlloHCT Group**

Figure 3. GAM modeling: Chemotherapy only group. Note: panel b. intentionally left blank to facilitate comparison with Figure 2.
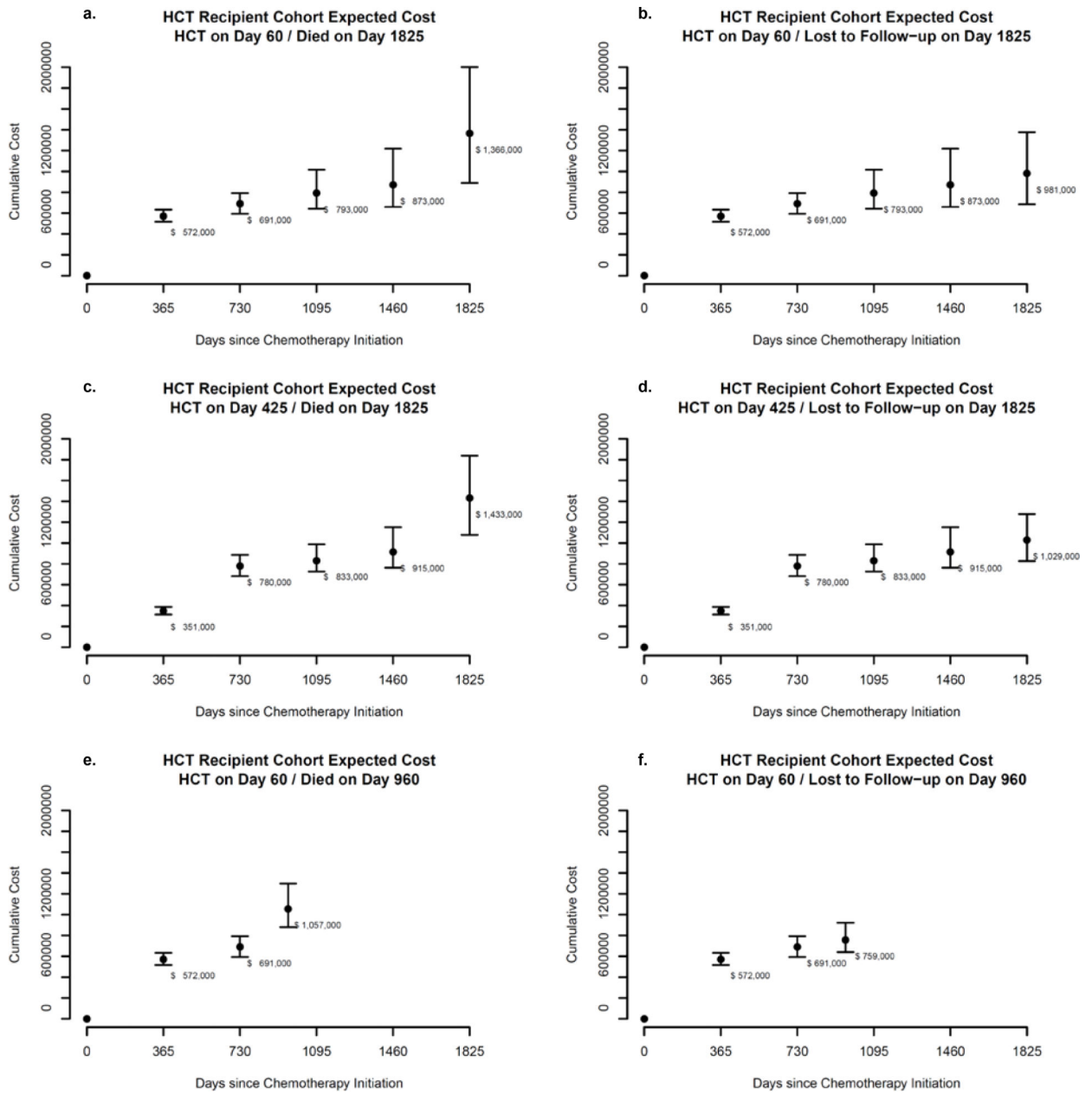
Figure 4. Expected costs since chemotherapy initiation: AlloHCT Group
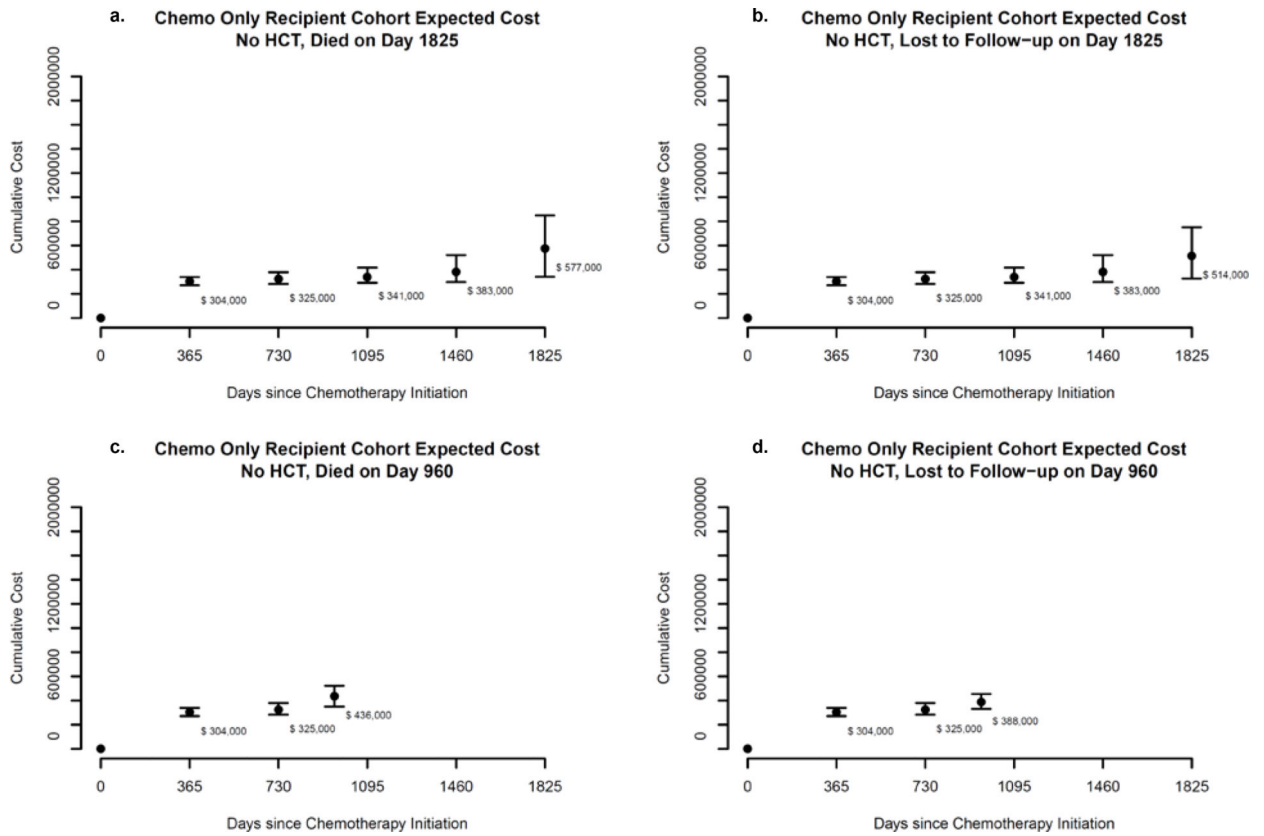
Figure 5. Expected costs since chemotherapy initiation: Chemotherapy only group