`threepm`: Estimating Censored Episode-of-Illness and Lifetime Costs in Stata

Robbie McGowan, MPH

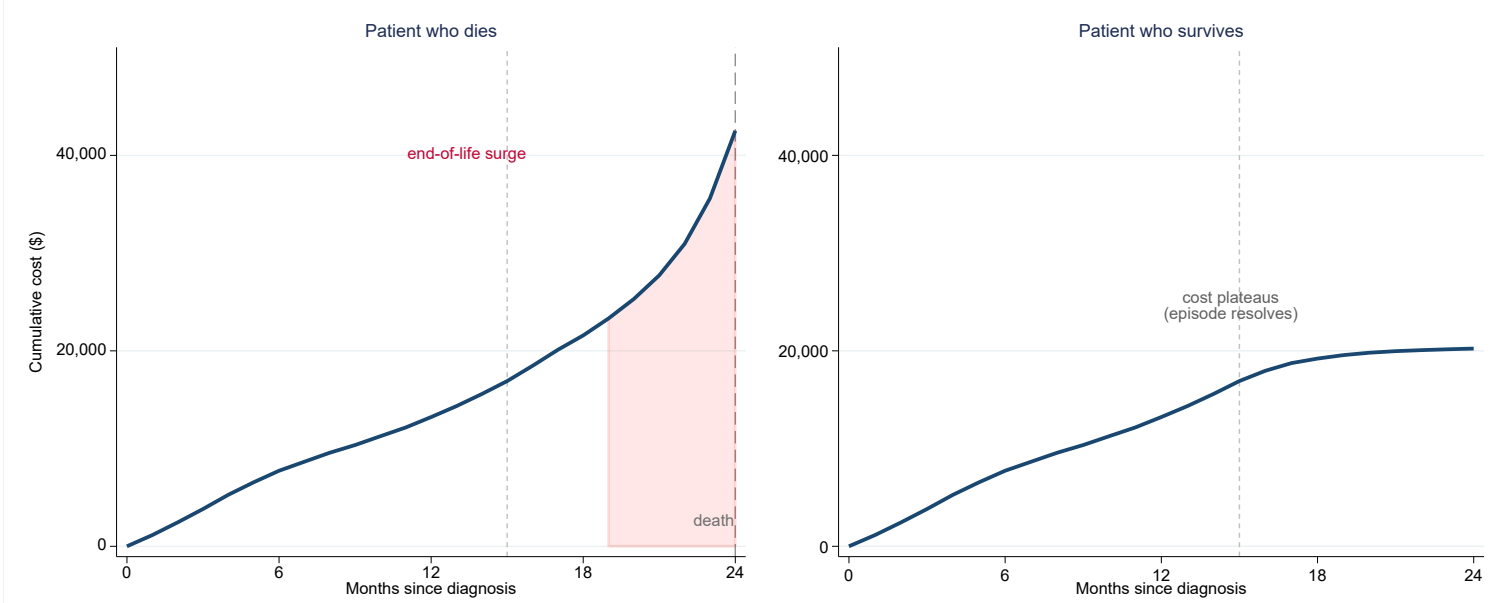


Table of Contents

- Motivation and existing methods
- The Basu–Manning three-part estimator
- threepm — the Stata command: syntax, options, output
- Example: out-of-pocket expenditures after a cancer diagnosis

What is the average cost of an illness or event?

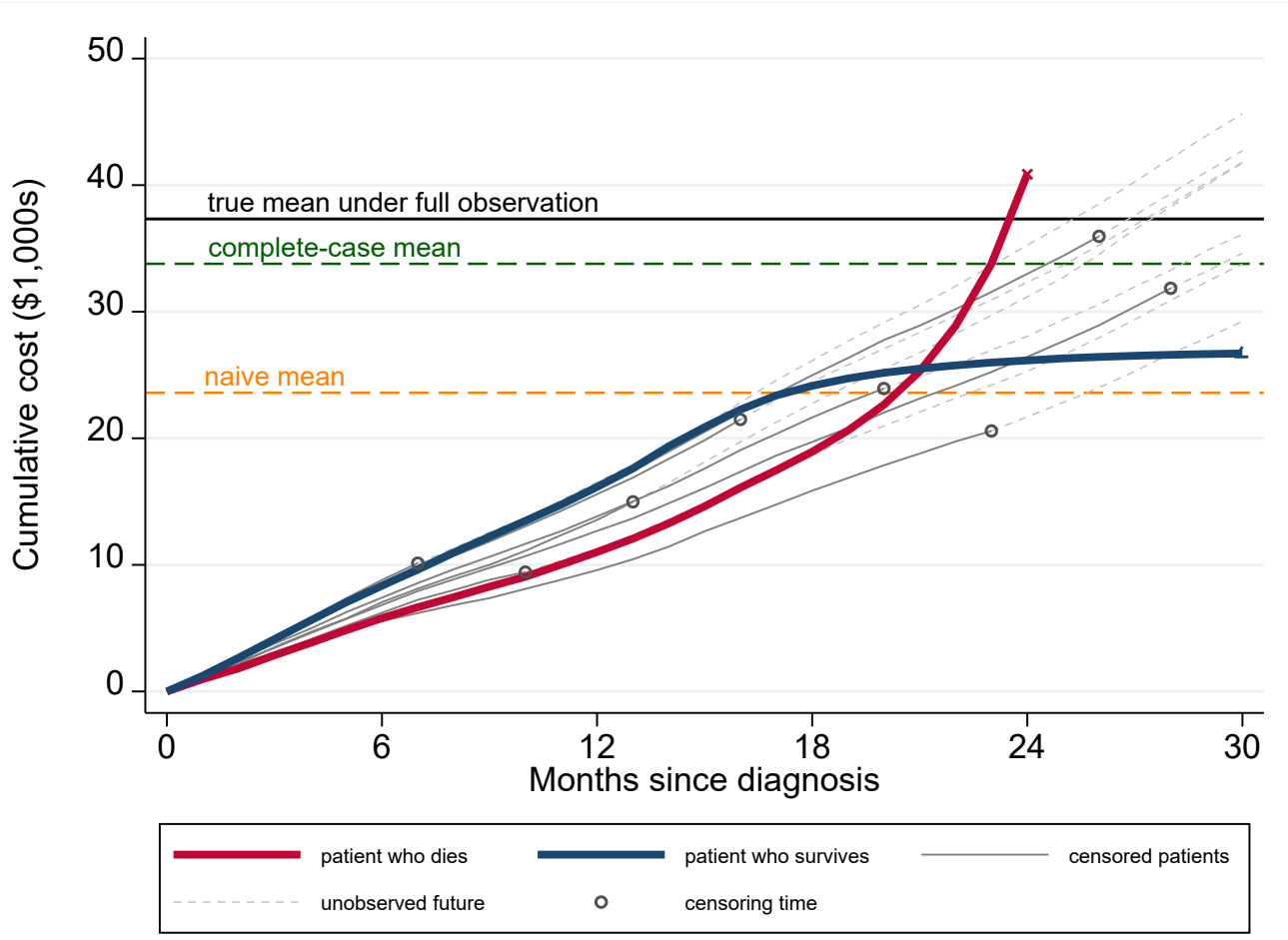
- Goal is to estimate the **mean cumulative cost** (expenditure) over an episode or lifetime from index event, like diagnosis
- Examples: lifetime cost of a cancer cohort, a Long-COVID episode, end-of-life spending
- Need an estimate for **all patients** who had the event regardless of their observation time
- Important for budgeting, cost-effectiveness evaluations, policy targeting affordability or disparities



University of Colorado
Anschutz Medical Campus

Right censoring of cost data

- **Right-censoring**: Observation can end before the end of an episode or death
- Censored persons' future costs are unobserved; downward bias
- Complete case analysis can also bias estimates if the censored observations were on a different trajectory



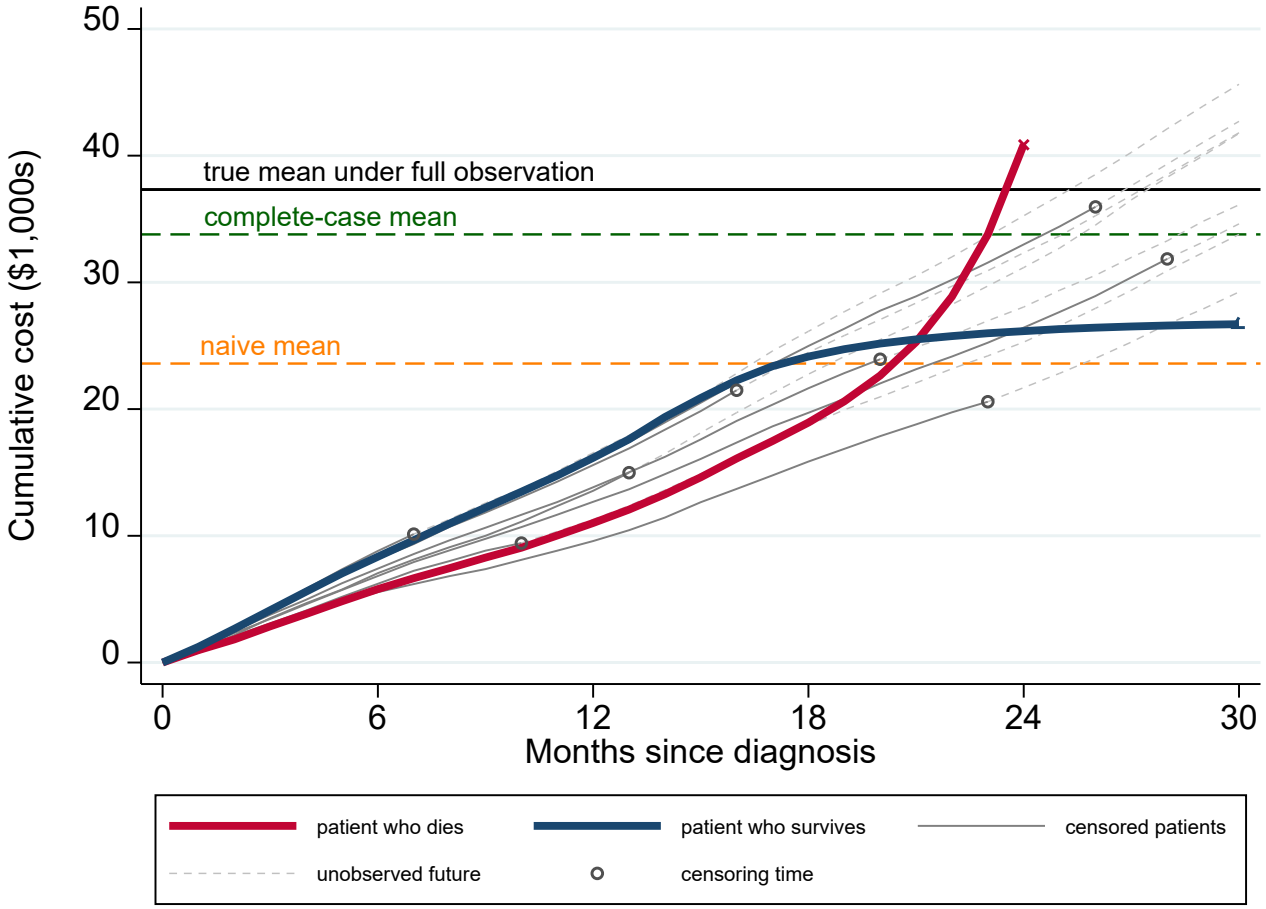
University of Colorado
Anschutz Medical Campus

Inverse probability weighting

- Bang and Tsiatis (2000) applied Inverse Probability Weighting (IPW) to censored cost, where an individual's cost in an interval is weighted by the inverse probability of being observed.
  - The "missing" data for censored individuals is "recovered" by upweighting the individuals who were observed in later intervals
    - In this case, post-death intervals are considered an "observed zero", so there are no costs to "recover"
  - Corrects censoring bias and is consistent for the mean, but treats total cost as a single reweighted quantity
    - Does not separately model cost in the death interval or within-interval death timing
  - Assumes non-informative censoring, or conditionally non-informative if censoring is modeled with covariates

The Basu and Manning Estimator

- To estimate the true mean when the unobserved costs are not known, we need to account for:
  - Costs in survived intervals
  - Costs in the interval of death
  - Survival
  - We essentially need to impute each of these for all censored person-intervals
- Basu & Manning (2010) described this method



The Basu and Manning Estimator

- For each interval, we calculate a survival probability-weighted average cost:

$$E[\,\text{cost}_j\,] \;=\; S(a_j)\,\mu_{1j} \;+\; [\,S(a_{j-1}) - S(a_j)\,]\mu_{2j}$$

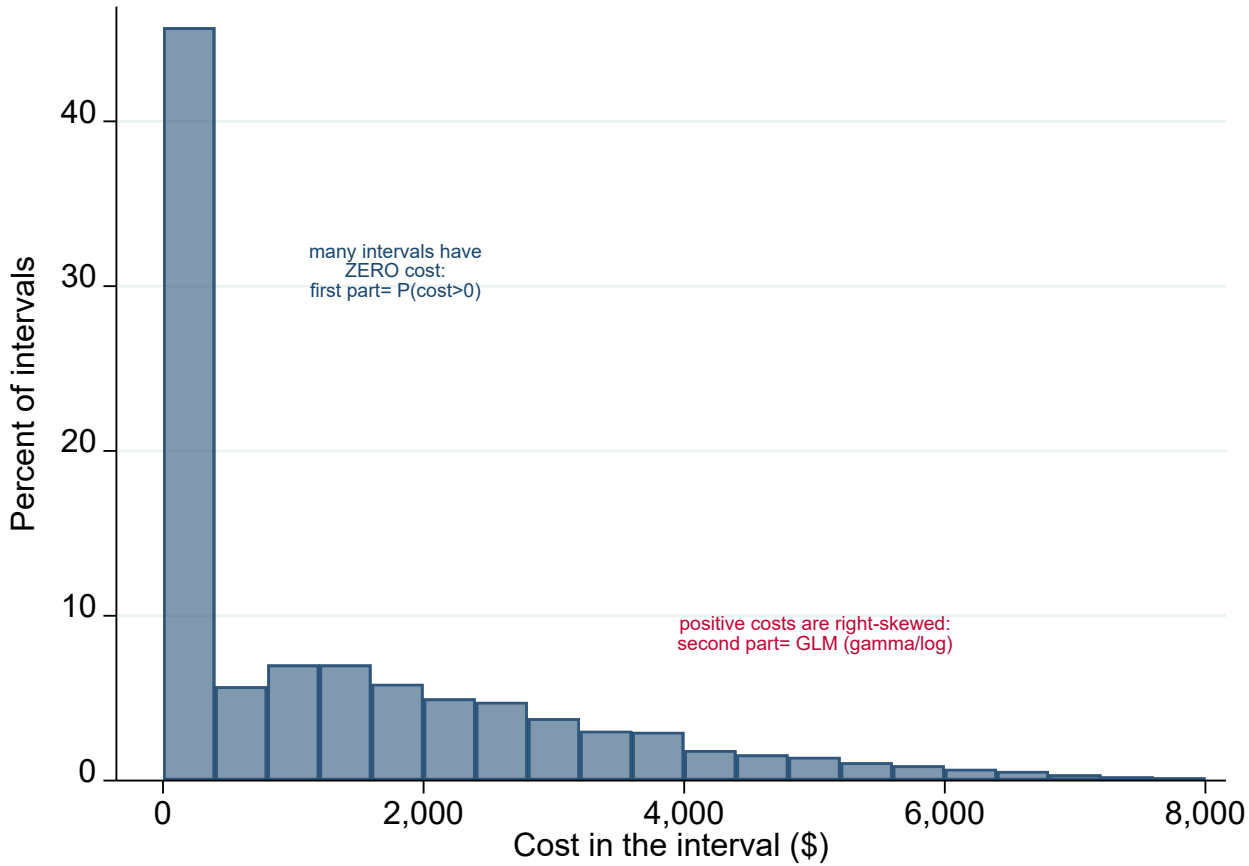
$S(a_j)$ = P(survive to end of interval j) $\rightarrow$ survived-interval cost μ_{1j}

$S(a_{j-1}) - S(a_j)$ = P(die during interval j) $\rightarrow$ death-interval cost μ_{2j}

- Summing all intervals yields cumulative mean costs for the entire episode

Parts 1 & 2 — cost models

- Part 1 — **survived**-interval cost: up to K observations per person, with person-clustered standard errors
  - All fully-observed intervals (no death or censoring) contribute
- Part 2 — **death**-interval cost: end-of-life spending, maximum 1 interval per person
  - Survivors and censored do not contribute
- Cost are generally skewed
  - GLMs with log links are typical
- If zeros are present, two-part models can be used
  - Pr(cost>0) × E[cost | cost>0]



University of Colorado
Anschutz Medical Campus

Within-interval death timing

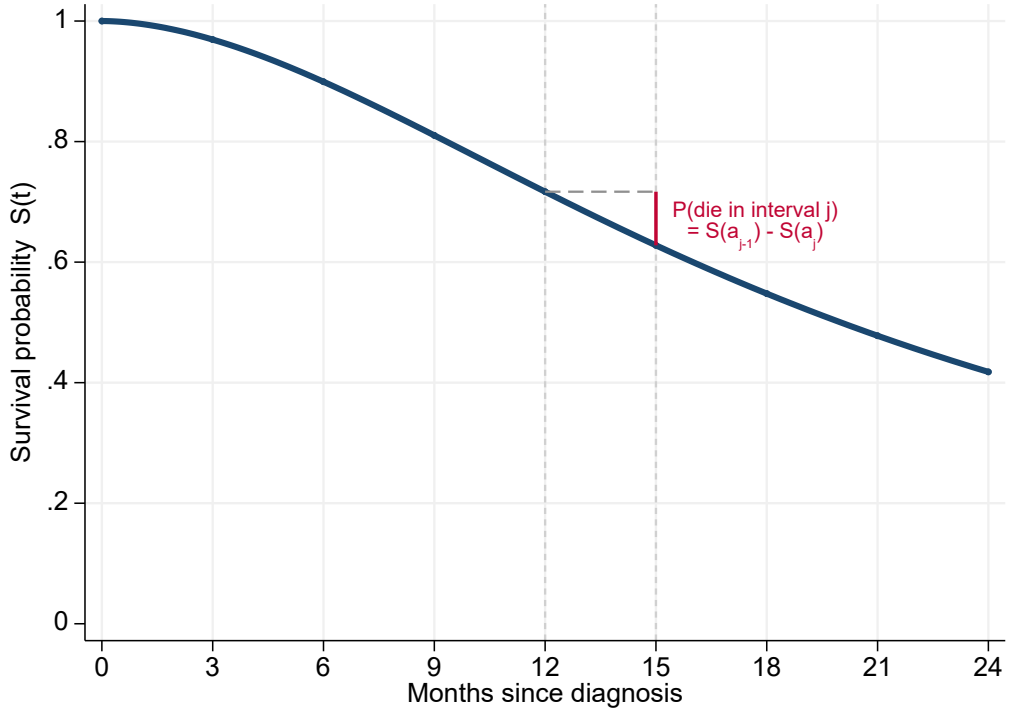
- In the interval of death, people die on different days, so we have different durations of cost accumulation
- For a person who dies on day 12 out of 30, the survival fraction is 12/30 = 0.4
- To account for “partial” observations, first sum death-interval cost over the empirical distribution of within-interval death time
- Then, compute a weighted average over the observed distribution of that fraction
  - Example: Only two death times. 40% occur on day 10; 60% on day 20

$$\hat{\mu}_{2j} \;=\; 0.40 \cdot \widehat{\text{cost}}(\text{frac} = 10/30) \;+\; 0.60 \cdot \widehat{\text{cost}}(\text{frac} = 20/30)$$

- This step is skipped if within-interval death timing is unknown because frac drops out of the model

Part 3 — Time-to-death survival model

- Accelerated failure time (AFT) model on continuous time-to-death
- Distributions: generalized gamma, Weibull, lognormal, log-logistic, exponential
- Yields S(a_j | X): probability of surviving to the end of interval j, for all person-intervals



University of Colorado
Anschutz Medical Campus

Final step: survival-weighted accumulation

- Mean cumulative cost is a sum over K of each interval's survival-weighted cost
- Average the predicted cost across the population (g-computation)

$$M(a_K) = \sum_{j=1}^{K} [\, S(a_j)\, \mu_{1j} + (S(a_{j-1}) - S(a_j))\mu_{2j} \,]$$

Decomposition: rate vs. survival

- Differences between groups are decomposed into two components: rate vs survival
- Rate — cost per period
  - Predict survival assuming everyone is in group A. Predict costs with everyone in group A and everyone in group B.
  - Accumulate under both cost scenarios, and take the difference
- Survival — difference in number of periods
  - Predict cost assuming everyone is in group A. Predict survival with everyone in group A and everyone in group B.
  - Accumulate under both cost scenarios, and take the difference
- Interpretation example: One group incurs higher cost per period, but the contrast is offset by shorter survival
- This is one advantage of the three-part model over other methods

threepm: the Stata command

- threepm is a Stata command that implements the three-part model
- Because the three parts are estimated separately, standard errors can only be obtained via bootstrap
- teffects-style syntax:
  - threepm (y survcost xvars, family() link() first() [twopart]) ///
    (deathcost xvars, ... twopart) ///
    (survcovs, distribution()), intlength() id() time() ...
- Example syntax from synthetic data based on newly-diagnosed cancer cohort in the Colorado All-Payer Claims Database:

```
local XC i.treat i.payer c.age i.time              // survived- and death-interval cost covariates
local XS i.treat i.payer c.age i.gender i.urban    // survival covariates
local OPT id(Member_Composite_ID) time(time) duration(bm_dur) intlength(30) ///
         deathtime(deathtime) centime(centime)
threepm (sum_allowed_amt `XC', first(logit) family(gamma) link(log) twopart) ///
        (`XC', first(logit) family(gamma) link(power 0.5) twopart) ///
        (`XS', distribution(weibull)), `OPT' group(treat) ///
        reps(100) seed(12345) bootopts(saving("reps_reduced.dta", replace))
```

Options, output, and companions

- Estimand: atpop(all|#) = ATE / ATT / ATC
- eol(#): end-of-life window
- Inference: reps() gives you a person-clustered bootstrap
- Output: accumulated cost by group + rate/survival decomposition
  - Full model outputs are optional, but recommended
- Companions: threepmspec (Park test, goodness-of-fit, overlap), threepmplot (plots over time with bootstrap CI bands), threepmmargins (margins / contrasts)

```
local XC i.treat i.payer c.age i.time              // survived- and death-interval cost covariates
local XS i.treat i.payer c.age i.gender i.urban    // survival covariates
local OPT id(Member_Composite_ID) time(time) duration(bm_dur) intlength(30) ///
          deathtime(deathtime) centime(centime)
threepm (sum_allowed_amt `XC', first(logit) family(gamma) link(log) twopart) ///
        (`XC', first(logit) family(gamma) link(power 0.5) twopart) ///
        (`XS', distribution(weibull)), `OPT' group(treat) ///
        reps(100) seed(12345) bootopts(saving("reps_reduced.dta", replace))
```

Synthetic data example

- We link the Colorado Central Cancer Registry to the Colorado All-Payer Claims Database to study cancer outcome disparities: out-of-pocket (OOP) cost, stage at diagnosis, treatment, and survival. Here, we focus on OOP.
- Monthly OOP cost has a large number of zeros, heavy right skew, and includes both death and disenrollment (censoring)
- Approach: the Basu–Manning three-part estimator (threepm)
- Fully synthetic data
  - 3,000 simulated patients with similar covariate distributions to the real data
- In the US, the type of insurance is a strong determinant of out-of-pocket expenditures, and more generally, access to care.

Medicare	Medicaid	Commercial

• How to qualify:
  • Age >= 65
  • ESRD
  • Qualified for SSDI
• Member out-of-pocket spending can be high due to deductibles, coinsurance, copays
	• How to qualify:
	  • Household income below threshold (varies by state)
	  • Some disabilities
	• Member out-of-pocket spending is generally lower than other types of insurance
		• How to qualify:
		  • Mostly employer-sponsored
		  • Individuals can also purchase their own plans
		• Member out-of-pocket spending varies widely due to differing plan designs

Data format

- ID = person-level identifier
- Time = interval
- **Complete** person-interval grid
- Deathtime = interval in which a person died
  - Missing if no death was observed
- Centime = interval in which a person was censored
  - Missing if never censored
- Duration = length of observation in the interval
  - Can be variable (days in a calendar month)
  - Represents the length of time (e.g. days) for which a person was alive AND observed
    - Person 2 died on day 11 of interval 3
    - Person 3 was censored on day 18 of interval
- Y = the outcome
  - Missing after censoring
  - 0 after death

id	time	y	duration	deathtime	centime
1	1	120	30	.	.
1	2	0	30	.	.
1	3	85	30	.	.
1	4	140	30	.	.
2	1	450	30	3	.
2	2	610	30	3	.
2	3	980	11	3	.
2	4	0	30	3	.
3	1	200	30	.	2
3	2	.	18	.	2
3	3	.	30	.	2
3	4	.	30	.	2

Threepmspec: pre-estimation specification tests

- APCD cancer cohort; synthetic dataset
- First two parts of the threepmspec output cover the two cost models

```
------------------------------------------------------------------------------
threepmspec: specification tests for the three Basu-Manning component models
(run before threepm; results inform family()/link()/first()/distribution())
------------------------------------------------------------------------------

--- Part 1: survived-interval cost ---
   intervals =    28,891   positive =    18,340   zeros =    10,551 (36.5%)
   (family & link are assessed on the positive-cost intervals; the zeros are modeled separately by the two-part logit)
   modified Park test: lambda =  1.99   (Var ~ mean^lambda; nearest GLM family: gamma)
   log-scale residual kurtosis =  5.28  (>3: heavy-tailed; gamma GLM less precise, Park test noisier)
   log-scale heteroscedasticity (Koenker BP): p = 0.291  (not detected at 5%)
   Pregibon link test (log): xb^2 p = 0.519  (log link not rejected at 5%)
   link comparison (family gamma; lower = better fit):
     link                AIC             BIC
     log                 244,324         244,480
     power 0.5           244,758         244,914
     power 0.333         244,528         244,685
     identity            245,597         245,754
   lowest AIC & BIC link (of those compared): log
   Hosmer-Lemeshow (decile mean residual = 0): F p = 0.939  (no decile lack-of-fit at 5%)
   Pearson corr(raw residual, predictor) =  0.008  (p = 0.419; no slope bias)
   zeros present: give twopart in this threepm cost equation (required)

--- Part 2: death-interval cost ---
   intervals =       729   positive =       521   zeros =       208 (28.5%)
   (family & link are assessed on the positive-cost intervals; the zeros are modeled separately by the two-part logit)
   modified Park test: lambda =  2.04   (Var ~ mean^lambda; nearest GLM family: gamma)
   log-scale residual kurtosis =  4.92  (>3: heavy-tailed; gamma GLM less precise, Park test noisier)
   log-scale heteroscedasticity (Koenker BP): p = 0.993  (not detected at 5%)
   Pregibon link test (log): xb^2 p = 0.313  (log link not rejected at 5%)
   link comparison (family gamma; lower = better fit):
     link                AIC             BIC
     log                   8,026           8,048
     power 0.5             8,019           8,040
     power 0.333           8,021           8,043
     identity              8,016           8,037
   lowest AIC & BIC link (of those compared): identity
   Hosmer-Lemeshow (decile mean residual = 0): F p = 0.078  (no decile lack-of-fit at 5%)
   Pearson corr(raw residual, predictor) = -0.005  (p = 0.885; no slope bias)
   zeros present: give twopart in this threepm cost equation (required)
```

University of Colorado
Anschutz Medical Campus

Threepmspec: pre-estimation specification tests

- APCD cancer cohort; synthetic dataset
- The third part of the threepmspec output

```
--- Part 3: survival (time to death), AFT distribution selection ---
   persons =     3,000   deaths =       729

   distribution            AIC            BIC
   ------------------------------------------------------------
   exponential          3797.9         3845.9
   weibull              3661.7         3715.8
   loglogistic          3639.9         3694.0
   lnormal              3669.7         3723.8
   ggamma               3646.3         3706.4
   ------------------------------------------------------------
   lowest AIC & BIC: distribution(loglogistic)

   generalized-gamma kappa =  0.53 [95% CI  0.34,  0.73]

Values above correspond to threepm's per-equation family()/link()/first()
sub-options and the survival equation's distribution().
```

Threepm component models

- APCD cancer cohort synthetic dataset
- No bootstrap standard errors and CIs
- Model output for component 1: survived-interval costs
  - We used a two-part model specification here, so the model on the right is the first part: a logistic regression estimate the probability of any costs
- We have standard errors for the model coefficients, but not the accumulated survival-weighted costs

```
. threepm $EQS, intlength(30) id(pid) time(time) duration(duration) ///
>     deathtime(deathtime) centime(centime) ///
>     group(payer_dx) display(models)
note: the death-interval cost equation contains no interval-time terms.

Component model 1 of 3: cost in intervals survived in full, gamma GLM (two-part)

Logistic regression                             Number of obs =  28,891
                                                Wald chi2(19) = 4864.13
                                                Prob > chi2   =  0.0000
Log pseudolikelihood = -16008.665               Pseudo R2     =  0.1558

                                (Std. err. adjusted for 2,952 clusters in pid)
```

_tp_pos	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
payer_dx						
Medicare FFS	1.752894	.0358207	48.94	0.000	1.682687	1.823102
Private	2.106694	.0353099	59.66	0.000	2.037488	2.1759
cancer_site						
Colorectal	.0812651	.0378292	2.15	0.032	.0071213	.1554089
Lung	.2931379	.0425041	6.90	0.000	.2098315	.3764443
Prostate	.0389357	.0557424	0.70	0.485	-.0703174	.1481888
agec	.0010703	.0012089	0.89	0.376	-.0012991	.0034398
1.female	.0319022	.0389174	0.82	0.412	-.0443745	.1081789
late_stage						
Late	.7457314	.036086	20.67	0.000	.6750042	.8164586
time						
2	-.3781432	.0724749	-5.22	0.000	-.5201914	-.236095
3	-.8262322	.0691255	-11.95	0.000	-.9617156	-.6907487
4	-1.209216	.0677054	-17.86	0.000	-1.341916	-1.076516
5	-1.261181	.0687964	-18.33	0.000	-1.39602	-1.126343
6	-1.339047	.069343	-19.31	0.000	-1.474957	-1.203137
7	-1.33382	.0712882	-18.71	0.000	-1.473542	-1.194098
8	-1.474986	.0695845	-21.20	0.000	-1.611369	-1.338603
9	-1.468658	.069725	-21.06	0.000	-1.605317	-1.332
10	-1.534112	.0713744	-21.49	0.000	-1.674003	-1.394221
11	-1.647686	.0714953	-23.05	0.000	-1.787814	-1.507558
12	-1.658211	.0732078	-22.65	0.000	-1.801696	-1.514727
_cons	-.013928	.0700632	-0.20	0.842	-.1512493	.1233933

Threepm component models

- APCD cancer cohort synthetic dataset
- No bootstrap standard errors and CIs
  - Model outputs report the standard cluster-robust analytic SEs
- Model output for component 1: survived-interval costs
  - We used a two-part model specification here, so the model on the right is the second part: a GLM with log link and gamma family
- We have standard errors for the model coefficients, but not the accumulated survival-weighted costs

```
Generalized linear models                         Number of obs   =     18,340
Optimization     : ML                             Residual df     =     18,320
                                                  Scale parameter =   1.280082
Deviance         =  27549.97448                   (1/df) Deviance =    1.50382
Pearson          =  23451.10694                   (1/df) Pearson  =   1.280082

Variance function: V(u) = u^2                     [Gamma]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =   13.32191
Log pseudolikelihood = -122141.9105               BIC             =  -152294.5
```

(Std. err. adjusted for **2,911** clusters in **pid**)

oop_mon	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
payer_dx						
Medicare FFS	1.413421	.0298594	47.34	0.000	1.354898	1.471945
Private	2.013594	.0287814	69.96	0.000	1.957183	2.070005
cancer_site						
Colorectal	-.0393133	.0246891	-1.59	0.111	-.087703	.0090763
Lung	.2593653	.0264735	9.80	0.000	.2074782	.3112524
Prostate	-.0029591	.0357193	-0.08	0.934	-.0729677	.0670494
agec	-.0007283	.0007882	-0.92	0.355	-.0022731	.0008164
1.female	-.0020031	.0242483	-0.08	0.934	-.0495289	.0455226
late_stage						
Late	.4968699	.0194888	25.50	0.000	.4586727	.5350672
time						
2	-.2284609	.0330035	-6.92	0.000	-.2931466	-.1637752
3	-.4799144	.0345396	-13.89	0.000	-.5476108	-.412218
4	-.7836281	.0354391	-22.11	0.000	-.8530876	-.7141687
5	-.8215428	.0369331	-22.24	0.000	-.8939305	-.7491552
6	-.7385421	.0365909	-20.18	0.000	-.810259	-.6668253
7	-.8217387	.0394464	-20.83	0.000	-.8990523	-.7444251
8	-.8151232	.0398097	-20.48	0.000	-.8931488	-.7370975
9	-.8208956	.0381264	-21.53	0.000	-.895622	-.7461692
10	-.7611055	.042797	-17.78	0.000	-.844986	-.6772249
11	-.7501435	.0422785	-17.74	0.000	-.8330079	-.6672791
12	-.8109065	.0427811	-18.95	0.000	-.894756	-.7270571
_cons	4.502713	.042996	104.72	0.000	4.418443	4.586984

University of Colorado
Anschutz Medical Campus

Threepm output: summary and group contrasts

- APCD cancer cohort synthetic dataset
- No bootstrap standard errors and CIs for the **cumulative costs**
- Threepm prints a summary of the data and model specifications
- It also includes group means and contrasts for the specified group() variable
- Contrasts are decomposed into rate vs survival, which sum to the total

```
In-sample calibration (predicted / observed total cost) = 0.9967
   sum of fitted =      7,094,346   sum observed =      7,117,499
   (near 1 by design; exact equality is not expected with non-canonical gamma links)

------------------------------------------------------------------------------
Basu-Manning three-part censored-cost estimator
------------------------------------------------------------------------------
Person-intervals       =     36,000        Persons           =      3,000
Intervals (K)          =         12        Interval length   = 30 (constant)
Death intervals        =        729        Censored persons  =        400
Fully-obs. intervals   =     28,891        (alive & uncensored)
Mean intervals at risk =       9.82        (until death/censoring/end)
Mean raw cost, death interval  =     700.34
Mean raw cost, full interval   =     228.69
U integration          = 20-point quantile grid of U/L (729 distinct values observed) [fast path]
Survival model         = loglogistic AFT
Death-int. cost        = gamma / log link (two-part, logit first part)
Survived cost          = gamma / log link (two-part, logit first part)
------------------------------------------------------------------------------

Mean cumulative cost through interval 12:
   payer_dx = 1                       253.06
   payer_dx = 2                      2101.21
   payer_dx = 3                      4156.98

Incremental effects (level b vs level a):
   contrast          total             rate             survival
   2v1              1848.15           1783.26              64.89
   3v1              3903.92           3764.45             139.48
   3v2              2055.77           2040.28              15.49

rate:  survival held at level a.  survival:  cost accumulation rate held at level b.  total = rate + survival.

Standard errors require a clustered bootstrap: add reps(500) to the threepm command (built-in, person-clustered)
idcluster(nid) reps(500): threepm ..., id(nid) ...
------------------------------------------------------------------------------
```

University of Colorado
Anschutz Medical Campus

Threepmmargins output, with bootstrap

- APCD cancer cohort synthetic dataset
- 200-rep bootstrap, reporting 95% percentile intervals
- Syntax based on Stata's native 'margins'
  - Accepts dydx(), at(), and interactions
  - Does not accept over(), subpop()
    - Can achieve same result with if() conditions
  - Does not accept vce() because it uses bootstrap SEs
  - Does not accept expression() or predict() because accumulated mean cost is the only estimand
- Threepmmargins also works without bootstrap, but the "[95% bootstrap pctile interval]" columns will be blank

```
. threepmmargins cancer_site using demo_reps
recomputing the margins in each of 200 bootstrap replicates...
    (coefficient bootstrap: each replicate's component coefficients are re-evalu
    and the standardization population fixed)
    50 of 200 replicates done
    100 of 200 replicates done
    150 of 200 replicates done
    200 of 200 replicates done

Predictive margins                                  Number of persons =     3,000
Expression: mean cumulative cost through interval K=12
CI: bootstrap percentile, 200 replicates, recomputed in demo_reps.dta
```

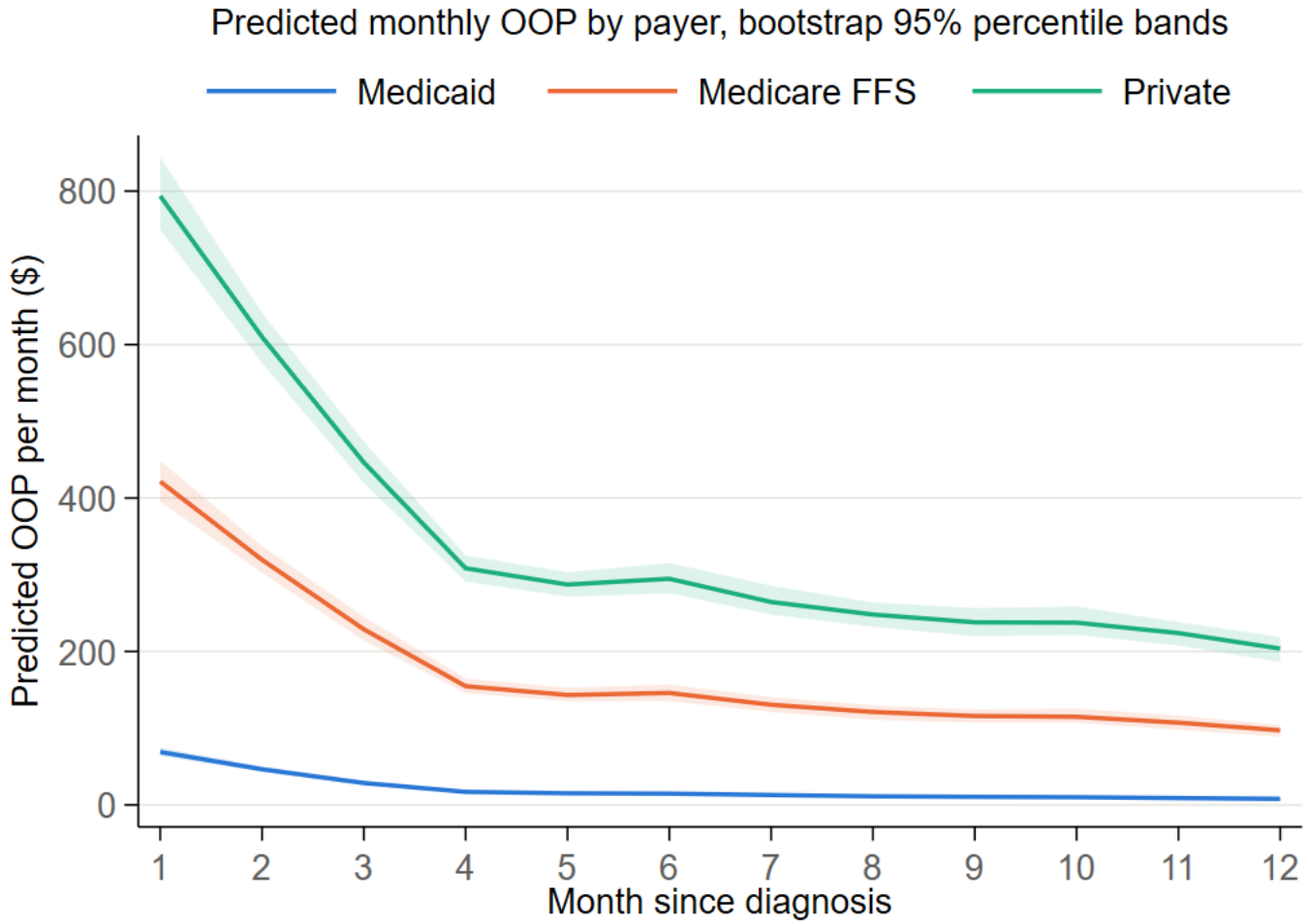
	Predictive Margin	[95% bootstrap pctile interval]	
cancer_site			
Breast	2367.66	2278.13	2437.34
Colorectal	2320.22	2237.89	2400.44
Lung	3002.48	2869.62	3122.20
Prostate	2371.44	2251.62	2520.01

Rate vs survival decomposition

cancer_site: 2 vs 1			
Total	-47.45	-168.74	54.72
Rate	-46.52	-173.79	54.77
Survival	-0.93	-19.83	16.66
cancer_site: 3 vs 1			
Total	634.82	491.08	780.64
Rate	839.61	696.60	974.67
Survival	-204.79	-252.25	-163.42
cancer_site: 4 vs 1			
Total	3.78	-156.80	162.09
Rate	13.05	-148.62	174.57
Survival	-9.27	-32.50	14.57

Threepmplot

- Built-in plotting command that uses stored bootstrap resamples to plot the outcome by interval
- With bootstrap results, the plotted line is accompanied by a shaded 95% bootstrap percentile band
- Options:
  - block() combines intervals (quarters, years, etc)
  - ‘effect’ plots the contrast between groups



Summary, availability, references

- threepm: a general Stata command for censored, covariate-adjusted episode/lifetime cost
- Extends the IPW method to account for elevated death costs and continuous distribution of death times
- Availability: version 1.0 is on GitHub: https://github.com/rmcg1992/threepm
- References:

  - Bang, H., & Tsiatis, A. (2000). Estimating medical costs with censored data. Biometrika, 87(2), 329–343. https://doi.org/10.1093/biomet/87.2.329
  - Basu A & Manning Wg. (2010). Estimating lifetime or episode-of-illness costs under censoring. Health Economics, 19(9). https://doi.org/10.1002/hec.1640

Acknowledgments

- Anirban Basu, PhD
  - Shared the original code implementing the method
- Marcelo Perraillon, PhD and Rick DeVoss, PhD
  - Adapted/rewrote the code
- Support from National Cancer Institute grant R01CA288259

Thank You!



Anschutz