5/5 Need to add some detail on models to clarify clustering component (e.g., time? hospital?).

Intermediate Progress Report

We can walk through this next week.

Asthma Treatment Pattern Longitudinal Analysis

Chong Kim

supervised by

Dr. Matthew Strand

November 9, 2016

The Department of Clinical Pharmacy (DOCP) has provided funding for this educational project but has not conducted the research or written this report. While the authors have worked on the best information available to them, neither DOBB/DOCP nor the authors shall in any event be liable for any loss, damage or injury howsoever suffered directly or indirectly in relation to the report or the research on which it is based.

Reference herein to trade names and proprietary products without stating that they are protected does not imply that they may be regarded as unprotected and thus free for general use. No endorsement of named products is intended nor is it any criticism implied of other alternative, but unnamed, products.

Contents

List of Figures

List of Tables

A Data

A.1 Description of Data

There are currently 3 data sets.

- Patient characteristic file
- Exposure file
- Outcome file

Study period: January 2006 Ðecember 2013

Population: All patients with asthma within most recent 10% random sample database housed by CePOR

Patient entry criteria:

1. evidence of asthma at least two recorded diagnosis for asthma (icd 9:493.xx) within 1 year.
2. having at least 24 months continuous eligibility after index date (follow-up period) and at least 6 months continuous eligibility before index date (pre period) environment.
3. Age 6 64 at index.

Patient exclusion criteria:

1. diagnosed with cystic fibrosis (ICD 9; 277.0x) or chronic obstructive pulmonary diseases (491.xx, 492.xx, 494.xx and 496.xx) or respiratory tract cancer (160.xx 164.xx or 231.xx) or bronchopulmonary dysplasia (770.7x) or respiratory distress syndrome (769.xx).
2. one of the following diagnoses: Addison disease (255.4x), glomerulonephritis (580.xx to 582.xx), multiple sclerosis (340.xx), polymyositis/dermatomyositis (710.3x, 710.4x), rheumatoid arthritis (714.xx), scleroderma (710.1x), Sjogren disease (710.2x), systemic lupus erythematosus (710.0x), uveitis (360.11, 363.20 364.3x), vitiligo (709.01), Wegener granulomatosis (446.4x), Primary systemic vasculitis (447.6x), Crohn?s disease (555.0x to 555.2x,

555.9x), Ulcerative colitis (556.0x to 556.6x, 556.8x, or 556.9x), Chronic eosinophilic pneumonia (518.3x), Idiopathic pulmonary fibrosis (516.3x or 515.xx), minimal change disease (581.3x), autoimmune hepatitis (571.42), Myasthenia Gravis (358.0x), Muscular dystrophy (359.0x, 359.1x, or 359.21), Still?s disease (714.2x), Churg Strauss syndrome (446.4x), Polymyalgia rheumatica (725.xx)

Index date: The first date of asthma diagnosis with 6 months eligibility before the index date (NOTE: asthma diagnosis can occur prior to the index date, however, these diagnoses would not have 6 months eligibility prior to the dates.)

A.2 Variables

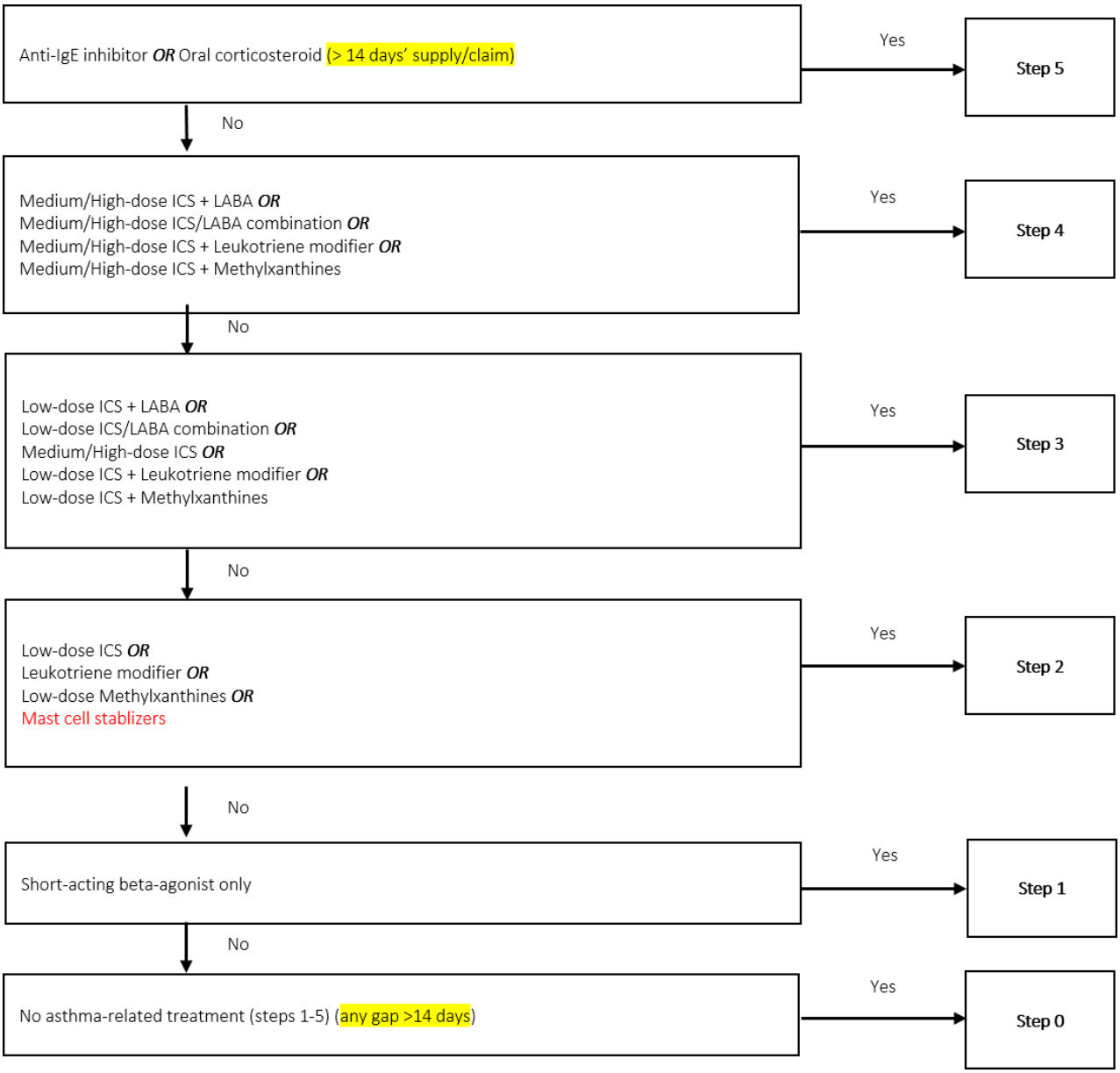
Outcomes: The primary outcome of interest is a nominal variable that includes:

1. No event.
2. Asthma-related hospitalization.
3. Asthma-related emergency department visit.
4. Asthma-related outpatient exacerbation.
5. Outpatient visit for lower respiratory tract infection treated with antibiotics.

(See codes and definition in excel) For patients who have at least one event, we would like to have data on type of event in numeric format and number of days after the index date in numeric format. For those patients with at least one event, only the event category that are observed need to be included i.e. patients who have hospitalization and outpatient exacerbation would only have event 1 and 3 included in the outcome file. For patient who do NOT have any events during the follow-up time, we would like to include them and have event = 0 and date = missing.

Exposure Variable: The primary exposure variable is treatment steps which is an ordinal variable characterized as:

Figure 1


Anti-IgE inhibitor OR Oral corticosteroid (> 14 days' supply/claim)
Yes
Step 5
No
Medium/High-dose ICS + LABA OR
Medium/High-dose ICS/LABA combination OR
Medium/High-dose ICS + Leukotriene modifier OR
Medium/High-dose ICS + Methylxanthines
Yes
Step 4
No
Low-dose ICS + LABA OR
Low-dose ICS/LABA combination OR
Medium/High-dose ICS OR
Low-dose ICS + Leukotriene modifier OR
Low-dose ICS + Methylxanthines
Yes
Step 3
No
Low-dose ICS OR
Leukotriene modifier OR
Low-dose Methylxanthines OR
Mast cell stablizers
Yes
Step 2
No
Short-acting beta-agonist only
Yes
Step 1
No
No asthma-related treatment (steps 1-5) (any gap >14 days)
Yes
Step 0

B Hypotheses

Hypothesis 1: We hypothesize that the likelihood of asthma treatment step will be different for different patient characteristics.

Hypothesis 1a: *The likelihood of asthma treatment step (ordinal variable) will be different for different patient characteristics.*

Hypothesis 2: We hypothesize that the likelihood of asthma exacerbation event will be different for different treatment patterns.

Hypothesis 2a: *The likelihood of asthma exacerbation event (yes/no) will be different for different treatment patterns.*

Hypothesis 2b: *The likelihood of asthma exacerbation event (nominal variable) will be different for different treatment patterns.*

C Statistical Methods

C.1 Hypothesis 1

Various treatment step patterns will be associated with patient characteristics as described in the covariates section using multinomial regression methods that account for the longitudinal (correlated) data structure of the treatment step patterns. Specifically, the regression model will be built using a generalized linear mixed model with a cumulative logit link and with multinomial distribution. There will be minimal model selection as we are going to include all of the covariates mentioned in the covariates section.

C.1.1 Model 1: Model with All Covariates

Model with all covariates(baseline age, gender, chronic comorbidity index, time, baseline cost, and treatment step). Results should indicate the relationship between the baseline characteristics of subjects and the outcome, asthma treatment step. Repeated measure $\mathbf{R_i}$ side random effects

What is clustering component and where is it in the model? I do not readily see it.

are not supported for multinomial distribution.

$$logit(P(Y_i < j)) = \theta_j - \beta_1 baseage_i - \beta_2 cci_i - \beta_3 time_{il} - \beta_4 basecost_i - gender_k \quad (1)$$
$$i = 1, .., n; j = 1, ..., J-1 \ where \ J = 6 \ steps; k = 1, 2; l = 1, ..., r_i;$$
$$\theta_j = Threshold \ variable;$$

Make sure to define indices on all models.

C.1.2 Model 2: Polynomial Trend

Predictor variables(baseline age, gender, chronic comorbidity index, time, baseline cost). Results should indicate the relationship between the baseline characteristics of subjects and the outcome, asthma treatment step.

$$logit(P(Y_i < j)) = \theta_j - \beta_1 baseage_i - \beta_2 cci_i - \beta_3 time_{il} - \beta_4 basecost_i - gender_k \quad (2)$$
$$i = 1, .., n; j = 1, ..., J-1 \ where \ J = 6 \ steps; k = 1, 2; l = 1, ..., r_i;$$
$$\theta_j = Threshold \ variable;$$

C.2 Hypothesis 2a

Asthma exacerbation events will be associated primarily with the asthma treatment steps along with additional covariates of interest using multinomial regression methods that account for the longitudinal (correlated) data structure of the asthma exacerbation event (yes/no). Specifically, the regression model will be built using a generalized linear mixed model with a logit link and with binomial distribution.

C.2.1 Model 1:Simple Model with all covariates

Model with all covariates(baseline age, gender, chronic comorbidity index, time, baseline cost, and treatment step). Results should indicate the relationship between the baseline characteristics of subjects and the outcome, asthma exacerbation events. Repeated measure $\mathbf{R_i}$ side random

effects are not supported for multinomial distribution.

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 baseage_i + \beta_2 cci_i + \beta_3 basecost_i + \beta_4 time_{ij} + gender_k \quad (3)$$
$$\beta_5 TS_{ij} + \beta_6 TS_{ij} * time_{ij} \qquad i = 1, .., n; k = 1, 2; j = 1, ..., r_i;$$

C.2.2 Model 2: Modeling polynomial trend for time

Model 2 + Polynomial variables for time. Results should indicate the relationship between the baseline characteristics of subjects and the outcome, asthma treatment step.

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 baseage_i + \beta_2 cci_i + \beta_3 basecost_i + \beta_4 time_{il} + \beta_5 time_{il}^2 + \quad (4)$$
$$gender_k + \beta_5 TS_{ij} + \beta_6 TS_{ij} * time_{ij}$$
$$i = 1, .., n; k = 1, 2; j = 1, ..., r_i;$$

C.3 Hypothesis 2b

Asthma exacerbation events will be associated primarily with the asthma treatment steps along with additional covariates of interest using multinomial regression methods that account for the longitudinal (correlated) data structure of the asthma exacerbation event (yes/no). Specifically, the regression model will be built using a generalized linear mixed model with a generalized logit link and with multinomial distribution.

C.3.1 Model 1: Model with all covariates

Predictor variables(baseline age, gender, chronic comorbidity index, time, baseline cost, and treatment step). Results should indicate the relationship between the baseline characteristics of subjects and the outcome, asthma treatment step.

$$\log(\frac{p_{ijc}}{p_{ij1}}) = \beta_0 + \beta_1 baseage_i + \beta_2 cci_i + \beta_3 basecost_i + \beta_4 time_{ij} + gender_k \quad (5)$$
$$\beta_5 TS_{ij} + \beta_6 TS_{ij} * time_{ij} \ for \ c = 2, 3, .., 5 \ i = `, .., n; k = 1, 2; j = 1, ..., r_i;$$

C.3.2 Model 2: Modeling polynomial trend for time

Model 1 + Polynomial variables for time. Results should indicate the relationship between the baseline characteristics of subjects and the outcome, asthma treatment step.

$$\log(\frac{p_{ijc}}{p_{ij1}}) = \beta_0 + \beta_1 baseage_i + \beta_2 cci_i + \beta_3 basecost_i + \beta_4 time_{il} + \beta_5 time_{il}^2 + \tag{6}$$

$$gender_k + \beta_5 TS_{ij} + \beta_6 TS_{ij} * time_{ij}$$

$$for\ c = 2, 3, .., 5;\ i = `, .., n; k = 1, 2; l = 1, ..., r_i;)$$

D Initial Descriptives

	Baseline Treatment Steps							
	0	1	2	3	4	5	p	test
n	781	112	39	38	26	4		
age (mean (sd))	31.42 (17.95)	28.00 (18.32)	22.26 (16.32)	29.89 (17.87)	33.19 (18.94)	42.00 (19.13)	0.013	
pat_region (%)							0.206	
E	219 (28.0)	25 (22.3)	6 (15.4)	10 (26.3)	9 (34.6)	2 (50.0)		
MW	254 (32.5)	40 (35.7)	13 (33.3)	15 (39.5)	9 (34.6)	1 (25.0)		
S	229 (29.3)	28 (25.0)	15 (38.5)	5 (13.2)	5 (19.2)	1 (25.0)		
W	79 (10.1)	19 (17.0)	5 (12.8)	8 (21.1)	3 (11.5)	0 (0.0)		
Insurance (%)							0.191	
1	665 (85.1)	90 (80.4)	34 (87.2)	29 (76.3)	24 (92.3)	3 (75.0)		
2	3 (0.4)	1 (0.9)	0 (0.0)	1 (2.6)	1 (3.8)	0 (0.0)		
3	20 (2.6)	8 (7.1)	2 (5.1)	1 (2.6)	0 (0.0)	1 (25.0)		
4	3 (0.4)	1 (0.9)	0 (0.0)	1 (2.6)	0 (0.0)	0 (0.0)		
5	86 (11.0)	11 (9.8)	3 (7.7)	6 (15.8)	1 (3.8)	0 (0.0)		
6	4 (0.5)	1 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
baseline_cost (mean (sd))	2361.08 (7009.85)	1506.44 (2521.98)	1142.77 (1332.55)	886.63 (1378.18)	1734.34 (3036.80)	1940.89 (504.53)	0.455	
CCI (mean (sd))	0.30 (0.62)	0.18 (0.71)	0.18 (0.56)	0.05 (0.23)	0.19 (0.49)	0.25 (0.50)	0.074	

E Graphical Visualization

Figure 2: Correlation Matrix of Continuous Variables

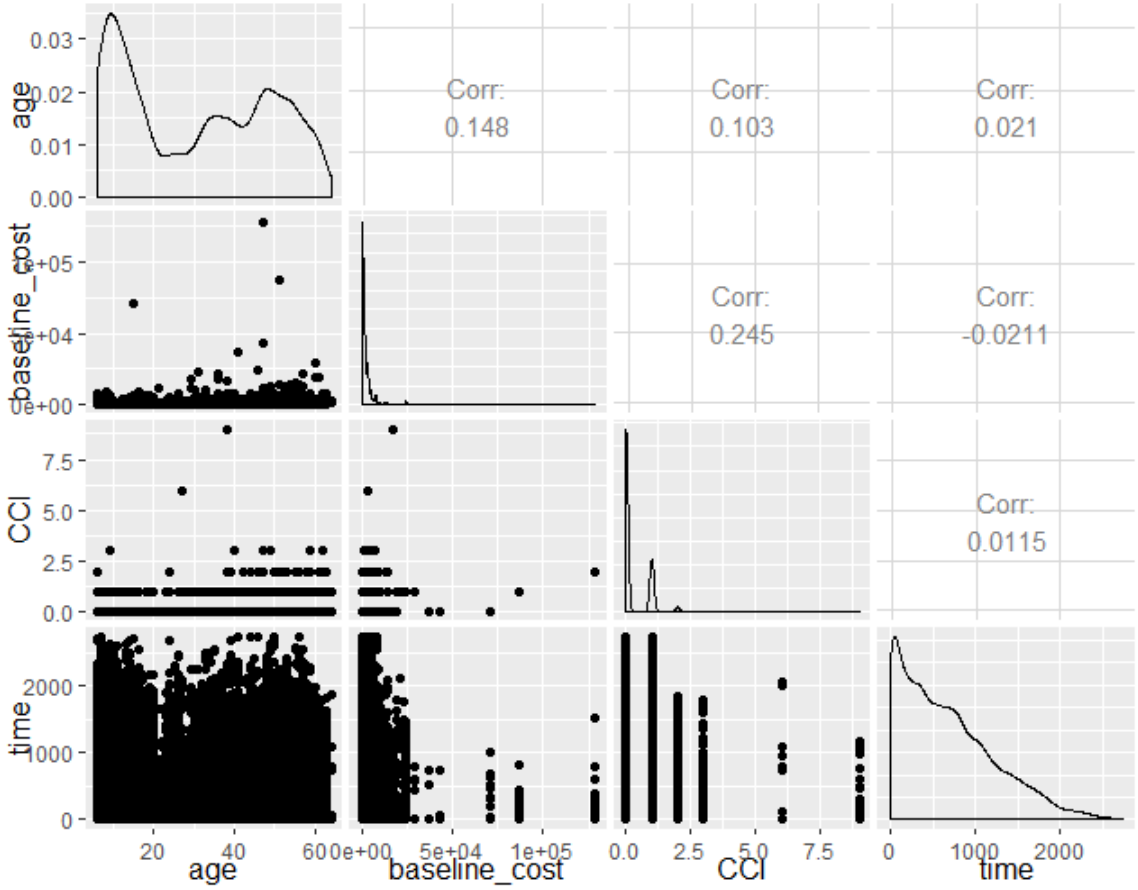


Figure 3: Histogram of Continuous Variables by Event as Binary

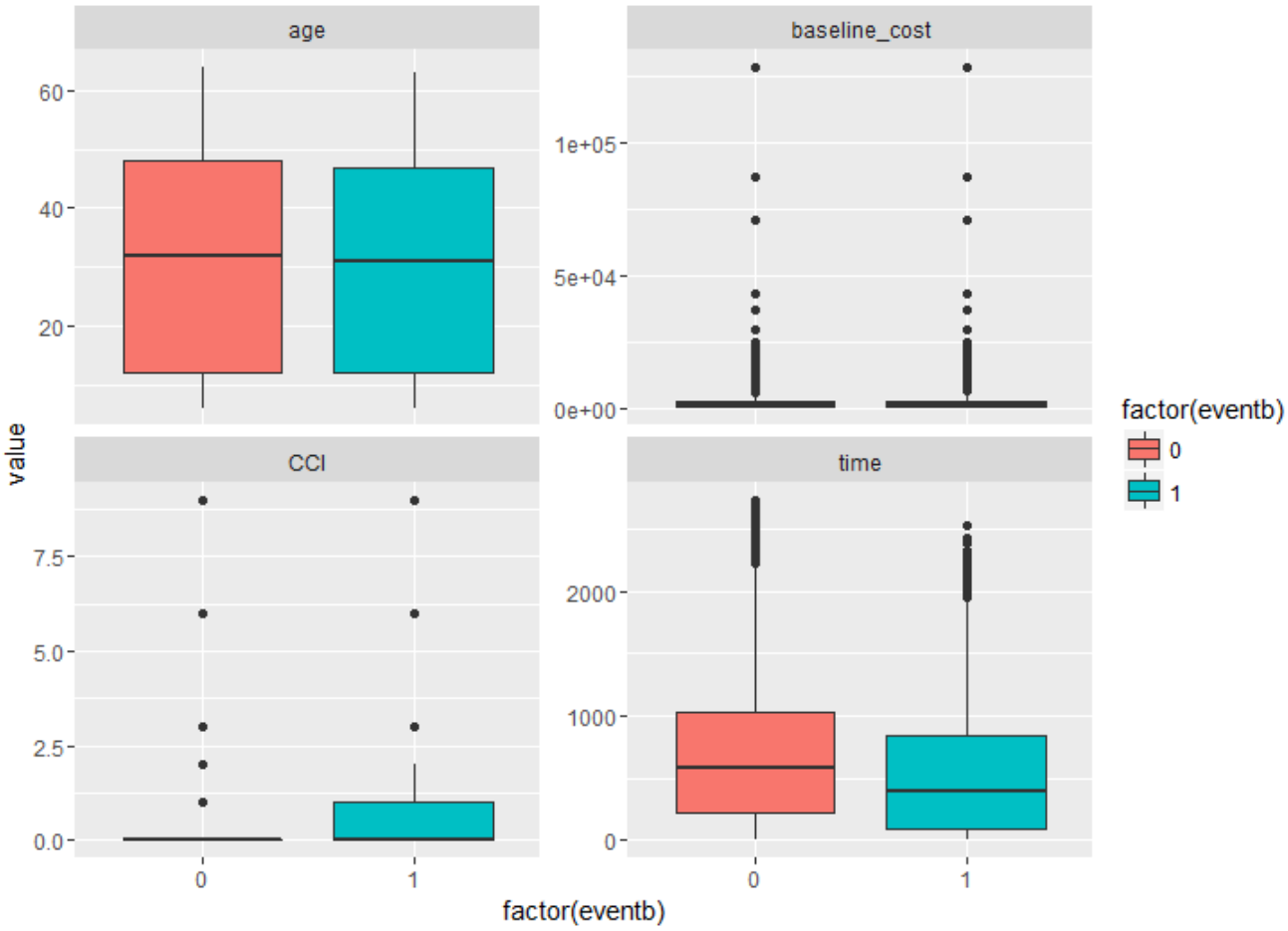


Figure 4: Bubble Diagram of Treatment Step by Event as Binary

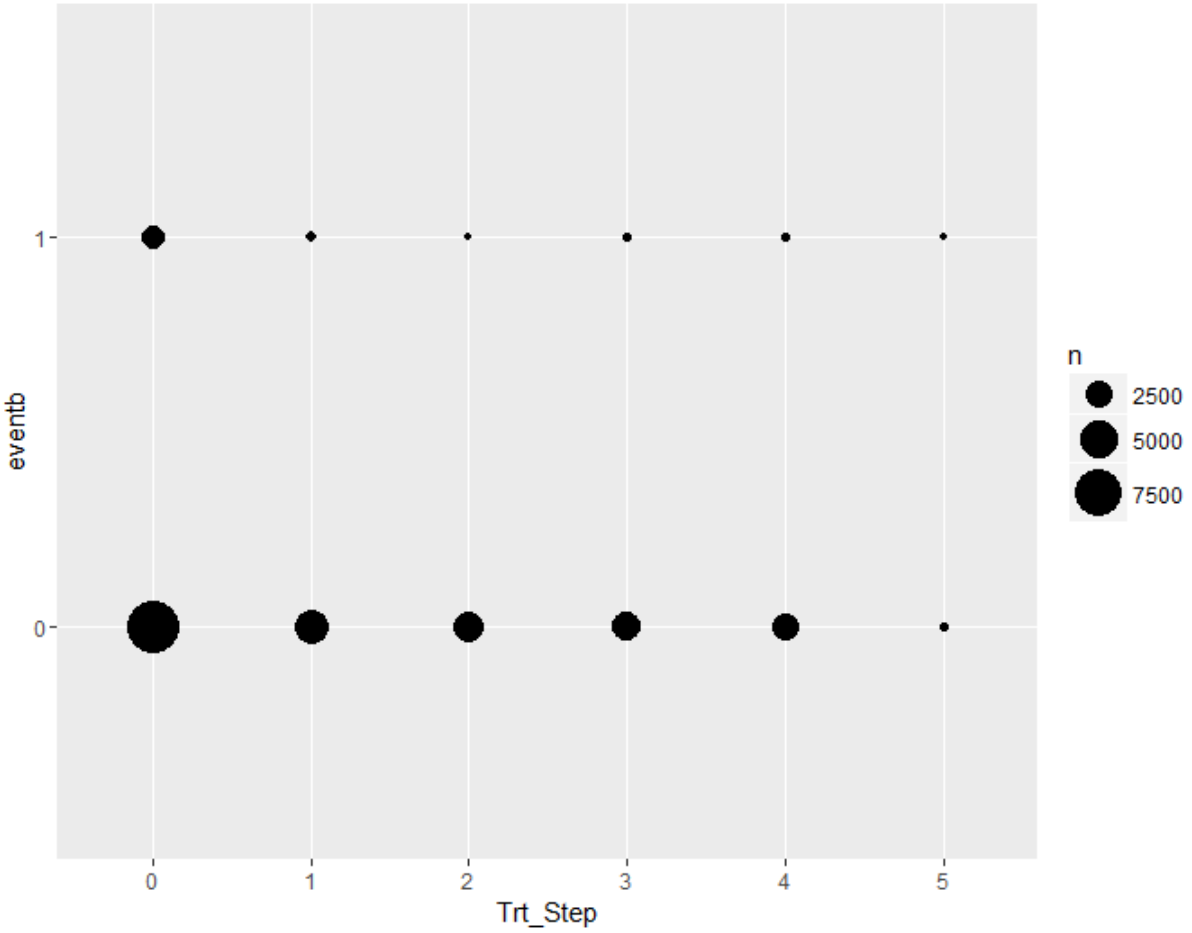
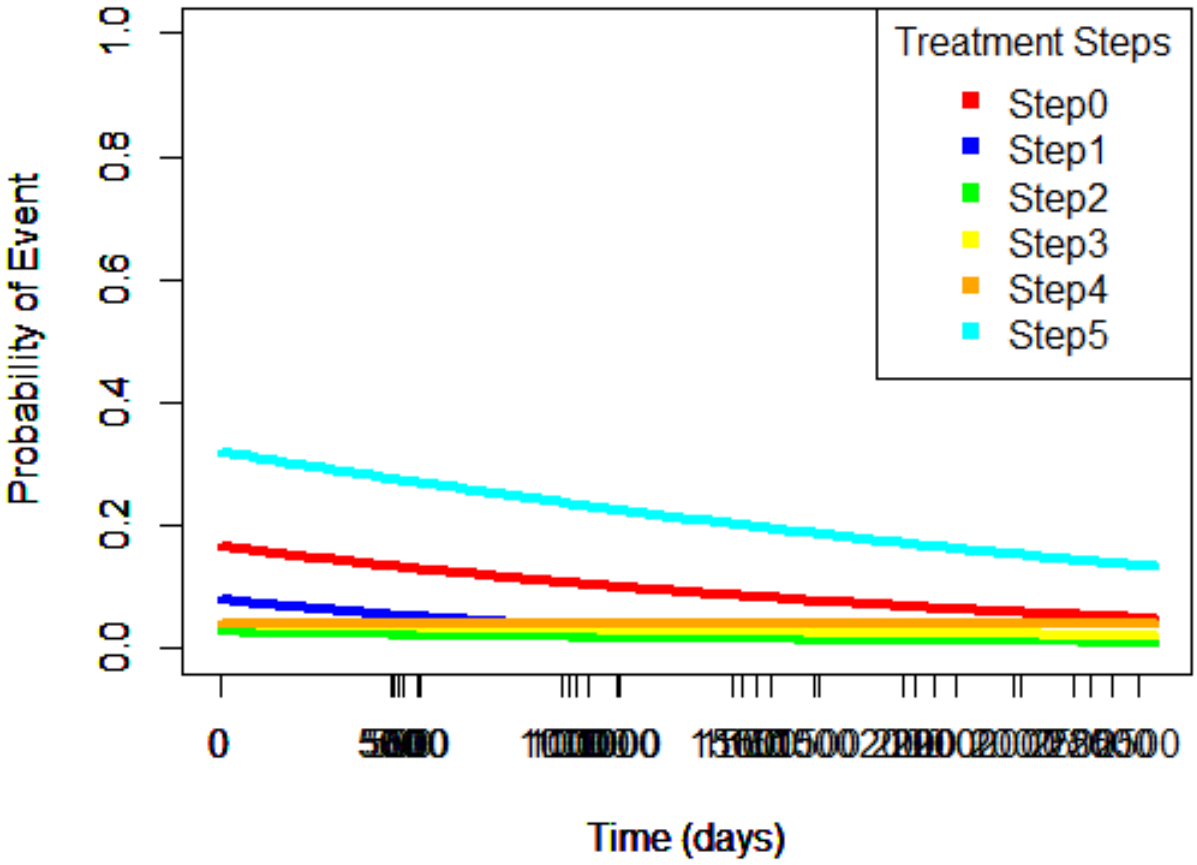


Figure 5: Probability of Event by Time



F Preliminary Results

* All models with additional random intercept for subjects did not converge.

F.1 Hypothesis 1

F.1.1 Model 1:Model with All Covariates

The GLIMMIX Procedure

Fit Statistics	
-2 Log Likelihood	74546.75
AIC (smaller is better)	74582.75
AICC (smaller is better)	74582.77
BIC (smaller is better)	74729.24
CAIC (smaller is better)	74747.24
HQIC (smaller is better)	74630.13

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	1	25283	9.65	0.0019
age	1	25283	6.02	0.0141
gender	1	25283	0.51	0.4762
region	3	25283	2.35	0.0708
CCI	1	25283	0.38	0.5379
Insurance	5	25283	2.98	0.0107
baseline_cost	1	25283	2.18	0.1399

The GLIMMIX Procedure

Fit Statistics	
-2 Log Likelihood	74407.25
AIC (smaller is better)	74449.25
AICC (smaller is better)	74449.28
BIC (smaller is better)	74620.16
CAIC (smaller is better)	74641.16
HQIC (smaller is better)	74504.53

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	1	25280	100.07	<.0001
time*time	1	25280	78.00	<.0001
time*time*time	1	25280	59.71	<.0001
time*time*time*time	1	25280	49.79	<.0001
age	1	25280	6.56	0.0104
gender	1	25280	0.95	0.3296
region	3	25280	1.85	0.1353
CCI	1	25280	0.42	0.5182
Insurance	5	25280	3.00	0.0104
baseline_cost	1	25280	1.84	0.1745

F.1.2 Model 2: Polynomial Trend

F.2 Hypothesis 2a

F.2.1 Model 1:Model with All Covariates

The GLIMMIX Procedure

Fit Statistics	
-2 Log Likelihood	16437.14
AIC (smaller is better)	16467.14
AICC (smaller is better)	16467.16
BIC (smaller is better)	16589.22
CAIC (smaller is better)	16604.22
HQIC (smaller is better)	16506.63

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	1	25286	132.28	<.0001
Trt_Step	1	25286	193.26	<.0001
age	1	25286	6.11	0.0135
gender	1	25286	23.06	<.0001
region	3	25286	10.18	<.0001
CCI	1	25286	50.49	<.0001
Insurance	5	25286	10.04	<.0001
baseline_cost	1	25286	0.76	0.3832

The GLIMMIX Procedure

Fit Statistics	
-2 Log Likelihood	15976.86
AIC (smaller is better)	16028.86
AICC (smaller is better)	16028.92
BIC (smaller is better)	16240.46
CAIC (smaller is better)	16266.46
HQIC (smaller is better)	16097.31

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	1	25275	59.00	<.0001
time*time	1	25275	23.83	<.0001
time*time*time	1	25275	13.19	0.0003
trt_stepn	5	25275	48.59	<.0001
time*trt_stepn	5	25275	2.83	0.0148
age	1	25275	12.64	0.0004
gender	1	25275	20.49	<.0001
region	3	25275	8.19	<.0001
CCI	1	25275	36.38	<.0001
Insurance	5	25275	8.53	<.0001
baseline_cost	1	25275	1.51	0.2192

F.2.2 Model 2:Polynomial Trend

F.3 Hypothesis 2b

F.3.1 Model 1:Model with All Covariates

The GLIMMIX Procedure

Fit Statistics	
-2 Log Likelihood	22306.84
AIC (smaller is better)	22458.84
AICC (smaller is better)	22459.31
BIC (smaller is better)	23077.38
CAIC (smaller is better)	23153.38
HQIC (smaller is better)	22658.93

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	4	25225	44.64	<.0001
trt_stepn	20	25225	37.66	<.0001
age	4	25225	9.49	<.0001
gender	4	25225	8.05	<.0001
region	12	25225	3.40	<.0001
CCI	4	25225	10.91	<.0001
Insurance	20	25225	2.80	<.0001
baseline_cost	4	25225	3.31	0.0101

F.3.2 Model 2: Polynomial Trend

The GLIMMIX Procedure

Fit Statistics	
-2 Log Likelihood	22191.27
AIC (smaller is better)	22375.27
AICC (smaller is better)	22375.95
BIC (smaller is better)	23124.02
CAIC (smaller is better)	23216.02
HQIC (smaller is better)	22617.48

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	4	25209	19.74	<.0001
time*time	4	25209	12.07	<.0001
time*time*time	4	25209	9.25	<.0001
time*time*time*time	4	25209	7.77	<.0001
tim*tim*tim*tim*time	4	25209	6.80	<.0001
trt_stepn	20	25209	37.10	<.0001
age	4	25209	9.46	<.0001
gender	4	25209	7.90	<.0001
region	12	25209	3.14	0.0002
CCI	4	25209	11.03	<.0001
Insurance	20	25209	2.84	<.0001
baseline_cost	4	25209	3.33	0.0097