BMI and Dementia: A Longitudinal Study Using GEE and Survival Analysis

Younghwan Cho

December 15, 2022

Abstract

BMI, long-term or short-term weight change have been suggested and shown to be linked to various health conditions and diseases including diabetes, high blood pressure, cardiovascular disease, and many others. However, relationship between BMI/weight change between dementia is not so straightforward. A higher BMI at ages 40–49 years was associated with increased dementia risk if BMI was above 30. After age of 50 years, low BMI ($<$ 22) was associated with increased risk of dementia, while higher BMI appeared to reduce the risk of dementia if BMI was $>$ 22. However, the nonlinear associations were not significant. Being overweight was significantly associated with lower risk of dementia in the age group of $\geq$ 70 years [1]. This suggests that BMI doesn't have strictly positive relationship between cause of the dementia, as opposed to most of other diseases. We analyze relationship between BMI and dementia extensively and longitudinally along with considerations of other factors including gender and race.

1 Data description

The data is downloaded from the Health and Retirement Study at Michigan's Social Research Institute. Details can be found at https://hrs.isr.umich.edu. Information of 38558 individuals have been collected. Information consists of gender, race, weight, height, BMI, and dementia status. This info has been collected every 2 years starting from 1996 until 2018, resulting in 12 longitudinal data for each individual. To determine the status of dementia of an individual, cognitive testing has been conducted on every biannual trials. Following is the demographic of 38558 individuals in the data.

2 Methodology

Generalized Estimating Equations (GEE) and survival analysis will be mainly used to analyze the effect of BMI, and additionally, gender and race on onset of dementia. GEE will show the direct relationship between BMI and dementia,

not accounting for over the period effect. Survival analysis can additionally account for the time variant aspect of this study. Here, brief introduction of the methodologies will be presented.

2.1 Introduction to GEE

We start with the assumption that mean of the response variable is associated with the covariates by the following formula:

$$g\left(\mu_{ij}\right) = x'_{ij}\beta$$

where μ_{ij} is mean of $_{ij}$. g is called the link function. Since, status of dementia (which is our response variable) is binary, we use logistic link. In this case, $g\left(\mu_{ij}\right) = \log\left[\frac{\mu_{ij}}{1-\mu ij}\right]$. GEE is defined as follows:

$$\sum_{i=1}^{n}\sum_{t=1}^{n_i}\frac{y_{it}-\mu\left(x_{it}^{\mathrm{T}}\beta\right)}{\tilde{\sigma}^2\left(x_{it},\beta\right)}\frac{\partial\mu\left(x_{it}^{\mathrm{T}}\beta\right)}{\partial\beta} = 0$$

Purpose of GEE is to find optimal β solving the equation. Specific to our case for logistic regression, GEE becomes:

$$\sum_{i=1}^{n}\sum_{t=1}^{n_i}x_{it}\left\{y_{it}-\pi\left(x_{it},\beta\right)\right\} = 0$$

where $\pi\left(x_{it},\beta\right) = e^{x_{it}^{\mathrm{T}}\beta}/\left(1+e^{x_{it}^{\mathrm{T}}\beta}\right)$. To find the covariance of our estimate $\hat{\beta}$, we define:

$$\begin{aligned} D_i(\beta) &= \left(\pi\left(x_{i1},\beta\right)\left\{1-\pi\left(x_{i1},\beta\right)\right\}x_{i1},\ldots,\pi\left(x_{in_i},\beta\right)\left\{1-\pi\left(x_{in_i},\beta\right)\right\}x_{in_i}\right) \\ &= X_i^{\mathrm{T}}\tilde{V}\left(X_i,\beta\right), \end{aligned}$$

with $\tilde{V}\left(X_i,\beta\right) = \mathrm{diag}\left\{\pi\left(x_{it},\beta\right)\left\{1-\pi\left(x_{it},\beta\right)\right\}\right\}_{t=1}^{n_i}$, we can verify that

$$\hat{B} = n^{-1}\sum_{i=1}^{n}X_i^{\mathrm{T}}\hat{V}_iX_i, \quad \hat{M} = n^{-1}\sum_{i=1}^{n}X_i^{\mathrm{T}}\hat{\varepsilon}_i\hat{\varepsilon}_i^{\mathrm{T}}X_i$$

where $\hat{\varepsilon}_i = \left(\hat{\varepsilon}_{i1},\ldots,\hat{\varepsilon}_{in_i}\right)^{\mathrm{T}}$ with residual $\hat{\varepsilon}_{it} = y_{it} - e^{x_{it}\hat{\beta}}/\left(1+e^{x_{it}\hat{\beta}}\right)$, and $\hat{V}_i = \mathrm{diag}\left\{\pi\left(x_{it},\hat{\beta}\right)\left\{1-\pi\left(x_{it},\hat{\beta}\right)\right\}\right\}_{t=1}^{n_i}$. So the cluster-robust covariance estimator for logistic regression is

$$\mathrm{cov}(\hat{\beta}) = \left(\sum_{i=1}^{n}X_i^{\mathrm{T}}\hat{V}_iX_i\right)^{-1}\sum_{i=1}^{n}X_i^{\mathrm{T}}\hat{\varepsilon}_i\hat{\varepsilon}_i^{\mathrm{T}}X_i\left(\sum_{i=1}^{n}X_i^{\mathrm{T}}\hat{V}_iX_i\right)^{-1}$$

2.2 Survival Analysis

2.2.1 Time-to-event data

Let $T \geq 0$ denote the outcome of interest. We can characterize a non-negative continuous T using its density $f(t)$, distribution function $F(t)$, survival function $S(t) = 1 - F(t) = \text{pr}(T > t)$, and hazard function

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \text{pr}(t \leq T < t + \Delta t \mid T \geq t)/\Delta t.$$

Within a small time interval $[t, t + \Delta t]$, we have approximation

$$\text{pr}(t \leq T < t + \Delta t \mid T \geq t) \cong \lambda(t)\Delta t,$$

so the hazard function denotes the death rate within a small interval conditioning on surviving up to time t. Both the survival and hazard functions are commonly used to describe a positive random variable. First, the survival function has a simple relationship with the expectation.

2.2.2 Kaplan-Meier survival curve

(S1) $t_1, \ldots, t_K$ are the death times, and $d_1, \ldots, d_K$ are the corresponding number of deaths; (S2) $r_1, \ldots, r_K$ are the number of patients at risk, that is, r_1 patients are not dead or censored right before time t_1, and so on; (S3) $c_1, \ldots, c_K$ are the number of censored patients within interval $[t_1, t_2), \ldots, [d_K, \infty)$. Kaplan and Meier (1958) proposed the following simple estimator for the survival function.

Definition 1 (Kaplan-Meier curve) First estimate the discrete hazard function at the failure times $\{t_1, \ldots, t_K\}$ as $\hat{\lambda}_k = d_k/r_k (k = 1, \ldots, K)$ and then estimate the survival function as

$$\hat{S}(t) = \prod_{k:t_k \leq t} \left(1 - \hat{\lambda}_k\right).$$

The $\hat{S}(t)$ in Definition 1 is also called the product-limit estimator of the survival function due to its mathematical form.

At each failure time t_k, we view d_k as the result of r_k Bernoulli trials with probability λ_k. So $\hat{\lambda}_k = d_k/r_k$ has variance $\lambda_k (1 - \lambda_k)/r_k$ which can be estimated by

$$\text{var}\left(\hat{\lambda}_k\right) = \hat{\lambda}_k \left(1 - \hat{\lambda}_k\right)/r_k.$$

We can estimate the variance of the survival function using the delta method. We can approximate the variance of

$$\log \hat{S}(t) = \sum_{k:t_k \leq t} \log\left(1 - \hat{\lambda}_k\right) \cong \sum_{k:t_k \leq t} \log\left(1 - \lambda_k\right) - \sum_{k:t_k \leq t} \left(1 - \lambda_k\right)^{-1} \left(\hat{\lambda}_k - \lambda_k\right)$$

by

$$\begin{aligned}\operatorname{var}\{\log \hat{S}(t)\} &= \sum_{k:t_k \le t} (1-\lambda_k)^{-2} \operatorname{var}\left(\hat{\lambda}_k\right) \\ &= \sum_{k:t_k \le t} \left(1-\hat{\lambda}_k\right)^{-2} \hat{\lambda}_k \left(1-\hat{\lambda}_k\right)/r_k \\ &= \sum_{k:t_k \le t} \frac{d_k}{r_k\left(r_k-d_k\right)},\end{aligned}$$

2.2.3 Cox model

We assume that the conditional hazard ratio function has the form

$$\lambda(t \mid x) = \lambda_0(t) \exp\left(x^{\mathrm{T}}\beta\right),$$

where β is an unknown parameter and $\lambda_0(\cdot)$ is an unknown function. Likelihood function to estimate β :

$$L(\beta) = \prod_{k=1}^{K} \frac{\exp\left(x_k^{\mathrm{T}}\beta\right)}{\sum_{l\in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right)},$$

We have:

$$\log L(\beta) = \sum_{k=1}^{K} \left\{ x_k^{\mathrm{T}}\beta - \log \sum_{l\in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) \right\}$$

and the score function is

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{k=1}^{K} \left\{ x_k - \frac{\sum_{l\in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) x_l}{\sum_{l\in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right)} \right\}.$$

Define

$$\pi_\beta\left(l \mid R_k\right) = \exp\left(x_l^{\mathrm{T}}\beta\right) / \sum_{l\in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right), \quad \left(l \in R\left(t_k\right)\right)$$

which sum to one, so they induce a probability measure leading to expectation $E_\beta\left(\cdot \mid R_k\right)$ and covariance $\operatorname{cov}_\beta\left(\cdot \mid R_k\right)$. With this notation, the score function simplifies to

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{k=1}^{K} \left\{ x_k - E_\beta\left(x \mid R_k\right) \right\},$$

where $E_\beta\left(x \mid R_k\right) = \sum_{l\in R(t_k)} \pi_l\left(\beta \mid R_k\right) x_l$; the Hessian matrix simplifies to

$$\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^{\mathrm{T}}} = -\sum_{k=1}^{K} \operatorname{cov}_\beta\left(x \mid R_k\right) \preceq 0,$$

where

$$
\begin{aligned}
&\operatorname{cov}_\beta\left(x \mid R_k\right) \\
&= \begin{pmatrix} \sum_{l \in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) x_l x_l^{\mathrm{T}} \sum_{l \in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) \\ -\sum_{l \in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) x_l \sum_{l \in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) x_l^{\mathrm{T}} \end{pmatrix} \Big/ \left\{ \sum_{l \in R(t_k)} \exp\left(x_l^{\mathrm{T}}\beta\right) \right\}^2 \\
&= \sum_{l \in R(t_k)} \pi_\beta\left(l \mid R_k\right) x_l x_l^{\mathrm{T}} - \sum_{l \in R(t_k)} \pi_\beta\left(l \mid R_k\right) x_l \sum_{l \in R(t_k)} \pi_\beta\left(l \mid R_k\right) x_l^{\mathrm{T}}.
\end{aligned}
$$

3 Results

3.1 data demogrpahic

Gender	Race	Ratio	Total
Male Male Male pt¡ -Male pt¿	White	31.2%	12014
	African American	7.6%	2928
	Other	3.9%	1522
Female Female Female pt¡ -Female pt¿	White	40.4%	15582
	African American	10.7%	4414
	Other	4.6%	1791
NA		0.524%	202

3.2 GEE

We first conduct analysis with the most recent (2018) data. Then, for consistency, we will replicate the analysis on 2014 data.

2018 Dementia vs 2018 BMI			
Coefficients	Estimate	Robust Std.error	Pr(> |w|)
(Intercept)	-1.83343	0.42153	$1.36e-05$ **
BMI18	-0.03520	0.01554	0.0235 *

Table1 : 2018 Dementia against 2018 BMI (corstr="independence")

2018 Dementia vs 2018 BMI			
Coefficients	Estimate	Robust Std.error	Pr(> |w|)
(Intercept)	-1.83343	0.66930	0.00616 **
BMI18	-0.03520	0.02499	0.15892

Table2 : 2018 Dementia against 2018 BMI (corstr="exchangeable")

With two different assumption on correlation structure, we have different standard errors with different p-value for the coefficient estimates. With independence assumption, BMI was significant factor. On the other hand, with ex-

changeable correlation assumption, BMI was no longer significant factor.

2018 Dementia vs 2018,2016 and 2014 BMIs			
Coefficients	Estimate	Robust Std.error	Pr($>\|w\|$)
(Intercept)	-1.3986	0.4778	$0.0034 **$
BMI18	-0.0637	0.0550	0.2468
BMI16	0.1198	0.0657	0.0680
BMI14	-0.1077	0.0330	0.0011 **

Table3 : 2018 Dementia against recent 3 BMIs

With 3 recent BMIs before the trial in consideration, we found that past BMI, especially 2 year before the trial, had more significant effect on the status of dementia than the current BMI.

2018 Dementia vs BMI diff in 2018 and 2016			
Coefficients	Estimate	Robust Std.error	Pr($>\|w\|$)
(Intercept)	-2.8031	0.0851	$< 2e-16 ***$
diff	-0.0469	0.0539	0.38

Table4 : 2018 Dementia against (2018 BMI - 2016 BMI)
Above result shows that 2018 dementia status was not affected significantly by recent 2 years' BMI change.

2018 Dementia vs increase in BMI between 2014 and 2018			
Coefficients	Estimate	Robust Std.error	Pr($>\|w\|$)
(Intercept)	-3.2067	0.1800	$< 2e-16 ***$
gain	0.1828	0.0564	0.0012 **

Table5: 2018 Dementia against BMI increase

2018 Dementia vs decrease in BMI between 2014 and 2018			
Coefficients	Estimate	Robust Std.error	Pr($>\|w\|$)
(Intercept)	-2.9390	0.1411	$< 2e-16 ***$
loss	-0.1094	0.0375	0.0036 **

Table5: 2018 Dementia against BMI decrease

To further analyze effect of BMI change on dementia, I have partitioned the data into BMI gain and BMI loss. Former are the cases where BMI was higher in 2018 compared to 2014, latter are the cases where BMI was lower in 2018 compared to 2014. We have found that BMI gain has no effect on dementia. However, BMI loss has significant effect on dementia.

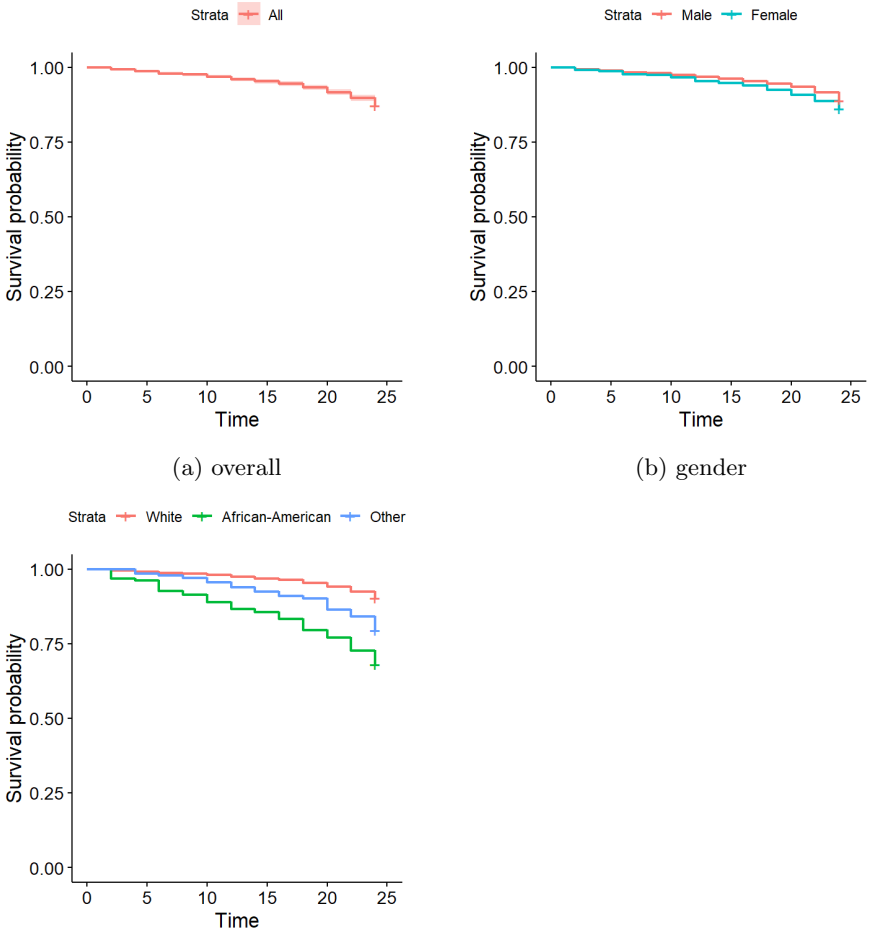
I have replicated the same analysis on 2014 data. The result was similar and as was follows: 2014 BMI was not significant with p-value 0.29. With past 3 BMIs in consideration, only 2014 and 2010 BMIs were significant with p-values 0.011 and 0.027. BMI difference (2014 BMI - 2010 BMI) was significant factor with p-value 0.021. BMI gain was not significant with p-value 0.11. On the

other hand, BMI loss was highly significant with p-value 7.6e-07.

We conclude that effect of current BMI on the current status of dementia is minimal and neglectable. However, either BMI decrease or BMI increase over the several years has significant effect on future onset of dementia. Especially, BMI decrease has shown consistently significant effect on onset of dementia.

3.3 Survival analysis

3.3.1 Kaplan-Meier plot



(a) overall

(b) gender

(c) race

Figure 1: Kaplan Meier plot

Figure (a) is Kaplan-Meier plot based on overall data. The survival proba-

bility didn't drop below 0.8, which is consistent with the data where majority of the subjects didn't get dementia until the end of the study. Figure (b) is Kaplan-Meier plots stratified based on gender, which shows female subjects are slightly more vulnerable to dementia even though the difference was not that visually significant. Figure (c) is Kaplan-Meier plots stratified based on race, which shows that African-Americans were significantly most vulnerable to dementia followed by whites and others.

3.3.2 Cox model

Model1 coxph(Surv(Time, Survival.fail) $\sim$ All past BMI + gender + race)

Cox hazard model on dementia				
Coefficients	Estimate	e^{Estimate}	Robust SE	Pr($>|w|$)
2014 BMI	-0.090645	0.913342	0.038506	0.019 *
gender	0.212175	1.236364	0.102857	0.039 *
race	0.706421	2.026725	0.064872	$< 2e-16$ ***
logrank test= 205 on 14 df, p$< 2e-16$				

Table6 : Cox hazard model with BMI,gender and race

We see that race was the most significant factor in Cox model. Gender was also significant factor with p-value close the 0.05 and one of the past BMI, specifically 2014 BMI, was significant. All other past BMIs were found to be insignificant covariates and I omit here.
Logrank test is performed to analyze following hypothesis: $\beta = 0 \iff \lambda_1(t) = \lambda_0(t) \iff S_1(t) = S_0(t)$. Recall model assumption: $\lambda_1(t) = \lambda_0(t)e^{\beta}$.
p-value for logrank test suggests that $\beta \neq 0$, i.e. model assumption is significant.

With only considering 2014 BMI as a covariate, we have the following result:
Model2 coxph(Surv(Time, Survival.fail) $\sim$ 2014 BMI))

Cox hazard model on dementia				
Coefficients	Estimate	e^{Estimate}	Robust SE	Pr($>|w|$)
2014 BMI	-0.014333	0.985769	0.008984	0.111
Score (logrank) test = 2.63 on 1 df, p=0.1, Robust = 2.7 p=0.1				

Table7: Cox hazard model with 2014 BMI

We see that 2014 BMI is not a significant factor in dementia in cox model.

We further analyze the model on 2014 BMI with gender and race as a strata respectively (stratified cox model).
Model3 coxph(Surv(Time, Survival.fail) $\sim$ 2014 BMI + strata(gender))

Cox hazard model on dementia				
Coefficients	Estimate	e^{Estimate}	Robust SE	Pr(> |w|)
2014 BMI	-0.013111	0.986975	0.008776	0.135
Score (logrank) test = 2.26 on 1 df, p=0.1, Robust = 2.37 p=0.1				

Table8: Cox hazard model with 2014 BMI with gender as strata

We see that 2014 BMI is still not a significant factor in dementia in cox model when stratified as gender. Similar result holds for stratifying on race. We conclude that BMI is not a significant factor in cox hazard model.

We now consider BMI diff in 2014 and 2010 as an additional factor.

Model4 coxph(Surv(Time, Survival.fail) $\sim$ 2014 BMI + BMIdiff + strata(gender))

Cox hazard model on dementia				
Coefficients	Estimate	e^{Estimate}	Robust SE	Pr(> |w|)
2014 BMI	-0.007174	0.992851	0.008535	0.401
BMIdiff	-0.066941	0.935250	0.014620	4.67e-06 ***
Score (logrank) test = 21.45 on 2 df, p=2e-05, Robust = 10.62 p=0.005				

Table9: Cox hazard model with 2014 BMI + BMIdiff with gender as strata

Model5 coxph(Surv(Time, Survival.fail) $\sim$ 2014 BMI + BMIdiff + strata(race))

Cox hazard model on dementia				
Coefficients	Estimate	e^{Estimate}	Robust SE	Pr(> |w|)
2014 BMI	-0.023131	0.977134	0.008692	0.00778 **
BMIdiff	-0.054861	0.946617	0.012530	1.2e-05 ***
Score (logrank) test = 24.17 on 2 df, p=6e-06, Robust = 17.45 p=2e-04				

Table10: Cox hazard model with 2014 BMI + BMIdiff with race as strata

We found that BMI diff has highly significant effect on dementia in both stratified cases on gender and race. In conclusion, both in GEE and survival analysis, BMI change has shown to be significant factor on the onset of dementia.

By considering all BMIs analogously to all subjects, we are missing time-variant and subject-variant aspects of the data. We can further extend the analysis only considering the effect of BMI on the year of first onset of dementia for each subject. We can also calculate BMI change in past 2 years before the onset of dementia, and see if BMI change is significant factor. For cases where no onset of dementia occured, mean of 12 BMIs were calculated.

References

[1] Li J, Joshi P, Ang TFA, Liu C, Auerbach S, Devine S, Au R. Mid- to Late-Life Body Mass Index and Dementia Risk: 38 Years of Follow-up of the Framingham Study. Am J Epidemiol. 2021 Dec 1;190(12):2503-2510. doi: 10.1093/aje/kwab096. PMID: 33831181; PMCID: PMC8796797.

[2] Shen J, Chen H, Zhou T, Zhang S, Huang L, Lv X, Ma Y, Zheng Y, Yuan C. Long-term weight change and its temporal relation to later-life dementia in the Health and Retirement Study. The Journal of Clinical Endocrinology Metabolism. 2022 Apr 14

[3] Karlsson IK, Lehto K, Gatz M, Reynolds CA, Dahl Aslan AK. Age-dependent effects of body mass index across the adult life span on the risk of dementia: a cohort study with a genetic approach. BMC medicine. 2020 Dec;18(1):1-1

[4] Qu Y, Hu HY, Ou YN, Shen XN, Xu W, Wang ZT, Dong Q, Tan L, Yu JT. Association of body mass index with risk of cognitive impairment and dementia: a systematic review and meta-analysis of prospective studies. Neuroscience Biobehavioral Reviews. 2020 Aug 1;115:189-98

[5] Ma Y, Ajnakina O, Steptoe A, Cadar D. Higher risk of dementia in English older individuals who are overweight or obese. International journal of epidemiology. 2020 Aug;49(4):1353-65

[6] Kang SY, Kim YJ, Jang W, Son KY, Park HS, Kim YS. Body mass index trajectories and the risk for Alzheimer's disease among older adults. Scientific reports. 2021 Feb 4;11(1):1-0

[7] Cao, H., Liu, W. and Zhou, Z. (2018) Simultaneous non-parametric regression analysis of sparse longitudinal data. Bernoulli, 24, 3013-3038