The Effects of Depression on Daily Cigarette Smoking

The National Longitudinal Study of Adolescent to Adult Health (ADD)

By John Semerdjian

Abstract

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year. The Add Health cohort has been followed into young adulthood with four in-home interviews. Add Health combines longitudinal survey data across multiple domains, including substance use and depression.

A subset of the the Add Health data was used to study the effect of depression, sex, and age on daily cigarette use over time. A continuous measure of depression was based on a subset of 9 items from Center for Epidemiological Studies Depression Scale (CES-D) was used. Subjects assessed their depressive feelings from the past 7 days and daily cigarette smoking from the last 30 days at each wave.

Age and sex were both associated with an increase in the odds of daily cigarette smoking. The longitudinal effect of CES-D was also associated with an increase in the odds of smoking, while the baseline effect of CES-D resulted in non-significant results from a Simple Random Effects model. However, GEE models supported a population-based effect from baseline CES-D scores.

About the Study

Study Design: Observational study of daily cigarette smoking and depression score as measured by a subset of 9 items from Center for Epidemiological Studies Depression Scale (CES-D).

Objective: To estimate the effect depression, age, and sex on daily cigarette smoking.

Methods: This study focuses on a subset of 748 13-14 year olds from wave 1 that were interviewed at each of the 4 study waves. The first wave's depression data was used as baseline, thus only 3 waves worth of data was analyzed. The National Longitudinal Study of Adolescent to Adult Health administers a subject of 9 depression survey questions on a scale 0-3 that were totaled and used as a continuous measure of depression. Baseline depression scores were compared to subsequent measures of depression at each wave. Respondents were also asked if they smoke cigarettes (dichotomized for daily smoking in the past 30 days).

Data: Publically available data was obtained from the Odum Institute at UNC online.

Research Question

———

How are the time-varying and initial depression levels (as measured by CES Depression Scale) related to the propensity to smoking cigarettes daily among survey respondents?

Study Description

- 4 wave longitudinal study of adolescents (wave 1 = baseline)
- Respondents surveyed in their homes
- Study focused on a subset of 748 13-14 year old subjects from wave 1 that were subsequently interviewed at each following wave (only complete cases were considered)

Wave 1	Wave 2	Wave 3	Wave 4
1994-1995 (baseline)	1996	2001-2002	2008

Figure 1 - Daily Cigarette Smoking at Wave 1 Group Comparison

Stratifying the data by daily smoker status at wave 1 (baseline), figure 1 shows a modest difference in depression scores over time. The variation in scores warrants a subject-specific random effect.

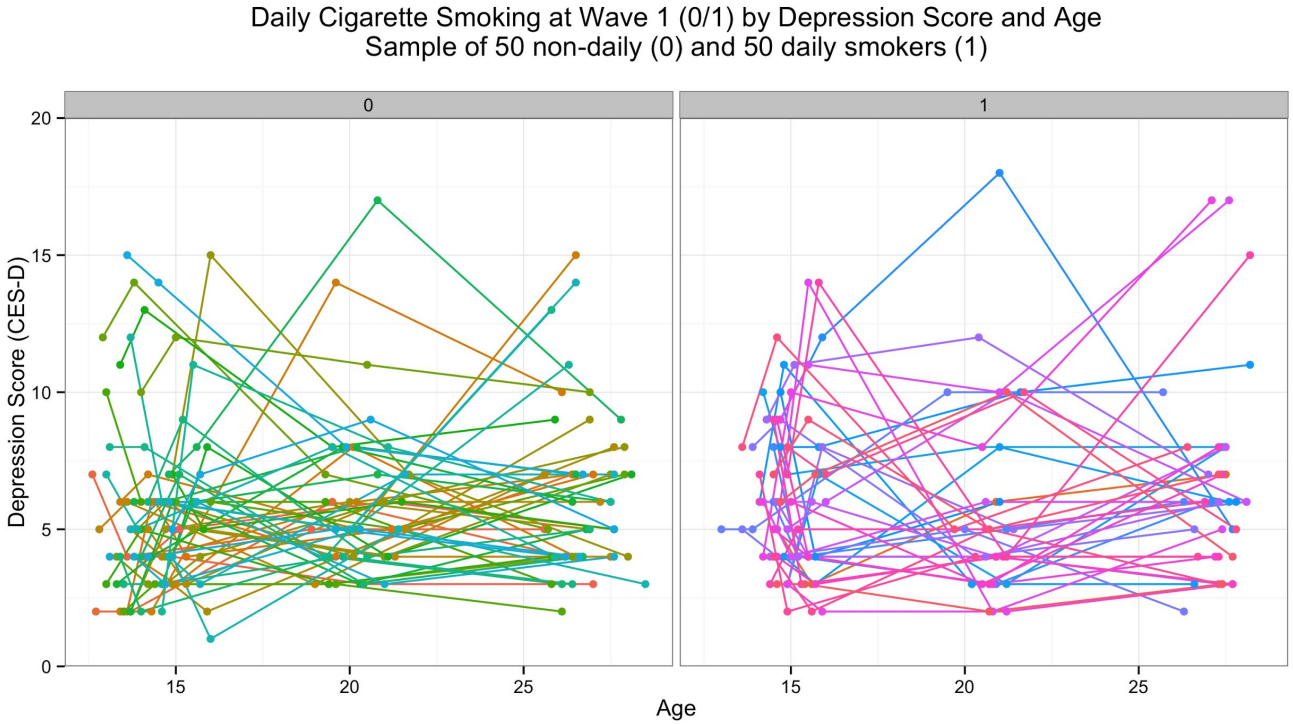


Figure 2 - Depression by Age, Sex, and Daily Smoker Status

When viewed cross-sectionally, there appears to be a difference in depression scores between males and females. There also seems to be a difference between daily smokers and non-daily smokers based on time-varying measures of smoking status.

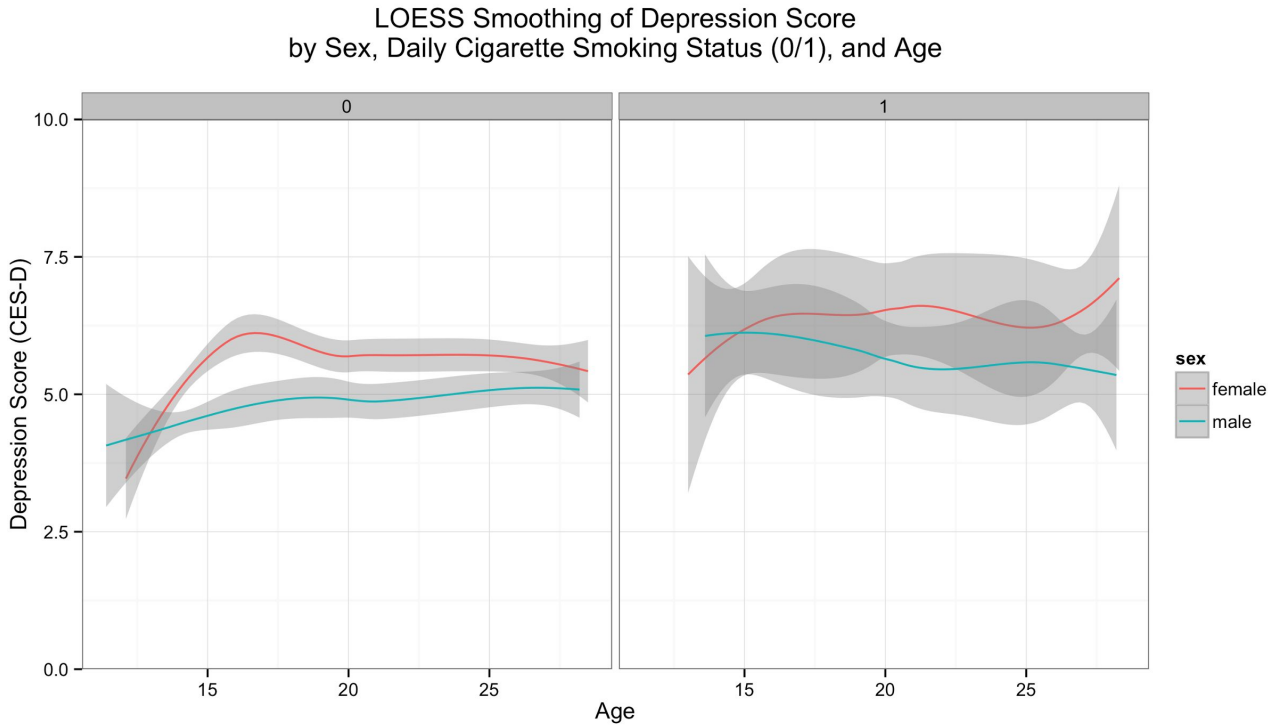


Figure 3 - Distribution of Depression Scores by Study Wave

Depression scores appear to be higher during wave 1 and 2 for daily cigarette smokers. Depression scores appear to be constant across wave for non-daily smokers. Naturally, study wave may serve as a proxy for subject age.

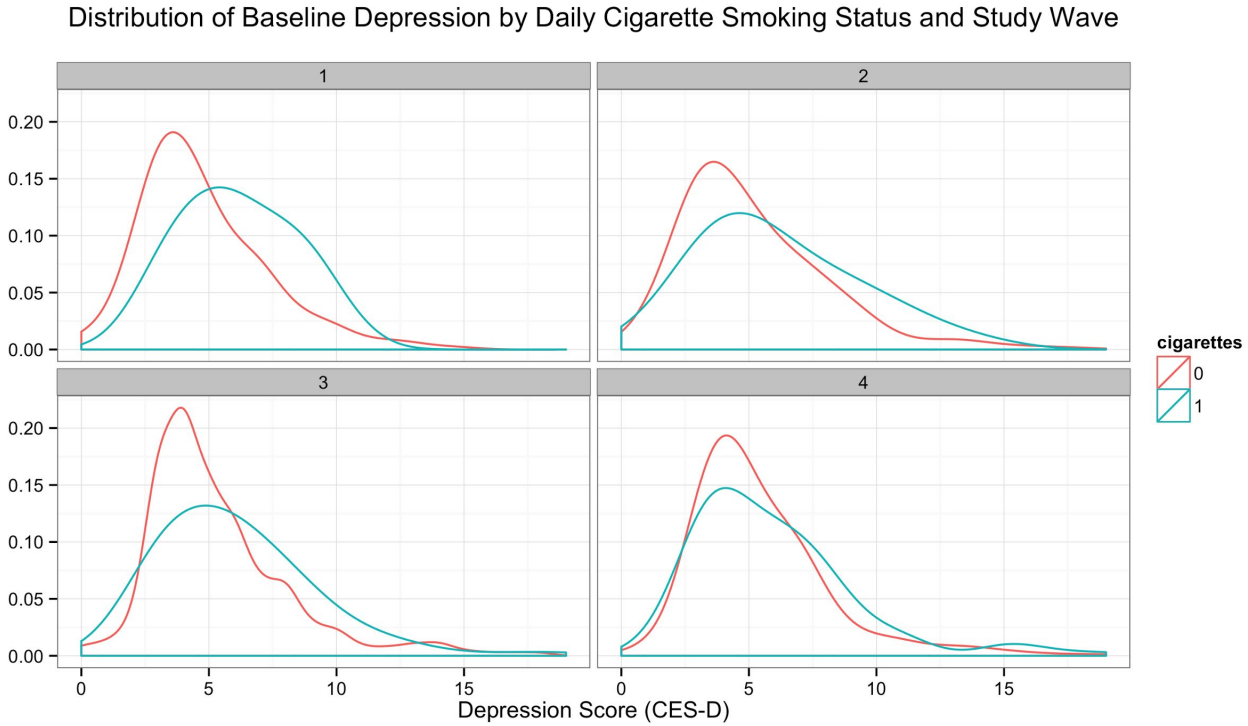
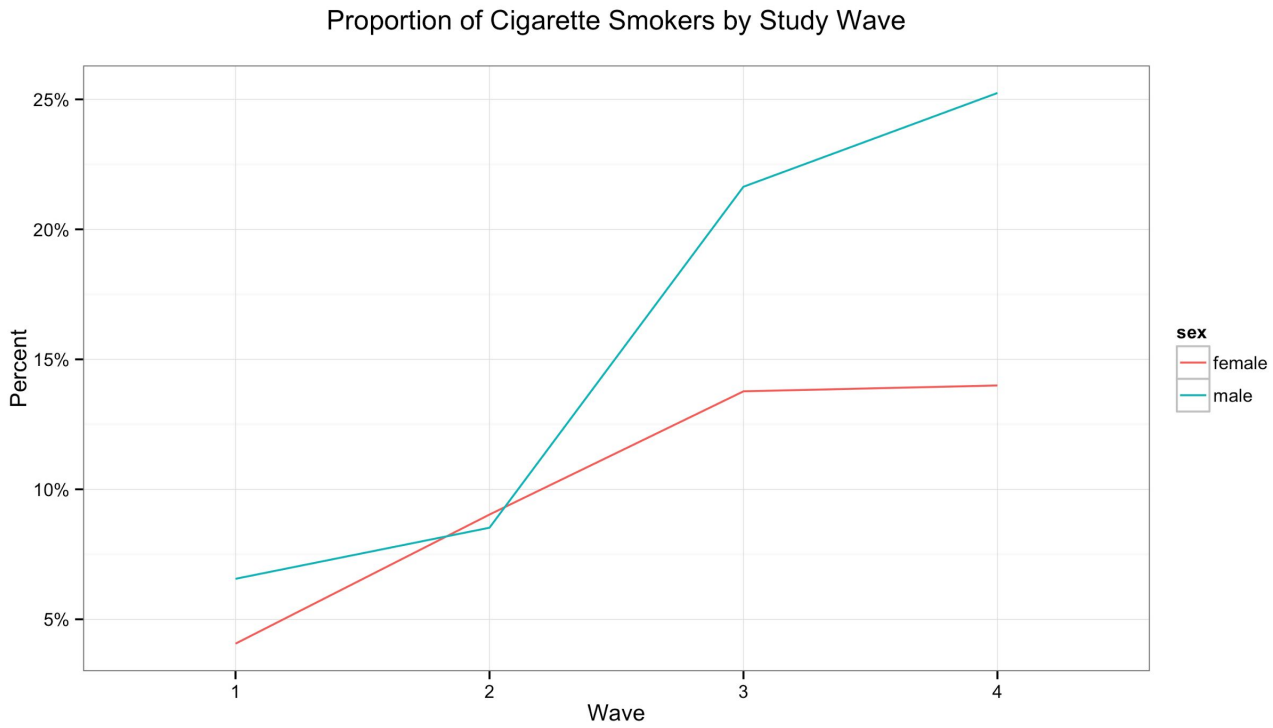


Figure 4 - Proportion of Daily Smokers by Study Wave

The fraction of daily cigarette smokers has increased over each wave for both males and females, though the rate increased faster for males after the 2nd wave.



Variables

Baseline

- Sex (male/female)
- Depression score (CES-D) at wave 1

Time-varying (one obs. at wave 2, 3, and 4)

- Age subtracted by 13 years (continuous)
- Depression score (CES-D)

Outcome

- Daily cigarette smoking:

  “During the past 30 days, on how many days did you smoke cigarettes?”

  Results dichotomized to indicate whether individuals smoked on all of the past 30 days

Depression Score Calculation

Respondents rated each item on a scale of 0-3, representing whether an individual experienced the symptom "never/rarely" to "most/all of the time" during the last 7 days from interview date.

Scale ranges from 0-27

Questions:

- Bothered by things that usually don't bother you
- Could not shakeoff the blues
- Felt that you were just as good as other people [reverse coded]
- Had trouble keeping your mind on what you were doing
- Were depressed
- Were too tired to do things
- Enjoyed life [reverse coded]
- Were sad
- Felt that people disliked you

Statistical Models

1. Logistic Regression & GEE

$$\text{logit}\{P(Y_{ij} = 1 | X_{i1} = x_{i1}, X_{ij} = x_i, z_i, v_{ij})\} = \beta_0 + \beta_1 x_{i1} + \beta_2 (x_{ij} - x_{i1}) + \beta_3 z_i + \beta_4 v_{ij}$$

2. Simple Random Effects

$$\text{logit}\{P(Y_{ij} = 1 | \beta_{0i}, X_{i1} = x_{i1}, X_{ij} = x_{ij}, z_i, v_{ij})\} = (\beta_0 + \beta_{0i}) + \beta_1 x_{i1} + \beta_2 (x_{ij} - x_{i1}) + \beta_3 z_i + \beta_4 v_{ij}$$

$i \in \{1, \ldots, 1752\}$, subject ID
$j \in \{2, 3, 4\}$, wave
$j = 1$ (baseline)

Methods

1. Logistic regression using the glm function. Ignore correlation structure between repeated measures and random per-subject intercept. Estimate both cross-sectional CES-D measure at baseline and longitudinal CES-D measure.

2. GEE using independent and exchangeable correlation structures. Apply same strategy as logistic regression. Used geeglm function from geepack library.

3. Simple random effects using same strategy and adding a random per-subject intercept. Used glmer function from lme4 library.

* All analyses done in R

Odds Ratios	Logistic Regression	GEE (independent)	GEE (exchangeable)	Simple Random Effects
OR, β_0 Intercept Naive SE (Robust) P-value	0.047 0.0097 (0.0111) < 0.001	0.047 0.0097 (0.0111) < 0.001	0.049 0.0116 (0.0111) < 0.001	0.0001 < 0.0001 < 0.001
OR, β_1 CES-D baseline Naive SE (Robust) P-value	1.093 0.0290 (0.0364) < 0.001	1.093 0.0290 (0.0364) 0.007	1.084 0.0363 (0.0350) 0.012	1.141 0.0961 0.117
OR, β_2 CES-D longitudinal Naive SE (Robust) P-value	1.092 0.0227 (0.0236) < 0.001	1.092 0.0228 (0.0235) < 0.001	1.080 0.0215 (0.0198) < 0.001	1.173 0.0528 < 0.001
OR, β_3 Sex=male Naive SE (Robust) P-value	1.750 0.2144 (0.2892) < 0.001	1.750 0.2147 (0.2888) < 0.001	1.868 0.3016 (0.3114) < 0.001	2.311 0.9453 0.041
OR, β_4 Age-13yrs Naive SE (Robust) P-value	1.068 0.0132 (0.0104) < 0.001	1.068 0.0133 (0.0103) < 0.001	1.067 0.0105 (0.0104) < 0.001	1.177 0.0263 < 0.001

Discussion

Despite its assumption of independence across observations, the logistic regression model provided odds ratios (OR) estimates that were similar to both GEE models. Robust and naive standard errors (SE) for logistic regression and GEE model assuming an independent correlation structure were nearly identical.

Between the GEE models, $\beta_1^{\text{CES-D baseline}}$ and $\beta_2^{\text{CES-D longitudinal}}$ estimates and SE were smaller for the exchangeable model, though $\beta_3^{\text{Sex=male}}$ and its SE were larger. The estimate of $\beta_4^{\text{Age-13yrs}}$ between the two was very similar. The exchangeable model's robust SE were relatively close to their naive counterparts.

The simple random effects (SRE) model returned noticeably larger odds ratios compared to the marginal models, particularly for $\beta_3^{\text{Sex=male}}$ where the OR went from 1.868 (GEE exc.) to 2.311. The SE of the SRE model estimates also significantly grew in size. Since this model only has a random intercept term and no random slopes, an exchangeable correlation structure is implied.

The p-value for the $\beta_3^{\text{CES-D baseline}}$ estimate was no longer < 0.05 in the SRE model (P = 0.117).

Discussion (cont.)

The among-subject standard deviation of the intercept (β_{0i}) is 6.3 on the log odds scale, while the fixed effect for the intercept (β_{0i}) term was -9.117 (SE: 0.766) on the log odds scale. The variation in the intercept between subjects was very large. The ICC for subjects was 0.924, which suggests that within subject effects are highly correlated, and our analysis would benefit from a SRE model. In other words, most of the variation is between-subject variation. If we desire a model that would describe individual effects and allow us to assign each individual with their own effect, than SRE would be an ideal choice. However, the GEE model with exchangeable correlation would also give us a population-based estimate.

The exchangeable GEE model can be interpreted as follows:

- The intercept for subjects is the OR of daily cigarette smoking for a 13 year-old female with a depression score of 0 (OR = 0.049) at baseline and no longitudinal change in depression.
- The marginal (or cross-sectional) effect of depression at baseline is that for every 1 unit increase in baseline depression, the mean OR increases by a factor of 1.084.
- The longitudinal effect of depression is that for every 1 unit increase in the change of depression from baseline, the mean OR increases by a factor of 1.080.
- Males have an increased OR of a factor of 1.868 when compared to females.
- For every 1 year increase in age, the mean OR of the response increases by a factor of 1.177.
- There is also an implied correlation structure across repeated measures for subjects.

Conclusion

The cross-sectional and longitudinal estimates from the SRE model was larger than the population average effect estimated by GEE in the marginal models, though the effect of baseline depression was no longer significant.

The SRE model helped explain that much of the variation of the data is between subjects.

Between the marginal models, the GEE model with exchangeable correlation is a better fit to the data, and also returned nearly equivalent robust and naive SE.

Future work may help address possible confounders in the data that may affect the response. A mixed effects model with a random slope could also add additional flexibility in the model and help explain variance in the data.

In the SRE model, an increase in the longitudinal effect of depression increased the OR, while a baseline effect was no longer significant on daily smoking at the individual-level. However, at the population-level, the GEE models did show an increase in the OR for the cross-sectional effect.

Works Cited - Data Source

Harris, Kathleen Mullan; Udry, Richard J., "National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave I, 1994-1995", doi:10.15139/S3/11900 UNF:5:ibLuHikTfzPYUvsYuDR49g== Odum Institute for Research in Social Science [Distributor] V3 [Version]

Harris, Kathleen Mullan; Udry, Richard J., "National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave II, 1996", http://dx.doi.org/10.15139/S3/11917 UNF:5:SXjreDd7s+9mPEFevnn37w== Odum Institute for Research in Social Science [Distributor] V4 [Version]

Harris, Kathleen Mullan; Udry, Richard J., "National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave III, 2001-2002", http://dx.doi.org/10.15139/S3/11918 UNF:5:oEDvveWmy0GDTy6G0iIfSA== Odum Institute for Research in Social Science [Distributor] V3 [Version]

Harris, Kathleen Mullan; Udry, Richard J., "National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave IV, 2008", http://dx.doi.org/10.15139/S3/11920 UNF:5:ZvaaD83oTDL8DJa7kZ7nrQ== Odum Institute for Research in Social Science [Distributor] V3 [Version]