Longitudinal Analysis of Young Adolescents’ Relationships with Family Members

WELLESLEY

Clara Kim ‘23, Data Science Major Capstone

Background and Research Questions

Although adolescents were generally at a lower risk of serious health complications from COVID-19, spending extensive time in isolation during the pandemic may impact their social skills and sense of identity [1]. During a pandemic, family is an especially important sphere of social interactions for adolescents, as most family members stay at home. Research has shown that intermediate or poor relationships with family members is correlated with a higher likelihood of premature death in adolescents [2] whereas positive relationships with lower likelihood of depression throughout adulthood [3]. Thus, understanding how the adolescents’ relationship with their family members changed throughout the recent pandemic is important and interesting. It may elucidate suggestions for parents to build healthy relationships with their adolescent children.

This capstone asks: **did the pandemic negatively affect the teenagers’ relationship with their family members? If so, what were the main contributing factors?**

Data

Description

This data set is from an ongoing longitudinal survey study conducted by the Youth, Media & Wellbeing (YMW) Lab at the Wellesley Centers for Women [4]. The study subjects are students in 6th–9th grade in two school districts in Northeastern United States (n=968) with a mean age of 13.02 (SD = 1.43) [5]. I used 26 survey questions and two demographic variables from Wave 4 of the study and extracted 135 variables, collected over 5 waves from 2017 - 2021.

	Data Collection Period
Wave 1	October 2017 - December 2017
Wave 2	October 2018 - December 2018
Wave 3	October 2019 - December 2019
Wave 4	October 2020 - December 2020
Wave 5	October 2021 - December 2021

Cleaning

The initial survey data contained 135 columns and 968 rows, with missing values in multiple fields. However, we did not impute these missing values due to ethical concerns regarding some predictors (e.g. “queer” or “mood.score”). Instead, the data cleaning process involved

(1) recoding the response variable,
(2) recoding predictor variables, and
(3) removing columns and rows that were missing too many fields.

I recoded the remaining variables with the objective of (1) combining similar survey questions under one variable for easier interpretation and (2) reducing the feature space. The recoding process did sacrifice some information contained in the original data but yielded cleaner and more interpretable predictors. In the recoded data, 335 out of 968 observations contained at least one missing value:

- First, I removed rows missing the response variable.
- Next, I removed columns that were missing in at least 10% of the rows.
- Finally, I removed the remaining rows that still contained missing values.

These 3 steps removed a total of 89 (~9.19%) rows and 12 columns. The final clean data contained 879 rows and 26 columns, which still provided sufficient statistical power and predictors with which to fit a regression model.

Data Modeling: Logistic Regression (Wave 4)

Model Fitting

As a final verdict, we conducted a likelihood ratio test comparing the two nested models (AIC vs. BIC model). This test yielded a p-value of ~0.001 and concluded that the AIC model, indeed, had a significantly better fit than the BIC model. Further model refinement upon the AIC model was considered but found to be insignificant in performance improvement.

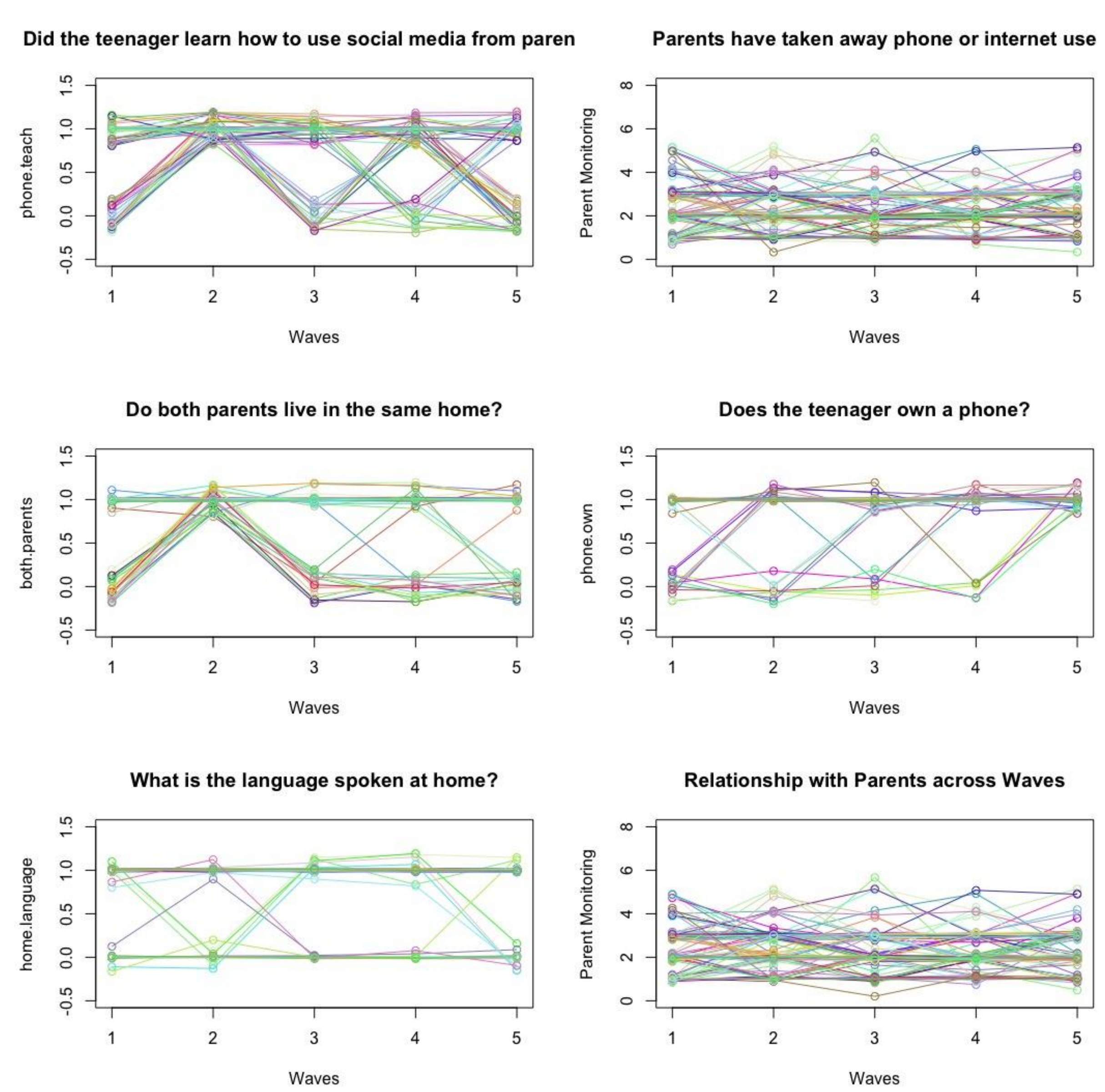
Model Evaluation & Selection

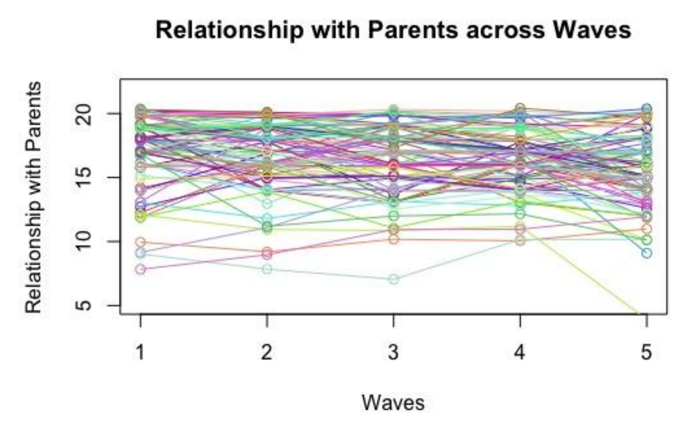
Due to the aforementioned imbalance in the response variable, a conventional discriminatory threshold of 0.5 to convert the predicted probabilities did not yield the best classification performance. We found an alternative threshold value for each of the two models based on the F-measure, which combines precision and recall, to optimize overall model performance. Rather, the AIC model performed optimally at a threshold of 0.25 and the BIC model at a threshold of 0.20 based on the results from k-fold cross-validation (k=10). With these new thresholds, both models fit the data as well as a saturated model but the AIC model performed in every performance metric except BIC and sensitivity.

As a final verdict, we conducted a likelihood ratio test comparing the two nested models (AIC vs. BIC model). This test yielded a p-value of ~0.001 and concluded that the AIC model, indeed, had a significantly better fit than the BIC model. Further model refinement upon the AIC model was considered but found to be insignificant in performance improvement.

Predictor	Est. Coefficient	Est. Odds	Associated Question
home.lang	0.529	1.698	Is English the main language spoken at home?
both.parents	0.491	1.635	Do both parents live in the same home?
rel.parents	-0.314	0.730	Relationship with parents (higher is stronger relationship).
phone.own	-0.917	0.379	Do you own a phone?
parmon3	0.434	1.545	My parents have taken away my cell phone or internet use.
phone.teach	-0.533	0.587	My parents taught me about social media.
iep	-0.416	0.660	I have an Individualized Educational Plan (IEP) or need special assistance during the school day.
hisp	-0.561	0.570	I am of Hispanic origin.

Data Modeling: Time Series Analysis (Waves 1-5)





There were significantly less participants (n=108) that participated in all 5 waves of the survey due to attrition. I investigated the overall patterns of the predictors of interest (identified in my final logistic regression model). The following jittered line plots display how the variables of interest change over time. There are not many meaningful conclusions I drew from this initial analysis yet but I plan to continue investigating.

Results

The logistic regression model finds that the eight diverse predictors found in the table are particularly interesting in answering our research question. The binary categorical predictors were if English was the main language spoken at home, both parents lived in the same home, the teenager owned a phone, parents had taken away cell phone or internet use, parents taught them about social media, having an IEP, and being of Hispanic origin. The quantitative variable was the self-evaluation of relationship with parents, with a larger number indicating a stronger relationship.

Data Ethics and Limitations

Due to ethical concerns, I chose not to impute any of the missing data but some rows of the data set as a result. For variables such as `mood.score` (a quantitative score calculated from the CESD-10 Depression Scale questions), imputing missing values would mean that we are guessing whether an adolescent had no, mild, moderate, or severe depression symptoms. For identity variables like `gender`, the missing values may not be missing at random (i.e. genderqueer individuals may not feel comfortable sharing their gender identity on a survey, so they may be more likely to leave this question blank).

References

[1] Volkin, S. (2020, May 11). The impact of the COVID-19 pandemic on adolescents. The Hub. Retrieved November 22, 2022, from https://hub.jhu.edu/2020/05/11/covid-19-and-adolescents/

[2] Chen, P and Mullan Harris, H. Association of positive family relationships with mental health trajectories from adolescence to midlife. JAMA Pediatrics.2019. doi:10.1001/jamapediatrics.2019.3336

[3] Alm, S., Brolin Låftman, S., & Bohman, H. (2019). Poor Family Relationships in Adolescence and the Risk of Premature Death: Findings from the Stockholm Birth Cohort Study. International journal of environmental research and public health, 16(10), 1690. https://doi.org/10.3390/ijerph16101690

[4] Charmaraman, L. (2018). Early Adolescent Social Technology Use and Parental Monitoring: Implications for Psychosocial and Behavioral Health. Youth, Media & Wellbeing Research Lab. https://www.wcwonline.org/Active-Projects/early-adolescents-social-technology-use

[5] Charmaraman, L., Lynch, A. D., Richer, A. M., & Zhai, E. (2022). Examining early adolescent positive and negative social technology behaviors and wellbeing during the COVID-19 pandemic. Technology, Mind, & Behavior, 3(1). https://doi.org/10.1037/tmb0000062