Mississippi State University

Data Analysis One

Emily Hubbard

December 11, 2025

Final Report

1. Executive Summary

A. Introduction:
Within the Bipolar I community, suicide is a real risk and safety concern. Compared to the general population, bipolar patients have a 20 times higher suicide risk, experiencing higher suicidal ideation, planning, and attempts (Izadi et al., 2023). Simultaneously, religion and spirituality can play a prominent role in reducing suicide risk within the public (Lawrence et al., 2016). However, some studies add complexity to this sentiment. For example, negative religious coping has been linked to worse outcomes in a sample of psychotic patients (Rosmarin et al., 2013).

So, despite the clinical significance of these research areas, there is little known about the overlap or relationship between religiosity and suicidality components (ideation, planning, and attempts) within the Bipolar I community.

With this study, my objective was to test whether a standardized religiosity index that combined importance of religion, prayer frequency, self-rated religiosity, and service attendance was associated with lifetime suicidality within the NSAL Bipolar I subgroup. In addition, I tested models that adjusted for sex, age, and the combination of the two. If a significant association is found, treatment options and recommended self-care practices could be informed and improved. If there is no association, those providing treatment can assess suicidal and religious struggles directly and separately if need be.

Our final analysis set includes BP-I respondents (N=40) with suicidality and all four religiosity items and 22 event occurrences. With a relatively small N and event occurences, it is important to be cautious when creating models. To keep the model reliable, I implemented a bias-reduced logistic regression (Firth), combined four religiosity items into a single standardized index, and limited covariates to a small set (age and sex) to avoid overfitting.

B. Data
The flow chart summarizes the data source, step-wise inclusion filters, and the final analytic

cohort: CPES merged file → NSAL → lifetime Bipolar I → respondents with both suicidality and religiosity items.

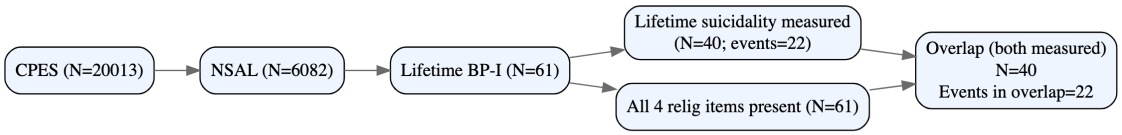

Figure 0.1: Data Flow Chart

The table below lists all variables with short descriptions, measurement type, and links to the appendix for more information.

CPES / NSAL Bipolar & Suicidality: Variable Cheat Sheet

Cohort / IDs

Variable	Description
CPESPROJ	CPES component study (1 = NCS-R; 2 = NLAAS; **3 = NSAL**). Filter to CPESPROJ == 3.
CPESCASE	Respondent ID (unique within CPES merged file).

Bipolar Filter

Variable	Description	Type	Appendix
V07842	**Lifetime Bipolar I**. 1 = endorsed; 5 = not endorsed (special codes → NA). Keep V07842 == 1.	Binary (Yes/No)	*Figure A.1*

Lifetime Suicidality

Variable	Description	Type	Appendix
V00876	Severe depression episode — **ever** thought about suicide (1 = Yes; 5 = No; specials → NA).	Binary (Yes/No)	*Figure A.2*
V00880	Severe depression episode — **ever** made a suicide plan (1 = Yes; 5 = No).	Binary (Yes/No)	*Figure A.3*
V00882	Severe depression episode — **ever** attempted suicide (1 = Yes; 5 = No).	Binary (Yes/No)	*Figure A.4*
V02044	**Ever** attempted suicide (lifetime; outside MDE block) (1 = Yes; 5 = No).	Binary (Yes/No)	*Figure A.5*

Religiosity Predictors

Variable	Description	Type	Appendix
V06618	Importance of religion in your life (1 = Very; 2 = Fairly; 3 = Not too; 4 = Not at all). Collapsed to High vs Low.	Ordinal (4) → Binary	*Figure A.6*
V06614	How often do you pray (1 = Nearly everyday; 2 = At least once a week; 3 = Few times a month; 4 = At least once a month; 5 = Few times a year; 6 = Never). Collapsed to Weekly+ / Yearly–Monthly / Never.	Ordinal (6) → 3-level	*Figure A.7*
V06621	How religious are you (1 = Very; 2 = Fairly; 3 = Not too; 4 = Not at all). Used as ordinal (or collapsed if needed).	Ordinal (4)	*Figure A.8*
V06593	Ever attended church services since age 18 (1 = Yes; 5 = No).	Binary (Yes/No)	*Figure A.9*
relig_index	**Composite index (0–1).** Mean of available scored items (require ≥3 answered); higher=more religious.	Continuous (0–1)	
relig_z	**Standardized index.** z-score of relig_index computed within the analysis set.	Continuous (z)	

Covariates: Age & Sex

Variable	Description	Type	Appendix
V07306	Respondent **age in years** at interview. CPES special missings $(-9, -8, -7, 97, 98, 99)$ set to NA. In analysis we standardize to age_z.	Continuous (years) → z-score	*Figure A.10*
V09036	Respondent **sex** (typically 1=Male, 2=Female). In analysis we recode to sex_male (Male=1, Female=0).	Binary (Male/Female)	*Figure A.11*

C. Summary Information

Preliminary results show that, in NSAL participants with lifetime Bipolar I who had both religiosity and suicidality measured (N=40; events=22), religiosity was not associated with lifetime suicidality. This was made evident using models with Firth correction as well as standard MLE (Maximum Likelihood Estimation) procedures. Even after adjusting for age and sex, no association was found between the two.

Below is a table of results from all fitted models that will be covered in this report.

Model	Predictor	OR	95% CI
Firth logit	Religiosity z	1.14	0.62–2.11
MLE logit	Religiosity z	1.15	0.62–2.15
Firth logit	Religiosity z + Age z	1.16	0.62–2.24
Firth logit	Religiosity z + Male	1.03	0.52–2.03
Firth logit	Religiosity z + Age z + Male	1.00	0.51–1.98

Table 0.6: Primary models (unweighted).

2. Data Analysis

A. Preliminary Visuals

To start, I took a look at the raw counts of suicidality outcomes within the NSAL Bipolar I subgroup.

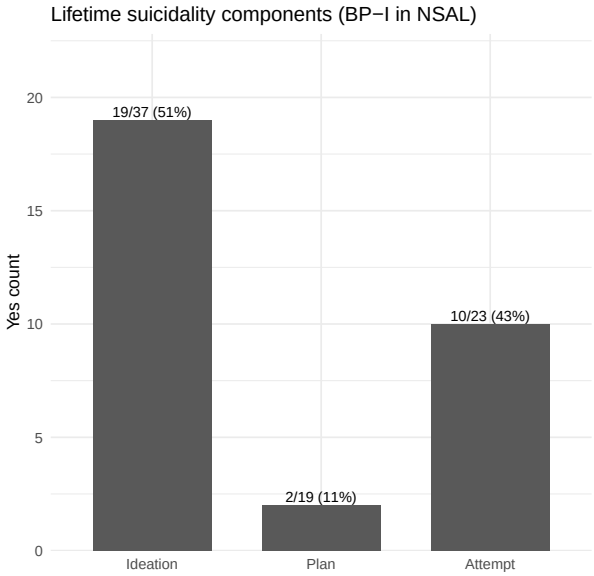

Figure 0.2: Lifetime Suicidality Components within Bipolar I subgroup

This figure shows the breakdown of each variable that contributes to the "lifetime suicidality" outcome. Note that they are not mutually exclusive and respondents can choose more than one.

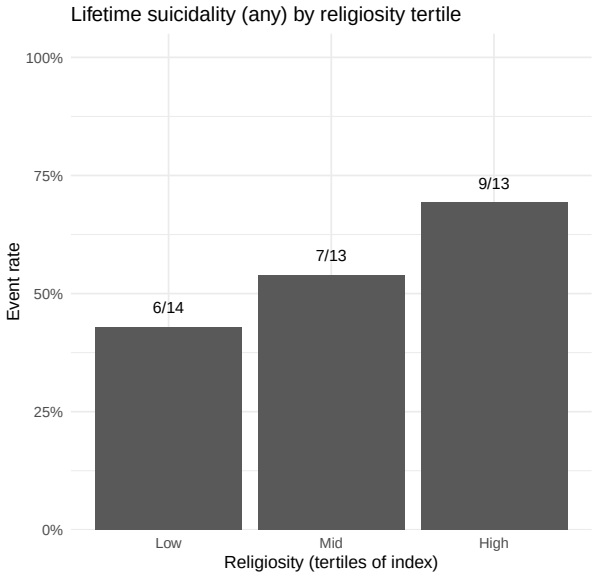

Figure 0.3: Counts of Lifetime Suicidality by Religious Tertile

This figure shows event counts for any lifetime suicidality (ideation, planning, or attempt) split by religiosity tertiles. Tertiles are based on the standardized religiosity index (z-score). Higher Score = Higher Religiosity. 'Low' is the bottom third of the index ($\leq$ 33rd percentile), 'Mid' is the middle third (33rd–67th percentiles), and 'High' is the top third ($>$ 67th percentile) among the analysis sample.

Although this figure does show a patterned increase by tertile, it is important to remember this is a small sample and the trend could be explained by simple variation. As such, a figure like this is simply exploratory in nature.

B. Model Fitting

Now, to begin model construction. Using a Firth model and the standardized religiosity measure to predict lifetime suicidality outcome, the following results were produced.

```
Firth logistic (penalized) on overlap:
logistf::logistf(formula = suic_life_any ~ relig_z, data = anal)

Model fitted by Penalized ML
Coefficients:
                coef  se(coef) lower 0.95 upper 0.95     Chisq         p method
(Intercept) 0.1880911 0.3107632 -0.4202123  0.8076984 0.3678082 0.5442014      2
relig_z     0.1312430 0.3041131 -0.4732444  0.7483473 0.1866186 0.6657459      2

Firth ORs (95% CI):
# A tibble: 2 × 4
  term           OR   low  high
  <chr>       <dbl> <dbl> <dbl>
1 (Intercept)  1.21 0.657  2.24
2 relig_z      1.14 0.623  2.11
```

Figure 0.4: Firth Model Results

Model Formula:

$$\text{logit}\{\Pr(\texttt{suic\_life\_any} = 1 \mid \texttt{relig\_z})\} = \beta_0 + \beta_1 \texttt{relig\_z}$$

With estimates $\widehat{\beta}_0 = 0.188$ and $\widehat{\beta}_1 = 0.131$,

$$\widehat{\text{logit}} = 0.19 + 0.13 \cdot \texttt{relig\_z}$$

Equivalently, the odds ratio for a one–standard deviation increase in religiosity is

$$\exp(\widehat{\beta}_1) = \exp(0.131) = 1.14 \text{ (95\% CI: 0.62, 2.11).}$$

What this odds ratio means is that for a 1 unit increase in the religiosity index , there is a 14% increase in the odds of lifetime suicidality. However, the confidence interval includes 1 which means the findings are insignificant. This can also be concluded from the large p-value (0.67) associated with the religiosity index.

Next, we will look at a standard MLE model to confirm our findings.

```
MLE logistic on overlap:

Call:
glm(formula = suic_life_any ~ relig_z, family = binomial(), data = anal)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.1941     0.3189   0.609    0.543
relig_z       0.1392     0.3189   0.437    0.662
```

Figure 0.5: MLE Model Results

Model Formula:

$$\widehat{\text{logit}} = 0.19 + 0.14 \cdot \texttt{relig\_z}$$

Equivalently, the odds ratio for a one–standard deviation increase in religiosity is

$$\exp(\widehat{\beta}_1) = \exp(0.139) = 1.15 \text{ (95\% CI: 0.62, 2.15).}$$

As with the Firth model, the confidence interval and large p-value (0.66) indicate insignificant association between religiosity and lifetime suicidality.

As a final measure, we will look at models that adjust for age, sex, and the combination of the two.

Here are the results of those three models:

Model A (Religiosity + Age)

```
logistf::logistf(formula = suic_life_any ~ relig_z + age_z, data = dat_A)

Model fitted by Penalized ML
Coefficients:
                  coef  se(coef) lower 0.95 upper 0.95     Chisq         p method
(Intercept)  0.1895842 0.3096040 -0.4171830  0.8083562 0.3755960 0.5399697      2
relig_z      0.1517088 0.3049296 -0.4549605  0.7698701 0.2476313 0.6187479      2
age_z       -0.2522519 0.3084065 -0.9357480  0.3411377 0.6926318 0.4052707      2
```

Model B (Religiosity + Sex)

```
Model fitted by Penalized ML
Coefficients:
                  coef  se(coef) lower 0.95 upper 0.95       Chisq          p method
(Intercept)  0.65506776 0.4261450 -0.1557474  1.5476318 2.488612566 0.11467277      2
relig_z     -0.02678976 0.3262564 -0.7232809  0.6088292 0.006677087 0.93487459      2
sex_male    -1.13094417 0.6749247 -2.5217398  0.1637007 2.922754788 0.08733864      2
```

Model C (Religiosity + Age +Sex)

```
Model fitted by Penalized ML
Coefficients:
                   coef  se(coef) lower 0.95 upper 0.95        Chisq         p method
(Intercept)  0.636181551 0.4234993 -0.1713462  1.5232169 2.3654560102 0.1240477      2
relig_z     -0.002898039 0.3258079 -0.6967431  0.6331145 0.0000780553 0.9929509      2
age_z       -0.231773024 0.3262408 -0.9755311  0.3818506 0.5164253300 0.4723705      2
sex_male    -1.082966027 0.6698836 -2.4639125  0.2067300 2.7003535087 0.1003260      2
```

Figure 0.6: Adjusted Firth Models

Model Formulas:

$$\textbf{A: } \widehat{\text{logit}} = 0.19 + 0.15 \cdot \texttt{relig\_z} - 0.25 \cdot \texttt{age\_z}$$

$$\textbf{B: } \widehat{\text{logit}} = 0.66 - 0.03 \cdot \texttt{relig\_z} - 1.13 \cdot \texttt{sex\_male}$$

$$\textbf{C: } \widehat{\text{logit}} = 0.64 - 0.00 \cdot \texttt{relig\_z} - 0.23 \cdot \texttt{age\_z} - 1.08 \cdot \texttt{sex\_male}$$

Model	Predictor	OR	95% CI	P-values
Firth logit	Religiosity z + Age z	1.16	0.62–2.24	0.62, 0.41
Firth logit	Religiosity z + Male	1.03	0.52–2.03	0.94, 0.09
Firth logit	Religiosity z + Age z + Male	1.00	0.51–1.98	0.99, 0.47, 0.10

Table 0.7: Primary models (unweighted).

Adjusting for age and sex leaves us with the same conclusion, there is no significant relationship between religiosity and lifetime suicidality.

C. Results

In the NSAL Bipolar I subset with complete measures in religiosity and lifetime suicidality (N=40; events=22), the standardized religiosity index was not associated with lifetime suicidality: Firth logistic odds ratio per 1-standard deviation = 1.14 (95% CI 0.62–2.11, p=0.67). Adjusted models produced similar results, adding age, sex, or both. Descriptively, in Figure 0.3, event rates were higher in the "higher religiosity" tertile, but the small N means the figure is not significantly reliable.

Within this particular dataset, higher religiosity did not increase or decrease lifetime suicidality among individuals with Bipolar I. In practice, this means clinicians should assess suicide risk directly without assuming protection or harm from religious leaning.

D. Limitations

Small analysis set (N=40) with 22 events; cross-sectional measurements (some lifetime, some current); composite religiosity index based on equal weights could potentially be inaccurate in capturing religiosity.

APPENDIX

A. Variables Explained:

V07842	DSM-IV Bi-Polar I (Lifetime)
Location:	11652-11653 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	DSM-IV Bi-Polar I (Lifetime)

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	0	0.0 %	-
-8 (M)	DONT KNOW	0	0.0 %	-
1	ENDORSED	162	0.8 %	1.1%
5	NOT ENDORSED	15009	75.0 %	98.9%
. (M)	-	4842	24.2 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
15171	4842	1.00	5.00	4.96	5.00	0.41

Figure A.1: V07842 (Lifetime Bipolar I)

V00876	Severe dep episode-thought about suicide
Location:	652-654 (width: 3; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	D26cc F (RB, PG 6-7 FOR EACH ITEM ENDORSED, ASK R TO MARK IT IN THE RB.) Did you think about committing suicide?

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	6	0.0 %	-
-8 (M)	DONT KNOW	5	0.0 %	-
1	YES	1268	6.3 %	31.2%
5	NO	2792	14.0 %	68.8%
. (M)	-	15942	79.7 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
4060	15953	1.00	5.00	3.75	5.00	1.85

Figure A.2: V00876 (Suicidal Thought)

V00880	Severe dep episode-made suicide plan
Location:	655-657 (width: 3; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	D26dd F (RB, PG 6-7 FOR EACH ITEM ENDORSED, ASK R TO MARK IT IN THE RB.) Did you make a suicide plan?

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	0	0.0 %	-
-8 (M)	DONT KNOW	3	0.0 %	-
1	YES	451	2.3 %	35.7%
5	NO	814	4.1 %	64.3%
. (M)	-	18745	93.7 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
1265	18748	1.00	5.00	3.57	5.00	1.92

Figure A.3: V00880 (Suicidal Plan)

V00882	Severe dep episode-attempted suicide
Location:	658-659 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	D26ee F (RB, PG 6-7 FOR EACH ITEM ENDORSED, ASK R TO MARK IT IN THE RB.) Did you make a suicide attempt?

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	0	0.0 %	-
-8 (M)	DONT KNOW	0	0.0 %	-
1	YES	386	1.9 %	30.4%
5	NO	882	4.4 %	69.6%
. (M)	-	18745	93.7 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
1268	18745	1.00	5.00	3.78	5.00	1.84

Figure A.4: V00882 (Suicidal Attempt During Depression)

V02044	Ever attempted suicide
Location:	2575-2577 (width: 3; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	SD6 F (RB, PG 21) Now look at the third of the three experiences on the list, Experience C. Did experience C ever happen to you ? EXPERIENCE C IS 'YOU ATTEMPTED SUICIDE'

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	1	0.0 %	-
-8 (M)	DONT KNOW	2	0.0 %	-
1	YES	703	3.5 %	33.7%
5	NO	1385	6.9 %	66.3%
. (M)	-	17922	89.6 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
2088	17925	1.00	5.00	3.65	5.00	1.89

Figure A.5: V02044 (Suicidal Attempt)

V06618 — **Importance of religion in your life**

Location: 8161-8163 (width: 3; decimal: 0)
Variable Type: numeric (ISO)
Interval: discrete
Range of Missing Values (M): -9 , -8 , .
Question: B17c. How important is religion in your life? (Would you say that it is very important, fairly important, not too important or not important at all?)

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	3	0.0 %	-
-8 (M)	DONT KNOW	9	0.0 %	-
1	VERY IMPORTANT	4586	22.9 %	75.6%
2	FAIRLY IMPORTANT	991	5.0 %	16.3%
3	NOT TOO IMPORTANT	335	1.7 %	5.5%
4	NOT IMPORTANT AT ALL	158	0.8 %	2.6%
. (M)	-	13931	69.6 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
6070	13943	1.00	4.00	1.35	1.00	0.70

Figure A.6: V06618 (Religious Importance)

V06614 — **Frequency of praying**

Location: 8149-8151 (width: 3; decimal: 0)
Variable Type: numeric (ISO)
Interval: discrete
Range of Missing Values (M): -9 , -8 , .
Question: B16d. F (RB, PG 7) How often do you pray? (Would you say nearly every day, at least once a week, a few times a month, at least once a month, a few times a year or never?)

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	5	0.0 %	-
-8 (M)	DONT KNOW	1	0.0 %	-
1	NEARLY EVERYDAY	4959	24.8 %	81.6%
2	AT LEAST ONCE A WEEK	417	2.1 %	6.9%
3	FEW TIMES A MTH	241	1.2 %	4.0%
4	AT LEAST ONCE A MONTH	100	0.5 %	1.6%
5	FEW TIMES A YR	192	1.0 %	3.2%
6	NEVER	167	0.8 %	2.7%
. (M)	-	13931	69.6 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
6076	13937	1.00	6.00	1.46	1.00	1.16

Figure A.7: V06614 (Prayer Frequency)

V06621	How religious are you
Location:	8170-8172 (width: 3; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	B18. How religious would you say you are - very religious, fairly religious, not too religious, or not religious at all?

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	7	0.0 %	-
-8 (M)	DONT KNOW	17	0.1 %	-
1	VERY RELIGIOUS	1892	9.5 %	31.2%
2	FAIRLY RELIGIOUS	3113	15.6 %	51.4%
3	NOT TOO RELIGIOUS	777	3.9 %	12.8%
4	NOT RELIGIOUS AT ALL	276	1.4 %	4.6%
. (M)	-	13931	69.6 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
6058	13955	1.00	4.00	1.91	2.00	0.78

Figure A.8: V06621 (Religious Self-Rating)

V06593	Ever attended church services since 18 years or older
Location:	8087-8089 (width: 3; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8 , .
Question:	B2. Other than for weddings or funerals, have you attended services at a church or other place of worship since you were 18 years old?

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	2	0.0 %	-
-8 (M)	DONT KNOW	1	0.0 %	-
1	YES	5580	27.9 %	91.8%
5	NO	499	2.5 %	8.2%
. (M)	-	13931	69.6 %	-

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
6079	13934	1.00	5.00	1.33	1.00	1.10

Figure A.9: V06593 (Church Attendance)

V07306	Age
Location:	7127-7128 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8
Question:	Age

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	0	0.0 %	-
-8 (M)	DONT KNOW	0	0.0 %	-
18	BOTTOM CODED AT 18	329	1.6 %	1.6%
19	-	395	2.0 %	2.0%
20	-	398	2.0 %	2.0%
21	-	403	2.0 %	2.0%
22	-	427	2.1 %	2.1%
23	-	398	2.0 %	2.0%
24	-	385	1.9 %	1.9%
25	-	383	1.9 %	1.9%
26	-	357	1.8 %	1.8%
27	-	433	2.2 %	2.2%
28	-	422	2.1 %	2.1%

Figure A.10: V07306 (Age)

V09036	Sex
Location:	7125-7126 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Range of Missing Values (M):	-9 , -8
Question:	Sex

Value	*Label*	*Frequency*	*%*	*Valid %*
-9 (M)	REFUSED	0	0.0 %	-
-8 (M)	DONT KNOW	0	0.0 %	-
1	MALE	8550	42.7 %	42.7%
2	FEMALE	11463	57.3 %	57.3%

Valid	*Invalid*	*Min*	*Max*	*Mean*	*Median*	*Stdev*
20013	0	1.00	2.00	1.57	2.00	0.49

Figure A.11: V09036 (Sex)

B. References

[Izadi et al., 2023] Izadi, N., Mitchell, R. H. B., Giacobbe, P., Nestor, D., Steinberg, R., Sinyor, M., & Schaffer, A. (2023). Suicide assessment and prevention in bipolar disorder: How current evidence can inform clinical practice. *Focus*, 21(4), 380–388. https://doi.org/10.1176/appi.focus.20230011

[Lawrence et al., 2016] Lawrence, R. E., Oquendo, M. A., & Stanley, B. (2016). Religion and suicide risk: A systematic review. *Archives of Suicide Research*, 20(1), 1–21. https://doi.org/10.1080/13811118.2015.1004494

[Rosmarin et al., 2013] Rosmarin, D. H., Bigda-Peyton, J. S., Öngur, D., Pargament, K. I., & Björgvinsson, T. (2013). Religious coping among psychotic patients: Relevance to suicidality and treatment outcomes. *Psychiatry Research*, 210(1), 182–187. https://doi.org/10.1016/j.psychres.2013.03.023

C. R Code

```
#### Load Packages ####
suppressPackageStartupMessages({
  library(haven)
  library(dplyr)
  library(stringr)
  library(tidyr)
  library(purrr)
})

#### Load Overall Data ####
path <- "20240-0001-Data.dta"   #Merged CPES file where NSAL-> CPESPROJ
    == 3
dat  <- read_dta(path)

# Sanity check that we're in the merged file
stopifnot("CPESPROJ" %in% names(dat))   # 1=NCS-R, 2=NLAAS, 3=NSAL

#Setting CPES special missing codes to NA
special_na <- function(x) replace(x, x %in% c(-9, -8, -7, 97, 98, 99),
   NA)

# Label from a haven-labelled column
get_label <- function(x) {
  lab <- attr(x, "label"); if (is.null(lab)) "" else as.character(lab)
}

#### Restrict to NSAL & Lifetime Bipolar One Respondents ####
#NSAL
nsal <- dat %>% filter(CPESPROJ == 3) #NSAL respondents only

#Bipolar One
nsal <- nsal %>%
  mutate(V07842 = special_na(V07842),
         bpi_life = case_when(V07842 == 1 ~ 1L,
                              V07842 == 5 ~ 0L,
                              TRUE ~ NA_integer_)) %>%
  filter(bpi_life == 1L). # V07842: 1=lifetime BP-I, 5=not lifetime BP-
     I (others NA)

#Prints rows left: 61
cat("Rows after NSAL + lifetime BP-I restriction:", nrow(nsal), "\n")

#### Search for Lifetime Suicidality Variables ####
#Var/label lookup within subset
lbl_tbl <- tibble::tibble(
  var   = names(nsal),
  label = vapply(nsal, get_label, FUN.VALUE = character(1))
)

#Inspect lifetime/ever suicidality candidates (by label search)
life_candidates <- lbl_tbl %>%
  mutate(label_low = str_to_lower(label)) %>%
  filter(
    str_detect(label_low, "suicid"),
    str_detect(label_low, "ever|lifetime|times attempted|ever attempted
       |ever thought|ever made|ever plan")
```

```
  ) %>%
  mutate(non_missing = map_int(var, ~ sum(!is.na(nsal[[.x]])))) %>%
  arrange(desc(non_missing))

#Prints all lifetime suicidality variables:
cat("\nLifetime suicidality candidates (sorted by coverage):\n")
print(life_candidates, n = 50)

#### Search for 12-month Suicidality Variables ####

#Find the 12-month suicidality items (by label search)
suic_candidates <- lbl_tbl %>%
  mutate(label_low = str_to_lower(label)) %>%
  filter(
    str_detect(label_low, "suicid") &
      (str_detect(label_low, "12") | str_detect(label_low, "past 12") |
         str_detect(label_low, "last 12") | str_detect(label_low, "12 m
           "))
  ) %>%
  mutate(non_missing = map_int(var, ~ sum(!is.na(nsal[[.x]])))) %>%
  arrange(desc(non_missing))

cat("\n--- Possible 12-month suicidality variables (merged file) ---\n"
   )
print(suic_candidates, n = 50)

#Build the 12-month suicidality outcome inside NSAL + BP-I
suic12_vars <- c("V01995",  # thought in past 12 mths
                 "V01999",  # plan in past 12 mths
                 "V02004")  # attempt in past 12 mths

# clean specials and build component/combined outcomes
nsal <- nsal %>%
  mutate(
    across(all_of(suic12_vars),
           ~ replace(., . %in% c(-9, -8, 97, 98, 99), NA_integer_)),
    suic12_thought = as.integer(.data[[suic12_vars[1]]] == 1L),
    suic12_plan    = as.integer(.data[[suic12_vars[2]]] == 1L),
    suic12_attempt = as.integer(.data[[suic12_vars[3]]] == 1L)
  ) %>%
  mutate(
    any_suic12 = as.integer(
      (suic12_thought == 1L) | (suic12_plan == 1L) | (suic12_attempt ==
          1L)
    )
  )

cat("\nCoverage per item (non-missing):\n")
print(sapply(suic12_vars, function(v) sum(!is.na(nsal[[v]]))))

cat("\n12-month suicidality (any):\n").
print(table(nsal$any_suic12, useNA = "ifany"))

cat("\nBreakdown by item (yes counts):\n")
print(colSums(nsal[, c("suic12_thought","suic12_plan","suic12_attempt")
   ] == 1, na.rm = TRUE))

cat("\nExample rows with any 12m suicidality:\n")
```

```
nsal %>%
  filter(any_suic12 == 1) %>%
  select(CPESCASE, any_suic12, suic12_thought, suic12_plan, suic12_
     attempt) %>%
  head(10) %>% print()

#### Realized there were not enough 12-month suicidality outcomes to
    create a stable model. Fell back to lifetime variables with larger N
    .

#Lifetime suicidality composites
life_items <- c("V00876", "V00880", "V00882", "V02044")  # ideation,
    plan, attempt-in-MDE, ever attempt
life_items <- intersect(life_items, names(nsal))

nsal <- nsal %>%
  mutate(across(all_of(life_items),
                ~ replace(., . %in% c(-9, -8, 97, 98, 99), NA_integer_)
                   ))

#Recode components as 0/1 with NAs preserved
yes1 <- function(x) as.integer(x == 1L)

comp_mat <- nsal %>%
  transmute(
    suic_life_ideation   = if ("V00876" %in% life_items) yes1(.data[["
       V00876"]]) else NA_integer_,
    suic_life_plan       = if ("V00880" %in% life_items) yes1(.data[["
       V00880"]]) else NA_integer_,
    suic_life_attemptMDE = if ("V00882" %in% life_items) yes1(.data[["
       V00882"]]) else NA_integer_,
    suic_life_attemptEver= if ("V02044" %in% life_items) yes1(.data[["
       V02044"]]) else NA_integer_
  )

# any lifetime & attempt lifetime with proper NA when all components
    missing
any_life <- {
  n_obs <- rowSums(!is.na(comp_mat))
  has_any <- as.integer(rowSums(comp_mat, na.rm = TRUE) > 0)
  has_any[ n_obs == 0 ] <- NA_integer_
  has_any
}
attempt_life <- {
  att_mat <- comp_mat[, c("suic_life_attemptMDE","suic_life_attemptEver
     ")]
  n_obs <- rowSums(!is.na(att_mat))
  has_any <- as.integer(rowSums(att_mat, na.rm = TRUE) > 0)
  has_any[ n_obs == 0 ] <- NA_integer_
  has_any
}

nsal <- bind_cols(nsal, comp_mat) %>%
  mutate(
    suic_life_any     = any_life,
    suic_life_attempt = attempt_life
  )
```

```
cat("\nCoverage␣(non-missing)␣per␣lifetime␣item:\n")
print(sapply(life_items, function(v) sum(!is.na(nsal[[v]]))))

cat("\nLifetime␣suicidality␣(any):\n")
print(table(nsal$suic_life_any, useNA = "ifany"))

cat("\nLifetime␣suicide␣attempt␣(composite):\n")
print(table(nsal$suic_life_attempt, useNA = "ifany"))

cat("\nBreakdown␣of␣'any'␣by␣lifetime␣components␣(yes␣counts):\n")
print(colSums(comp_mat == 1, na.rm = TRUE))

#### Religiosity items & Coverage/Overlap ####
relig4 <- c("V06618","V06614","V06621","V06593")

cov_tbl <- tibble::tibble(
  var         = relig4,
  label       = map_chr(relig4, ~ {lab <- attr(nsal[[.x]], "label"); if
      (is.null(lab)) .x else as.character(lab)}),
  non_missing = map_int(relig4, ~ sum(!is.na(nsal[[.x]]))),
  pct         = round(non_missing / nrow(nsal) * 100, 1)
)
cat("\nReligiosity␣coverage␣in␣NSAL␣+␣BP-I␣subset:\n")
print(cov_tbl)

nsal <- nsal %>%
  mutate(
    relig_all4     = rowSums(across(all_of(relig4), ~ !is.na(.))) == 4,
    relig_atleast2 = rowSums(across(all_of(relig4), ~ !is.na(.))) >= 2
  )

cat("\nOverlap␣with␣outcomes␣(primary=lifetime␣any;␣secondary=lifetime␣
    attempt):\n")
print(with(nsal, table(relig_all4,     suic_life_any,     useNA = "
    ifany")))
print(with(nsal, table(relig_atleast2, suic_life_any,     useNA = "
    ifany")))
print(with(nsal, table(relig_all4,     suic_life_attempt, useNA = "
    ifany")))
print(with(nsal, table(relig_atleast2, suic_life_attempt, useNA = "
    ifany")))

cat("\nRows␣that␣have␣lifetime-any=1␣AND␣all␣4␣relig␣items␣present␣(
    peek):\n")
nsal %>%
  filter(suic_life_any == 1, relig_all4) %>%
  transmute(
    CPESCASE,
    V06618 = haven::as_factor(V06618),
    V06614 = haven::as_factor(V06614),
    V06621 = haven::as_factor(V06621),
    V06593 = haven::as_factor(V06593)
  ) %>% print(n = Inf)


##### Flow Chart Visual ####
#Counts from your current objects
library(dplyr)
```

```
N_cpes   <- nrow(dat)
N_nsal   <- sum(dat$CPESPROJ == 3, na.rm = TRUE)
N_bpi    <- nrow(nsal)

nsal$suic_life_any <- ifelse(is.infinite(nsal$suic_life_any), NA, nsal$
    suic_life_any)

N_suic_nonmiss   <- sum(!is.na(nsal$suic_life_any))
N_suic_events    <- sum(nsal$suic_life_any == 1, na.rm = TRUE)
N_relig_all4     <- sum(nsal$relig_all4, na.rm = TRUE)
N_overlap_any    <- sum(nsal$relig_all4 & !is.na(nsal$suic_life_any),
    na.rm = TRUE)
N_overlap_events <- sum(nsal$relig_all4 & nsal$suic_life_any == 1, na.
    rm = TRUE)

#Flowchart
library(DiagrammeR); library(glue)

grViz(glue('
digraph␣flow␣{{
␣␣graph␣[rankdir=LR,␣fontsize=10]
␣␣node␣␣[shape=box,␣style="rounded,filled",␣fillcolor="#EEF5FF"]
␣␣edge␣␣[color="#6b6b6b"]

␣␣CPES␣[label="CPES␣(N={N_cpes})"]
␣␣NSAL␣[label="NSAL␣(N={N_nsal})"]
␣␣BPI␣␣[label="Lifetime␣BP-I␣(N={N_bpi})"]
␣␣SUIC␣[label="Lifetime␣suicidality␣measured\\n(N={N_suic_nonmiss};␣
    events={N_suic_events})"]
␣␣REL␣␣[label="All␣4␣relig␣items␣present␣(N={N_relig_all4})"]
␣␣OVER␣[label="Overlap␣(both␣measured)\\nN={N_overlap_any}\\nEvents␣in␣
    overlap={N_overlap_events}"]

␣␣CPES␣->␣NSAL␣->␣BPI
␣␣BPI␣␣->␣SUIC
␣␣BPI␣␣->␣REL
␣␣SUIC␣->␣OVER
␣␣REL␣␣->␣OVER
}}'))


#### Religiosity Index & Visuals ####
suppressPackageStartupMessages({
  library(ggplot2)
})

# Build a single religiosity index
# Mapping:
# V06618: 1=very, 2=fairly, 3=not too, 4=not at all  (High ->1, Low->0)
# V06614: 1=nearly every day, 2= 1 /wk, 3=few/m, 4= 1 /m, 5=few/yr,
    6=never (more ->1)
# V06621: 1=very religious, 2=fairly, 3=not too, 4=not at all (more
    ->1)
# V06593: 1=yes attended since 18, 5=no (yes ->1)
nsal <- nsal %>%
  mutate(
    relig_imp_num = case_when(V06618 %in% c(1,2) ~ 1,
```

```
                              V06618 %in% c(3,4) ~ 0,
                              TRUE ~ NA_real_),
    pray_num      = case_when(V06614 %in% c(1,2) ~ 1,
                              V06614 %in% c(3,4,5) ~ 0.5,
                              V06614 == 6          ~ 0,
                              TRUE ~ NA_real_),
    self_rel_num  = case_when(V06621 == 1 ~ 1,
                              V06621 == 2 ~ 2/3,
                              V06621 == 3 ~ 1/3,
                              V06621 == 4 ~ 0,
                              TRUE ~ NA_real_),
    attend_num    = case_when(V06593 == 1 ~ 1,
                              V06593 == 5 ~ 0,
                              TRUE ~ NA_real_),
    relig_items_answered = rowSums(across(c(relig_imp_num, pray_num,
       self_rel_num, attend_num),
                                          ~ !is.na(.))),
    relig_index = ifelse(relig_items_answered >= 3,
                         rowMeans(across(c(relig_imp_num, pray_num,
                            self_rel_num, attend_num)),
                                  na.rm = TRUE),
                         NA_real_),
    relig_z = as.numeric(scale(relig_index))
  )

#Analytic cohort:
anal <- nsal %>%
  filter(!is.na(suic_life_any), relig_all4, !is.na(relig_index)) %>%
  mutate(
    r_tertile = factor(
      dplyr::ntile(relig_index + runif(dplyr::n(), -1e-9, 1e-9), 3),
      labels = c("Low","Mid","High")
    )
  )


cat("\nAnalysis␣N␣(overlap):", nrow(anal),
    "␣|␣events:", sum(anal$suic_life_any == 1, na.rm = TRUE), "\n")

#Visuals:

#Component bar of lifetime suicidality (in the whole BP-I subset where
   measured)
comp_long <- nsal %>%
  filter(!is.na(suic_life_any)) %>%
  transmute(
    ideation = suic_life_ideation,
    plan     = suic_life_plan,
    attempt  = suic_life_attempt
  ) |>
  tidyr::pivot_longer(everything(), names_to = "component", values_to =
      "yes") |>
  dplyr::summarise(n_yes = sum(yes == 1, na.rm = TRUE), .by = component
     )

p1 <- ggplot(comp_long, aes(x = component, y = n_yes)) +
  geom_col() +
  labs(title = "Lifetime␣suicidality␣components␣(BP-I␣in␣NSAL)",
```

```
       x = NULL, y = "Yes count") +
  theme_minimal(base_size = 12)

print(p1)

library(dplyr)
library(tidyr)
library(ggplot2)
library(scales)

comp_long <- nsal %>%
  filter(!is.na(suic_life_any)) %>%
  transmute(
    ideation = suic_life_ideation,
    plan     = suic_life_plan,
    attempt  = suic_life_attempt
  ) %>%
  pivot_longer(everything(), names_to = "component", values_to = "yes")
      %>%
  # summarize counts + denominators for each component
  summarise(
    n      = sum(!is.na(yes)),
    n_yes  = sum(yes == 1, na.rm = TRUE),
    prop   = n_yes / n,
    .by    = component
  ) %>%
  # nice labels and ordering
  mutate(
    component = factor(component,
                       levels = c("ideation","plan","attempt"),
                       labels = c("Ideation","Plan","Attempt")),
    label_counts   = as.character(n_yes),
                                        # just counts
    label_full     = paste0(n_yes, "/", n, " (", percent(prop, 1), ")")
         # counts + %
  )

#Visual:
p1_counts <- ggplot(comp_long, aes(x = component, y = n_yes)) +
  geom_col(width = 0.7) +
  geom_text(aes(label = label_counts), vjust = -0.3, size = 4) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.15))) +
  labs(title = "Lifetime suicidality components (BP-I in NSAL)",
       x = NULL, y = "Yes count") +
  theme_minimal(base_size = 12)
print(p1_counts)

p1_full <- ggplot(comp_long, aes(x = component, y = n_yes)) +
  geom_col(width = 0.7) +
  geom_text(aes(label = label_full), vjust = -0.3, size = 3.6) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.20))) +
  labs(title = "Lifetime suicidality components (BP-I in NSAL)",
       x = NULL, y = "Yes count") +
  theme_minimal(base_size = 12)
print(p1_full)

#Event rate by religiosity tertile (in analysis overlap  N 40 )
rate_tbl <- anal |>
```

```
  dplyr::summarise(
    n      = dplyr::n(),
    events = sum(suic_life_any == 1, na.rm = TRUE),
    prop   = events / n,
    label  = paste0(events, "/", n),
    .by    = r_tertile
  )

p2 <- ggplot(rate_tbl, aes(x = r_tertile, y = prop)) +
  geom_col() +
  # put the counts just above each bar:
  geom_text(aes(y = prop + 0.03, label = label), vjust = 0, size = 4) +
  scale_y_continuous(
    labels = scales::percent_format(accuracy = 1),
    limits = c(0, 1),
    expand = expansion(mult = c(0, 0.05))
  ) +
  labs(
    title = "Lifetime␣suicidality␣(any)␣by␣religiosity␣tertile",
    x = "Religiosity", y = "Event␣rate"
  ) +
  theme_minimal(base_size = 12)

print(p2)



#### First Model: Firth Logistic & MLE for Comparison ####
#Helper to print ORs nicely
or_table <- function(fit, use_profile = FALSE) {
  cf <- coef(fit);  # vector
  SE <- sqrt(diag(vcov(fit)))
  if (!use_profile) {
    ci <- cbind(cf - 1.96*SE, cf + 1.96*SE)
  } else {
    ci <- confint(fit)  # slow, profile CI
  }
  tibble::tibble(
    term = names(cf),
    OR   = exp(cf),
    low  = exp(ci[,1]),
    high = exp(ci[,2])
  )
}


library(logistf)
#Primary predictor = relig_z (per +1 SD religiosity)
#Firth
if (requireNamespace("logistf", quietly = TRUE)) {
  fit_firth <- logistf::logistf(suic_life_any ~ relig_z, data = anal)
  cat("\nFirth␣logistic␣(penalized)␣on␣overlap:\n"); print(summary(fit_
     firth))
  firth_or <- tibble::tibble(
    term = names(fit_firth$coefficients),
    OR   = exp(fit_firth$coefficients),
    low  = exp(fit_firth$ci.lower),
    high = exp(fit_firth$ci.upper)
```

```
  )
  cat("\nFirth␣ORs␣(95%␣CI):\n"); print(firth_or)
} else {
  cat("\nPackage␣'logistf'␣not␣installed␣   ␣running␣plain␣MLE␣glm()␣
     only.\n")
}

#Plain MLE (for reference)
fit_mle <- glm(suic_life_any ~ relig_z, data = anal, family = binomial
    ())
cat("\nMLE␣logistic␣on␣overlap:\n"); print(summary(fit_mle))
cat("\nMLE␣ORs␣(Wald␣95%␣CI):\n"); print(or_table(fit_mle))


#### Running Model to include Age and Sex ####
#Clean covariates (Age = V07306, Sex = V09036) and add to analytic
    cohort
anal <- anal %>%
  mutate(
    age_raw  = special_na(V07306),
    age_num  = suppressWarnings(as.numeric(age_raw)),
    age_z    = as.numeric(scale(age_num)),           # z-score (mean 0,
        sd 1)
    sex_raw  = special_na(V09036),
    # CPES/NSAL convention is typically 1=Male, 2=Female
    sex_male = dplyr::case_when(sex_raw == 1L ~ 1L,
                                sex_raw == 2L ~ 0L,
                                TRUE          ~ NA_integer_)
  )

# Quick coverage check inside the overlap set  y o u ll  model
cat("\nCoverage␣in␣overlap␣set:\n")
print(tibble::tibble(
  N_overlap = nrow(anal),
  N_events  = sum(anal$suic_life_any == 1, na.rm = TRUE),
  N_age     = sum(!is.na(anal$age_z)),
  N_sex     = sum(!is.na(anal$sex_male)),
  N_both    = sum(!is.na(anal$age_z) & !is.na(anal$sex_male))
))

#Helper to fit Firth and return a tidy OR table
or_table_firth <- function(fit) {
  ci <- suppressMessages(confint(fit))
  tibble::tibble(
    term = names(coef(fit)),
    OR   = exp(coef(fit)),
    low  = exp(ci[, 1]),
    high = exp(ci[, 2])
  )
}

#Fit Adjusted Firth models:

# (A) relig_z + age
dat_A <- anal %>% dplyr::filter(!is.na(suic_life_any), !is.na(relig_z),
    !is.na(age_z))
fit_A <- logistf::logistf(suic_life_any ~ relig_z + age_z, data = dat_A
   )
```

```
cat("\nFirth: suic_life_any ~ relig_z + age_z  (n=", nrow(dat_A),
    ", events=", sum(dat_A$suic_life_any==1), ")\n", sep = "")
print(or_table_firth(fit_A))

# (B) relig_z + sex
dat_B <- anal %>% dplyr::filter(!is.na(suic_life_any), !is.na(relig_z),
    !is.na(sex_male))
fit_B <- logistf::logistf(suic_life_any ~ relig_z + sex_male, data =
   dat_B)
cat("\nFirth: suic_life_any ~ relig_z + sex_male  (n=", nrow(dat_B),
    ", events=", sum(dat_B$suic_life_any==1), ")\n", sep = "")
print(or_table_firth(fit_B))

# (C) relig_z + age + sex
dat_C <- anal %>% dplyr::filter(!is.na(suic_life_any), !is.na(relig_z),
                                !is.na(age_z), !is.na(sex_male))
fit_C <- logistf::logistf(suic_life_any ~ relig_z + age_z + sex_male,
   data = dat_C)
cat("\nFirth: suic_life_any ~ relig_z + age_z + sex_male  (n=", nrow(
   dat_C),
    ", events=", sum(dat_C$suic_life_any==1), ")\n", sep = "")
print(or_table_firth(fit_C))

# (A) relig_z + age
summary(fit_A)
# (B) relig_z + sex
summary(fit_B)
# (C) relig_z + age + sex
summary(fit_C)
```