**Predict**

Lapse (EMA\_1). Values 1=No, 2=Yes. This is reported 4x per day.

We will code and predict a daily lapse label for each participant such that it is YES, if there is any report of a lapse during the 24 hour period of each day and no if none of the EMA reports indicate a lapse.

IMPORTANT: If EMA\_1 =2 (lapse) than the participant reports the date (EMA\_1.1) and time (EMA\_1.2; nearest hour in military time in America/Chicago time zone) of the actual start of the lapse. This date/time may not be on the day that the lapse reported. Therefore the reports of lapse EMA\_1 need to be adjusted for this information. For a subset of early participants, this date/time information is missing. In these cases, we will assume that the lapse occurred on the day that it is reported.

**Features**

**General issues**

Consider which second order effects to include

**Features available 1X**

Each of these signals is measured only once at the screening or intake session when the participant first starts the study

Age: DEM\_1

Sex: DEM\_2

Race: DEM\_3 [Categorical with values 1-6; Use 5 dummy codes)

Ethnicity: DEM\_4 [Categorical with values of 1-5; Collapse to two categories: 1 vs. 2-5)

Education (scored ordinally): DEM\_5

Employment: DEM\_6 [Categorical with values of 1-8; Use 7 dummy codes]

Income: DEM\_7

Marital status: DEM\_8 [Categorical with values of 1-5; Use 4 dummy codes]

Number Living parents: DEM2\_2

Number Deceased parents: DEM2\_4

Number Living Children: DEM2\_6

Number Deceased Children: DEM2\_8

Alcohol use history: AUH\_1, AUH\_2, AUH\_3, AUH\_4, AUH\_5, AUH\_6\_1, AUH\_6\_2, AUH\_6\_3, AUH\_6\_4, AUH\_6\_5, AUH\_6\_6, AUH\_6\_7,AUH\_7, AUH\_9, AUH\_10, AUH\_11, AUH\_12, AUH\_13, AUH\_14

AUH date of last quit: AUH\_8\_Month AUH\_8\_Day AUH\_8\_Year [Date variable. Convert to numeric days to current day]

DSM5 total score: DSM5\_Tot

WHO Assist: ASSIST\_1\_1, ASSIST\_1\_2, ASSIST\_1\_3, ASSIST\_1\_4, ASSIST\_1\_5, ASSIST\_1\_6, ASSIST\_1\_7, ASSIST\_1\_8, ASSIST\_4\_1, ASSIST\_4\_2, ASSIST\_4\_3, ASSIST\_4\_4, ASSIST\_4\_5, ASSIST\_4\_6, ASSIST\_4\_7, ASSIST\_4\_8, ASSIST\_5\_1, ASSIST\_5\_2, ASSIST\_5\_3, ASSIST\_5\_4, ASSIST\_5\_5, ASSIST\_5\_6, ASSIST\_5\_7, ASSIST\_5\_8, ASSIST\_6\_1, ASSIST\_6\_2, ASSIST\_6\_3, ASSIST\_6\_4, ASSIST\_6\_5, ASSIST\_6\_6, ASSIST\_6\_7, ASSIST\_6\_8, ASSIST\_7\_1, ASSIST\_7\_2, ASSIST\_7\_3, ASSIST\_7\_4, ASSIST\_7\_5, ASSIST\_7\_6, ASSIST\_7\_7, ASSIST\_7\_8, ASSIST\_8

Alcohol problems lifetime: YAP\_Life

Alcohol problems past year: YAP\_Year

SCL90 questionnaire: SCL90\_Som, SCL90\_OC, SCL90\_Sens, SCL90\_Dep, SCL90\_Anx, SCL90\_Ang, SCL90\_Pho, SCL90\_Par, SCL90\_Psy

MPQ questionnaire: MPS\_WB, MPS\_SP, MPS\_AC, MPS\_SC, MPS\_SR, MPS\_AL, MPS\_AG, MPS\_CT, MPS\_HA, MPS\_TD, MPS\_AB, MPS\_UV

IUS questionnaire: IUS\_Tot

ASI3 questionnaire: ASI3\_PC, ASI3\_CC, ASI3\_SC

DTS questionnaire: DTS\_Toler, DTS\_Absorb, DTS\_Appraise, DTS\_Regulate

FAD questionnaire: FAD\_Prob, FAD\_Comm, FAD\_Role, FAD\_Resp, FAD\_Inv, FAD\_Beh, FAD\_Gen

**Features available 3X**

Each of these signals in this category are available up to 3X per subject (approximately once per month).

Include each of the available 3X signals as a feature

Include most recent signal – 1 previous

Include most recent signal – mean of past two previous

WHO Assist: ASSIST\_2\_1, ASSIST\_2\_2, ASSIST\_2\_3, ASSIST\_2\_4, ASSIST\_2\_5, ASSIST\_2\_6, ASSIST\_2\_7, ASSIST\_2\_8, ASSIST\_3\_1, ASSIST\_3\_2, ASSIST\_3\_3, ASSIST\_3\_4, ASSIST\_3\_5, ASSIST\_3\_6, ASSIST\_3\_7, ASSIST\_3\_8

PACS\_Tot

DASS21\_Anx, DASS21\_Dep, DASS21\_Str

PSS10\_Tot

QOL\_Tot

MSPSS\_Fam, MSPSS\_Fri, MSPSS\_SO

DAS\_0, DAS\_Con, DAS\_Sat, DAS\_Coh, DAS\_Aff

AASE\_ConNA, AASE\_ConSoc, AASE\_ConPhys, AASE\_ConCrav

MAM\_2, MAM\_3, MAM\_4, MAM\_5, MAM\_6, MAM\_7, MAM\_10, MAM\_11, MAM\_13,MAM\_14,MAM\_15,MAM\_16, MAM\_17, MAM\_19, MAM\_21,MAM\_22, MAM\_23

**Categorical 3X variables**

MAM\_1 – 4 levels

**Ecological Momentary Assessments 4X daily**

Most recent Previous signal

Mean/Max of signals within the a) previous day, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Linear function over time for all signals to date

Most recent previous signal - mean of a) previous day, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Mean/Max of signals over previous day – mean of a) previous 24 hours, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Mean/Max of signals over past 3 days – mean of a) previous 24 hours, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Mean/Max of signals over past week – mean of a) previous 24 hours, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Standard deviation of signals over a) previous day, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Standard deviation of over past 24 hours – standard deviation of signals a) previous 24 hours, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Standard deviation of over past 3 days – standard deviation of signals a) previous 24 hours, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

Standard deviation of over past week – standard deviation of signals a) previous 24 hours, b) previous 3 days, c) previous week, d) previous month, e) all previous signals

EMA\_2, EMA\_3, EMA\_4, EMA\_5, EMA\_6, EMA\_7, EMA\_8, EMA\_9, EMA\_10

**GPS**

**SMS meta data**

**Voice call meta data**

**SMS Natural language processing**

**Sleep**

**Missing data?**