Table of Contents

[Overview 2](#_Toc476749694)

[Collaborators and Other Important People 3](#_Toc476749695)

[Summary of Measures 4](#_Toc476749696)

[Inclusion/Exclusion Criteria 6](#_Toc476749697)

[Raw Data Files and Structure 7](#_Toc476749698)

[Empatica (physiology) 9](#_Toc476749699)

[Data Access and Storage 9](#_Toc476749700)

[Compliance 9](#_Toc476749701)

[Moves (GPS) 10](#_Toc476749702)

[Data Access and Storage 10](#_Toc476749703)

[Additional Notes 10](#_Toc476749704)

[Beddit (sleep) 11](#_Toc476749705)

[Ecological Momentary Assessment 15](#_Toc476749706)

[Data Access and Storage 15](#_Toc476749707)

[Compliance 15](#_Toc476749708)

[Audio Messages 16](#_Toc476749709)

[Data Access and Storage 16](#_Toc476749710)

[Sending iMessage Recordings to the Transcription Company 17](#_Toc476749711)

[Saving Data from Transcription Company-Will add greater detail to section once audio recordings from pilot data have been received from Datagain 18](#_Toc476749712)

[SMS 19](#_Toc476749713)

[Data Access and Storage 19](#_Toc476749714)

[Additional Notes 22](#_Toc476749715)

[Voice 24](#_Toc476749716)

[Data Access and Storage 24](#_Toc476749717)

[Additional Notes 27](#_Toc476749718)

[Qualtrics ID Batteries 28](#_Toc476749719)

[Data Access and Storage 28](#_Toc476749720)

[Interview Data 29](#_Toc476749721)

[Wake/Sleep times 32](#_Toc476749722)

[Mongo DB 33](#_Toc476749723)

# Overview

# Collaborators and Other Important People

# Summary of Measures

**Screening Session**

* Demographics
* Alcohol Use History
* DSM-5 Checklist
* Young Adult Alcohol Problems Test
* WHO-The Alcohol, Smoking and Substance Involvement Screening Test
* Symptom Checklist-90-Revised
* Intolerance of Uncertainty Scale
* Anxiety Sensitivity Index
* Distress Tolerance Questionnaire
* McMaster Family Assessment Device
* Multidimensional Personality Questionnaire Brief Form

**Intake**

* Recovery Environment Interview
* Penn Alcohol Craving Scale
* Alcohol Abstinence Self-Efficacy Scale
* Monthly Addiction Monitor
* Depression Anxiety Stress Scale-21
* Perceived Stress Scale
* Quality of Life Questions
* Dyadic Adjustment Scale
* Multidimensional Scale of Perceived Social Support

**Follow-up visit #1**

* Recovery Environment Interview-Check for Updates
* Penn Alcohol Craving Scale
* Alcohol Abstinence Self-Efficacy Scale
* Monthly Addiction Monitor
* WHO-The Alcohol, Smoking and Substance Involvement Screening Test
* Depression Anxiety Stress Scale-21
* Perceived Stress Scale
* Quality of Life Questions
* Dyadic Adjustment Scale
* Multidimensional Scale of Perceived Social Support
* Burden measure

**Follow-up visit #2**

* Recovery Environment Interview-Check for Updates
* Penn Alcohol Craving Scale
* Alcohol Abstinence Self-Efficacy Scale
* Monthly Addiction Monitor
* WHO-The Alcohol, Smoking and Substance Involvement Screening Test
* Depression Anxiety Stress Scale-21
* Perceived Stress Scale
* Quality of Life Questions
* Dyadic Adjustment Scale
* Multidimensional Scale of Perceived Social Support
* Burden measure

**Final visit #3**

* Recovery Environment Interview-Check for Updates (Unreported Contacts and Locations Only)
* Burden Measure

**Real-time Data**

* Morning Daily Ecological Momentary Assessment Survey
* 3x Daily Ecological Momentary Assessment Survey
* 1x Daily 15-30 sec. Audio Survey via voice note messaging
* Phone Call Logs (Incoming and Outgoing)
* Text Message Logs (Incoming and Outgoing)
* Text Message Content (Incoming and Outgoing)
* GPS Location via MOVES Activity Diary Smartphone Application
* Sleep Quality via Beddit 3 Sleep Tracker
* Movement (Accelerometer) via Empatica MOVES Activity Diary Smartphone Application
* Heart Rate via Beddit Sleep Sensor

# Inclusion/Exclusion Criteria

* 18 or older
* Must be able to read and write in English
* Abstinent from alcohol for at least 1 week and no longer than 2 months
* Meet criteria for Alcohol use disorder with at least moderate severity (> 4 DSM-5 criteria by self-report on DSM-5 Checklist). This will be confirmed at the screening session.
* No current severe symptoms of psychosis (< 2.24) or paranoia (< 2.82) confirmed by SCL-90 at screening session
* Agree to use personal smart phone as their primary phone while enrolled in the study.

# Raw Data Files and Structure

All data files are CSV files. Data types for data streams in these CSV files are either numeric (e.g., integer, real) or character strings (of varying length). Character strings are enclosed in double quotes. Missing values for numeric data are indicated with NA. Missing values for strings are empty strings (“”).

The root data folder for raw data is \RawData on Private drive on server. The root data folder for clean data (for input into MongoDB) is E:\Database\. File names and folder structure are the same across the raw and clean data folders.

CSV files for data streams that include all subjects in one CSV file are saved directly at root. These CSV files are for self-report batteries at screening (Screen.csv), intake (Intake.csv), follow-up (Followup12.csv), final session (FinalVisit.csv), morning EMA (EMAMorning.csv), later EMA (EMALater.csv), audio messages (Audio.csv), sleep times (SleepTimes.csv), and subject details from three sessions (DETAILS1.csv; DETAILS2.csv; DETAILS.csv).

These CSV files contain a header which indicates the name of each data stream in each column. The first column in all of these CSV files is called UTC and it contains integer unix time stamps in UTC for the data in each row. The second column in these CSV files is called SubID and it contains the subject ID (‘001’ – ‘200’) associated with the data in each row as a string. The data streams in all of the remaining columns will either be numeric or character strings (of varying length). The MongoDB code is flexible enough to accommodate changes in the number of columns of data, the names of these data streams (from the header), and their type (numeric or string). Although I provide filenames above, the MongoDB code should be flexible enough to obtain data streams from all CSV files in this root folder and its subfolders regardless of their filenames. My R code for cleaning the raw data (mak\_CleanData.R) is NOT flexible. Any changes to filenames or locations (or even column names) will require updates to this R script.

The root folder will contain 200(ish) subfolders (1 for each subject). Each subject’s subfolder will be named ### (e.g., \ROOT\001\, \ROOT\002, . . . , \ROOT\200\ ). CSV files that contain data streams for only one subject will be saved in these subject subfolders. Each file in these folders with have an ###\_ prepended to the filename.

Similar to the CSV files described above, the CSV files in the subject folders will contain a header that indicates the name of each data stream. The first column will again be called UTC and contain an integer unix time stamp in UTC. There is not a column for SubID. The SubID is determined by the subfolder name and a prepended ###\_ on all filenames. The remaining data streams will either be numeric or string. The MongoDB code is flexible regarding data stream names, number of data streams (columns), data types (numeric or string), and overall number and name of CSV files in each subject’s subfolder. My R code is NOT.

These CSV files include beddit sleep data (###\_Sleep.csv), SMS (###\_SMS\_Handle.csv and ###\_SMS\_Data.csv), Voice (###\_Voice.csv).

There are some files from the interview that contain data that span multiple columns. For these files ……. These files are contacts (###\_Contacts.csv), locations (###\_Locations.csv), risk dates (###\_Dates.csv), and risk times (###\_Times.csv).

GPS (from Moves app) is saved as a ZIP file for raw data (###\_GPS.zip) but is converted to two CSV files for clean data (###\_GPSPoints.csv; ###\_GPSEvents.csv).

PHYSIOLOGY files are saved in batches as ZIP files for raw data outside of the subject folder in /RawData/Physiology/. The file naming convention for the zipped archives is as follows: [Device SerialNumber] [yymmdd-HHmmss]. The date is UTC time stamped. There is also a file called EmpaticaLog.xlxs in this raw data folder that records device id and start/end dates for each subject. PHYSIOLOGY is converted to CSV files by physiology measure and saved in the subject folders for clean data (###\_EDA\_TTTTTTTTTT.csv; ###\_HR\_TTTTTTTTTT.csv; ###\_TEMP\_TTTTTTTTTT.csv; ###\_ACC\_TTTTTTTTTT.csv, where TTTTTTTTTT is the unix time stamp for the start time for the specific physiology data stream).

## Moves (GPS)

### Data Access and Storage

Using the Downloader App

1. Participant Moves data should be downloaded via the Downloader webapp at least once a week and before any scheduled Follow Up visits.
2. Go to the Downloader app page at <https://psych.wisc.edu/MovesDownloader/> (You will need to copy this link into your web browser. It does NOT work to click it)
3. Sign in using your Net ID and password (if prompted)
4. Click **Download**.
5. Select the **Start Date** and **End Date** from the drop down boxes.
6. Under the column titled **Download .GPX File** click on the  to the right of the appropriate SubID.
7. Open the GPX file that was downloaded.
8. Save the GPX file as: SubID\_GPS (ex: 001\_GPS). Overwrite any previously downloaded GPS file. Always check to ensure that all of the study dates are accounted for prior to deleting the older file. If RA does not have time to verify that all the data is accounted for at the time of download, move the oldest file to the folder entitled :OLDER in the participant’s raw data folder so that only one GPS file remains in the participant’s raw data folder.
9. Save data file in participant’s study file-P:\Study Data\RISK\RawData

Additional Notes

Each subject will have their own account in Moves using UW[ARCRisk+###@gmail.com](mailto:ARCRisk+), with their own unique password.

These three CSV files are saved in \RawData\###\Moves\. They are:

**###\_storylineTrackpoints.csv**. This file contains all GPS location samples. We use only three columns: Date Created, Latitude, Longitude. All other columns can be ignored. Date Created is in CT (CDT or CST) but there is no indicate of the time zone. This appears to be lost when I use ExpertGPS to convert from gpx to csv with Lat/Long

**###\_storyline.csv**: This file contains processed data. We are interested in the following columns (Name, Start, End, Duration). Start and End are time stamps in CT (CDT or CST) with time zone indicated as -5 (CDT) or -6 (CST)

**###\_places.csv**: This file contains additional information about the “Place” entries in the Name column from ###\_storyline.csv. There is one row in ###\_places.csv for every “Place in” entry in the Name column of ###\_storyline.csv. ###\_places.csv provides the lat/lon values for each of these Place in entries. These lat/long values should be added to the information saved from ###\_storyline.csv. There is no additional information for the other entry times in Names (walking, cycling, transport, etc). Start and End are time stamps in CT (CDT or CST) with time zone indicated as -5 (CDT) or -6 (CST). These are not needed but can be used to match entries across ###\_storyline.csv and ###\_places.csv

## Beddit (sleep)

TBD

**Data Access and Storage**

1. Each subject will have their own account with Beddit using ARCRisk+###@gmail.com, with their own unique password. This will need to be edited into W:\Open\bedditdata.php
2. Afterwards, open the file from the web - <http://dionysus.psych.wisc.edu/Open/bedditdata.php> outputs the code as json format
3. Copy the output and paste the data into a Notepad file
4. Name downloaded data file: Sleep\_#SubID
5. Save this file as a .JSON file
6. Save file here: P:\Study Data\RISK\RawData
7. Each subject will have their own account with Beddit using ARCRisk+###@gmail.com, with their own unique password. This will need to be edited into W:\Open\bedditdata.php
8. Afterwards, open the file from the web - <http://dionysus.psych.wisc.edu/Open/bedditdata.php> outputs the code as json format, if you want to call it directly from R
9. Or, to get a download file, visit <https://json-csv.com/> and paste in the URL above (<http://dionysus.psych.wisc.edu/Open/bedditdata.php>) - it will automatically start the transformation.
10. Hit **Download** when you see the results.
11. Name downloaded data file: TBD
12. Save this file as .json file or as raw txt
13. Save file here: P:\Study Data\RISK\RawData

**QS Access App for Collecting Beddit Sleep Data from iPhones ONLY**

1. Open the QS Access app.
2. Scroll down until you see the Sleep Analysis option. Click on the Sleep Analysis option.
3. This will bring you to this screen. Click on Tabulate sleep analysis. You will see a table with the sleep analysis output from Healthkit.
4. Click on the icon in the upper right corner of the screen to send the csv via email.
5. Click on the Mail icon and send the table to the study email [arl4@psych.wisc.edu](mailto:arl4@psych.wisc.edu). In the Subject put QS Sleep data.
6. Download the file from the email, rename the file (SubID)\_QSSleepData and cut and paste the file into the subject’s raw data folder. Once the file has been transferred you delete the email.
7. At Follow up 2 and 3 check the existing data file against the new file before overwriting the existing file. The data should match, but if it doesn’t check with the participant and keep the old file in the Older folder.

**Beddit API Codebook**

**User-Related Measures**

**Tags -** user-added comments about a sleep session (i.e., “too hot to sleep”)

**User -** user ID assigned by Beddit, should update to SubID

**Tips -** “Sleep Tips” presented to the user, selected from Beddit’s DB of tips for getting better sleep

**Time/Date Measures**

**Date**

**timezone**

**Updated -** The updated field contains the timestamp of **when the sleep object was put to the server**. If you want to periodically query new sleep objects, you can store the latest updated value, and use it in query.

**start\_timestamp = session\_range\_start**

**end\_timestamp = session\_range\_end**

**Time Tracked Measures – each of these should be a separate table**

**time\_value\_tracks\_\_| -** In the CSV, name code for the items below

**time\_value\_tracks\_\_|\_\_items\_\_001 -** In the CSV, timestamp for the items below

**time\_value\_tracks\_\_|\_\_items\_\_002 -** In the CSV, value for the items below

**Sleep stages -** This is the state transition sequence for the sleep stages. The **timestamp is the time when a state begins** and the value represents the new state. The gap represents a period of missing signal in the measurement period.

|  |  |
| --- | --- |
| **Value** | **Meaning** |
| 65 | Away from bed |
| 83 | Sleep |
| 82 | Restless sleep |
| 87 | Awake |
| 78 | No signal |
| 71 | Gap in measurement |

**Snoring episodes -** The timestamp represents the time of a snoring episode and the value is the length of the episode. The snoring episodes give a high-level view of snoring activity and have around **10-minute resolution**.

**Sleep cycles -** The timestamp and value represents the sleep depth at that time. The value is a floating point number between 0.0 and 1.0. **Value** **0.0 corresponds to lightest possible sleep and 1.0 to deepest possible sleep.** There will be **one datapoint every 2 minutes**, beginning from the moment the subject is deemed to be sleeping and ending when the subject ultimately wakes up.

**Actigram epochwise -** The **timestamp represents the start time of an epoch** and the **value is the number of detected movements** during the epoch. The movements are reported only for epochs during which the user is present on the bed**. The epoch is 60 s long**. This track is named activity events in sleeps created by newer versions of the app.

**Heart rate curve -** The timestamp and value pairs describe a smooth heart rate curve. **The time between curve data points is around 5 minutes**. A **gap in the curve is marked by a negative value**. A good way to display the curve is to split the time-value sequence into curve parts by the negative gap-values and display each curve part individually.

**Nap periods -** This represents the **starts and ends** of nap periods. **The value for start = 1 and for end = 2**.

**Presence value -** Intermediate presence results. The timestamp represents the time when the subject's presence state changes during the session. The timestamp is the time when a state begins, and the value represents the subsequent presence state.

The state **65=away** represents the start time of a period when the user is away.

The state **80=present** represents the start time of the period when the user is present.

The state **78=end** represent the start time of the period when the presence of the user is unknown. This currently occurs when there is no signal.

This track is an intermediate result. To have more accurate presence data, use the "sleep\_stages" track instead.

**Alarm event -** List of alarm clock time and event type values.

|  |  |
| --- | --- |
| **Value** | **Meaning** |
| 0 | Alarm started ringing |
| 1 | User snoozed alarm |
| 2 | User dismissed alarm |
| 3 | Alarm was ignored (stopped automatically after x minutes) |
| 4 | Alarm was automatically dismissed after user had ignored it several times |

**Properties (Main) Data Table**

|  |  |
| --- | --- |
| **Property** | **Meaning** |
| sleep\_time\_target | The user's personal sleep time target in seconds |
| resting\_heart\_rate | The resting heart rate reading for the night |
| average\_respiration\_rate | The average respiration rate reading over the night |
| sleep\_efficiency | The proportion of sleep in the measurement period. The time after waking up the last before ending the measurement is not incorporated in the calculation. If the subject did not fall asleep, the value is missing. |
| sleep\_latency | The time it takes to fall asleep, in seconds. If the subject did not fall asleep, the value is missing. |
| away\_episode\_count | Number of absences during the night |
| total\_snoring\_episode\_duration | Total amount of snoring, in seconds |
| stage\_duration\_A | Total time in away state in the measurement period, in seconds |
| stage\_duration\_S | Total time in sleep state in the measurement period |
| stage\_duration\_R | Total time in restless sleep state in the measurement period |
| stage\_duration\_W | Total time in wake state in the measurement period |
| stage\_duration\_N | Total amount of "no signal" time in the measurement period |
| stage\_duration\_G | Total amount of "gap" time in the measurement period |
| total\_nap\_duration | Total amount of sleep during periods marked as naps by the "nap\_periods" track |
| sensor\_status | The summarized sensor status based on the "sensor\_status" track in the associated sessions. |
| activity\_index | The number of movements per hour during measurement period. The time it takes to fall asleep and the time after waking up the last before ending the measurement are not incorporated in the calculation. |
| evening\_HRV\_index | A heart rate variability index measured during a 15 min time span during the first third of the night. The index is calculated as the root mean square of successive differences (RMSSD). |
| morning\_HRV\_index | A heart rate variability index measured during a 15 min time span during the last third of the night. The index is calculated as the RMSSD. |
| all\_night\_HRV\_index | The heart rate variability index measured during the whole night. The index is calculated as the RMSSD. |
| resting\_HRV\_index | The heart rate variability index measured during a 15 min time span centered at the instance of lowest heart rate (one minute mean) during the night. The index is calculated as the RMSSD. |
| total\_sleep\_score | Sleep score calculated based on total sleep time and parameters affecting the quality of sleep. The sleep score presents a measure of the physical (not perceived) goodness of sleep during one night. The sleep score has a value between 0 and 100. In version 1 of the sleep score the value can be over 100. |
| sleep\_score\_version | The version of the sleep score. The version is an integer which is incremented whenever the sleep score algorithm changes significantly. If the property is missing, that should be interpreted as the version being 1, which is the lowest possible version. The purpose of the version is to differentiate sleep scores with different interpretations. |
| score\_bed\_exits | Sleep score item for number of bed exits |
| score\_amount\_of\_sleep | Sleep score item for total amount of sleep |
| score\_snoring | Sleep score item for total duration of snoring episodes |
| score\_sleep\_latency | Sleep score item for time it took to fall asleep |
| score\_sleep\_efficiency | Sleep score item for sleep time vs time spent in bed |
| score\_awakenings | Sleep score item for number of awakenings during the night |

## Ecological Momentary Assessment

### Data Access and Storage

There are two CSV files for the EMA data stream. These two files will contain data for all subjects. They are saved at the root. These files are for the morning EMA report (EMAMorning.csv) and the other 3 later EMA reports (EMALater.csv).

EMAMorning.csv will contain data for all subjects with 90 reports per subject.

EMALater.csv will contain data for all subjects with 90 reports per time X 3 times per day per subject.

1. Participant data should be downloaded the morning of each scheduled follow-up visit. The downloading of the data is down via an R Script.
2. file

### Compliance

### Data analysis

NRT1 discovered a problem with EMA and resolved it thusly: Because the EMA are received as a series of texts, it’s easy for a participant to click the wrong text, even if they are responding at the correct time. Therefore in Qualtrics, the embedded data of sent time/date isn’t relevant. What you’re going to do is create epochs from one text to the next. An incoming response is associated with the more recent text, regardless of the embedded data. If multiple responses come in between texts, the more proximate response is recorded and the extra is discarded.

## Audio Messages

### Data Access and Storage

**Transfer Audio Files Via iMazing**  
Recordings should be listened to each day (Mon-Fri) so that they can be saved to the P drive and then immediately deleted from the lab iPhone:

1. In order for iMazing to work you must have iTunes installed on your computer. Free installation of iTunes is available at <https://www.apple.com/itunes/download/>
2. Install imazing on your computer by going to P:\Methods\Equipment\Software Manuals and Installers and entering the activation code: **id789289485709odr**
3. Connect 5 ACHESS STUDY phone to your computer using a lighting cable
4. Open iMazing, select the 5 ACHESS STUDY phone and then select “backup”
5. The backup will give you a list of options, select the following: Backup Encryption (Disabled), Backup Location (backed up on private ([\\dionysus](\\\\dionysus)) )P:)), Backup Archiving (Disabled), Wi-Fi Connection (Disabled). Once the appropriate settings have been selected, select Back Up.
6. Once the phone has been backed up, select Close Window in the bottom right hand corner of the Back up screen.
7. Go to **Messages** . The messages will be sorted by SubId. To transfer messages to the appropriate SubId folder, select a SubId from the list in iMazing and click on the message. Right click on the message and select Export Attachments. A window labeled “Browse For Folder” will pop up. Select This PC > StudyData > RISK > RawData > and select the correct Sub ID’s folder and select their Audio folder. Then press OK. If the Sub ID doesn’t have an Audio folder yet you can “Make New Folder.”
8. When the message has successfully copied from iMazing to the RawData folder you will see a message in iMazing that says “Audio Copied” in the Operations drop down 
9. Now, open the RISK raw data folder found here: **P:\StudyData\RISK\RawData.** You can use the iMazing backup or the study phone to check that the messages transferred.
10. If any of the messages transferred in a different format than .amr, you’ll need to convert the messages to .amr files before renaming them using the R script. Common file types other than .amr are 3GP and m4a. If the files are not converted before running the R script they will not be renamed. To see the format select the folder of the Sub ID you want to check, select the audio folder and look under the 3rd column labeled Type. If it says anything other than AMR File it will need to be converted to .amr. To convert audio files, follow the directions provided below. Once the file has been converted to .amr and saved in the correct Sub ID’s folder you can delete the original unconverted file from the Sub ID’s folder.
11. To rename the files open the file **P:\StudyData\RISK\Analysis\RISK**\**mak\_iMazing**. Select all of the code by clicking anywhere in the code box and hitting Ctrl + A on the keyboard. Then click Ctrl + Enter to run the code (or click run in the upper right corner of the code box). The script is done running when the  disappears.

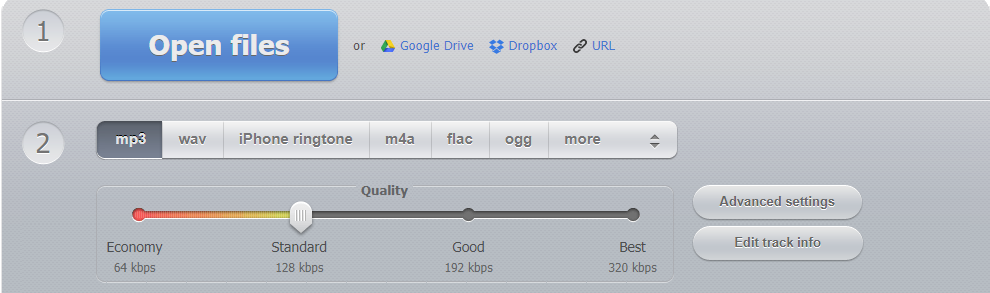
Each individual file will be labeled in the Sub ID’s folder with SubID\_year of recording\_date of recording\_time of recording (military time HH:MM) Ex: 002\_2016\_1128\_0730 – confirm that the relabeled files are correct.

1. Record the name of the files to be sent on the **Audio Transcript Check** spreadsheet, located in P:\ StudyData\RISK\Administration\Compliance\Audio Transcript Check.
2. Once file has been copied to the shared drive delete from iPhone in text message log and delete the backup in iMazing.
3. To delete files from iMazing click on and select Delete All. Once the backup is deleted you can close iMazing.

**To convert files that won’t play use this website** <http://audio-joiner.com/>

-Click on the  button in the top left-hand side of the screen

-Click on  and click on the participant’s audio file in the Raw Data folder



-Click on  and download the file

-Save downloaded file to the participant’s audio folder

**In the event that an audio message is sent in two parts use this website** <http://audio-joiner.com/> to join the two parts together before sending the audio file to the transcription company.

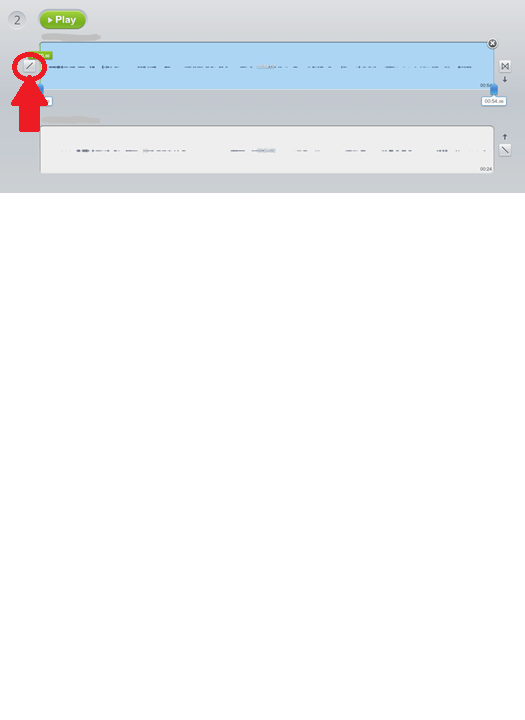
Instructions on how to use the website:

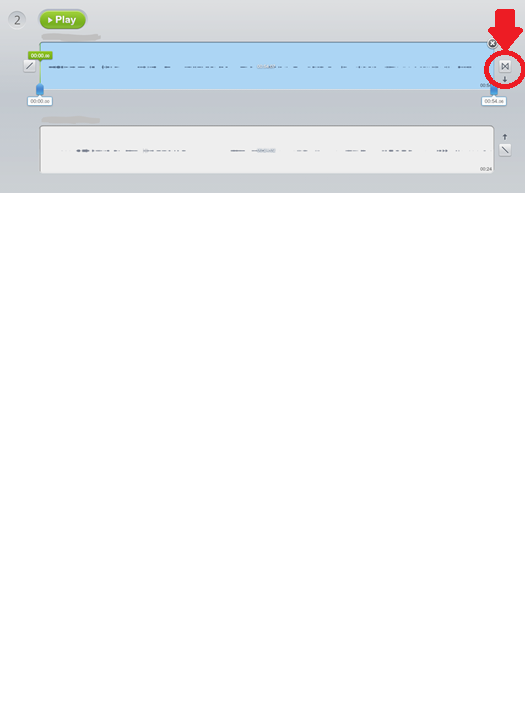
-Open <http://audio-joiner.com/>

-Add the two audio files that you want to join by clicking on the  button and selecting the files, one at a time, in sequential order.

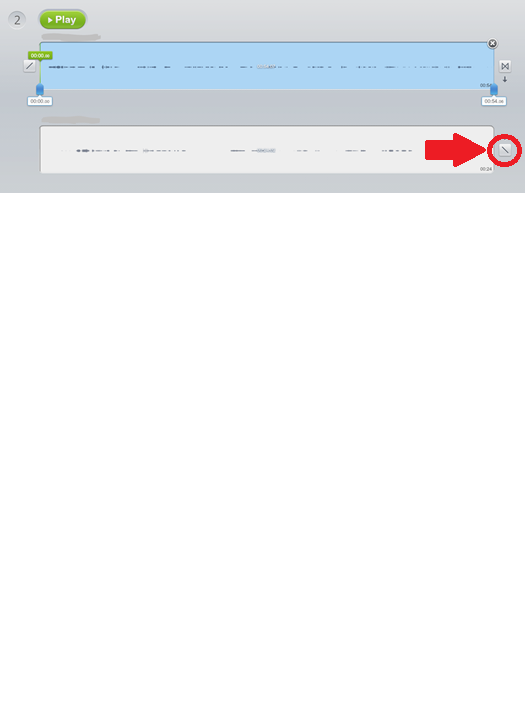
-You will see a screen that looks like this:

-Make sure that Fade-in is disabled by clicking this icon , it will appear like this when it is disabled 



-Make sure that crossfade is disabled by clicking this icon , it will appear like this when it is disabled 

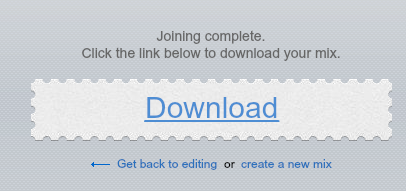
-Make sure that Fade-Out is disabled by clicking this icon , it will appear like this when it is disabled 



-Play the files together before you join them to make sure they are in sequential order

-Once you have listened to the files you need to join, and have verified that they are in sequential order, click 

-The joined file will then be available to download once you get this message:



-Click on Download

### Sending iMessage Recordings to the Transcription Company

1. Go to: <https://transcription.datagainservices.com/transcription/enterprise>
2. Select Log In
3. Log In with Username and Password (Find in LastPass in RISK folder)
4. **Select Files** to upload.
5. **Browse** to the files you want to upload. Then press **Upload**.
6. Make sure that under **# of Speakers** there is a **1**, and that under **Turnaround** it says **Standard**.
7. Check the files you wish to submit**,** or you can select all of them by checking **Filename**.
8. Please click on the **Submit Selected** button and make sure you receive a confirmation message saying the files were successfully submitted.
9. Record files sent on the **Audio Transcript Check** spreadsheet, located in P:\ StudyData\RISK\Administration\Compliance\Audio Transcript Check.
   1. The spreadsheet is organized by Sub ID and is used to keep track of when audio files sent to Datagain, who sent them, when we receive the transcribed files back, and who checked the received files.
   2. When you upload files to the Datagain transcription service you will record the date the file was sent, as well as the initials of the RA who sent the files.

### Saving Data from Transcription Company

1. Datagain (Pinky Shinde) will send a copy of the most recently transcribed files by the 15th of each month, containing transcripts from the prior month, to the study’s dropbox account. The information on how to access the drop box account can be found in last pass. The file will be downloaded and saved here: P:\StudyData\RISK\Adminsitartion\Datagain. The file will then be deleted from the dropbox account. No transcripts will be saved permanently in the drop box account.
2. The file will then be reviewed by study staff who will then correct inaudible sections when able.
3. The reviewed transcripts will then be added to the master excel file located here: spreadsheet here: P:\StudyData\RISK\RawData\Audio
4. Make sure to update the **Audio Transcript Check** spreadsheet with the date that the files were reviewed and make note of any issues that you found. Study coordinator will be notified if there are missing files or major issues with the transcribed audio files.

The data stream for the daily audio messages will be saved in Audio.xlsx. This file will contain data from all subjects. It will be saved in \RawData\

There will be 90 rows per subject. The first column is MessageName which includes the audio file name without the extension. The second column is Text which contains the message content as string

## SMS

### Data Access and Storage

For both device platforms, there will be 3 files created per participant across the life of the study. They are to be saved with the naming convention: SubID\_SMS\_Data\_#.csv, where # is the sequential file identifier. (\_1, \_2, \_3) Number corresponds with FILE, not with VISIT. If we do not have Visit 1 data, the identifier on visit2 will be \_1, etc.

#### iPhone

Steps to get SMS content & metadata

1. Log in as RA to one of the office laptops.
2. Plug iPhone into computer with lightening cable.
3. Open iTunes
   1. A dialogue box opens:
   2. iTunes will say “would you like to access this iPhone/ Please respond via the phone”.
   3. Go to the phone and hit **Trust** to allow iTunes to connect.
   4. When you see the device in the left-hand pane, right-click and select “**backup**”
      1. Don’t transfer purchases
      2. Don’t encrypt
   5. A message bar appears across the top of iTunes, wait until backup is complete.
   6. Make sure the backup files you need, exist:
      1. Browse to C:\Users\RA\AppData\Roaming\Apple Computer\MobileSync\Backup (where RA is the login you are using on that computer).
      2. Make sure you look in the backup that you just created, choose the backup with the most recent date/timestamp. Click into that.
      3. In the search box, type “**3d0d7e**”. If you see a single file in the search results, your file has been found (full name **3d0d7e5fb2ce288813306e4d4636395e047a3d28**.)
         1. You will need this file later, so it’s convenient to click, ctrl-c (copy) the file, then click ctrl-V (paste) on the desktop to put the file somewhere easy to get back to it
      4. When you see this file, you know the backup is complete and you can disconnect the iPhone and return it to the participant.
      5. Rename the backup file with the crazy name to 001\_SMS\_#.sql (where 001 is the subID of the person you are working with and # is the incremental file order at each backup)
      6. Copy the sql backup files to the server. NOTE – you will FIRST need to connect the laptop to the LAN with an ethernet cord to do this.
      7. Copy the file and drop into P:\Study Data\RISK\RawData\$SubID
      8. Delete the .sql files from the laptop desktop.
4. Cleaning:
   1. Remove all \r that are not \r\n with R code. (There are extra line breaks in OneReach messges, this removes all that are not at the end of a line.)
   2. The “handle” table should be matched with the data table matching the handle ID to the foreign key
5. Note: You will be able to detect deleted messages by inspecting the Row ID of the Message table.
6. SMS\_Data file headers: (https://www.theiphonewiki.com/wiki/Messages#message)

**ROWID** – ordered number of SMS

**guid** – unique object id

**text** – content of message. If there is no text content (ie “?” only), check **is\_audio\_message** or **cache\_has\_attachments**. If either is true, this was a picture or audio message.

**replace** – unknown, possible values: 0, 1, 2

**service\_center** – NULL

**handle\_id** – recipient or sender (depending on value of **is\_from\_me**)

**subject** – The subject of an iMessage/MMS-message, or NULL if it's an SMS or if subject is not used on a iMessage/MMS-message**country** – NULL or an ISO country code (eg: 'ch' for Switzerland)

**attributedBody** –blob, content unknown. The only strings in it are "JFIF" and "Exif", so this is probably meta-data.

**version** – OS version

**type** – 0, or Null if pre-iOS 5

**service** – SMS vs iMessage – not sure this matters

**account** – NULL if not an iMessage or 'p:' & own phone number or 'e:' & email registered for iMessage

**account\_guid** – unknown

**error** – empty string if message is pre-iOS 5.0 or 0 if after**date** – datetime value in iOS format UTC time.

**date\_read** – datetime value in iOS format UTC time.

**date\_delivered** – always zero in the examples I have as reference

**is\_delivered** – 1 for true

**is\_finished** – 1 for true, 0 for draft

**is\_emote** – 1 for true

**is\_from\_me** – 1 for true

**is\_empty** – 1 for true

**is\_delayed** – 1 for true

**is\_auto\_reply** – 1 for true

**is\_prepared** – always zero in the examples I have as reference

**is\_read** – 1 for true

**is\_system\_message** – 1 for true

**is\_sent** – 1 for true

**has\_dd\_results** –

**is\_service\_message** – 1 for true

**is\_forward** – 1 for true

**was\_downgraded** – 1 for true

**is\_archive** – 1 for true

**cache\_has\_attachments** – 1 for true

**cache\_roomnames** – unknown

**was\_data\_detected** – 1 for true

**was\_deduplicated** – 1 for true

**is\_audio\_message** – 1 for true

**is\_played** – 1 for true

**date\_played** – always zero in the examples I have as reference

**item\_type** – unknown

**other\_handle** – unknown

**group\_title** – unknown

**group\_action\_type** – unknown

**share\_status** – 1 for true

**share\_direction** – unknown

**is\_expirable** – 1 for true

**expire\_state** – unknown

**message\_action\_type** –

**message\_source** – unknown

**associated\_message\_guid** – unknown

**balloon\_bundle\_id** – unknown

**payload\_data** – unknown

**associated\_message\_type** – unknown

**expressive\_send\_style\_id** – unknown

**associated\_message\_range\_location** – unknown

**associated\_message\_range\_length** – unknown

**time\_expressive\_send\_played** – unknown

**message\_summary\_info** – unknown

Some additional column information, (ie default 0 or null) can be found at <https://translate.google.com/translate?hl=en&sl=ja&u=http://va2577.github.io/post/51/&prev=search>

1. SMS\_Handle file headers:

**ROWID** – ordered row identifier

**id** - 10 digit phone number (usually with leading 1)

**country** – country of the number

**service** – sending service used for texting, usually SMS or iMessage

**uncanonicalized\_id** – 10 digit phone number stripped of any leading 1’s

#### Android

App Installation

1. Search the app store for “SMS Backup & Restore” by Carbonite.
2. Follow installation instructions.
3. After installation, you will be prompted to select a backup location. Choose “Local backup” and leave “default app folder” selected as the location choice

General Access to Data Files (SMS)

Note: These were tested on a Verizon Samsung Galaxy. The exact location and names of menu items may differ between manufacturers or carriers.

1. Log in as RA to one of the office laptops.
2. Connect the phone to computer with a USB cord.
3. On the phone, enter the SMS Backup & Restore app.
   1. On the main menu click “Backup” to create a backup file. You will do this prior to every download. You may need to verify that the backup is local or hit “confirm local backup only”
   2. When it says Backup Complete, hit okay.
4. Swipe down (for most androids) to bring up the Windowshade menu
   1. One of the items is probably “USB Charging”. Select that option. It should give you other options for the USB interaction. Choose “File transfer”
   2. This should cause a File Explorer window to open on the computer. One of the Folders will be “Phone” or “<Name of Phone>”. Hit “open device to view files”
   3. Inside that folder will be many app folders. Look for the SMSBackupRestore folder.
   4. Inside this folder will be at least 4 files. Two have simple names (sms.xsl and calls.xsl) these are not the ones you want. The backup files are .xml files and they are named with a long number string consisting of the date and time the backup was made (sms-20170309120406.xml)
   5. Copy the files to a temporary location on your desktop.
   6. Rename the backup file to 001\_SMS\_#.xml (where 001 is the subID of the person you are working with and # is the incremental file order at each backup)
   7. Afterwards you can disconnect the participant phone and return to them.
5. Copy the raw backup files to the server. NOTE – you will FIRST need to connect the laptop to the LAN with an ethernet cord to do this.
   * 1. Copy the file and drop into P:\Study Data\RISK\RawData\$SubID
     2. Delete the .xml files from the laptop desktop.

SMS Data Fields

1. protocol – Messaging software protocol, irrelevant
2. address – phone number of the other party on the SMS
3. date – unix timedate SMS received by this phone, in UTC
4. type – indicates message direction
   1. 1 – Received (incoming)
   2. 2 – Sent (outgoing)
   3. 3 - Draft
5. subject – Some messaging software may allow titles to SMS?
6. body – the message itself
7. toa – unknown, blank in all my tests
8. sc\_toa – unknown, blank in all my tests
9. service\_center – unknown, blank in all my tests
10. read - 0 for undread 1 for read
11. status - – unknown, -1 in all my tests
12. locked – whether the message has been locked
13. date\_sent – for INCOMING messages, the send date (seems to differ slightly from the main date)
14. readable\_date - irrelevant
15. contact\_name - Name assigned to the number in the user’s phone Contacts (if any).

### Final focal variables

**SMS\_Phone**: The 10 digit (maybe more for international numbers) phone number

**SMS\_Type**: 1=incoming, 2=outgoing

**SMS\_Text**: the text message content

### Additional Notes

These data will be downloaded each month for the phone. We will create one aggregate file for full 90 days for meta data and content.

## Voice

### Data Access and Storage

For both device platforms, there will be 3 files created per participant across the life of the study. They are to be saved with the naming convention: SubID\_Voice\_#.csv, where # is the sequential file identifier. (\_1, \_2, \_3) Number corresponds with FILE, not with VISIT. If we do not have Visit 1 data, the identifier on visit2 will be \_1, etc.

iPhone

Computer requirements to access Voice:

1. Install iTunes
2. Install iMazing

Steps to get voice content & metadata

1. Log in as RA to one of the office laptops.
2. Plug iPhone into computer with lightening cable.
3. Open iTunes
   1. If prompted with the following: ‘would you like to access this iPhone/Please respond via the phone’
      1. Hit Trust on the phone
      2. Have the participant put in their passcode if prompted
   2. Click on the device in the left hand pane
      1. Select Back up ‘on this computer’
      2. Select Encrypt local backup
      3. Type in “password” for the password (we’ll be removing the password and encryption before the participant leaves)
      4. iTunes will automatically start backing up the phone after you set the password
      5. You will see a progress bar that will disappear when the backup is complete
4. Once the encrypted backup is complete, open iMazing on the computer
   1. Select Back Up
   2. In Options
      1. Make sure Automatic Backups are disabled
      2. you should see that the Backup is Encrypted
      3. The Backup location should be the default
      4. Make sure that Backup Archiving is disabled
      5. Make sure Wi-Fi Connection is disabled
      6. Additional options do not need to be changed
      7. Then press Back Up in the lower right corner
      8. Once the backup is complete click Close Window in the bottom right corner
   3. In the second pane scroll down to Export All Data
      1. Uncheck everything EXCEPT Call History
      2. Hit Export
      3. Save the file to the participant’s raw data folder
      4. Rename the file SubID\_Voice\_# (and keep the extension as a .cs Rename the file SubID\_Voice\_# (and keep the extension as a .csv file
   4. In iMazing delete the backup and forget the device
      1. On the left side of the window right click on the participant’s phone
      2. Scroll to Delete backup and follow any prompts to delete the backup
      3. Then scroll to Forget and click Forget – make sure that Remove pairing is checked and follow all prompts to remove the device
      4. Close iMazing
5. Go back to iTunes
   1. Make sure This Computer is selected under the Backups field
   2. Uncheck the Encrypt local backup and enter the password – this will reset the encryption password
   3. Backup the phone

Computer requirements to access Voice:

1. Install iTunes
2. Install SQLLite <http://sqlitestudio.pl/>
   1. Just download the zip file, unzip, and place the unzipped file in your C:/Program Files folder. No additional “installation” is required. Hit SQLiteStudio.exe to run.

Steps to get voice content & metadata

1. Log in as RA to one of the office laptops.
2. Plug iPhone into computer with lightening cable.
3. Open Itunes
   1. A dialogue box opens:
   2. iTunes will say “would you like to access this iPhone/ Please respond via the phone”.
   3. Go to the phone and hit **Trust** to allow iTunes to connect.
   4. When you see the device in the left-hand pane, right-click and select “**backup**”
      1. Don’t transfer purchases
      2. Don’t encrypt
   5. A message bar appears across the top of iTunes, wait until backup is complete.
   6. A message bar appears across the top of iTunes, wait until backup is complete.
   7. Make sure the backup files you need, exist:
      1. Browse to C:\Users\RA\AppData\Roaming\Apple Computer\MobileSync\Backup (where RA is the login you are using on that computer).
      2. Make sure you look in the backup that you just created, there will be multiple backup folders so you need to choose the one with the most recent date/timestamp. Click into that.
      3. In the search box, type “**5a4935**”. If you see a single file in the search results, your file has been found (full name **5a4935c78a5255723f707230a451d79c540d2741**.)
         1. You will need this file later, so it’s convenient to click, ctrl-c on the file, then click ctrl-V on the desktop to put the file somewhere easy to get back to it
      4. When you see this file, you know the backup is complete and you can disconnect the iPhone and return it to the participant.
      5. Rename the backup file with the crazy name to 001\_Voice\_#.sql (where 001 is the subID of the person you are working with and # is the incremental file order at each backup)
4. Copy the sql backup files to the server. NOTE – you will FIRST need to connect the laptop to the LAN with an ethernet cord to do this.
   * 1. Copy the file and drop into P:\Study Data\RISK\RawData\$SubID
     2. Delete the .sql files from the laptop desktop.
5. Manual Cleaning:
   1. Phone numbers may be stored in different formats depending on the iOS or how they were entered in the Contacts. R will need to be used to clean these.
6. Note: You will be able to detect deleted calls by inspecting the Row ID of the table.
7. Delete the temp file.
8. File headers:

**Z\_PK** – Ordered list of calls

**Z\_ENT** – The identifier of the table

**Z\_OPT** – Z\_OPT indicates the number of times an entity has been changed (starting with 1 when it is initially added to the database).

**ZANSWERED** – This indicates if the call has been accepted (1) or rejected (0).

**ZCALL\_CATEGORY** – uncertain, but not enough data to determine if other values might be useful.

**ZCALLTYPE**– 1 if standard phone call, 16 if FaceTime audio only, 8 if full Facetime A/V

**ZDISCONNECTED\_CAUSE** – enough data to determine if other values might be useful.

**ZFACE\_TIME\_DATA** –not enough data to determine if other values might be useful.

**ZHANDLE\_TYPE** – uncertain, but not enough data to determine if other values might be useful.

**ZNUMBER\_AVAILABILITY** – uncertain, but not enough data to determine if other values might be useful.

**ZORIGINATED** – 1 for outgoing calls, 0 for incoming calls (confirmed)

**ZREAD** – INTEGER DEFAULT '1' read=0 if the call is a missed callread=1 if the missed calls has been checked.

**ZDATE** – datetime value in iOS format UTC time.

**ZDURATION** – duration in seconds

**ZDEVICE\_ID** – unique device ID

**ZISO\_COUNTRY\_CODE**

**ZLOCATION** – unsure if needed

**ZNAME** – Unknown

**ZSERVICE\_PROVIDER** – Unknown

**ZUNIQUE\_ID** – Unknown

**ZADDRESS** – recipient or caller, depending on **ZOriginated**

#### Android

App Installation

1. Search the app store for “SMS Backup & Restore” by Carbonite.
2. Follow installation instructions.
3. After installation, you will be prompted to select a backup location. Choose “Local backup” and leave “default app folder” selected as the location choice

General Access to Data Files (Voice)

Note: These were tested on a Verizon Samsung Galaxy. The exact location and names of menu items may differ between manufacturers or carriers.

1. Enter the SMS Backup & Restore app.
2. On the main menu click “Backup” to create a backup file. You will do this prior to every download.
3. Log in as RA to one of the office laptops.
4. Connect the phone to a computer with a USB cord.
5. Swipe down (for most androids) to bring up the Windowshade menu
6. One of the items is probably “USB Charging”. Select that option. It should give you other options for the USB interaction. Choose “File transfer”
7. This should cause a File Explorer window to open on the computer. One of the Folders will be “Phone” or “<Name of Phone>”
8. Inside that folder will be many app folders. Look for the SMSBackupRestore folder.
9. Inside this folder will be 4 files. Two have simple names (sms.xsl and calls.xsl) these are not the ones you want. The backup file for Voice is an .xml files and is named with a long number string consisting of the date and time the backup was made (calls-20170309120406.xml)
10. Copy the files to a temporary location on your desktop.
    1. Rename the backup file to 001\_Voice\_#.xml (where 001 is the subID of the person you are working with and # is the incremental file order at each backup)
    2. Afterwards you can disconnect the participant phone and return to them.
11. Copy the raw backup files to the server. NOTE – you will FIRST need to connect the laptop to the LAN with an ethernet cord to do this.
    * 1. Copy the file and drop into P:\Study Data\RISK\RawData\$SubID
12. Delete the .xml files from the laptop desktop.

Voice Call Data Fields

1. \_number – the other party on the call
2. \_duration – in seconds
3. \_date – Unix time in UTC
4. \_type - indicates call direction/success status
   1. 1 - Incoming
   2. 2 - Outgoing
   3. 3 – Missed (incoming)
   4. 4 – Voicemail (incoming)
   5. 5 – Rejected (incoming, rejected by direct user action)
   6. 6 - Refused List (incoming, blocked automatically)
   7. 7 – Answered externally (simultaneous ring on multiple devices not answered on this device)
5. \_presentation – the number presentation rules (i.e., Caller ID) as set by the network. Possible values: 1 (allowed), 2 (restricted), 3 (unknown), 4 (payphone). See <https://developer.android.com/reference/android/provider/CallLog.Calls.html#NUMBER_PRESENTATION>.
6. \_readable\_date - irrelevant
7. \_contact\_name - Name assigned to the number in the user’s phone Contacts (if any).

### 

### Final Focal Variables

Voice\_Phone: 10 digit (or longer for international) phone number

Voice\_Duration: call duration in seconds

Voice\_Type: 1=incoming, 2=outgoing is definitive. Other codes are available for android (see table below) but not used right now. We “could” get these other codes for IOS as well but probably not worth it.

### <https://app.asana.com/0/0/480923759693421/f>:

### Android iOS

### Incoming/received 1 zOrg = 0; zAnswered =1, zRead = 1

### Incoming (Missed, Voicemail) 3, 4 zOrig = 0; zAnswered =0, zRead = 0, zOpt = 1

### OR zOrig = 0; zAnswered =0, zRead = 1, zOpt > 1

### Incoming (Rejected/Refused) 5, 6 zOrig = 0; zAnswered =0, zRead = 1 , zOpt = 1

### Outgoing 2 zOrig = 1; zAnswered =0, zRead = 1

## Qualtrics ID Batteries

There will be additional CSV files generated from Qualtrics surveys given at each visit. These files are: screening (Screen.csv), intake (Intake.csv), follow-up (Followup12.csv), final session (FinalVisit.csv), and subject details from three sessions (details1.csv; details2.csv; details3.csv). Each of these files will contain data from all subjects. They are saved at \RawData\. There will be one row per subject in each file except for follow-up which will have two rows per subject.

### Data Access and Storage

1. ID data should be downloaded each time a participant completes the study.
2. Login to Qualtrics
3. Select and download ID battery:

* ID Screen Battery-RISK
* ID Intake-Battery
* ID Followup 1 & 2 Battery-RISK
* ID Followup 3 Battery-RISK

1. Select **Data and Analysis** from the menu at the top of the screen
2. Select down arrow from **Export & Import icon** on the right side of the screen, and select **Export Data**
3. Choose **Export Data with Legacy Format**
4. CSV is the default. Check “Use legacy View Results Format.”
5. Hit **More Options**. Select “Use numeric values” radio button and **uncheck** all boxes except “Use Question Numbers”.
6. Hit download. The file is saved as a .csv file.
7. Save file here: P:\Study Data\RISK\RawData
8. Save downloaded ID batteries as follows (below) overwriting the existing file.

* Screening
* Intake
* Followup12
* FinalVisit

## Interview Data

These files will contain data obtained via a structured interview about the participant’s recovery environment that takes place during the participant’s intake study visit. At follow-up visit one and follow-up visit two, the data obtained at intake will be updated as needed to reflect any changes that have occurred in-between study visits.   
  
There will be several Excel files generated from the interview data. These files contain data that span multiple columns. Blank templates of the interview files can be found in: **P:/StudyData/RISK/Administration/Templates/Recovery Environment Interview**. The interview files can be found here: **P:/StudyData/RISK/RawData/**.   
Each file will begin with the participant’s subject ID\_file type:   
contacts (###\_Contacts), locations (###\_Locations), dates (###\_Dates), times (###\_Times), and vacations (###\_Vacations).

All files should be pasted into participant’s folder even if no locations or emotionally important dates are every reported

Once the participant is off study, it should be confirmed that all available data has been entered and NA should be entered for any missing data.

UTC Time Stamp  
All files will have a UTC column. The UTC can be obtained by following the steps below:

1. Go to [https://www.epochconverter.com/](https://www.epochconverter.com/" \t "_blank)

2. Enter in the month, day, year, and hour (in military time) the interview was completed. Select local time in the drop-down menu. Click on Human date to Timestamp.

3. The time stamp will appear next to -“Epoch timestamp: ##########”

4. Insert this time in the UTC column for all interview data files.

File Contents  
The locations file (####\_Locations) contains nine columns:

1. **UTC**: Obtained by following the steps above.
2. **StreetAddress, City, and State**: Enter in location address with single spaces between the street number, street name, and street type (ex: Avenue, Boulevard, Drive, Lane, etc). Spell out street type
   1. Spell out street numbers (Fifth) unless the number is greater than 99. i.e, 104th Street
   2. If a range of addresses is entered on the location log and these cannot be narrowed down, select the first address number
3. **City**
4. **State**
5. **Type:** Enter in the location type. Spell out location type exactly as it appears on the location log utilized to complete the interview. If it is the participant’s home address, type in “Home.”
6. **DrankHere:** Enter Yes or No to indicate if participant has ever drank alcohol at this location.
7. **AlcoholHere:** Enter Yes or No to indicate if alcohol is available in this location.
8. **LocationEmotion:** Enter Pleasant, Unpleasant, Mixed, or Neutral to indicate how participant’s experiences are when at this location.
9. **LocationRisk:** Enter in High, Medium, Low, or No to indicate participant’s risk level to begin drinking again when visiting this location. If the location is one of the participant’s risky locations, enter in “High,” in the LocationRisk column.
10. **RiskAvoid:** Enter in Yes or No to indicate if participant has identified this place as a place they are trying to avoid now that they are sober. The locations identified by participant as “risky locations,” will all be recorded as “Yes,” all other locations will be “No.”
11. **Vacation:** **If this address is a location that the participant visited while on vacation enter “Yes,” if not, enter “No.” This is something that the interviewer will have to note on the location log. If it is not noted, as the interviewer, then note it on the location log and enter in the appropriate response.**

**Important Note-When Updating Location Times at Follow-up #1 and #2**

* If a location that was previously recorded is mistakenly reported again, please ignore the additional entry. We are only concerned with the initial entry that was reported.

The contacts file (####\_Contacts) contains twelve columns:

1. **UTC**: Obtained by following the steps above.
2. **Always enter in the study participant’s cell phone number as the first contact**. For type enter in “Self.”
3. **HomePhone:** Enter in home phone number with no spaces between digits.
4. **CellPhone:** Enter in cell phone number with no spaces between digits. Default entry if cell or landline has not been specified and is not irrelevant/spam
5. **OtherPhone1:** To be utilized if contact has a number other than a home/cell number that participant uses to reach contact. Enter in number with no spaces between digits. Mostly used for irrelevant/spam entries
6. **OtherPhone2:** Same as #4 above
   * 1. NOTE: If a person has reported two numbers for the same individual at the same visit, they can be reported in the same row
     2. NOTE: If, however, the person reported a new number for a previously reported contact at a later visit, put the new number in a new row
7. **Type:** Enter type of relationship. Spell out relationship exactly as it appears on the contact log.. If it is a family member, indicate kin type instead of “Family.” The kin types are as follows: Parent, Sibling, Child, Aunt/Uncle, Cousin, Grandparent, or Family-Other (anything not in one of the previous noted kin categories). \*\*Please Note that a step mother/father should be entered in as “Parent,” and a step brother/sister should be entered in as “Sibling.” All in-laws should be entered in as “Family-Other.”
8. For any contacts listed as “Irrelevant/Spam ” record the phone number under OtherPhone1 and Type as Irrelevant/Spam,” no other information will be needed or available for these contacts. “Irrelevant/Spam”refer to contacts that will be seen on the participant’s phone who do not have an impact on their recovery (ex: school notifications, spam texts like store coupons and sales notices, delivery notices, business contacts (e.g., clients, customers, vendors, etc.), other automated messages, etc.).
9. **ContactDrankPast:** Enter in Never/Almost Never, Occasionally, or Almost Always/Always if participant has ever drank alcohol with this contact.
10. **DrinkStatus:** Enter in Drinker, NonDrinker, or Dont Know (enter don’t without an apostrophe) the drinking status of contact.
11. **ContactDrinkFuture:** Enter in Yes or No to indicate if participant thinks contact would drink alcohol in their presence.
12. **Recovery:** Enter Yes, No, or Dont Know (enter don’t without an apostrophe) to indicate if contact is in recovery from alcohol or other substances.
13. **SupportStatus:** Enter in Supportive, Unsupportive, Mixed, or Neutral to indicate the level of support the contact provides participant.
14. **ContactExperience:** Enter in Pleasant, Unpleasant, Mixed, or Neutral to indicate how participant identifies feeling with contact.
15. **Monthly Visit:** Enter in “yes” if the participant visits the contact in their home monthly. This column is only applicable for family and friends. Do not include monthly visits to the doctor, co-workers, etc.. Also, do not include “monthly” visits to family or friends that the participants lives with.

**Street Address, City, and State:** If the monthly visit column is “yes” please enter the home address of the contact. Enter in location address with single spaces between the street number, street name, and street type (ex: Avenue, Boulevard, Drive, Lane, etc). Spell out street type.

The time-periods file (####\_Times) contains six columns:

1. **UTC**: Obtained by following the steps above.
2. **Type:** Enter in type of time-period exactly as it appears in the interview form under the weekly time-periods section.
3. **RiskOrProtect:** Enter in Risky or Protective to indicate how participant views time-period.
4. **Day:** Enter in Day of week (Mo, Tu, We, Th, Fr, Sa, or Su) to indicate the day of the week the time-period occurs. If it occurs on multiple days, create a new row for each applicable day of the week.
5. **StartTime:** Enter the start time in military time (including colon to separate hour from minutes)
6. **EndTime:** Enter the end time in military time (including colon to separate hour from minutes)

* EndDate: Enter in the EndStudy date from the participant’s Excel Visit Dates file located in the participant’s raw data folder **unless otherwise indicated**.   
  \*\*\*Ex: If a participant completed an intake visit and reported that his work hours are protective from 9 am-5 pm Monday-Friday and then returns to his first follow-up appointment and reports he no longer has the same job; you will need to change the EndDate to the participant’s last date of employment instead of the EndStudy date. The last date of employment should be indicated on the participant’s follow-up interview form.  
    
  \*\*\*Important Note: When Updating Risky Times at Follow-up #1 and #2\*\*\*   
  If the EndDate needs to be changed from the EndStudy date please make sure to make the necessary changes. Employment, counseling sessions, IOP programs, etc..are times that may end while a participant is on-study and therefore need updating.You will include a separate entry (row) for each day for which this risk/protective time is applicable. For example, if they reported the hour after work each weekday, you would include 5 entries, one for each of the workdays of the week. You will also need a separate entry (row) for each risk/protective time that continues past midnight. For example, if they reported a risky time as Fridays from 9pm to 3am you would have an entry for Friday from 21:00 to 23:59 and one for Saturday from 0:00 to 3:00.

The Emotionally Important Days file (####\_Dates) contains 3 columns:

1. **UTC**: Obtained by following the steps above.
2. **Type**: Enter in type of day exactly as it appears under the Emotionally Important Days section on the interview form.
3. **Date:** Enter in Date mm/dd/yyyy

These days will typically be days that occur annually, however if there are dates that occur weekly or monthly, enter in a new row for each weekly or monthly date.

The Vacations file (###\_Vacations) contains 6 columns:

1. **UTC**: Obtained by following the steps above.
2. **In Town or Out of Town**: Enter either In Town or Out of Town based on which box is checked on the interview form.
3. **City**
4. **State**
5. **Start Date**: Enter in Date mm/dd/yyyy
6. **End Date**: Enter in Date mm/dd/yyyy

## Wake/Sleep times

1. Form is filled out by on paper by the RA during the session. Wake/Sleep times will be recorded in military time.
2. At the end of the intake visit, the RA enters the data into the Qualtrics form, “Sleep-Wake Times Battery – RISK.” The wake/sleep times will be recorded in military time within the “Sleep-Wake Times Battery – RISK.”
3. When a new participant’s record is entered, the aggregate battery results should be exported:
   1. Choose **Export Data with Legacy Format**
   2. CSV is the default. Check “Use legacy View Results Format.”
   3. Hit **More Options**. Select “Use numeric values” radio button and **uncheck** all boxes except “Use Question Numbers”.
   4. Hit download. The file is saved as a .csv file.
   5. Save file here: P:\Study Data\RISK\RawData\SleepSchedule.csv
4. This file becomes the data input file for generating the participant Rolodex of EMA times.

# Mongo DB

The database code is under version control using my AFS user name @ trilobite.cs.wisc.edu at

svn+ssh://jjcurtin@trilobite.cs.wisc.edu/p/zhu/10/svn/repositoryAlcohol

pw: Abcd1234\*

**Build database**

[First go to folder where this code lives]

MongoDBServer.m at matlab command line

BuildDBFromScratch(P:\StudyData\RISK\Database’) at matlab command line

A file will be created at E:\ that lists all files imported into the database

**Working with data streams.**

GetFeatureNames('SS') To get list of features that contain “SS’. This is case insensitive

function [ result,status ] = queryDB( PID, X1, X2, X3 )

%%

%X1 feature name

%X2 t1

%X3 t2 or k

[a,b]=queryDB(001,'SMS\_Text','2016-11-18,13:57:18',3)

Steps for feature engineering, model training and test

GetXY (uses ConstructFeatureVector): This produces X and Y (unprocessed)

Lasso Does imputations, normalization, drop zero variance, then CV and fit final training model

# 

Questions for Jerry

How to handle missing data on features – best missing data estimation procedure?

Mean vs. median vs. KNN?

Normalize,impute, identify nzv inside or outside of CV

AUC vs. cost to pick lambda