# Module 4: Set-Up of Dependent Variable

April 26, 2021

## Contents



**MODULE 4 GOAL:** By the end of this module, you will be able to:

- Learn the logic of constructing the Dependent Variable using scientific considerations (i.e., our motivating scientific question) and practical considerations (i.e., the data collection design).

## 1 Scientific Question

For convenience, let's display the scientific question we introduced in an earlier module.

**SCIENTIFIC QUESTION:** On average, is self-efficacy *at the current time point* associated with *the proximal occurrence of cigarette smoking* during the post-quit period?

Draw Figure Here.

## 2 Implementation

**Reference file with R code:** `module-04.R`

We will zoom in on specific sections within the implementation in `module-04.R`.

An outer loop goes through each participant. For each participant, the inner loop goes through each of their rows in `dat_analysis`. Recall that, at this point, only Random EMAs having *any* response

(i.e., `with_any_response=1`) are included in `dat_analysis`. Hence, for a particular participant, `total_random_ema` will be the total number of Random EMAs for which the participant provided *any* response.

```
for(i in 1:total_participant_ids){

  # More code here

  for(j in 1:total_random_ema){

    # More code here

  }
}
```

Draw Figure Here.

The inner loop calculates `count_within_bounds`, our Dependent Variable.

```
  for(j in 1:total_random_ema){
    current_lower_bound <- all_lower_bound[j]
    current_upper_bound <- all_upper_bound[j]

    # How many EMAs were launched between the two Random EMAs we
    # are looking at now?
    dat_within_bounds <- current_dat_smoking %>%
      # Note the use of '>' instead of '>=' when checking against left end point
      # We do not include the number of reported cigarettes smoked in the left end point
      # However, we will include the number of reported cigarettes smoked in the right end point
      filter((time_unixts > current_lower_bound) & (time_unixts <= current_upper_bound))

    number_within_bounds <- nrow(dat_within_bounds)
    # Only proceed with further calculations if we have at least one EMA
    if(number_within_bounds > 0){
      number_missing <- sum(is.na(dat_within_bounds$smoking_qty))
      # Only proceed with further calculations if there is no missing value in smoking_qty
      if(number_missing == 0){
        current_count_within_bounds <- sum(dat_within_bounds$smoking_qty)
      } # Mark end of IF STATEMENT
```

```
  } # Mark end of IF STATEMENT

  current_dat_analysis$count_within_bounds[j] <- current_count_within_bounds
} # Mark end of FOR LOOP over Random EMAs having with_any_response=1
```

**What are the bounds?**

Draw Figure Here.

**Which EMAs within these bounds should we *not* use?**

Draw Figure Here.

**When may missing values come about in our Dependent Variable `count_within_bounds`?**

Draw Figure Here.

BREAK: Any questions?