In the dictionary csv, you will see a column named “flag” on the right hand side. This will be the only column you edit. You will be deciding which vversions you think you should drop and keep for each variable within each vlist. The only values that will be kept when you run the 02.Output.R script are values equal to “include” in the flag column, everything else will be dropped. The flags are designed to be very sensitive. So I would recommend not bothering even checking the “include” values and just focusing on the rows that don’t say “include” to see if you want to change any of them to “include”. There are 4 different values the flag column can equal:

1. non.na - This means this particular vversion was flagged because of a significant difference between the number of existing values within this vversion and the weighted average number of existing values for all vversions of this same variable and vlist. Take a look at these columns to decide whether to include this vversion or not:

* exist - existing incidence of values for this vversion
* wm.exist - weighted mean of existing incidence of values across all vversions within this variable and vlist... weighted on sample size
* sample - total sample size for this vversion within this variable and vlist

1. missing – The flag column will contain this value when all vversions for this variable and vlist are made up of all missing values.
2. zero – This signifies a significant difference between the number of zeros for this vversion and the weighted average number of zeros for all vversions within this same variable and vlist. The idea behind this is to account for situations when an answer choice would be added later to a question in the survey. Sometimes Decipher will add in an answer choice later and all the previous values for this answer choice become zero instead of missing. To look out for this, look for rows where zero is significantly lower than wm.zero.
3. include – This row was not flagged as suspicious; it will be included.