

## **NAACCR Record Uniqueness – SAS Macro (v1.2)**

### **April 2007**

#### **Syntax**

```
filename UMacro 'physical location of the file named uniquemacro';  
%INCLUDE UMacro;
```

```
%Unique(DATAIN=SAS dataset name,  
        UNIQ_VARS= variable names,  
        UNIQ_FMT= variable name format name,  
        FREQ_YN=YES or NO,  
        FREQ_VARS=variable names,  
        FREQ_FMT= variable name format name,  
        TITLE1="title with double quotes",  
        TITLE2="title with double quotes",  
        TITLE3="title with double quotes",  
        TITLE4="title with double quotes",  
        TITLE5="title with double quotes"  
        );
```

#### **Description of Macro Parameters**

- |                  |   |
|------------------|---|
| <b>DATAIN</b>    | The name of the SAS dataset containing the variables to be analyzed   |
| <b>UNIQ_VARS</b> | The list of variables to be analyzed for uniqueness. Variables are separated by a space. Because of memory limitations, the maximum number of variables that can be examined is nine.   |
| <b>UNIQ_FMT</b>  | The list of variables and corresponding format names. By creating different variable formats, the macro can be executed multiple times without recreating the SAS data set. If there are no formats associated with any of the analysis variables, insert NONE.<br><br>Example:<br><pre>        ...<br/>        UNIQ_FMT  race race_fmt. Age age_fmt.,<br/>        ....</pre> |
| <b>FREQ_YN</b>   | Value can be YES or NO. A value of YES will generate frequency distributions for any variables listed in the parameter FREQ_VARS.   |
| <b>FREQ_FMT</b>  | The list of any variables and corresponding format names listed in the FREQ_VARS parameter. If there are no formats associated with any of the analysis variables, insert NONE.   |

## **NAACCR Record Uniqueness – SAS Macro (v1.2) Continued**

**TITLE1-TITLE5** These are optional titles that appear in the output. Titles must be enclosed with double quotes. If no additional titles are required, insert “ ” (a space enclosed with double quotes).

### **Output**

**Basic results** This report lists all combinations which have at least one unique record, or unique records which occur in groups of 5 or less. The report also shows the total records, the variable combination, and the total number of observations used for the analysis. The results are sorted by the number of unique records.

**Advanced results** This page gives the relative contribution of each variable in the Variable Set to the total number and proportion of unique records on the data file. If a user wants to decrease the number of unique records on the data file, the variable with the greatest weight should be selected and the values should be aggregated into larger categories. For example, aggregating primary site into SEER Site groups would be a logical iteration to affect the greatest reduction in unique records. This would be followed by aggregation of age into age groups, race into race groups, year of diagnosis, and finally sex. When re-aggregation of variable values into meaningful categories is insufficient in reducing the magnitude of unique records, then one should consider omitting the variable from data file to be released to a user.

### **Hints/Troubleshooting**

1. Keep analysis variables in discrete categories. The program executes faster with fewer categories used.
2. Make sure variable names are separated by a space. The end of a parameter line should end in a comma. The last parameter line in the macro (TITLE5) does not use a comma.
3. Check to make sure titles are enclosed with double quotes.
4. Make sure format names have a period.