Array Applications in Determining Periodontal Disease Measurement

Liang Wei, DB Consulting Group, Inc.; Laurie Barker, Division of Oral Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention; Paul Eke, Division of Population Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention

ABSTRACT

A SAS® array can be defined as a series of related variables under a single name. It can be created with the array statement within a data step and can provide a way for repetitively processing variables using a do-loop. SAS arrays can be used for creating or modifying a group of variables, reshaping a data set or comparing across observations. When SAS arrays and do-loops are properly used, they can become powerful data manipulation tools which make the SAS code more concise and more efficient. This paper introduces one-dimensional and two-dimensional array applications for manipulation of repeated measures data, using periodontal disease data for illustration. The coding approach can be particularly helpful for other researchers and data analysts working in oral health and other medical fields and can also be helpful for SAS programmers working on the similar repetitive task.

INTRODUCTION

One-dimensional and two-dimensional arrays can be used to calculate outcomes based on multiple variable measures. This coding approach can be helpful for researchers and data analysts working in oral health and other medical research fields where multiple measurements are used to calculate summary measures. The examples provided in this report are based on clinical periodontitis (gum disease) measurements and data structures from the National Health and Nutrition Examination Survey (NHANES), 2009-2010. This paper describes one- and two-dimensional array definitions, application of arrays to periodontitis measurements, SAS code with Do Loop applications and SAS code with more complicated calculations based on adjacent teeth.

Figure 1. Calculation of Attachment Loss

Case definitions for surveillance of periodontitis consider two types of clinical periodontal measurements: attachment loss (AL) and pocket depth (PD). Attachment loss, also called loss of attachment (LOA), clinical attachment loss or calculated attachment loss (CAL), is the detachment of the gum from the tooth root and loss of other supporting tissues, such as bone, that hold the teeth in place. When the gums pull away from the teeth pockets can form. Larger values of attachment loss and pocket depth, and greater numbers of teeth with affected sites, indicate more severe periodontal disease. To make these measurements, the examiner inserts a small probe with a rounded tip and
millimeter marks between the gum tissue and the tooth. Measurements are recorded at multiple sites on each tooth. Fractions of millimeters are not recorded – all measurements are recorded to the lowest whole millimeter.

Pocket depth is the distance in millimeters from the gum line, also called the free gingival margin (FGM), to the bottom of the pocket formed between the gum and tooth. Gingival recession is the distance in millimeters from the gum line (FGM) to the cemento-enamel junction (CEJ). The CEJ is where the smooth enamel of the tooth crown meets the tooth root’s rough-textured covering (cementum). To calculate attachment loss (Figure 1) subtract gingival recession from pocket depth. If the gum line (FGM) has receded past the CEJ onto the tooth root, gingival recession (the distance from the gum line (FGM) to the CEJ) will be a negative number.

The 2009-2010 National Health and Nutrition Examination Survey (NHANES) recorded pocket depth and gingival recession at six probing sites for each of 28 teeth (Figure 2). Probing sites for the third molars, also called wisdom teeth, were not included in the assessment because, in the U.S. population, these teeth are often extracted soon after eruption or surgically removed before eruption. Periodontal measurements are taken only where teeth are present, so data can be missing if the tooth is missing. Data may also be missing for other reasons that may have interfered with measurement at that site (presence of partial dentures, only the tooth root is present, etc.). People who are edentulous (have no natural teeth) are excluded from the assessment.

Figure 2. Periodontal Probing Sites

The three probing sites on the side of the teeth closest to the face or lips are called the facial or buccal sites, and the three on the side of the teeth closest to the tongue are called the lingual sites. On each side of the tooth, measurements are taken by probing at one site closer to the front or centerline of the mouth (mesial, mesio-), another site towards the back of the mouth (distal, disto-), and a final site about halfway between the mesial and distal sites (mid-). The six probing sites are called mesio-facial, mid-facial, disto-facial, mesio-lingual, mid-lingual and disto-lingual (Figure 2).

The mesial and distal sites are also called interproximal sites, because they are near the space between two adjacent teeth. Attachment loss, if present, is expected to be greater at these four sites than at the mid-facial and mid-lingual sites.

The pocket depth and gingival recession measurements were recorded directly into an NHANES oral health data management program that instantly calculated AL as the difference between pocket depth and gingival recession. All three measurements are available from the NHANES 2009-2010 public use data set.
We use arrays and do loops to calculate periodontitis severity for two case definitions, the Centers for Disease Control and Prevention/American Academy of Periodontology definition (CDC/AAP)\(^3,4\) used in the U.S. and the European Workshop on Periodontology definition\(^5\) used in Europe.

First, we categorized survey participants into one of three mutually exclusive levels of severity of periodontitis (i.e., severe, moderate and mild periodontitis) following the CDC/AAP case definitions:

- **Severe periodontitis** was defined as the presence of 2 or more interproximal sites with ≥ 6 mm AL (not on the same tooth) and 1 or more interproximal site(s) with ≥ 5 mm PD.
- **Moderate periodontitis**, among those who did not meet the severe periodontitis case definition, was defined as 2 or more interproximal sites with ≥ 4 mm clinical AL (not on the same tooth) or 2 or more interproximal sites with PD ≥ 5 mm, also not on the same tooth.
- **Mild periodontitis**, among those who met neither the severe nor moderate periodontitis case definition, was defined as ≥2 interproximal sites with ≥3 mm AL and ≥2 interproximal sites with ≥4 mm PD (not on the same tooth) or 1 site with ≥5 mm PD.

Two additional categorizations were calculated:
- any (or total) periodontitis, defined as those with severe, moderate, or mild periodontitis, and
- moderate/severe periodontitis, defined as those with either moderate or severe periodontitis.

Second, we classified survey participants into two mutually exclusive groups, incipient or severe cases of periodontitis following the criteria of the European Workshop of Periodontology definition:

- **Severe periodontitis** was defined as having more than 30 percent of teeth with loss of attachment larger than five millimeters.
- **Incipient cases**, among those who were not classified as having severe periodontitis, with AL greater than three millimeters (AL >3 mm) on two non-adjacent teeth.
- Additionally, any (or total) periodontitis by the European definition was defined as those with either incipient or severe periodontitis.

**DATA STRUCTURE AND CODING FOR TWO PERIODONTITIS CASE DEFINITIONS**

1. ONE- AND TWO-DIMENSIONAL ARRAYS APPLIED TO TOOTH SURFACES AND WHOLE TEETH

1.1 Two-dimensional array

NHANES stores oral health examination measurements on teeth in variables numbered 1 through 32, beginning with the upper right 3rd molar, and ending with the lower right 3rd molar. Because the periodontal examination excludes 3rd molars (teeth numbered 1, 16, 17 and 32), the periodontal measurements are stored in variables numbered 2 through 15 for teeth in the upper arch, and 18 through 31 for teeth in the lower arch.

<table>
<thead>
<tr>
<th>Probing sites</th>
<th>Attachment loss for upper right 2nd molar</th>
<th>Pocket depth for upper right 2nd molar</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHANES variable name</td>
<td>Variable in array loa[6,28]</td>
<td>NHANES variable name</td>
</tr>
<tr>
<td>1 disto-facial</td>
<td>OHX02LAD</td>
<td>loadisf1</td>
</tr>
<tr>
<td>2 mesio-facial</td>
<td>OHX02LAS</td>
<td>loamesf1</td>
</tr>
<tr>
<td>3 mesio-lingual</td>
<td>OHX02LAA</td>
<td>loamesl1</td>
</tr>
<tr>
<td>4 disto-lingual</td>
<td>OHX02LAP</td>
<td>loadisl1</td>
</tr>
<tr>
<td>5 mid-facial</td>
<td>OHX02LAM</td>
<td>loamdfl1</td>
</tr>
<tr>
<td>6 mid-lingual</td>
<td>OHX02LAL</td>
<td>loamdfl1</td>
</tr>
</tbody>
</table>

**Table 1. Array Structure for Periodontal Measurement Variables**

The NHANES variables follow the naming convention OHX##PMS where OHX is the prefix for the oral health exam, ## is the 2 digit tooth number, PM is the periodontal measurement (LA for loss of attachment, and PC for Pocket Depth), and S is the probing site (D for disto-facial, S for mesio-facial, A for mesio-lingual, P for disto-lingual, M for mid-buccal, and L for mid-lingual). For each person there are 168 measurements, six sites on each of 28 teeth.

Because variables for only 28 teeth are used in the array, the array variables are numbered 1 through 28. Array variables begin with "loa" for loss of attachment or "pd" for pocket depth. The next four characters indicate the probing site, and the final 1-2 digits, the tooth number. For example, loss of attachment measured on the upper right molar at
the disto-facial site is stored in array variable, loadisf1, and the corresponding measurement on the lower right molar is stored in array variable, loadisf28.

Table 1 shows the NHANES variables used for recording the periodontal examination and the corresponding array variable names used in the code for tooth number 2, the upper right 2nd molar. The mid-facial and mid-lingual sites are placed at the end of the array because they are not used for the CDC/AAP definition.

A two-dimensional array loa[6,28] was defined for the calculated attachment loss variables. The first index is the number of sites per tooth (rows in Table 1), and the second is the number of teeth per person (columns in Table 1):

```plaintext
array loa[6,28] loadisf1-loadisf28 loamesf1-loamesf28
loames11-loames128 loadis11-loadis128
loamdf1-loamdf28 loamdl1-loamdl28;
```

Valid values for attachment loss range from zero to 24 millimeters.

Similarly, we define the array for the pocket depth measurements as:

```plaintext
array probe[6,28] pddisf1-pddisf28 pdmesf1-pdmesf28
pdmes11-pdmes128 pddis11-pddis128
pdmdf1-pdmdf28 pdmdl1-pdmdl28;
```

Valid values for pocket depth range from zero to 12 millimeters.

1.2 One-dimensional array

Four one-dimensional arrays were used to create the periodontal disease categorizations:

```plaintext
array s_missing[28] s_missing1-s_missing28;
array t_missing[28] t_missing1-t_missing28;
array al3_t[28] al3_t1-al3_t28;
array al5_t[28] al5_t1-al5_t28;
```

The array s_missing[28] stores the number of missing sites on 28 teeth (valid range from zero to 6). The array t_missing[28] indicates whether each tooth is present (1) or missing (0). The array al3_t[28] indicates which teeth have loss of attachment greater than three millimeters. The array al5_t[28] indicates which teeth have attachment loss greater than five millimeters.

2. SAS CODE FOR CDC/AAP PERIODONITIS DEFINITION

The CDC/AAP periodontitis categorizations were based on four interproximal sites including mesio-facial, disto-facial, mesio-lingual and disto-lingual sites.

The outcome variables of CDC/AAP periodontal definitions are obtained in three steps:

1. Define array variables required for calculation of periodontal measurements
2. Calculate periodontal measurements based on CDC/AAP periodontal definitions
3. Create outcome variables of CDC/AAP periodontal definitions

2.1 Define array variables required for calculation of periodontal measurements

The code begins with a two-dimensional array loa[6,28] for attachment loss and another two-dimensional array probe[6,28] for pocket depth. Then five one-dimensional arrays are defined for attachment loss ≥6mm (loa_ac), attachment loss ≥4mm (loa_bc), pocket depth ≥5mm (pdc); attachment loss ≥3mm (mild_ac) and pocket depth ≥4mm (mild_pd). The code also defines two variables to count the number of sites with valid values for attachment loss and pocket depth.

```plaintext
data test1;
set one;
array loa[6,28] loadisf1-loadisf28 loamesf1-loamesf28
loames11-loames128 loadis11-loadis128
loamdf1-loamdf28 loamdl1-loamdl28;
array probe[6,28] pddisf1-pddisf28 pdmesf1-pdmesf28
pdmes11-pdmes128 pddis11-pddis128
pdmdf1-pdmdf28 pdmdl1-pdmdl28;
array loa_ac[28] lac1-lac28;
array loa_bc[28] lbc1-lbc28;
array pdc[28] pdcl-pdc28;
```

2.2 Calculate periodontal measurements based on CDC/AAP periodontal definitions

2.3 Create outcome variables of CDC/AAP periodontal definitions
array mild_al[28] mild_all-mild_al28;
array mild_pd[28] mild_pd1-mild_pd28;
loa_valid=0;/*initialize counter for the number of sites with valid AL data*/
pd_valid=0;/*initialize counter for the number of sites with valid PD data*/

2.2 Calculate periodontal measurements based on CDC/AAP periodontal definitions

The do loops following the array definitions apply the two-dimensional arrays and one-dimensional arrays to populate indicator variables for sites and teeth with attachment loss greater than 3, 4, and 6 millimeters, and pocket depth greater than 4 and 5 millimeters, respectively. Nested do loops were used to process the two-dimensional array. The first index, for teeth, is the outer loop and the second index, for sites, is the inner loop.

\[
\begin{align*}
d &\gets 1 \text{ to } 28; \\
d &\gets 1 \text{ to } 4; \\
\text{if} \ loa[i,j] &\geq 0 \text{ and } loa[i,j] \leq 24 \text{ then } loa_valid+1; \\
\text{if} \ loa[i,j] &\geq 6 \text{ and } loa[i,j] \leq 24 \text{ then } loa_ac[j]=1; \\
\text{if} \ loa[i,j] &\geq 4 \text{ and } loa[i,j] \leq 24 \text{ then } loa_bc[j]=1; \\
\text{if} \ loa[i,j] &\geq 3 \text{ and } loa[i,j] \leq 24 \text{ then } mild_al[j]=1; \\
\text{if} \ probe[i,j] &\geq 0 \text{ and } probe[i,j] \leq 12 \text{ then } pd_valid+1; \\
\text{if} \ probe[i,j] &\geq 5 \text{ and } probe[i,j] \leq 12 \text{ then } pdc[j]=1; \\
\text{if} \ probe[i,j] &\geq 4 \text{ and } probe[i,j] \leq 12 \text{ then } mild_pd[j]=1;
\end{align*}
\]

end; end;

After the site level measurements of attachment loss and pocket depth were calculated, the measurements were summed over the number of teeth.

loasum = sum(of lac1-lac28);
lobsum = sum(of lbc1-lbc28);
pdsum = sum(of pdcl1-pdcl28);
mild_ac_sum = sum(of mild_all-mild_al28);
mild_pd_sum = sum(of mild_pd1-mild_pd28);

2.3 Create outcome variables of CDC/AAP periodontal definitions

The code first determines whether valid values exist for periodontal measurements. If there is no valid value, then the outcome variable is assigned a missing value. If there is a valid value, then outcome variables are assigned values corresponding to the levels of the CDC/AAP periodontitis case definition. The CDC/AAP definition categorizes periodontitis into three mutually exclusive categories of mild, moderate and severe. (In estimates and code presented here, if a person has only one tooth with valid AL or PD values, the person is considered a non-case and is included in the denominator of prevalence estimates.)

\[
\begin{align*}
/*if there are no valid AL or PD values then set the CDC/AAP 
definition variables to missing*/
\text{if} \ loa_valid=0 \text{ and } pd_valid=0 \text{ then } do; \\
\text{cdc_pd_s}=.; \\
\text{cdc_pd_m}=.; \\
\text{cdc_pd_t}=.; \\
\text{cdc_pd_mild}=.; \\
\text{cdc_pd_total}=.; \\
end; \\
/*if there are valid AL or PD values then assign the CDC/AAP categories*/
\text{else if} \ loa_valid>0 \text{ or } pd_valid>0 \text{ then } do; \\
/*cdc/aap severe*/
\text{if} \ (loasum>=2 \text{ and } pdsum>=1) \text{ then } cdc_pd_s=1; \\
\text{else} \ cdc_pd_s=0; \\
\text{if} \ cdc_pd_s=0 \text{ then } do; \\
/*cdc/aap moderate*/
\text{if} \ loasum>=2 \text{ or } pdsum>=2 \text{ then } cdc_pd_m=1; \\
\text{else} \ cdc_pd_m=0; \\
end; \\
\text{else if} \ cdc_pd_s=1 \text{ then } cdc_pd_m=0; \\
/* severe and moderate categories combined*/
\text{if} \ cdc_pd_s=1 \text{ or } cdc_pd_m=1 \text{ then } cdc_pd_t=1; \\
\text{else} \ cdc_pd_t=0; \\
/*cdc/aap mild periodontitis definition*/
\end{align*}
\]
if cdc_pd_s=1 or cdc_pd_m=1 then cdc_pd_mild=0;
else if cdc_pd_s ^=1 and cdc_pd_m ^=1 then
do;
/*limit mild category to those not categorized as severe or moderate*/
if mild_ac_sum>=2 and (mild_pd_sum>=2 or pdsum>=1) then cdc_pd_mild=1;
else cdc_pd_mild=0;
end;
/*cdc/aap total (any periodontitis): severe, moderate and mild
categories combined*/
if (cdc_pd_s=1 or cdc_pd_m=1 or cdc_pd_mild=1) then cdc_pd_total=1;
else cdc_pd_total=0;
end;
run;

The combination of arrays and do-loops is particularly efficient and fast in locating the variables used for the
calculations and computing values for the calculated variables.

3. SAS CODE FOR EUROPEAN PERIODONTITIS DEFINITIONS

The European Periodontitis case definitions are obtained in four steps:

1. Count the number of sites with loss of attachment greater than three and five millimeters
2. Count the number of teeth with loss of attachment greater than three and five millimeters
3. Exclude persons who have only two adjacent teeth with loss of attachment greater than 3 millimeters
4. Create variable for severe case definition

3.1 Count the number of sites with loss of attachment greater than three and five millimeters

The code begins with a two-dimensional array and four one-dimensional arrays (the array definitions were described
in the first part of this paper). The array t_missing indicates which teeth are missing. The array al_s3 counts the
number of sites on each tooth with attachment loss greater than three millimeters. The array al_s5 counts the
number of sites on each tooth with attachment loss greater than five millimeters.

data test2;
set one;
array loa[6,28] loadisbl-loadisb28 loamesbl-loamesb28
     loames11-loames128 loadis11-loadis128
     loamdb1-loamdb28 loamd11-loamd128;
array s_missing[28] s_missing1-s_missing28;
array t_missing[28] t_missing1-t_missing28;
array al3_t[28] al3_t1-al3_t28;
array al5_t[28] al5_t1-al5_t28;
/*initialize counters for attachment loss greater than three or five
millimeters*/
al_s3=0; al_s5=0;

do j=1 to 28;
s_missing[j]=0;
al3_t[j]=0;
al5_t[j]=0;
do i=1 to 6;
if ((loa[i,j]=.) or (loa[i,j] in (-1,-2)) then
do;
s_missing[j]+=1;
end;
else if loa[i,j]>=0 and loa[i,j]<24 then
do;
s_missing[j]+=0;
if loa[i,j]=3 then
do;
al_s3+=1;
al3_t[j]+=1;
end;
if loa[i,j]=4 then
do;
al_s3+1;
al3_t[j]=1;
end;
if loa[i,j]>=5 then
do;
al_s3+1;
al_s5+1;
al3_t[j]=1; al5_t[j]=1;
end;
end;/*if all six sites are missing data, then tooth is counted as missing*/
if s_missing[j]=6 then t_missing[j]=1;
else t_missing[j]=0;
end;
end;

3.2 Count the number of teeth with loss of attachment greater than three and five millimeters
The variables al3_t_sum and al5_t_sum sum the total number of teeth with attachment loss greater than three millimeters and five millimeters, respectively. The variable t_count records the total number of teeth present, by deducting the number of missing teeth from the total of 28 teeth. The variable t_miss records the total number of missing teeth.

al3_t_sum=sum(of al3_t1-al3_t28);
al5_t_sum=sum(of al5_t1-al5_t28);
t_miss=sum(of t_missing1-t_missing28);
t_count=28-t_miss;

3.3 Exclude persons who have only two adjacent teeth with loss of attachment greater than three millimeters
The do loops at the end of the code indicate which survey participants have attachment loss greater than three millimeters only on two adjacent teeth. Teeth are considered adjacent within the upper and lower arches, but not across arches. For example, the upper teeth (teeth numbered 1 to 14 in the arrays) can be adjacent to one another and the lower teeth (teeth 15 to 28) can be adjacent to one another. For example the upper right 1st and 2nd molars (teeth 2 and 1) are adjacent to each other, and the lower right 1st and 2nd molars (teeth 27 and 28) are adjacent to each other, but the upper right and lower right 1st molars (teeth 2 and 27) that are in contact when the mouth is closed are not considered to be adjacent to each other (Figure 2). If a survey participant has only two teeth with attachment loss greater than three millimeters and these two teeth are adjacent to one another, then this person would not be considered to have periodontitis; otherwise this person would be considered to have periodontitis. (In code and estimates presented here, this person would be considered a non-case and would be included in the denominator of prevalence estimates.)

if al3_t_sum=2 then
do;
t2_adj=0;
do j=1 to 14;
if j=1 then
do;
if al3_t[1]=1 then
do;
if al3_t[2]=1 then t2_adj=1;
else t2_adj=0;
end;
end;
else if (j>1 and j<14) then
do;
if al3_t[j]=1 then
do;
if (al3_t[j+1]=1 or al3_t[j-1]=1) then t2_adj=1;
else t2_adj=0;
end;
end;
do j=15 to 28;
if j=15 then
do;
if al3_t[15]=1 then
do;
if al3_t[16]=1 then t2_adj=1;

```r
else j=j+1;
end;
end;
else if (j>15 and j<28) then
do;
if al3_t[j]=1 then
do;
if (al3_t[j+1]=1 or al3_t[j-1]=1) then t2_adj=1;
else t2_adj=0;
end;
end;
if al3_t_sum>2 then euro_perio=1;
else if al3_t_sum=2 then
do;
if t2_adj=0 then euro_perio=1;
else euro_perio=0;
end;
else if al3_t_sum<2 then euro_perio=0;
al3_per=0; al5_per=0;
t_al3_per=0; t_al5_per=0;

3.4 Create variable for severe case definition

This final part of the code calculates the percentage of teeth with attachment loss greater than five millimeters and creates indicator variable, euro_severe, for European severe periodontitis. If the survey participant was found to have 30 percent of teeth with attachment loss greater than 5 millimeters, then this person was considered to have severe periodontitis.

```
weighted estimates of prevalence were provided for analysts who may wish to compare their calculations to these estimates. The weighted estimates are representative of U.S. adults 30 years of age and older, and the estimated standard errors accounted for the clustered sample design.

A recent publication reiterated the importance of measuring attachment loss as well as pocket depth in clinical studies of periodontal disease.7 Although we developed this code for use with surveillance data, it may also be of interest for analysis of periodontal disease data from clinical studies.

The two commonly used case definitions for periodontal disease surveillance are different and result in different prevalence estimates. The arrays and loops efficiently calculate both definitions.

REFERENCES


ACKNOWLEDGMENTS

This work was supported by the Dental, Oral and Craniofacial Data Resource Center, a joint project of the Centers for Disease Control and Prevention’s (CDC’s) Division of Oral Health and the National Institute of Health’s (NIH’s) National Institute of Dental and Craniofacial Research.

CONTACT INFORMATION

Liang Wei
DB Consulting Group, Inc.
4470 Buford Hwy, NE, MS F-80
Atlanta, GA 30341
Work Phone: (770) 488-6070
E-mail: lwei1@cdc.gov
Web: drc.hhs.gov

Laurie Barker
Surveillance, Investigations, and Research Team
Division of Oral Health
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention
4770 Buford Hwy, MS F-80
Atlanta, GA 30341
Work Phone: (770) 488-5961
E-mail: LBarker@cdc.gov
Fax: 770-488-6080

Paul Eke
Division of Population Health
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention
4770 Buford Hwy, MS F-80
Atlanta, GA 30341
Work Phone: (770) 488-6092
E-mail: PEke@cdc.gov
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.