

Bad Date: How to find true love with Partial Dates!

Namrata Pokhrel, Accenture Life Sciences, Berwyn, PA

ABSTRACT

This poster will discuss the difficulties encountered while trying to impute for partial dates for a specific ISS/ISE study with particularly dirty data. This study contained multiple types of partial dates within a data set. Character dates were in a smorgasbord of DATE9, YYMMDD10, YYYYMMDD, YYMMDD8, formats. Some dates had dashes, some had UNK or UK or nothing at all to represent the missing piece of the date. There were also many partial dates that were simply invalid. The poster will discuss the lessons learned while trying to impute for these partial dates in DATE9 and YYYYMMDD formats in particular and will provide recommendations for future programmers.

INTRODUCTION

Dates are important parts of data collected within Clinical Trials. They play important role in data domains such as Visit, Adverse Events, Concomitant Medication just to mention a few. The validity of clinical trials among other factors depend upon the valid and complete date values. However, in reality, the clinical trial datasets almost always contain date values that are either missing, partial or invalid.

WHAT IS A PARTIAL DATE ?

A partial date is the incomplete or invalid date with one or two date components missing. The missing component may be the day or month or year or any combination of two of these. Example below lists partial dates that will be converted into valid SAS dates in DATE9 format.

	Partial/Invalid Date
1	08AUG2010
2	8AUG2010
3	--FEB2010
4	UNFeb2012
5	AUG2010
6	--AUG-2010
7	UN-AUG-2010
8	2010
9	08AUG
10	08AUG----
11	08UNK2010
12	08-UNK-2010
13	UN-UNK-2010
14	08-JUN-2010
15	ABCDEFGHI

Bad Date: How to find true love with Partial dates!

Example below lists partial dates that will be converted into valid SAS dates in YYYYMMDD format.

	Partial/Invalid Date
1	20120808
2	2012-08-08
3	2010-02-
4	2010-02-UN
5	201002
6	2010-02---
7	2010
8	2010-UN-UN
9	---0808
10	UNK-0808
11	UN-0808
12	0808
13	2010UN08
14	2010-UN-08
15	ABCDEFGHJI
16	ABCD

The author has developed a macro program that converts the partial date types listed above. The algorithm is user friendly and provides the user flexibility to update the macro as needed for dates in formats other than DATE9 and YYYYMMDD.

INPUT DATASETS DATE1 AND DATE2

- **DATE1 – EXAMPLE OF A DATASET WITH DATE9 FORMATTED CHARACTER DATES**

```
data date1;
length dt $20;
dt='08AUG2010'; output;
dt='8AUG2010'; output;
dt='--FEB2010'; output;
dt='UNFeb2012'; output;
dt='AUG2010'; output;
dt='---AUG-2010'; output;
dt='UN-AUG-2010'; output;
dt='2010'; output;
dt='08AUG'; output;
dt='08AUG----'; output;
dt='08UNK2010'; output;
dt='08-UNK-2010'; output;
dt='UN-UNK-2010'; output;
dt='08-JUN-2010'; output;
dt='ABCDEFGHJI'; output;
run;
```

- **DATE2 – EXAMPLE OF A DATASET WITH YYYYMMDD FORMATTED CHARACTER DATES**

```
data date2;
length dt $20;
dt='20120808'; output;
dt='2012-08-08'; output;
dt='2010-02-'; output;
```

Bad Date: How to find true love with Partial dates!

```
dt='2010-02-UN'; output;  
dt='201002'; output;  
dt='2010-02---'; output;  
dt='2010'; output;  
dt='2010-UN-UN'; output;  
dt='----0808'; output;  
dt='UNK-0808'; output;  
dt='UN-0808'; output;  
dt='0808'; output;  
dt='2010UN08'; output;  
dt='2010-UN-08'; output;  
dt='ABCDEFGHI'; output;  
dt='ABCD'; output;  
run;
```

THE MACRO PROGRAM

MACRO 'PARTIAL' AND ITS PARAMETERS

```
%macro partial(din=, date_format=, char_date_var=, replacement_month=,  
replacement_day= );
```

EXPLANATION OF MACRO PARAMETERS

- DIN = Represents required Input dataset . For example: Date1 and Date2
- DATE_FORMAT = Character date has to be either in one of three formats: DATE9, YYYYMMDD (with or without dashes).
If this is not the case, the input dataset will need to be pre-processed in order for the macro to work properly. This is a required parameter.
- CHAR_DATE_VAR = Character date variable with partial dates that will need imputation.
- REPLACEMENT_MONTH = If month is missing, then which month should be imputed?
If this derivation is more complicated than a simple substitution, it is not within the scope of this macro. It is required that if parameter DATE_FORMAT is DATE9 then value must be one of JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC. It is required that if parameter DATE_FORMAT is YYYYMMDD, then value must be one of 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12.
- REPLACEMENT_DAY = If day is missing, then which day should be imputed?
If this derivation is more complicated than a simple substitution, it is not within the scope of this macro. It is required that the value must be either FIRST, LAST, or 15 where
FIRST = first day of month, LAST = last day of month, 15 = 15th of the month.

```
options missing='';
```

```
*****UPCASE MACRO VARIABLES BELOW IN ORDER TO STANDARDIZE USER PROVIDED  
VALUE *****;
```

```
%let date_format = %upcase(&date_format);  
%let replacement_month = %upcase(&replacement_month);  
%let replacement_day = %upcase(&replacement_day);
```

Bad Date: How to find true love with Partial dates!

```
data &dout;
  set &din;

***** REMOVE LEADING AND TRAILING BLANKS FROM MACRO VARIABLE CHAR_DATE_VAR AND
UPCASE ITS VALUES *****;

  tempdt=upcase(strip(&char_date_var));

***** CONDITIONAL LOGIC STATEMENTS USED FOR DATE9 FORMATTED CHARACTER DATES
*****;

  %if &date_format=DATE9 %then
  %do;
    *****REMOVES DASHES*****;

    tempdt=compress(tempdt,"-");

*****IF FIRST TWO LETTERS ARE UN, THEN REMOVE AND KEEP ONLY THE REST OF THE
COMPONENTS*****;

    if substr(tempdt,1,2)='UN' then tempdt=substr(tempdt,3);
    lngth=length(tempdt);

*****CREATE VARIABLE ALPHA TO INDICATE FIRST ENCOUNTER OF ALPHABETIC VALUE, IF
ANY *****;

    alpha=indexc(tempdt,'ABCDEFGHIJKLMNOPQRSTUVWXYZ');

    if lngth=9 and index(tempdt,'UNK') eq 0 then idtc=tempdt;

***** IMPUTATION OF INVALID MONTH COMPONENT 'UNK' IF THE LENGTH OF TEMPDT=9
*****;

    else if lngth=9 and index(tempdt,'UNK') ne 0 then
    do;
      mis_mnth='yes';
      idtc = tranwrd(tempdt,'UNK',"&replacement_month");
    end;

***** CONCATENATE '0' TO TEMPDT TO OBTAIN MISSING DAY COMPONENT IF THE LENGTH OF
TEMPDT=8 *****;

    else if lngth=8 and alpha=2 then idtc='0' || tempdt;

***** IMPUTATION OF INVALID MONTH 'UNK' AND MISSING DAY COMPONENT IF THE LENGTH
OF TEMPDT=7 *****;

    else if lngth=7 and alpha=1 then
    do;
      if index(tempdt,'UNK') gt 0 then
      do;
        mis_mnth='yes';
        tempdt = tranwrd(tempdt,'UNK',"&replacement_month");
      end;

      %if &replacement_day = FIRST %then
      %do;
        mis_day='yes';
        idtc='01' || tempdt;
      %end;
      %else %if &replacement_day = 15 %then
      %do;
```

Bad Date: How to find true love with Partial dates!

```
        mis_day='yes';
        idtc='15' || tempdt;
    %end;
    %else %if &replacement_day = LAST %then
    %do;
        mis_day='yes';
        tempdt2=input(('15' || tempdt),date9.);

    ***** USE INTNX FUNCTION TO OBTAIN FIRST DAY OF NEXT MONTH AND SUBTRACT ONE DAY
    TO OBTAIN LAST DAY OF THE MONTH*****;

        idtc=put(intnx('month',tempdt2,1)-1,date9.);
    %end;
end;

    ***** IMPUTATION OF MISSING MONTH AND DAY COMPONENTS IF THE LENGTH OF TEMPDT=4
    AND ALPHA=0*****;

    else if lngth=4 and alpha=0 then
    do;
        mis_mnth='yes';
        tempdt3=compress("&replacement_month" || tempdt);
        %if &replacement_day = FIRST %then
        %do;
            mis_day='yes';
            idtc='01' || tempdt3;
        %end;
        %else %if &replacement_day = 15 %then
        %do;
            mis_day='yes';
            idtc='15' || tempdt3;
        %end;
        %else %if &replacement_day = LAST %then
        %do;
            mis_day='yes';
            tempdt4=input(('15' || tempdt3),date9.);

    ***** USE INTNX FUNCTION TO OBTAIN FIRST DAY OF NEXT MONTH AND SUBTRACT ONE DAY
    TO OBTAIN LAST DAY OF THE MONTH*****;

            idtc=put(intnx('month',tempdt4,1)-1,date9.);
        %end;
    end;

    ***** DATE COULD NOT BE IMPUTED IF THE YEAR IS MISSING*****;

    else if lngth=5 and alpha=3 then
    do;
        put "ALERT: Date could not be imputed because year is missing "
        (&char_date_var) (=);
        mis_year='yes';
    end;
    idtn=input(idtc,??date9.);

    ***** MACRO COULD NOT IMPUTE DATE - INVESTIGATE*****;

    if idtn=. then put "ALERT: Macro could not impute date, please investigate. "
    (&char_date_var) (=);

    format idtn date9.;

    %end;
```

Bad Date: How to find true love with Partial dates!

```
***** CONDITIONAL LOGIC STATEMENTS USED FOR YYYYMMDD FORMATTED CHARACTER DATES
*****;
```

```
  %if &date_format=YYYYMMDD %then
  %do;
```

```
***** DATE COULD NOT BE IMPUTED IF THE YEAR IS MISSING or INVALID*****;
```

```
  if substr(tempdt,1,3)='UNK' or substr(tempdt,1,2)='UN' or
  substr(tempdt,1,2)='--' then
  do;
    put "ALERT: Date could not be imputed because year is missing "
    (&char_date_var) (=);
    mis_year='yes';
  end;
```

```
  else
  do;
```

```
    ***** REMOVES DASHES*****;
```

```
    tempdt=compress(tempdt,"-");
    lngth=length(tempdt);
```

```
*****CREATE VARIABLE ALPHA TO INDICATE FIRST ENCOUNTER OF ALPHABETIC VALUE, IF
ANY *****;
```

```
    alpha=indexc(tempdt,'ABCDEFGHIJKLMNOPQRSTUVWXYZ');
```

```
***** IMPUTATION OF INVALID DAY AND MONTH COMPONENTS *****;
```

```
  if substr(tempdt,5,4)='UNUN' then
  do;
    mis_mnth='yes';
    mis_day='yes';
    tempdt2=substr(tempdt,1,4) || "&replacement_month";
```

```
    %if &replacement_day = FIRST %then
    %do;
```

```
      idtc=strip(tempdt2) || '01';
```

```
    %end;
```

```
    %else %if &replacement_day = 15 %then
```

```
    %do;
```

```
      idtc=strip(tempdt2) || '15';
```

```
    %end;
```

```
    %else %if &replacement_day = LAST %then
```

```
    %do;
```

```
      tempdt3=input((strip(tempdt2) || '15'),yymmdd10.);
```

```
***** USE INTNX FUNCTION TO OBTAIN FIRST DAY OF NEXT MONTH AND SUBTRACT ONE DAY
TO OBTAIN LAST DAY OF THE MONTH*****;
```

```
      idtc=put(intnx('month',tempdt3,1)-1,yymmdd10.);
```

```
    %end;
```

```
  end;
```

Bad Date: How to find true love with Partial dates!

```
***** IMPUTATION OF INVALID DAY COMPONENT*****;

else if substr(tempdt,7,2)='UN' then
do;
  mis_day='yes';
  tempdt2=substr(tempdt,1,6);

  %if &replacement_day = FIRST %then
  %do;
    idtc=strip(tempdt2) || '01';
  %end;
  %else %if &replacement_day = 15 %then
  %do;
    idtc=strip(tempdt2) || '15';
  %end;
  %else %if &replacement_day = LAST %then
  %do;
    tempdt3=input((strip(tempdt2) || '15'),yymmdd10.);

***** USE INTNX FUNCTION TO OBTAIN FIRST DAY OF NEXT MONTH AND SUBTRACT ONE DAY
TO OBTAIN LAST DAY OF THE MONTH*****;

    idtc=put(intnx('month',tempdt3,1)-1,yymmdd10.);
  %end;
end;

else if lngth=8 then idtc = put(input(tempdt,??yymmdd10.),yymmdd10.);

***** IMPUTATION OF MISSING DAY COMPONENT *****;

else if lngth=6 then
do;
  mis_day = 'yes';
  %if &replacement_day = FIRST %then
  %do;
    idtc=strip(tempdt) || '01';
  %end;
  %else %if &replacement_day = 15 %then
  %do;
    idtc=strip(tempdt) || '15';
  %end;
  %else %if &replacement_day = LAST %then
  %do;
    tempdtn2=input((strip(tempdt) || '15'),yymmdd10.);

***** USE INTNX FUNCTION TO OBTAIN FIRST DAY OF NEXT MONTH AND SUBTRACT ONE DAY
TO OBTAIN LAST DAY OF THE MONTH*****;

    idtc=put(intnx('month',tempdtn2,1)-1,yymmdd10.);
  %end;
end;

***** IMPUTATION OF MISSING DAY AND MONTH COMPONENTS *****;

else if lngth=4 then
do;
  temp=input(tempdt,?? best.);
```

Bad Date: How to find true love with Partial dates!

```
if 1900 <= temp <= 2020 then
do;
tempdt2=substr(tempdt,1,4) || "&replacement_month";
mis_mnth='yes';
mis_day='yes';
%if &replacement_day = FIRST %then
%do;
idtc=strip(tempdt2) || '01';
%end;
%else %if &replacement_day = 15 %then
%do;
idtc=strip(tempdt2) || '15';
%end;
%else %if &replacement_day = LAST %then
%do;
tempdt3=input((strip(tempdt2) || '15'),yymmdd10.);

***** USE INTNX FUNCTION TO OBTAIN FIRST DAY OF NEXT MONTH AND SUBTRACT ONE DAY
TO OBTAIN LAST DAY OF THE MONTH*****;

idtc=put(intnx('month',tempdt3,1)-1,yymmdd10.);
%end;
end;

***** IMPUTATION NOT POSSIBLE DUE TO NONSENSICAL YEAR- PUT ALERT IN LOG
*****;

else put "ALERT: Date could not be imputed because year is nonsensical or
is not between 1900 and 2020 " (&char_date_var) (=);
end;

***** IMPUTATION OF INVALID MONTH COMPONENT IF IDTC= ' ' AND LENGTH OF TEMPDT=8
*****;

if idtc=' ' and lngth=8 and substr(tempdt,5,2)='UN' then
do;
idtc = tranwrd(tempdt,'UN',"&replacement_month");
mis_mnth='yes';
end;
end;

idtc = put(input(idtc,yymmdd10.),yymmdd10.);
idtn = input(idtc,?? yymmdd10.);

***** MACRO NOT ABLE TO IMPUTE DATE IF IDTN =. *****;

if idtn=. then put "ALERT: Macro could not impute date, please investigate. "
(&char_date_var) (=);

format idtn yymmdd10.;

%end;

run;

%mend partial;
```

Bad Date: How to find true love with Partial dates!

```

***** MACRO CALL EXAMPLES*****;

%partial(din=date1, dout=DT_NINE, date_format=date9, char_date_var=dt,
replacement_month=jan, replacement_day = last);

%partial(din=date2, dout=DT_TEN, date_format=yyyymmdd, char_date_var=dt,
replacement_month=01, replacement_day = last);

***** CREATE VARIABLE LABELS AND DROP UNNECESSARY VARIABLES FROM FINAL DATASET
DT_NINE *****;

data DT_NINE;
    set DT_NINE ;
    label
        dt='Partial Date'
        tempdt='Compressed Partial Date'
        lngth='Length of Compressed Input Date'
        alpha='Position of Alphabets'
        mis_mnth='Missing Month'
        mis_day='Missing Day'
        mis_year='Missing Year'
        idtc='Imputed Character Date'
        idtn='Imputed Numeric Date'
        ;
    drop tempdt4;
run;

***** DT_NINE DATASET WITH IMPUTED DATE VARIABLE VALUES *****;

```

	Partial Date	Compressed Partial Date	Length of Compressed Input Date	Position of Alphabets	Missing Day	Missing Month	Missing Year	Imputed Character Date	Imputed Numeric Date
1	08AUG2010	08AUG2010	9	3				08AUG2010	08AUG2010
2	8AUG2010	8AUG2010	8	2				08AUG2010	08AUG2010
3	--FEB2010	FEB2010	7	1	yes			28FEB2010	28FEB2010
4	UNFeb2012	FEB2012	7	1	yes			29FEB2012	29FEB2012
5	AUG2010	AUG2010	7	1	yes			31AUG2010	31AUG2010
6	---AUG-2010	AUG2010	7	1	yes			31AUG2010	31AUG2010
7	UN-AUG-2010	AUG2010	7	1	yes			31AUG2010	31AUG2010
8	2010	2010	4	0	yes	yes		31JAN2010	31JAN2010
9	08AUG	08AUG	5	3			yes		
10	08AUG----	08AUG	5	3			yes		
11	08UNK2010	08UNK2010	9	3		yes		08JAN2010	08JAN2010
12	08-UNK-2010	08UNK2010	9	3		yes		08JAN2010	08JAN2010
13	UN-UNK-2010	JAN2010	7	1	yes	yes		31JAN2010	31JAN2010
14	08-JUN-2010	08JUN2010	9	3				08JUN2010	08JUN2010
15	ABCDEFGHI	ABCDEFGHI	9	1				ABCDEFGHI	

Bad Date: How to find true love with Partial dates!

```
***** CREATE VARIABLE LABELS AND DROP UNNECESSARY VARIABLES FROM FINAL DATASETS
DT_TEN *****;
```

```
data DT_TEN;
  set DT_TEN ;
  label
    dt='Partial Date'
    tempdt='Compressed Partial Date'
    lngth='Length of Compressed Input Date'
    alpha='Position of Alphabets'
    mis_mnth='Missing Month'
    mis_day='Missing Day'
    mis_year='Missing Year'
    idtc='Imputed Character Date'
    idtn='Imputed Numeric Date'
    ;
  drop temp tempdtn2;
run;
```

```
***** DT_TEN DATASET WITH IMPUTED DATE VARIABLE VALUES *****;
```

	Partial Date	Compressed Partial Date	Length of Compressed Input Date	Position of Alphabets	Missing Day	Missing Month	Missing Year	Imputed Character Date	Imputed Numeric Date
1	20120808	20120808	8	0				2012-08-08	2012-08-08
2	2012-08-08	20120808	8	0				2012-08-08	2012-08-08
3	2010-02-	201002	6	0	yes			2010-02-28	2010-02-28
4	2010-02-UN	201002UN	8	7	yes			2010-02-28	2010-02-28
5	201002	201002	6	0	yes			2010-02-28	2010-02-28
6	2010-02---	201002	6	0	yes			2010-02-28	2010-02-28
7	2010	2010	4	0	yes	yes		2010-01-31	2010-01-31
8	2010-UN-UN	2010UNUN	8	5	yes	yes		2010-01-31	2010-01-31
9	----0808	----0808					yes		
10	UNK-0808	UNK-0808					yes		
11	UN-0808	UN-0808					yes		
12	0808	0808	4	0					
13	2010UN08	2010UN08	8	5		yes		2010-01-08	2010-01-08
14	2010-UN-08	2010UN08	8	5		yes		2010-01-08	2010-01-08
15	ABCDEFGH	ABCDEFGH	9	1					
16	ABCD	ABCD	4	1					

Limitation-This macro does not impute the partial dates that have

- A missing year component.
- Invalid day, month and year components.

Conclusion – This poster summarizes the usage of Macro to impute Partial dates in DATE9 and YYYYMMDD formats.

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.

Bad Date: How to find true love with Partial dates!

Author Name: Namrata Pokhrel
Company: Accenture
Address: 1160 W. Swedesford Road, Berwyn, PA
Work Phone: (610) 407-7587
Email: namrata.pokhrel@accenture.com
Web: www.accenture.com

Confirmation Number: 165