## Exercise 6: Dealing with incomplete dates

At the end of this exercise you should be able to:

  a. Approximate dates when day and/or month is missing

  b. Create a date from three component fields (day, month, year)

  c. Make use of temporary variables in the check file to make calculations.

Date information is commonly collected, most commonly perhaps for the calculation of intervals or another purpose. However, information on the date is often incomplete. For instance, if a date of symptom onset is asked from a patient, the patient may remember only the year or only the year and the month. The date of onset is thus unknown, and an interval between the current doctor's visit (which might be known exactly) and date of symptom onset cannot be calculated and must remain unknown. Nevertheless, if for instance year and month of symptom onset are known (but not the day) and the date of the actual visit is known exactly, something is known, and an approximate interval could be provided. This exercise will offer a way to make such approximations.

As we should not leave it up to the data entry person to enter incorrect, approximated dates, with possibly arbitrary and changing rules, we must avoid requiring to enter a date, but rather split it into its components year, month, and day, and then reassemble it.

Let us assume that we wish to calculate the variables:

```
regexct       Exact registration date <dd/mm/yyyy>
regappr Approximate registration date <dd/mm/yyyy>
```

from the fields:

```
regdd           Day of registration ##
regmm         Month of registration ##
regyy          Year of registration ####
```

The field REGEXCT will be an existing date within the legal range currently defined if all three components of the date are known else its value will be set to 01/01/1800. The field REGAPPR will be equal to REGEXCT if the latter is within the legal range. If the day only is unknown, the day value for REGAPPR will be 15 (mid-month), if both month and day are unknown, the month will be 07 and the day 01 (mid-year). If all three components are unknown, the value of REGAPPR will be set to 01/01/1800.

We will impose a hierarchical structure, i.e., once the month is unknown, it does not matter even if you know the day – even if entered as such – it does not make sense, and if the year is unknown, neither a known month nor known day will affect the calculation.

Thus, we will replace the current field REGDATE with the above five fields, three of which (REGDD, REGMM, and REGYY) are entered (MUSTENTER fields) while the other two (REGEXCT and REGAPPR) are calculated from the former three (and will thus be NOENTER Fields).

In addition, we will add another variable that measures the quality of the calculated date information, and that we might name:

```
regqual    Quality of registration date #
```

This field can take the values 0, 1, 2, and 3, where 0 (zero) indicates that none is known, and 3 that all three date components are known.

For this task, we use temporary variables that we have briefly introduced in an earlier exercise. Although for this specific example, it would be possible to solve it without such variables, they make things easier and more transparent. Temporary variables can have a field length of 16 and they do not appear in the QES file and will not become part of the dataset (see our earlier exercise). They will be used in the Check file for internal calculations in a BEFORE FILE block as in:

```
BEFORE FILE
  DEFINE regddTemp ##
  DEFINE regmmTemp ##
  DEFINE regyyTemp ####
END
```

BEFORE FILE means exactly what it says: Before anything is entered into the file, these variables are created: we have used this in an earlier exercise. There is also a CHK file command BEFORE RECORD (look up in the Help file what these commands do).

To create a date from the three component fields, the CHK file command is:

```
fulldate=DATE(varday,varmonth,varyear)
```

*Tasks:*

o   *Create an* A_EX06.* *triplet (using the* A_EX05.* *files as the starting point). The questionnaire should display all the calculated variables.*

o   *Try to preserve the data you have (in* A_EX05.REC*), accepting that you lose the information in the field REGDATE, which is easier to update than recreating the entire file.*

o   *Edit the* A_EX06.CHK *file to make the calculations. Note that you will need to define temporary variables for this task.*

o   *Update the data file to check the functionality*