Solution to Exercise 1: A data documentation sheet for a simple questionnaire

Key Point(s):

• It is good practice to write a data documentation sheet before you make your actual data entry form with EpiData Manager.

• You should always define a value if no answer was provided to a question.

• "Date" is a reserved name in EpiData and cannot be used as a field name.

Task:

o *Complete the data documentation sheet for all fields in the questionnaire. Note that you should always define a value if no answer was provided to a question.*

o *Think of the most efficient ways to code reason for examination and results of microscopic examination.*

Solution:

There are many different solutions, but for the sake of uniformity, we will be using the following (but later revise some components of it) as shown on the next page.

Field name	Question [Field label]	Field type	Field length	Field values	Value labels	Notes [Comment]
serno	Laboratory serial number	I	4	1-9000 9001, 9002,		*[Serial number starting with 1 each year]* Enter 9001, 9002,… if serial number is *not unique* or *missing*, and write a data entry note
regdate	Registration date	D	10	01/01/2000-31/12/2005		
sex	Examinee's sex	I	1	1 2 9	Female sex Male sex Sex not recorded	
age	Examinee's age in years	I	3	0-125, 999		*[Range of legal age years]* Enter 999 if age not recorded
reason	Examination reason	I	1	0 1 2 3 4 5 6 7 8 9	Diagnosis Follow-up at 1 month Follow-up at 2 months Follow-up at 3 months Follow-up at 4 months Follow-up at 5 months Follow-up at 6 months Follow-up at 7 months or later Follow-up, month not stated Reason not recorded	
res1	Result of specimen 1	I	1	0 1 2 3 4 5 6 9	Negative 1+ positive 2+ positive 3+ positive Positive, not quantified Scanty, not quantified Scanty, quantified Result not recorded	*[If the entered value is other than 6, then bypass next field and bypass it]*
res1sc	Result of specimen 1 scanty	I	1	1 2 3 4 5 6 7 8 9	1 AFB per 100 OIF 2 AFB per 100 OIF 3 AFB per 100 OIF 4 AFB per 100 OIF 5 AFB per 100 OIF 6 AFB per 100 OIF 7 AFB per 100 OIF 8 AFB per 100 OIF 9 AFB per 100 OIF	
res2	Result of specimen 2	I	1	0 1	Negative 1+ positive	*[If the entered value is other than 6, then bypass next field and bypass it]*

				2	2+ positive	
				3	3+ positive	
				4	Positive, not quantified	
				5	Scanty, not quantified	
				6	Scanty, quantified	
				9	Result not recorded	
res2sc	Result of specimen 2 scanty	I	1	1	1 AFB per 100 OIF	
				2	2 AFB per 100 OIF	
				3	3 AFB per 100 OIF	
				4	4 AFB per 100 OIF	
				5	5 AFB per 100 OIF	
				6	6 AFB per 100 OIF	
				7	7 AFB per 100 OIF	
				8	8 AFB per 100 OIF	
				9	9 AFB per 100 OIF	
res3	Result of specimen 3	I	1	0	Negative	[*If the entered value is other than 6, then bypass next field and bypass it*]
				1	1+ positive	
				2	2+ positive	
				3	3+ positive	
				4	Positive, not quantified	
				5	Scanty, not quantified	
				6	Scanty, quantified	
				9	Result not recorded	
res3sc	Result of specimen 3 scanty	I	1	1	1 AFB per 100 OIF	
				2	2 AFB per 100 OIF	
				3	3 AFB per 100 OIF	
				4	4 AFB per 100 OIF	
				5	5 AFB per 100 OIF	
				6	6 AFB per 100 OIF	
				7	7 AFB per 100 OIF	
				8	8 AFB per 100 OIF	
				9	9 AFB per 100 OIF	

Note the following here. For an unknown laboratory date (REGDATE), we must enter a legally existing (valid) date and we chose a legal but practically impossible date in the past, i.e. 01/01/1800. EpiData will not accept a date 99/99/9999 nor for that matter 29/02/2001. It is a personal preference of us to usually use 9 or 99.9 or the like to define unknown values, be this for text or numeric variables. We also introduced a "legal range" for some variables like REGDATE and AGE. We did this a bit arbitrarily, but still tried to keep it within what might be expected.

We made two fields for each result. There are 17 possibilities for a result, and therefore a length of 2 is the minimum required, but even with that, the values might not be intuitive, but they should be. An alternative version uses a float of length three to get for instance:

```
0.0     Negative
1.0     1+ Positive
2.0     2+ positive
…
0.1     Scanty, 1 AFB per 100 OIF
0.2     Scanty, 2 AFB per 100 AFB
```

Scanty results are relatively rare among positives (perhaps some 10% among diagnostic and some 20% among follow-up examinations in quality-assured high-burden country laboratories), and positives themselves are relatively rare among all (perhaps 10% to 20% among patients coming for diagnostic evaluation). Thus, scanty positive results might be only 1% of all results. We therefore chose to use integer variables and bypass the field scanty, unless the first field defines the result as quantified scanty.