

Paper PO-02

A Visual Approach to Monitoring Case Report Form Submission During Clinical Trials

Rebecca Horney, VA Cooperative Studies Program Coordinating Center, Perry Point, MD

Karen Jones, VA Cooperative Studies Program Coordinating Center, Perry Point, MD

Annette Wiseman, VA Cooperative Studies Program Coordinating Center, Perry Point, MD

ABSTRACT

Clinical Trials data management requires close monitoring of Case Report Form (CRF) submission, particularly for trials involving a complex mixture of hundreds or thousands of subjects, multiple recruitment sites and many different rating periods. Monitoring must be performed on a routine, almost “real-time”, basis to assure timeliness and accuracy of data submission as well as the overall integrity and validity of the trial.

Using common SAS® code and procedures, we have developed a tabular method of presenting Clinical Trial data submission on a subject-by-subject basis over the course of the trial. This report, run in tandem with the usual monitoring and progress reports, allows us to quickly scan and visually detect several common types of errors and inconsistencies, including:

- Adherence to the visit schedule or alternately, the emergence of what we’ve termed “visit creep”
- Patterns of missed visits
- Erroneous dates
- Incorrect subject numbers or identifiers
- Incorrect visit numbers
- Duplicate CRF data received for multiple visits (potential fraud detection)

Aside from its usefulness to the data coordinating center, this report can be used for monitoring and review by other stakeholders, such as the study sponsor or chairperson; scientific and data review bodies; management groups and on-site clinical monitors. Though this report was created in a Clinical Trials context, we believe its applicability could extend to other fields and disciplines such as epidemiology, banking and educational settings.

INTRODUCTION

As data management personnel at a clinical trials data coordinating center, we find it essential to track the completion and submission of case report forms (CRFs) on as close to a “real-time” basis as possible. This close and timely monitoring helps us to assure the accuracy of the data submitted to the study database as well as the integrity of the trial overall. As a result of this need, we wrote a SAS program that uses a trial participant’s dates of enrollment (randomization) into and termination from the trial to determine a unique schedule of forms expected and due dates for each participant. After merging actual participant data with the schedule of forms expected and due dates, a report is produced which visually illustrates how well each participant is adhering to the schedule of visits and highlights possible errors in data completion. We describe the individual steps of the SAS program below which end ultimately with the production of a final report.

STEP 1. DETERMINE SCHEDULE OF CASE REPORT FORMS DUE

A template (data set) is created, which consists of regularly scheduled clinic visits (observations/rows) and CRFs due (variables/columns). Each time a CRF is expected at a given visit, the code “M” is placed in that position in the template as a placeholder; otherwise, a “.” in a position indicates that the CRF is not due at that particular visit. Due to space constraints in the final report, the template currently only includes those CRFs that are expected to be received throughout the course of the study. Those CRFs completed on an “as needed” basis could easily be added to the template.

```
data forms_expected; *M= expected, .=not expected;
input
week F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14 F15 f16 f17 f18 f19 f20 f21 f22 f23 f24 f25 f26 f27 f28 f29 f30 f31 f36;
CARDS;
0 . M M M M M M M M M M . M M M M M M M M M M M M M M M M M M .
1 . . . . . . . . . . M . . . . . . . . . M M M . . . . M .
2 . . . . . . . . . . M . . . . . . . . . M M M . . . . M .
4 . . . . M . . . . . M . . . . . . . . . M M M . . . . M .
6 . . . . . . . . . . M M M M M M M M M M M M M M M M M .
8 . . . . M . . . . . M . . . . . . . . . M M . . . . M .
12 . . . M M . . . . . M M M M M M M M M M M M M M M M M .
16 . . . M . . . . . . M . . . . . . . . . M M M . . . . M .
20 . . . M . . . . . . M . . . . . . . . . M M M . . . . M .
24 . . M M M . . . . . M M M M M M M M M M M M M M M M M . M
;
```

STEP 2. PARTICIPANT STATUS

The status of each participant is determined in a separate process, either manually or via another SAS program (not shown here). The status (i.e. a participant completed the trial, left the trial earlier than expected or is currently active in the trial) of a participant is used to determine which of the clinic visits will be displayed in the report for that participant.

STEP 3. CALCULATE THE EXPECTED VISIT DATES FOR EACH PARTICIPANT IN THE TRIAL BASED ON THE DATE OF ENROLLMENT (OR RANDOMIZATION)

A data set is created which contains the expected clinic visit dates for each participant based on the participant's date of enrollment (or randomization) into the trial. The expected clinic visit dates are calculated by adding the appropriate number of days (weeks, etc. as the case may be) to the participant's date of randomization. In the example below, clinic visits occurred at various weeks throughout the course of the study; the due dates for the visits were calculated by adding various multiples of 7 (corresponding to the weeks when visits were scheduled) to a participant's date of randomization. Also included in this dataset are the variables gender, term_date (participant's end date of the trial) and reason_term (participant's reason for ending the trial), which are used later on in the program to determine whether or not certain visits or CRFs are expected for a given participant.

```
data randa;
merge DataIn.random(keep=center patient ref_code randate in=inr)
      DataIn.form01(keep=center patient gender)
      DataIn.form36(keep=center patient term_date reason_term);
by center patient;
if inr;
week=0; due=randate; output randa;

do week=1,2,4,6,8,12,16,20,24;
  due=randate+(week*7);
  output randa;
end;
format due mmddyy10.;
run;
```

Example: Center 1 Patient 1 was randomized on Jan 1, 2011.

Center	Patient	Week	Due
1 1	0	1/1/11	date of randomization
1 1	1	1/8/11	
1 1	2	1/15/11	
1 1	4	1/29/11	
1 1	6	2/12/11	
1 1	8	2/26/11	
1 1	12	3/26/11	
1 1	16	4/23/11	
1 1	20	5/21/11	
1 1	24	6/18/11	

STEP 4. MERGE TEMPLATE OF FORMS EXPECTED WITH THE DATA SET CONTAINING DUE DATES

The data set containing the actual expected visit dates for each participant created in Step 3 is now merged with the file that includes participant status (completer, early terminator or currently active in trial) described in Step 2. The resulting data set is in turn merged with the template of forms expected (from Step 1) to create a unique template for each participant which includes the participant's status and due dates for each visit.

```
data rand;
  merge randa (in=inr)
        status(in=in1);
  by center patient;
  if inr and in1;
  run;

data both;
  merge rand
        forms_expected;
  by week;
  run;
```

STEP 5. READ THE CASE REPORT FORM FILES

The CRF files are now read into SAS; these data sets are comprised of actual data received from participants and include the following variables: center number, participant number, visit (week) number and actual visit (assessment) date. In the example below, some CRFs are collected only at baseline, prior to randomization. For these CRFs, week is set to 0 in the data set; for all other CRFs, week is read from the CRF file itself.

```
data CRF1;
  set DataIn.form01(keep=center patient assess date);
  week=0;
  rename assess_date=d1;
  run;

data CRF2;
  set DataIn.form02(keep=center patient assess_date);
  week=0;
  rename assess_date=d2;
  run;

data CRF3;
  set DataIn.form03(keep=center patient assess_date week);
  rename assess_date=d3;
  run;

data CRF4;
  set DataIn.form04(keep=center patient assess_date week);
  rename assess_date=d4;
  run;

data CRF5;
  set DataIn.form05(keep=center patient assess_date week);
  rename assess_date=d5;
  run;

.
.
.
```

STEP 6. MERGE THE CASE REPORT FORM FILES WITH VISIT DUE DATES AND FORMS EXPECTED

The unique template for each participant containing due dates (from Step 4) is merged with CRF data sets (from Step 5) by center, participant and week. For each CRF received, the actual visit dates (variables d1-d31 from Step 5) are used to calculate the number of days before or after the expected visit dates (variables f1-f31 in example below). If the CRF is expected but missing, the 'M' read in (Step 1) is kept. If the form has been received, the 'M' is replaced by the number of days.

```

data final;
  merge both(in=inb)   CRF1(in=in1)   CRF2(in=in2)   CRF3(in=in3)   CRF4(in=in4)
        CRF5(in=in5)   CRF6(in=in6)   CRF7(in=in7)   CRF8(in=in8)   CRF9(in=in9)
        CRF10(in=in10) CRF11(in=in11) CRF12(in=in12) CRF13(in=in13) CRF14(in=in14)
        CRF15(in=in15) CRF16(in=in16) CRF17(in=in17) CRF18(in=in18) CRF19(in=in19)
        CRF20(in=in20) CRF21(in=in21) CRF22(in=in22) CRF23(in=in23) CRF24(in=in24)
        CRF25(in=in25) CRF26(in=in26) CRF27(in=in27) CRF28(in=in28) CRF29(in=in29)
        CRF30(in=in30) CRF31(in=in31) CRF36(in=in36);
  by center patient week;
  if inb;

  array f {31} f1-f31;          *M or . from Step 1, calculate days;
  array d {31} d1-d31;          *Visit date from CRF (Step 5);
  if gender=1 then f5=.;        *CRF 5 is not expected for Males;

  do i=1 to 31;
    *Calculate the number of days before or after the due date;
    if d{i} gt 0 then f{i} = d{i} - due;

    *Participant was not interviewed for this visit. Only form 31 is expected;
    if ptinterview=2 and f{i}=.M then f{i}=.;

    *CRFs due within 7 days of the report date are within the window (W);
    If due gt &duedate -7 and f{i}=.M then f{i}=.W;

    *CRFs after date of termination are not expected;
    if due gt term_date and term_date ne . and f{i}=.M and reason_term ne 1
      then f{i}=.T;
  end;

```

STEP 7. PRODUCE THE REPORT

To produce the report for each participant in the trial, the number of days before or after the due date that a CRF is actually collected (calculated in Step 6) is what is displayed in the report for each visit/CRF combination (see output examples at end of paper). Only observations from the final data set created in Step 6 that correspond to “good” visits (those visits due on or before the day the report is generated or those visits for which a CRF has been received) are kept. Any observations that represent visits due after the day the report is generated and for which no forms have been received are not shown in the report.

As is usually the case in clinical trials, visit windows are established before a trial begins to indicate the number of days before or after a visit due date that a CRF can be completed and still be legitimately considered belonging to that visit. For the example trial being used in this paper, the following visit windows were established before the trial began:

Baseline visit – CRFs completed up to 30 days prior to the date of randomization are considered within the window.

Visits 1, 2 & 4 – CRFs completed up to 2 days prior to or 2 days after the due date are considered within the window.

Visits 6, 8, 12, 16, 20 & 24 – CRFs completed up to 5 days prior to or 5 days after the due date are considered within the window.

Alpha codes are used to represent missing CRFs in the report and formats (colors) are used to indicate the degree to which a CRF is out of the submission window. These are described below:

M = CRF is missing and is past the allowable submission window

W = CRF is expected but not yet received; however, it is still within the allowable submission window

```

proc format;
  *the format base is used for the baseline visit where the window is -30 days;
  value base
    low - -31    ='Red'
    1    - high  ='Red'
    other        ='White';
  *the format two is used for visits 1, 2 & 4 where the window is ± 2 days;
  value two
    low - -5     ='Red'
    -4 - -3     ='Yellow'

```

```

3 - 4 = 'Yellow'
5 - high = 'Red'
.M = 'LightBlue'
other = 'White';

*the format seven is used for visits 6, 8, 12, 16, 20 & 24 where the window is ± 7
days;
value seven
  low - -15 = 'Red'
-14 - -8 = 'Yellow'
  8 - 14 = 'Yellow'
  15 - high = 'Red'
.M = 'LightBlue'
other = 'White';

run;
```

As can be seen from these formats, a CRF that is within its submission window has no background coloring at all in the report. However, to highlight them in the report, a CRF that is slightly out of its submission window is colored yellow and a CRF that is more severely out of its submission window is colored red. A CRF that is missing and is past the allowable submission window (M) is colored light blue and a CRF that is expected and not yet submitted but is still within the allowable submission window (W) has no background coloring.

CONCLUSION

The use of a data collection and submission monitoring report such as the one described in this paper is invaluable to us in our roles in a clinical trials data coordinating center. The earlier this report is created and implemented, the sooner we will be alerted to problems and patterns in the data collection process. In addition, regular and thorough review of the report results in a more accurate trial database, a shorter time between the end of a trial and a final database and, ultimately, a higher quality trial. Though we have used this report exclusively in a clinical trials setting, we believe it may also have applicability in banking, educational and various business settings.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Rebecca A. Horney
 VA Maryland Health Care System
 CSPCC (151E)
 PO Box 1010
 Perry Point, MD 21902
 Work Phone: 410-642-2411 ext 5298
 Fax: 410-642-1129
 Email: Rebecca.Horney@va.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.

Output Examples

Example 1. All forms were completed within the expected windows.

25

Week	due	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10	F 11	F 12	F 13	F 14	F 15	F 16	F 17	F 18	F 19	F 20	F 21	F 22	F 23	F 24	F 25	F 26	F 27	F 28	F 29	F 30	F 31	F 36		
100-1001 Completer Week 24 CAPS=Yes																																			
0	05/01/07	-4	0	0	0		0	0	-4	-4	-4	-4		-4	-4	-4	-4	-4	-4	-4	0	-4	-4	-4	0	0	0	0	-4	-4	0				
1	05/08/07													0											0	0	0						0		
2	05/15/07													0											0	0	0						0		
4	05/29/07													0											0	0	0						0		
6	06/12/07												1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1			
8	06/26/07													0											0	0	0						0		
12	07/24/07				2								2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		2	2		2		
16	08/21/07													0											0	0	0						0		
20	09/18/07													0											0	0	0						0		
24	10/16/07		2	2									2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		

Example 2. Week 4 visit was 3 days past the expected due date (the window for the visit is ± 2 days). CRF 14 at visit 24 appears to have an error in the visit date; CRF 23 at visit 20 is 28 days past the expected due date representing a possible error in visit number.

5

Week	due	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10	F 11	F 12	F 13	F 14	F 15	F 16	F 17	F 18	F 19	F 20	F 21	F 22	F 23	F 24	F 25	F 26	F 27	F 28	F 29	F 30	F 31	F 36		
101-1003 Completer Week 24 CAPS=Yes																																			
0	06/25/07	-5	-3	0	0		0	0	-5	-5	-5	-3		-5	-3	-3	0	0	-5	0	-3	-5	-5	-5	-3	-3	-3	-3	-5	-5	0				
1	07/02/07													0											0	0	0						0		
2	07/09/07													0											0	0	0						0		
4	07/23/07													3											3	3	3						3		
6	08/06/07												3	3	3	3	3	3	3	3	3	3	3	3	1	3	3		3	3		3			
8	08/20/07													3											3	3	3						3		
12	09/17/07				0								3	3	3	3	3	3	3	3	3	3	3	3	3	3	3		3	3		3			
16	10/15/07													3											3	3	3						3		
20	11/12/07													3											28	3	3	3					3		
24	12/10/07		-4	-4									-4	-4	-73 1	-4	-4	-4	-4	-4	-4	-4	-4	-4	M	-4	-4	-4	-4	-4	-4	-4	-4		

Example 3. Missing CRFs (M) for visits 6, 8 and 12; CRFs were missing for visit 16 but the date of the report was within the submission window for the visit (W).

Week	due	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10	F 11	F 12	F 13	F 14	F 15	F 16	F 17	F 18	F 19	F 20	F 21	F 22	F 23	F 24	F 25	F 26	F 27	F 28	F 29	F 30	F 31	F 36				
104-1005 Unknown																																					
0	06/13/08	-1	-1	-17	-17		-1	-17	-1	-1	-1	-8		-1	-8	-1	-8	-8	-1	-8	-1	-1	-1	M	-17	-17	-17	-8	-1	-1	0						
1	06/20/08													-1											-1	-1	-1						-1				
2	06/27/08													-2											-2	-2	-2						-2				
4	07/11/08													-1											-1	-1	-1						-1				
6	07/25/08												M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M		M	M		M				
8	08/08/08												M												M	M	M						M				
12	09/05/08				M								M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M		M	M		M				
16	10/03/08													W											W	W	W						W				

Example 4. Beginning with visit 1, all CRFs were out of the submission windows for the respective visits; this illustrates how a participant can get “off schedule” early in a trial and continue the pattern throughout the trial (“visit creep”).

Week	due	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10	F 11	F 12	F 13	F 14	F 15	F 16	F 17	F 18	F 19	F 20	F 21	F 22	F 23	F 24	F 25	F 26	F 27	F 28	F 29	F 30	F 31	F 36				
100-1003 Completer Week 24 CAPS=Yes																																					
0	01/21/09	-13	-13	-7	-13		-13	-9	-13	-13	-9	-13		-9	-7	-9	-9	-7	-7	-1	-1	-1	-1	-1	-7	-7	-7	-1	-1	-1	0						
1	01/28/09													5												5	5	5					5				
2	02/04/09													6												6	6	6					6				
4	02/18/09													8												8	8	8					8				
6	03/04/09												8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8		8	8		8				
8	03/18/09													15												15	15	15					15				
12	04/15/09				15								15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15		15	15		15				
16	05/13/09													15												15	15	15					15				
20	06/10/09													14												14	14	14					14				
24	07/08/09			14	14								14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14			

Example 5. Case-Control Study. This is an example of a report from a paired case control study. The variable caseid is used to identify the pairs. The number of days before or after the enrollment date is listed for the enrollment and lab forms. For the 60 day follow-up, the number of days before or after the expected date of the 60 day follow-up visit is listed. The additional information shown includes enrollment status (E=Y), age, and two eligibility questions (Q1 and Q2).

Eligibility and Enrollment												Lab Forms						60 Day Followup					
CASEID	Child ID	E	Enroll Date	F2	Age	Q1	Q2	F3	F4	F5	F6	Lab ID	F7	F8	F9	F10	F11	F12	Day 60	F13	F14	Status	
1010001	1010001	Y	10DEC11	0	27	Y	Y	0	0	0	0	1009	0	3	43		25	185	08FEB12	69	69	Comp	
	1010002	Y	13DEC11		30			0	0		0	1004	0	2	40		22	183	11FEB12	66	66	Comp	
1020015	1010015	Y	10DEC11	0	11	Y	Y	0	0	0	0	1001	0	3	43		25	185	08FEB12	8	8	Comp	
	1010018	Y	13DEC11		11			0	0		0	1016	0	2	40		22	183	11FEB12	5	5	Comp	
1010027	1010027	Y	11DEC11	0	49	Y	Y	0	0	0	0	1010	0	3	42		24	184	09FEB12	10	10	Comp	
	1010030	Y	14DEC11		49			0	0		0	1006	0	2			21	182	12FEB12	7	7	Comp	
1010038	1010038	Y	13DEC11	0	9	Y	Y	0	0	0	0	1008	0	2	40		22	183	11FEB12	5	5	Comp	
	1010041	Y	19DEC11		7			0	0		0	1003	0	4			16	177	17FEB12	-1	-1	Comp	