How to Generate Questionnaire Data Compliance Report

Phaneendhar R. Gondesi, Blueprint Medicines Corporation

ABSTRACT

Achieving acceptable compliance is an important factor for the success of any clinical trial. Periodic monitoring of study data is key to ensure the expected compliance. Although the acceptable compliance rate varies from study to study, from assessment to assessment, the basic concept of calculating compliance remains the same. In this paper, we are going to discuss how to programmatically generate compliance report for most questionnaire (QS) data.

INTRODUCTION

Generally, a patient's QS information is collected either through paper case report form (CRF) (on-site) or electronic device (on-site or at home) and later entered into the clinical database. This data is then extracted periodically for various purposes like interim analysis, data cleaning, data monitoring etc.

For illustration purposes, we are going to consider an open-label, 2 period study where study drug is administered in a 28-day cycles. In period 1, patient is given treatment A and in period 2, patient is given treatment A and B. On day 1 of each cycle, patient is given medication onsite and for reminder of the cycle, dispensed medication is taken at home by the patient. Per protocol, on day 1, site collects the QS data. On other days of the cycle, all patients are required to answer the QS directly in an electronic device. QS data is collected prior to dosing and until the end of treatment or withdrawal of consent or end of study (whichever comes first). Study drug administration (SDA) data is collected at the site during each patient visit.

During the study conduct, the study team would like to check for compliance rate at each of the below categories. We are going to discuss how to programmatically generate QS compliance report for each category.

- QS data Compliance per patient per cycle
- QS data Compliance per patient per study period
- Overall QS data compliance per patient for the entire study
- Overall QS data Compliance by Site (including all the patients at each site)
- Overall QS data Compliance for the entire study (including all the patients at any study site)

SAMPLE QS FORM

Please continue on the next page...

SAMPLE QS DATA EXTRACT

	STUDYID	SITEID	SUBJECT	PERIOD	VISIT	DATE	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	STUDY1	1001	2023-001	1	CYCLE 1 DAY 1	2023-01-01	7	6	2	9	7	5	4	5	9	5
2	STUDY1	1001	2023-002	1	CYCLE 1 DAY 1	2023-01-01	7	5	6	7	8	6	4	3	9	3
3	STUDY1	1001	2023-002	1	CYCLE 1 DAY 2	2023-01-02	9	8	6	8	7	5	4	5	9	7
4	STUDY1	1001	2023-002	1	CYCLE 1 DAY 3	2023-01-03	5	6	3	5	6	4	4	5	4	3
5	STUDY1	1001	2023-002	2	CYCLE 1 DAY 1	2023-01-06	5	6	1	5	4	4	4	5	9	1
6	STUDY1	1002	2023-004	1	CYCLE 1 DAY 1	2023-01-01	7	2	6	9	7	9	4	5	9	4
7	STUDY1	1003	2023-005	1	CYCLE 1 DAY 1	2023-01-01	7	6	6	9	7	9	8	8	8	7

SAMPLE SDA DATA EXTRACT

	STUDYID	SITEID	SUBJECT	PERIOD	SDASTDT	SDAENDT	EXDOSE	MODTYPE
1	STUDY1	1001	2023-001	1	2023-01-01		25	
2	STUDY1	1001	2023-002	1	2023-01-01	2023-01-28	25	
3	STUDY1	1001	2023-002	1	2023-01-29	2023-02-25	25	
4	STUDY1	1001	2023-002	2	2023-02-26	2023-03-25	0	Dose interrupted
5	STUDY1	1001	2023-002	2	2023-03-26		25	
6	STUDY1	1002	2023-004	1	2023-01-05	2023-02-02	25	
7	STUDY1	1003	2023-005	1	2023-01-05	2023-01-30	25	

WHAT IS COMPLIANCE?

Compliance can be simply defined as being in accordance with the established guidelines or processes. In a clinical trial, it is adhering to the established protocol schedule underlined in schedule of assessments section. In this paper, we are discussing about compliance of QS data, where we are looking at the number of actual QS entries vs the expected number of QS entries. Finally, the compliance rate is calculated as number of actual QS entries per category divided by expected number of QS entries for the same category, multiplied with 100.

COMPLIANCE RATE FORMULA:

Compliance rate = actual number of QS entries per category (variable X) / expected number of QS entries per category (variable Y) *100.

Table1 summarizes X and Y variables for all categories:

Accurately defining the upper and lower boundaries of each category is important in identifying the accurate number of entries. Factors like protocol amendments, reconsent or even the pandemic disruptions could affect the boundaries. X, Y variables along with generic boundaries of each category are discussed in more detail in the following sections.

Number	Category	Variable X	Variable Y
1	QS data compliance per patient per cycle	Number of days QS data is collected in each cycle per patient.	Number of days in a cycle per patient.
2	QS data compliance per patient per study period	Number of days QS data is collected in each study period per patient.	Number of days the patient participated in each study period.
3	Overall QS data compliance per patient for the entire study	Number of days QS data is collected for a patient throughout the study.	Number of days a patient has participated in the study.
4	Overall QS data compliance by Site	See description in the corresponding section.	See description in the corresponding section.
5	Overall QS data compliance for the entire study	See description in the corresponding section.	See description in the corresponding section.

Table 1. Description of X and Y variables for each category

CATEGORY 1: QS DATA COMPLIANCE PER PATIENT PER CYCLE

PREPARATION

Variable Y:

In SDA data, identify the first (CzD1) and last day of each cycle i.e., (Cz+1D1)-1. Variable Y is then calculated as cycle end date – cycle start date +1.

- For all cycles but last,

  Y = ((Cz+1D1) – CzD1), where C is the cycle, z is the cycle number and D is the day of the visit.

Note:

Above formula can be used for all cycles but last when the patient is still on treatment in either period 1 or period 2. This formula assumes that patient has taken dose on each day of the cycle. For more accurate compliance rate, subtract the number of days where patient has missed treatment or had dose interruption etc. essentially eliminating records where actual dose taken is zero. However, if QS data is collected on the days when the patient hasn't received any dose, it could result in compliance rate greater than 100%, which would then need to be queried for data entry accuracy. If cycle information is not collected in SDA page, then Subject visits (SV) page can be used.

- For last cycle,

  o If a patient has crossed over to period 2, then period 1 number of last cycle (y) days are calculated as

    Y = Period 2 C1D1 – Period 1 CyD1

Note:

Above formula assumes that patient has taken dose everyday prior to Period 1 C1D1. This maybe not be always possible in a study, due to washout or dose interruptions etc. For more accurate results, subtract number of days where actual dose taken is zero.

  o If a patient did not cross over to period 2,

    ▪ if a patient has missing last record SDA end date and missing end of treatment (EOT) date (i.e., still on treatment), then

      Y = Data extraction date – CyD1+1

    ▪ if a patient has non-missing last record SDA end date, then

      Y = SDA last dose date – CyD1+1

    ▪ if a patient has missing last SDA end date and non-missing EOT date, then

      Y = {minimum (EOT date, last known SDA date) – CyD1} +1

Variable X:

Merge cycle start and end dates described above with QS data and identify the entries that occur between a specific cycle start and end dates (inclusive). Then count number of entries with non-missing QS data within a cycle to get variable X.

Apply compliance rate FORMULA: to obtain the compliance rate for each patient per cycle.

Below is a sample program to generate category 1 compliance report:

```
*** Cycle information is not collected in SDA as shown in sample SDA data
extract, we are using SV dataset to identify cycle start and end dates ***
;

* Date of extraction ;
```

```
proc sql noprint;
    select strip(put(datepart(crdate),is8601da.)) into :svcreated
    from dictionary.tables
    where libname='RAW' and memname='SV'
    ;
quit;

%put &svcreated ;

* Merge SV with EOT ;
data sv_eot;
merge raw.sv raw.eot(keep=subject eotdt);
by subject;
run;

proc sort data =sv_eot;
by subject period avisitn astdt;

* Identify cycle start and end dates;
data sv_edc_days(rename=(astdt_l=astdt avisit_l=avisit avisitn_l=avisitn
period_l=period));
set sv_eot;
by subject period avisitn ;

    avisit_l    =    lag(avisit);
    avisitn_l   =    lag(avisitn);
    period_l    =    lag(period);
    astdt_l     =    lag(astdt);

    if first.subject then do; avisit_l='' ; period_l=''; astdt_l=.;end;
    cyc_start = astdt_l;
    if astdt ne . then cyc_end = astdt-1;
    if subject=lag(subject) and astdt_l ne . then cycdays=astdt-astdt_l;

    if cycdays ne .;

    keep siteid subject astdt_l avisit_l cycdays cyc_start cyc_end
avisitn_l period_l eotdt;
    label cycdays= "Days from CxD1 to Cx+1D1 - 1";
    format astdt_l cyc_start cyc_end date9. ;
run;

proc sort;
by subject period avisitn astdt;
run;

* Correct last cycle end date;
data sv_final;
set sv_edc_days ;
by subject period avisitn astdt;
    if last.avisitn then do;
        if cyc_start=. then cyc_start=astdt;
        if cyc_end=. and eotdt eq . then
cyc_end=input("&svcreated.",yymmdd10.);*set extract date as last date of
last cycle when EOT date and cycle end dates are missing;
        if cyc_end=. and eotdt ne . then cyc_end=eotdt;*set EOT date as
last date of last cycle when both cycle end date is missing;
```

```
    end;
run;

* Merge SV with QS data ;
proc sql;
create table edc_tm as
select a.*
    ,cyc_end-cyc_start+1 as Y
    ,b.cyc_start
    ,b.cyc_end
    ,b.avisitn
    ,b.period
    ,eotdt
    ,case
        when cyc_start <= a.astdt <= cyc_end then b.avisit
        else ''
    end as avisit
from qs as a left join sv_final as b
    on a.subject=b.subject
    having avisit ne ''
    order by a.subject, period, avisitn, a.astdt
;quit;


* Calculate compliance by Cycle per patient ;
data Patient_Cycle(keep=subject cyc_start cyc_end comp_cy_c compliance
period avisit);
set edc_tm;
by subject period avisitn;
    retain tm_start;
    if first.avisitn then tm_start = astdt;
    tm_end = astdt;

    if first.avisitn then X= 0;
    X+1;

    comp_cy_c = strip(put(X,best.))||"/"||strip(put(Y,best.));
    if Y gt 0 then compliance = round((entries*100)/cycdays);

    if cyc_start > cyc_end then do;comp_cy_c=''; compliance
=.;tm_start=.;tm_end=.; end;

    if last.avisitn then output;

    format tm_: date9.;
    label
        Y="Total no of days in the period by cycle"
        X="Total Entries in period by cycle"
        Cyc_start="Cycle start date"
        Cyc_end="Cycle end date"
        Compliance="Compliance by cycle"
        Comp_cy_c="Actual number of entries per cycle/Expected number of
    entries per cycle"
    ;
run;

* Generate report;
```

```
proc export data=patient_cycle outfile="testpath/compliance_report.xlsx"
dbms=xlsx replace label;
sheet='patient_bycycle';
run;
```

CATEGORY 2: QS DATA COMPLIANCE PER PATIENT PER STUDY PERIOD

PREPARATION

Variable Y:

In SDA data, identify the first (C1D1) and last visit date for each period.

1) For period 1,

    a) If the patient has crossed over to period 2, then

$$Y = \text{Period 2 C1D1} - \text{Period 1 C1D1}$$

Note:

Above formula assumes that patient has taken dose everyday prior to Period 1 C1D1. This maybe not be always possible in a study, due to washout or dose interruptions etc. For more accurate results, subtract number of days where actual dose taken is zero.

    b) If the patient didn't cross over to period 2 and

        (i) Still on the study treatment with non-missing last SDA record end date, then

$$Y = \text{Last SDA end date} - \text{C1D1}$$

        (ii) Still on study treatment with missing last SDA record end date, then

$$Y = \text{Date of data extraction} - \text{C1D1}$$

        (iii) If the patient has ended treatment, then

$$Y = \text{minimum (Period 1 last known SDA date, EOT date)} - \text{C1D1}$$

2) For period 2,

    a) If the patient is still on study treatment and

        i) With non-missing last SDA record end date, then

$$Y = \text{Last SDA end date} - \text{Period 2 C1D1}$$

        ii) With missing last SDA record end date, then

$$Y = \text{Date of data extraction} - \text{Period 2 C1D1}$$

    b) If the patient has ended treatment, then

$$Y = \{\text{minimum (Period 2 SDA last known date, EOT date)}\} - \text{Period 2 C1D1}$$

Variable X:

Merge periods start and end dates described above with QS data and identify the entries that occur between a specific period start and end dates (inclusive). Then count number of entries with non-missing QS data within a period to get variable X.

Apply compliance rate FORMULA: to obtain the compliance rate for each patient per period.

Below is a sample program to generate category 2 compliance report:

```
** First dose date in Period 1 **;
data sda_ft;
set raw.sda;
```

```sas
where period=1 and dose > 0 and modtype = '' and sdastdt ^= .;
    keep subject sdastdt;
run;

proc sort data = sda_ft;
by subject sdastdt ;
run;

data sda_ftf;
set sda_ft;
by subject ;
    if first.subject;
    keep subject sdastdt;
    rename sdastdt = p1d1;
run;

** Last dose date in Period 1 **;
data sda_lt;
set raw.sda;
where period=1 and dose > 0;
    * for ongoing patients;
    if sdaendt = . and sdastdt ^= . then sdaendt = sdastdt;
    keep subject dose sdastdt sdaendt ;
run;

proc sort data = sda_lt;
by subject sdaendt dose;
run;

data sda_ltf;
set sda_lt;
by subject ;
    if last.subject;
    keep subject sdaendt;
    rename sdaendt=p1dL;
run;

** First dose date in Period 2 **;
data sda_ft2;
set raw.sda;
where period=2 and dose > 0 and modtype = '' and sdastdt ^= .;
    keep subject sdastdt ;
run;

proc sort data = sda_ft2;
by subject sdastdt;
run;

data sda_ft2f;
set sda_ft2;
by subject sdastdt;
    if first.subject;
    rename sdastdt=p2d1;
run;

** Last dose date in Period 2 **;
data sda_lt2;
```

```
set raw.sda;
where period=2 and dose > 0;
    *  for ongoing patients;
    if sdaendt = . and sdastdt ^= . then sdaendt = sdastdt;
    keep subject dose sdastdt sdaendt ;
run;

proc sort data = sda_lt2;
by subject sdaendt dose;
run;

data sda_lt2f;
set sda_lt2;
by subject ;
    if last.subject;
    keep subject sdaendt;
    rename sdaendt=p2dL;
run;

** End of treatment;
data eot;
set raw.eot;
keep subject eotdt;
run;

proc sort data = eot;
by subject;
run;

** all patients - period 1 and 2 - start and end dates ;
data all_sda;
merge sda_ftf sda_ft2f sda_ltf sda_lt2f eot;
by subject;
    * if patient crossed over;
    if nmiss(p2d1,p2dL)<2 then do;
          if p1dL = . and p2d1 ^= . then p1dL=p2d1-1;
          if p2dL = . then do;
                if p2dL = . and eotdt ^= . then p2dL=eotdt;
                else if nmiss(p2dL,eotdt)=2 then p2dL = &svcreated.;
                else if nmiss(p2dL,eotdt)=0 then p2dL = min(p2dL,eotdt);
          end;
    end;
    * for patient's in period 1 and not crossed over to period 2;
    else if nmiss(p2d1,p2dL)=2 then do;
          if p1dL = . and eotdt ^= . then p1dL=eotdt;
          else if nmiss(p1dL,eotdt)=2 then p1dL = &svcreated.;
          else if nmiss(p1dL,eotdt)=0 then p1dL = min(p1dL,eotdt);
    end;
run;

** Calculate compliance by Period per patient ;
proc sort data = qs;
by subject;
run;

data qs_sda;
merge qs all_sda;
```

```
by subject;
    if date ^= . then do;
            if datepart(p1d1) <= date <=datepart(p1dL) then period =1;
            if datepart(p2d1) <= date <=datepart(p2dL) then period =2;
    end;
    if period ^= .;*only consider QS records after Period 1 C1D1;
run;

proc sort;
by subject period;

data patient_period;
set qs_sda;
by subject period;
    * X - count the qs entries;
    if first.period then X=1;
    else X+1;

    * Y - expected qs entries;
    if nmiss(p1d1,p1dL)=0 and period=1 then Y = datepart(p1dL)-
datepart(p1d1)+1;
    if nmiss(p2d1,p2dL)=0 and period=2 then Y = datepart(p2dL)-
datepart(p2d1)+1;

    * compliance rate;
    compliance = X*100/Y;

    if last.period;

    keep subject compliance period p1d1 p1dL p2d1 p2dL;

    label
            subject='Subject identifier'
            compliance='Compliance by period'
            period ='Study period'
            p1d1='Period 1 first dose date'
            p1dL='Period 1 Last dose date'
            p2d1='Period 2 first dose date'
            p2dL='Period 2 Last dose date'
            X='Actual number of QS entries'
            Y='Expected number of QS entries'
    ;
run;

* Generate report;
proc export data=patient_period outfile="testpath/compliance_report.xlsx"
dbms=xlsx replace label;
sheet='patient_byperiod';
run;
```

CATEGORY 3: OVERALL QS DATA COMPLIANCE PER PATIENT FOR THE ENTIRE STUDY

PREPARATION

Variable Y:

1) If the patient is still on study treatment

    a) with non-missing last SDA end date, then

    $$Y = \text{Last SDA end date} - \text{Period 1 C1D1} + 1$$

    b) with missing last SDA end date, then

    $$Y = \text{Date of data extraction} - \text{Period 1 C1D1} + 1$$

2) If the patient has ended treatment, then

    $$Y = \{\text{minimum (SDA last known date, EOT date)}\} - \text{Period 1 C1D1} + 1$$

Variable X:

Merge study start and end dates described above with QS data and identify the entries that occur between the start and end dates. Then count number of entries with non-missing QS data within the study to get variable X.

Apply compliance rate FORMULA: to obtain overall compliance rate for each patient in the study.

Below is a sample program to generate category 3 compliance report:

```
* Calculate compliance by patient for entire study;
data patient_study;
set qs_sda;*from category 3 code;
by subject ;
    * X- count the qs entries;
    if first.subject then X=1;
    else X+1;

    * expected qs entries;
    if nmiss(p1d1,p2dL)=0 and p2d1 ^= . then Y = datepart(p2dL)-
datepart(p1d1)+1;
    if nmiss(p1d1,p1dL)=0 and p2d1 = . then Y = datepart(p1dL)-
datepart(p1d1)+1;

    * compliance rate;
    compliance = X*100/Y;

    if last.subject;

    keep subject compliance p1d1 p1dL p2d1 p2dL ;

    label
        subject='Subject identifier'
        compliance='Compliance by study'
        p1d1='Period 1 first dose date'
        p1dL='Period 1 Last dose date'
        p2d1='Period 2 first dose date'
        p2dL='Period 2 Last dose date'
        actual='Actual number of QS entries'
        expected='Expected number of QS entries'
    ;
run;

* Generate report;
proc export data=patient_study outfile="testpath/compliance_report.xlsx"
dbms=xlsx replace label;
```

```
sheet='patient_bystudy';
run;
```

CATEGORY 4: OVERALL QS DATA COMPLIANCE BY SITE

For individual site compliance, calculate the average of all the patient's overall compliance rate from the category 3: Overall QS data compliance per patient for the entire study at the corresponding site.

Below is a sample program to generate category 4 compliance report:

```
proc univariate data = patient_study noprint;
by studyid siteid;
var compliance;
output out=stat mean=mean;
run;

proc export data=stat outfile="testpath/compliance_report.xlsx" dbms=xlsx
replace label;
sheet='compliance_bysite';
run;
```

CATEGORY 5: OVERALL QS DATA COMPLIANCE FOR THE ENTIRE STUDY

For individual study compliance, calculate the average of all the study patient's overall compliance rate from category 3: Overall QS data compliance per patient for the entire study.

Below is a sample program to generate category 5 compliance report:

```
proc univariate data = patient_study noprint;
by studyid;
var compliance;
output out=stat_study mean=mean;
run;

proc export data=stat_study outfile="testpath/compliance_report.xlsx"
dbms=xlsx replace label;
sheet='compliance_bystudy';
run;
```

CONCLUSION

In this paper, we have discussed what compliance is and how to calculate compliance rate for the mentioned sample QS data. There are many ways to achieve the goal, and this is just one of them. Above method can be applied to calculate compliance rate of most scheduled assessments. The important aspect to remember is to accurately identify the boundaries of the given category based on study protocol and consider all possible scenarios while applying the formula. The approach discussed in this paper only provided a general idea and raw data anomalies should be considered while applying the logic.

ACKNOWLEDGEMENTS

The examples discussed in this paper are solely for illustration purposes and not in any form representative of actual clinical trial conducted at Blueprint Medicines Corporation.

RECOMMENDED READING

- *Base SAS® Procedures Guide*

- *SAS® For Dummies®*

CONTACT INFORMATION

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

Phaneendhar R. Gondesi
Programmer analyst,
Blueprint Medicines Corporation
pgondesi@blueprintmedicines.com

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