SAS® GLOBAL FORUM 2021

Analyzing Data Missing Due to COVID-19: from Raw Data to ADaM and Beyond

Yuliia Bahatska, Roche

ABSTRACT

The COVID-19 pandemic has a tremendous impact on everyone - and on the pharmaceutical industry. Naturally, those who are developing drugs took the first hit trying to treat it; however, statistical programmers who are working on any ongoing study had to reconsider their activities and timelines as well. New analyses need to be performed and the existing ones need to be amended taking lots of missing and incomplete data into account. CDISC has published the "Interim User Guide for COVID-19" [1], which describes how to represent COVID-19 related data with CDASH and SDTM. In my paper, I would like to go one step further and walk you through an example of preparing this data for analysis from mapping it to the VE SDTM [2] domain introduced by CDISC, to creation of the corresponding ADaM dataset, and then to the R shiny app supporting the analyses team was required to perform.

INTRODUCTION

It is hard to think of an area of our lives that has not been affected by the pandemic. Multiple variations of restrictions, lockdowns and general uncertainty is what we are dealing with on a daily basis. And this is even more the case for the patients who are participating in the clinical trials. Many sites all over the world have been temporarily closed for several months and could not perform any procedures or collect any information, and therefore could not enter anything in the clinical trials databases. As many programmers, I have recently faced the challenge of having a significant amount of data missing for the primary study endpoint and together with other stakeholders needed to make a decision whether this would impact the study results and whether a protocol amendment was required.

STUDY SPECIFIC INFORMATION

There are numerous ways of how to assess the impact of missing data on the study results so before choosing one or the other one needs to understand whether it fits to their particular study so I would like to provide a brief overview of the study I am working on. It is a phase III, randomized, double-controlled, double-blind, multicenter clinical study to evaluate efficacy, safety, PK and biomarker effects of intrathecally administered Tominersen in patients with manifest HD. The primary efficacy endpoints were the change from baseline in composite Unified Huntington's Disease Rating Scale and Total Functional Capacity Scale score at week 101. After the start of pandemic when the sites were closed, so many visits were replaced with the phone calls and information on adverse events and concomitant medications was collected through these calls. The composite UHDRS is a multidimensional measure of progression in Huntington's disease (HD) and TFC is one of its components. For both of these measures it is possible to obtain the result remotely so once the sites were no longer able to collect them, it was decided to switch to remote method of data collection. In order to estimate the impact of the missing data on the study results the following analyses were planned:

- Summary of missed visits due to any COVID-19 related reasons
- Summary of missed doses due to any COVID-19 related reasons
- Summary of missed cUHDRS assessments due to any COVID-19 related reasons
- Summary of the duration of not being treated due to any COVID-19 related reasons

Also, the primary efficacy outcome data collected remotely was of interest so the programmers needed to distinguish between remote and onsite visits in order to be able to perform a sensitivity analysis.

FROM RAW DATA TO SDTM

The project team decided to follow the CDISC recommendation for ongoing studies and create a VE domain as a starting point. Below you can see a brief overview of the domain:

ve.xpt

Row	STUDYID	DOMAIN	USUBJID	VESEQ	VETERM	VEDECOD	VEPRESP	VEOCCUR	VISITNUM	VISIT	VISTDY	VEDTC	VESTDTC	VEDY	VESTDY
1	CVD-3	VE	301	1	ONSITE VISIT	PLANNED VISIT	Y	Y	1	BASELINE	1	2020-03-04	2020-03-04	1	1
2	CVD-3	VE	301	2	ONSITE VISIT	PLANNED VISIT	Y	Y	2	WEEK 1	8	2020-03-11	2020-03-11	8	8
3	CVD-3	VE	301	3	REPEAT VISIT FOR ABNORMAL LAB	UNSCHEDULED VISIT			2.5			2020-03-13	2020-03-13	10	10
4	CVD-3	VE	301	4	ONSITE VISIT	PLANNED VISIT	Y	Y	3	WEEK 2	15	2020-03-19	2020-03-19	16	16
5	CVD-3	VE	301	5	ONSITE VISIT	PLANNED VISIT	Y	N	4	WEEK 3	22	2020-03-25		22	
6	CVD-3	VE	301	6	ONSITE VISIT	PLANNED VISIT	Y	N	5	WEEK 4	29	2020-04-01		29	
7	CVD-3	VE	301	7	ONSITE VISIT	PLANNED VISIT	Y	N	6	WEEK 5	26	2020-04-08		36	
8	CVD-3	VE	301	8	VIRTUAL VISIT	PLANNED VISIT	Y	Y	7	WEEK 6	43	2020-04-15	2020-04-15	43	43

Table 1: VE domain structure

Basically, it can contain the information whether visit occurred onsite or remotely (stored in VETERM), whether the visit was planned or not (contained in VEDECOD and VEPRESP), and whether or not it occurred (VEOCCUR). This was almost sufficient for our needs. The only thing missing was how to determine if an assessment at the visit was missing due to COVID or any other reason. Since when the eCRF was designed no one planned for a pandemic, there was no option envisaged to cover that scenario and instead of doing a migration the team agreed to the rule of thumb: search for a COVID related protocol deviation at the visit and if there was one, assume the missing assessments were not recorded because of COVID-19.

After the pandemic started all the sites were instructed to record COVID-19 related protocol deviations with the description starting with "COVID-19". This data is linked to a certain visit so checking whether the deviation description contains the text "COVID-19" allowed us to see whether the visit was affected by the pandemic. We then decided to use one of the variables suggested by CDISC: VEEPCHGI, which is labelled as 'Epi/Pandemic Related Change Indicator' and is a flag to store this information. Therefore, the simplified version of code looks like below:

```
data COVIDFL;
  set CTMS;
  length VISIT $200;
  …
  *study specific visit derivations;
  …
  if(find(DVDESC,'COVID','i') then COVIDFL="Y";
  if COVIDFL="Y";
```

```
    keep USUBJID VISIT COVIDFL;
run;

proc sort data = COVIDFL nodupkey;
  by USUBJID VISIT;
run;

data PRE_VE;
  merge VISIT(in=vis) COVIDFL(in=cov);
  by USUBJID VISIT;
  if vis;
  if cov then VEEPCHGI="Y";
  else VEEPCHGI="N";
run;
```

The second step of derivations is required in order to identify the remote visits. It was agreed that such visits would be recorded as "NOT DONE" in RAVE but then would have the assessment results recorded in the non-CRF data. Therefore, if the visit was marked as "DONE" in RAVE it was treated as an onsite visit, if it was marked as "NOT DONE" but there was a record in the non-CRF data, the visit is treated as remote, and otherwise we conclude that the visit did not occur at all:

```
data VE;
    merge VISIT NON_CRF(in=non_crf);
    by USUBJID VISIT;
    if VISIT = "UNSCHEDULED" then VETERM = "UNSCHEDULED VISIT";
    else if NOTDN = 0 then VETERM   = "ONSITE VISIT";
    else if NOTDN = 1 and non_crf = 1 then VETERM   = "REMOTE VISIT";
    else put "War" "ning: VETERM isn't populated" USUBJID = VISIT =;

  if VISIT = "UNSCHEDULED" then do;
     VEDECOD = "UNSCHEDULED VISIT";
     VEPRESP = " ";
  end;
  else do;
     VEDECOD = "PLANNED VISIT";
     VEPRESP = "Y";
     if NOTDN = 0 or (NOTDN = 1 and non_crf = 1) then VEOCCUR = "Y";
     else VEOCCUR = "N";
  end;
run;
```

FROM SDTM TO ADAM

ADMS, EXPECTED PARAMETERS

Once we created VE, we need to proceed to ADaM that would fit our analysis needs. The dataset is called ADMS, which stands for "Missed assessments analysis dataset". We decided to use BDS structure and have one record per subject per visit per parameter. There are three categories of parameters: expected parameters, completed parameters and duration of dosing missed due to COVID-19.

First, we need to know which visits are expected for every subject and which assessments are expected on the certain visits. Luckily, all the information about scheduled visits can be found in TV (trial visits) SDTM domain. Using it and ADSL (subject level analysis dataset) we are able to create a starting point for expected parameters derivations:

```
proc sql noprint;
```

```
    create table EXPECTED_VISITS as select a.*, b.VISIT, b.VISITNUM,
        b.VISITDY from ADSL a, TV b order by a.USUBJID, b.VISITNUM;
quit;
```

This is a Cartesian product of all study subjects and all scheduled study visits. Note, that this simple code assumes that all subjects in the study have the same visit schedule. If the visit schedule varies, for example, for different treatment arms, this code will need to be modified.

The next step is to determine whether certain visit was expected for a certain subject. For example, if a subject discontinued study prior to the visit, the visit is obviously not expected. Neither is it expected if the subject is still on study but has not yet reached the visit at the point of the snapshot. In order to be able to identify such visits we need to derive planned date (PDT) of visit first. We use date of Day 1 visit as a starting point and then add to it the planned day of visit. In case the visit is on or after Day 1, we substitute one day. In case visit occurred, we just derive analysis date (ADT):

```
data DAY1;
    set SV(where=(VISIT="DAY 1"));
    if not missing(SVSTDTC) then DAY1DT=input(SVSTDTC,yymmdd10.);
    keep USUBJID DAY1DT;
run;

data EXPECTED_VISITS;
    merge EXPECTED_VISITS VE;
     by USUBJID VISITNUM;
run;

data ALL_VISITS;
    merge EXPECTED_VISITS DAY1;
    by USUBJID;
    if VEOCCUR="Y" then ADT=input(VESTDTC,yymmdd10);
    else PDT=DAY1DT+VISITDY-(VISITDY>0);
run;
```

Now we are able to relate visits to important study dates. We can compare the planned date to the date of program run to identify visits that are not reached and to date of study discontinuation to identify the visits that will never occur. In the first case, we set analysis value (AVAL) for the parameter to 0 and reason not done ARSND to "VISIT NOT REACHED". In the second case, we also set analysis value (AVAL) for the parameter to 0 and reason not done ARSND to the reason for study discontinuation:

```
if PDT>today() and EOSSTT="ONGOING" then do;
    AVAL=0;
    ARSND="VISIT NOT REACHED";
end;
else if PDT>EOSDT and EOSSTT="DISCONTINUED" then do;
    AVAL=0;
    ARSND=DCSREAS;
end;
else AVAL=1;
```

Similarly, it is possible to derive expected parameters for all study endpoints of interest. The only difference is that since not all endpoints are collected at every visit, we need to check the protocol to identify the visits where a specific endpoint is collected, and subset the dataset EXPECTED_VISITS accordingly.

ADMS, COMPLETED PARAMETERS

Afterwards we are able to proceed to the second category of parameters. For each parameter that is expected (ADMS.AVAL=1 at a given record) we need to know whether the assessment was performed and if not, why. For this, we create a simple macro:

```
%macro exp_compl(expected=, actual=,cond=1,paramcd=,param=);
    proc sort data=&actual(where=(&cond)) out=ACTUAL nodupkey;
        by USUBJID VISITNUM VISIT;
    run;

    data &paramcd;
        length PARAM ARSND $200 PARAMCD $8;
        merge &expected
        ACTUAL(in=in1 where=(VISIT ne "UNSCHEDULED"));
        by USUBJID VISITNUM;
        PARAMCD="&paramcd.E";
        PARAM="&param Expected";
        output;
        if AVAL=1 then do;
                PARAMCD="&paramcd.C";
                PARAM="&param Completed";
                %if &paramcd=DOSE %then %do;
                        if in1 and ECOCCUR="Y" then AVAL=1;
                        else do;
                                AVAL=0;
                                if VEEPCHGI="Y" then ARSND="COVID-19";
                                else ARSND=coalescec(ECREASOC,"NOT SPECIFIED");
                        end;
                %end;
                %else %do;
                        if in1 then AVAL=1;
                        else do;
                                AVAL=0;
                                if VEEPCHGI="Y" then ARSND="COVID-19";
                                else ARSND="NOT SPECIFIED";
                        end;
                %end;
                output;
        end;
    run;
%mend;

%exp_compl(expected=expected_doses, actual=EC, paramcd=DOSE, param=Dose);
%exp_compl(expected=expected_TFC, actual=QS, cond=%str(missing(QSSTAT) and
    QSSCAT = "FUNCTIONAL CAPACITY"), paramcd=TFC, param=%str(Total
    Functional Capacity));
```

The purpose of this macro is to compare visits where the parameter was expected to the subset of respective SDTM dataset with results of assessment or information about doses or visits. Note, that doses are treated separately. This is because for all other assessments when the assessment was not done, the reason why it was not done is recorded as free text and we are not interested in it. For doses we use EC domain (exposure as collected) and on the eCRF the reason has a codelist and therefore it is possible to analyze it properly.

Table 2 shows an example of a subject who is still on study. This subject has completed all the visits up to week 77 and the next ones are not expected because their planned date is in the feature (compared to the date of program run).

USUBJID	VISIT	VISITNUM	VEDECOD	VEEPCHGI	PARAM	ADT (format...	PDT (for...	AVAL	ARSND
BN40423-001	DAY 1	1	PLANNED VISIT	N	Visit Overall Expected	15JUL2019	.	1	
BN40423-001	SCREENING	-1	PLANNED VISIT	N	Visit Overall Expected	24JUN2019	.	1	
BN40423-001	WEEK 13	4	PLANNED VISIT	N	Visit Overall Expected	30SEP2019	.	1	
BN40423-001	WEEK 21	6	PLANNED VISIT	N	Visit Overall Expected	03DEC2019	.	1	
BN40423-001	WEEK 29	8	PLANNED VISIT	N	Visit Overall Expected	28JAN2020	.	1	
BN40423-001	WEEK 37	10	PLANNED VISIT	Y	Visit Overall Expected	10MAR2020	.	1	
BN40423-001	WEEK 45	12	PLANNED VISIT	N	Visit Overall Expected	19MAY2020	.	1	
BN40423-001	WEEK 5	2	PLANNED VISIT	N	Visit Overall Expected	12AUG2019	.	1	
BN40423-001	WEEK 53	14	PLANNED VISIT	N	Visit Overall Expected	14JUL2020	.	1	
BN40423-001	WEEK 61	16	PLANNED VISIT	N	Visit Overall Expected	15SEP2020	.	1	
BN40423-001	WEEK 69	18	PLANNED VISIT	N	Visit Overall Expected	03NOV2020	.	1	
BN40423-001	WEEK 77	20	PLANNED VISIT	N	Visit Overall Expected	04JAN2021	.	1	
BN40423-001	WEEK 85	22			Visit Overall Expected	.	22FEB2021	0	VISIT NOT REACHED
BN40423-001	WEEK 93	24			Visit Overall Expected	.	19APR2021	0	VISIT NOT REACHED
BN40423-001	DAY 1	1	PLANNED VISIT	N	Visit Overall Completed	15JUL2019	.	1	
BN40423-001	SCREENING	-1	PLANNED VISIT	N	Visit Overall Completed	24JUN2019	.	1	
BN40423-001	WEEK 13	4	PLANNED VISIT	N	Visit Overall Completed	30SEP2019	.	1	
BN40423-001	WEEK 21	6	PLANNED VISIT	N	Visit Overall Completed	03DEC2019	.	1	
BN40423-001	WEEK 29	8	PLANNED VISIT	N	Visit Overall Completed	28JAN2020	.	1	
BN40423-001	WEEK 37	10	PLANNED VISIT	Y	Visit Overall Completed	10MAR2020	.	1	
BN40423-001	WEEK 45	12	PLANNED VISIT	N	Visit Overall Completed	19MAY2020	.	1	
BN40423-001	WEEK 5	2	PLANNED VISIT	N	Visit Overall Completed	12AUG2019	.	1	
BN40423-001	WEEK 53	14	PLANNED VISIT	N	Visit Overall Completed	14JUL2020	.	1	
BN40423-001	WEEK 61	16	PLANNED VISIT	N	Visit Overall Completed	15SEP2020	.	1	
BN40423-001	WEEK 69	18	PLANNED VISIT	N	Visit Overall Completed	03NOV2020	.	1	
BN40423-001	WEEK 77	20	PLANNED VISIT	N	Visit Overall Completed	04JAN2021	.	1	

Table 2: Ongoing subject

Table 3 illustrates the situation where the subject discontinued study due to withdrawal of consent. Here we can see that doses after week 21 are not expected because subject is no longer in the study. We can also see that the subject received dosing up to week 13 and then on week 21 there was no dose but there is no record in EC for this dose so the reason why dose was not received is set to "NOT SPECIFIED".

USUBJID	VISIT	VISITNUM	VEDECOD	VEEPCHGI	PARAM	AVAL	ADT (format...	PDT (format...	ARSND
BN40423-002	DAY 1	1	PLANNED VISIT	N	Dose Expected	1	16JUL2019	.	
BN40423-002	WEEK 5	2	PLANNED VISIT	N	Dose Expected	1	13AUG2019	.	
BN40423-002	WEEK 13	4	PLANNED VISIT	N	Dose Expected	1	15OCT2019	.	
BN40423-002	WEEK 21	6	PLANNED VISIT	N	Dose Expected	1	.	02DEC2019	
BN40423-002	WEEK 29	8			Dose Expected	0	.	27JAN2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 37	10			Dose Expected	0	.	23MAR2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 45	12			Dose Expected	0	.	18MAY2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 53	14			Dose Expected	0	.	13JUL2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 61	16			Dose Expected	0	.	07SEP2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 69	18			Dose Expected	0	.	02NOV2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 77	20			Dose Expected	0	.	28DEC2020	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 85	22			Dose Expected	0	.	22FEB2021	WITHDRAWAL BY SUBJECT
BN40423-002	WEEK 93	24			Dose Expected	0	.	19APR2021	WITHDRAWAL BY SUBJECT
BN40423-002	DAY 1	1	PLANNED VISIT	N	Dose Completed	1	16JUL2019	.	
BN40423-002	WEEK 5	2	PLANNED VISIT	N	Dose Completed	1	13AUG2019	.	
BN40423-002	WEEK 13	4	PLANNED VISIT	N	Dose Completed	1	15OCT2019	.	
BN40423-002	WEEK 21	6	PLANNED VISIT	N	Dose Completed	0	.	02DEC2019	NOT SPECIFIED

Table 3: subject discontinued study

ADMS, DURATION OF DOSING MISSED DUE TO COVID-19

The last parameter we derive for ADMS is duration of dosing missed due to COVID-19. We define it as difference in days between actual date or planned date of dose of the last dose prior to missed dose due to COVID-19 and the first dose after dose missed due to COVID-19 or the planned date of the dose missed not due to COVID-19. In case the dosing is never resumed and the last planned dose is missed due to COVID, we use either the study discontinuation date, if available or the date of the program run if study discontinuation date is not available.

```
proc sql noprint;
    create table FOR_COVID as select distinct USUBJID from DOSE where
        find(ARSND,"COVID-19");
quit;

data DOSE;
    merge DOSE FOR_COVID (in=in1);
    by USUBJID;
    if in1;
run;

data DOSE_DATES NOT_RESUMED;
    set DOSE;
    retain DATE_PREV;
    by USUBJID VISITNUM;
    DATE_TEMP1=lag(ADT);
    DATE_TEMP2=lag(PDT);
    DATE_TEMP=max(DATE_TEMP1, DATE_TEMP2);
    REAS_PREV=lag(ARSND);
    if first.USUBJID then call missing(of DATE_TEMP: DATE_PREV REAS_PREV);

    *start of the interval of dosing missed due to COVID;
    if ARSND="COVID-19" and (REAS_PREV ne "COVID-19") then
        DATE_PREV=DATE_TEMP;

    *stop of the interval of missed dosing;
    if REAS_PREV ne "COVID-19" and ARSND ne "COVID-19" then call
        missing(DATE_PREV);

    output DOSE_DATES;
    if last.USUBJID and ARSND="COVID-19" then output NOT_RESUMED;
run;
```

USUBJID	VISIT ▲	ADT (for... ▲	AVAL	ARSND ▲	REAS_PREV	DATE_PRE...	DATE_TEMP (f...
BN40423-001	WEEK 29	28JAN2020	1			.	03DEC2019
BN40423-001	WEEK 37	10MAR2020	0	COVID-19		28JAN2020	28JAN2020
BN40423-001	WEEK 45	19MAY2020	1		COVID-19	28JAN2020	10MAR2020
BN40423-001	WEEK 53	14JUL2020	1			.	19MAY2020

Table 4

USUBJID	VISIT	ADT (format...	AVAL	ARSND	REAS_PREV	DATE_PREV (...	DATE_TEMP (f...
BN40423-003	WEEK 29	19FEB2020	0	DRUG HOLIDAY		.	06JAN2020
BN40423-003	WEEK 37	16APR2020	0	COVID-19	DRUG HOLIDAY	19FEB2020	19FEB2020
BN40423-003	WEEK 45	09JUN2020	0	COVID-19	COVID-19	19FEB2020	16APR2020
BN40423-003	WEEK 53	11AUG2020	0	DRUG HOLIDAY	COVID-19	19FEB2020	09JUN2020
BN40423-003	WEEK 61	13OCT2020	0	DRUG HOLIDAY	DRUG HOLIDAY	.	11AUG2020
BN40423-003	WEEK 69	30NOV2020	0	DRUG HOLIDAY	DRUG HOLIDAY	.	13OCT2020
BN40423-003	WEEK 77	25JAN2021	0	DRUG HOLIDAY	DRUG HOLIDAY	.	30NOV2020

Table 5

Table 4 and Table 5 illustrate different scenarios of missed dosing. In the first one, the subject has only missed one dose due to COVID-19 and in the second one there are more missed doses but only two of them were missed due to COVID. Then we only need to subset for those records where the dosing was either resumed or missed due to some other reasons, calculate the durations of the intervals and sum them up:

```
data COVID_DOSE;
    set DOSE_DATES(rename=AVAL=AVAL_OLD);
    if not missing(DATE_PREV) and (AVAL_OLD=1 or ARSND ne "COVID-19");
    _AVAL=coalesce(ADT,PDT)-DATE_PREV-(coalesce(ADT,PDT)>=DATE_PREV);
    keep USUBJID _AVAL;
run;

data NOT_RESUMED;
    set NOT_RESUMED;
    if not missing(EOSDT) then _AVAL=EOSDT-DATE_PREV-(EOSDT>=DATE_PREV);
    else _AVAL=today()-DATE_PREV-(today()>=DATE_PREV);
    keep USUBJID _AVAL;
run;

data COVID_DOSE;
    set COVID_DOSE NOT_RESUMED;
run;

proc sql noprint;
    create table COVID_DOSE_SUM as select distinct USUBJID, sum(_AVAL) as
            AVAL from COVID_DOSE group by USUBJID;
quit;
```

This is the final part of ADMS derivations.

ANALYSES

Once the ADaM is finalized, one can proceed to the analysis steps. For this study the team produced frequency tables for missed assessments and also descriptive statistics tables for the scores that were collected remotely. The SAP also envisages multiple imputations analysis and although it was not the original purpose of ADMS to support this kind of analysis, it is helpful to have the information on why the assessments are missing because the type of imputation and values used for it depend on the reason for the value being missing.

After the first lockdown started, the stakeholders wanted to be able to monitor the missed assessments. At Roche, we have a shiny-based interactive exploration framework that allows the users to create their own shiny apps based on standard modules so the team

decided to create an app that would give an overview of missed visits and doses. Below is a screenshot from the one of the most illustrative tabs. The app modules enable the users to distinguish between visits and doses, select the reasons of interest and add subgroups, i.e. geographic region, country, site etc. to the analyses. Around week 29-week 37 in Figure 2 (March-April 2020 Figure 3) one can clearly see the peak of the endpoints missing due to COVID-19 (Figure 2). This corresponds to the start of the lockdown period.

The data closer to the current date shows also the increase of missed assessments, however this should be interpreted with caution and the reviewers should keep in mind that not all the data might be entered in the database (see the peak of "NOT SPECIFIED" in Figure 2)
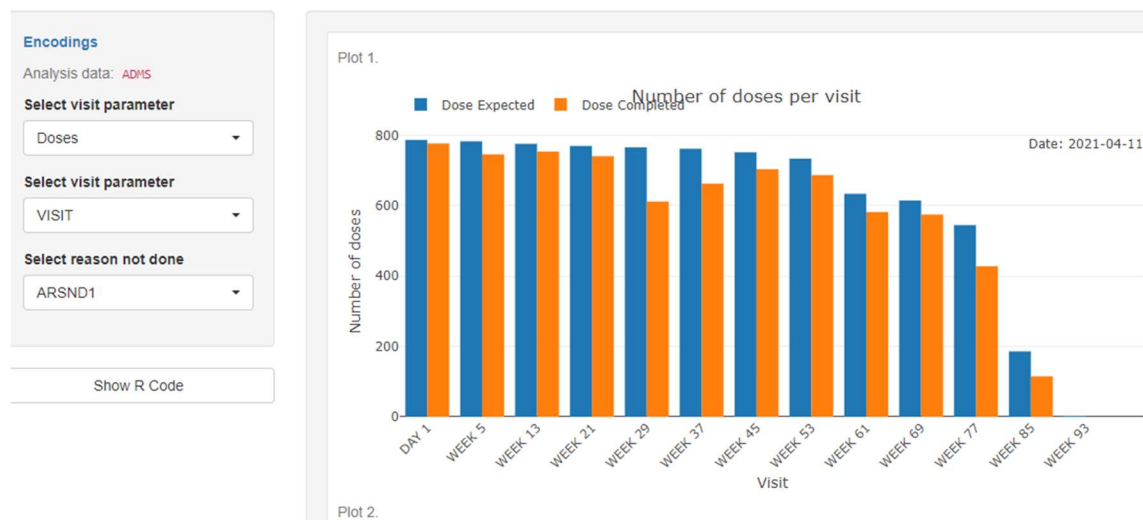
Figure 1: Number of expected doses vs completed doses per visit
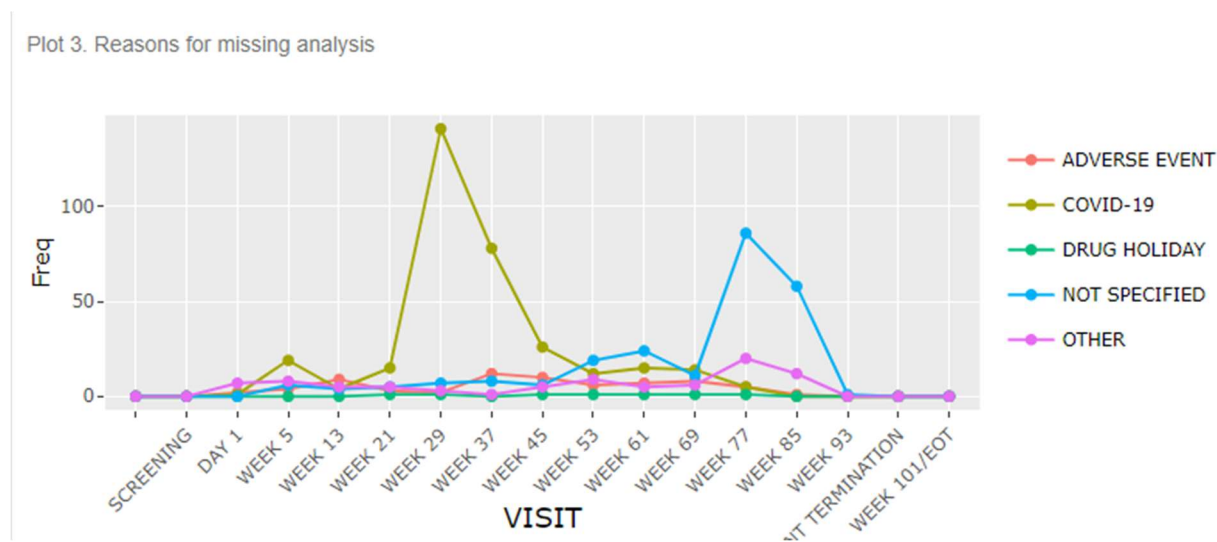
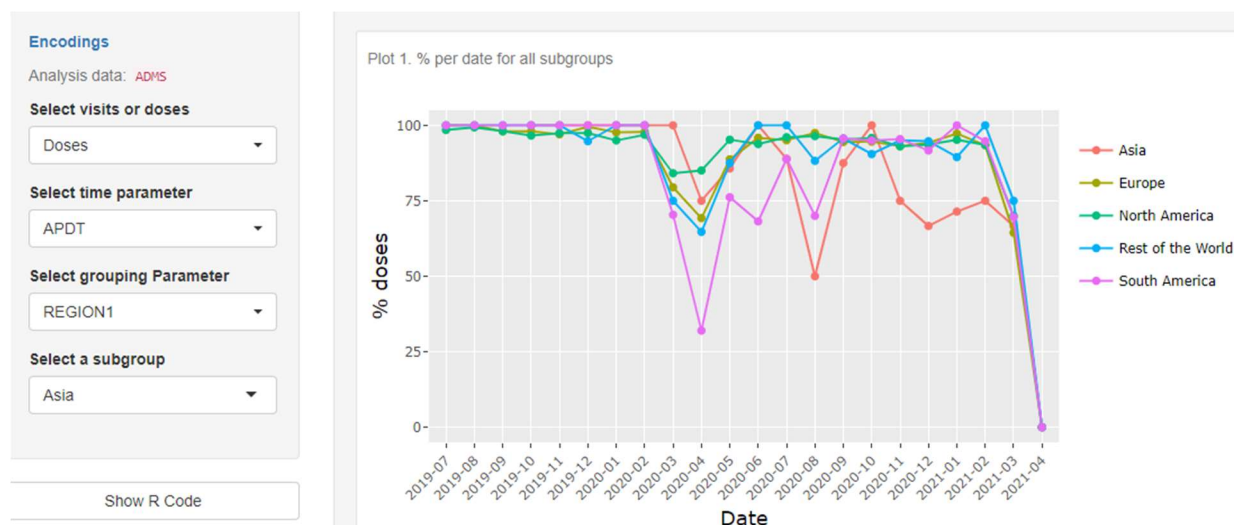**Figure 2: Reasons for missed doses by visit**

Figure 3: % of received doses by date

CONCLUSIONS

In this paper, I have described the way we dealt with data missing due to COVID-19. It was possible to implement this approach was because of the collaboration of the study stakeholders throughout the end-to-end process. The proper instructions were given to the sites, the analysis concepts were discussed beforehand with statisticians and scientists and sometimes instead of searching for a perfect way to deal with the data the team rather agreed on a rule of thumb. All this enabled the programmers in the end to provide easily interpretable results.

The dataset structure we have chosen has proved to be very flexible and extensible. For example, it enables the team to add further parameters if some other study endpoints would become of interest. The combination of ADT and PDT variables allows to subset the data for a certain interval, i.e. for the first wave of the pandemic.

This paper presents a simplified version of the code and I have omitted study specific details that are not relevant for the data structure. For example, after treatment discontinuation the study subjects were required to return for two more visits to collect the data so this needed to be accounted for when dealing with the visits schedule. Another complication was the data cleaning, since the inconsistencies between raw RAVE and non-CRF data resulted in the issues at the programming stage.

REFERENCES

[1] Guidance for Ongoing Studies Disrupted by COVID-19 Pandemic Version 1.0. Available at: https://www.cdisc.org/standards/therapeutic-areas/covid-19/interim-user-guide-covid-19

[2] Clinical Data Interchange Standards (CDISC). Study Data Tabulation Model Implementation Guide (SDTMIG) (version referenced – 3.2/Nov. 2013). Available at: https://www.cdisc.org/standards/foundational/sdtmig

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Yuliia Bahatska
Roche
yuliia.bahatska@roche.com