PharmaSUG China 2023 - Paper AP-119

Using Centralization Process to Streamline Programming Work on Multiple Late Phase Oncology Studies

Shuang Zhang, Yaoyu Zhu, Sanofi

ABSTRACT

In order to optimize and improve programming framework to manage multiple studies more efficiently, this paper introduces a new concept named "centralization process" in Sanofi biostatistics and programming, which is focused on the centralized programs and centralized documents, with the purpose of unifying the entire statistical analysis and programming as much as possible.

From the perspective of statistical programmer, centralized programs mean to have the one central program and one unique SDTM/ADaM metadata for multiple studies that are feasible to be centralized. Even though certain differences may exist across studies, centralized programs will consider all the study specificities. They are not only implemented at the level of SDTM and ADaM programming, but also implemented at the level of statistical analysis reports programming. From the perspective of statistician, centralized documents like master table shell also target to create and maintain one unique document for different studies annotated by different marks, taking into consideration of all the specificities across multiple indications or multiple cohorts within an indication.

Centralization process can not only realize the homogenization and simplification to the greatest extent in both biostatistics and programming, but also enable closer collaboration between statisticians and statistical programmers, as well as with other relevant departments like Data Management team.

This paper will be presented around the topic of centralization process, to analysis and summarize based on our current experience of development and implementation in multiple clinical trials, mainly describe from four perspectives on "Introduction", "Program Structure Development", "Process Management" and "Benefit and Risk", aiming to improve efficiency in programming and the collaboration between different function departments, and at the same time providing a reference of programming design for the biotech industry.

INTRODUCTION

Centralize program is aimed, to share one central program and one unique SDTM/ADaM metadata for multiple studies, and these programs will both include the common points across different studies and will also take into account on all the unique content of each study specificities. That means we will follow the same definition and convention as much as possible in different but similar studies. Usually, we will perform the centralization process under the same compound level, as defining the efficacy endpoint or safety endpoint within the compound will be mutually referential. When we implement centralization, because it is under the same compound level, the decision made by the entire compound level will always follow the same definition and convention in most cases. At the same time, centralization process can also simplify our programming process and optimize the consistency of statistical documents, because we only need to make changes in one set of central programs or documents when decisions are made for the compound, which could avoid multiple times update on each study to change the same content

PROGRAM STRUCTURE DEVELOPMENT

SCOPE

At the beginning of the establishment of the centralization process, we first need to measure the scope of the centralization we need to do, which needs to be discussed by all programmers at the entire compound level. Centralization's scope mainly includes two aspects:

1. The first aspect is the scope of the study, that is, we have to consider which studies should be

included in the centralize, which ones can be included, which ones are to be determined, and which ones do not conform to the centralize process, we all need to consider a few points:

- First of all, the first one is that the types of studies are the same, for example, they are all unify studies, and there is no out-source study. At the same time, the indications of the studies are also very similar, so we design the CRF, including metadata and Table shells, we all want to have as little difference as possible. This is also necessary. Compound level PPL is to be controlled macroscopically, and at the same time, it has a close relationship with the data management department. We try our best to provide the data management department with a study that can be referred to. When the CRF was initially established, the structure was as similar as possible.
- In addition, suppose a certain pivotal study has a regular milestone, so if a centralize is established, this pivotal study can be a good benchmark, and other studies can benefit from it, and the pivotal study can be prioritized Phase III project, or Phase II project with submission plan.
- Then, if some two or more studies start at almost the same time period, from the time of protocol available, to the multiple reviews of CRF, to the estimated time of First patient in, they are all almost at the same time , so at the level of statistical analysis, we will try our best to keep these items unified.
- Finally, when most of the study programmers work base in the same site, it will be more convenient and smooth to communicate, which is also one of the factors that need to be considered in the centralization process.
- Of course, if the study has some unavoidable particularities, if:

  Study A: A non-unify study (out-sourced) that does not comply with the CDISC standard. Currently, we directly implement ADS from raw data for this study.

  Study B: It is a unify study, which has been done by the Japan team and started very early, so the program structure of this study is very mature and stable, and all its programs are based on the older Based on the metadata version.

If these two studies have to be integrated into the centralization process, it may take more time and energy to change the structure of these two studies. Therefore, for this inevitable particularity, we do not need to use this type of study Put it in the scope of centralization.

2. The second aspect is the scope of the program, which programs need to be centralized. Ideally, all the programs from SDTM to ADaM and then to TLF should be centralized, but the specific decision should be made according to the different conditions of the compound, because for example, a compound that is more complicated, or a compound that incorporates a lot of studies, there are It may be easier to centralize on the dataset, but it is not easy to use a set of programs to run it on TLF, because the more studies are integrated, the more centralize needs to be adjusted on the formatting of the table shell, and, If you want to centralize the TLF program, then all milestone types in the study must also be taken into consideration. For example, for the milestone DMC, the report needs to have both an open report and a close report similar to these two, so Both need to be written in the centralize program of TLF.

RESPONSIBILITY

For responsibility, there are basically two ways to implement centralization:

1. The first is that a programmer is responsible for three or four datasets, but it must include all studies. For example, one person is in charge of ADAE, ADCM, and ADMH, so it is necessary to write all the studies, including the particularity of the studies, into these three programs. This kind of division of labor may be used more when doing pooling, or when there are fewer projects, programmers can see this division of labor. In addition, or, although there may be relatively more projects, the first patient in is almost at the same time, so one person is responsible for several data sets, and the battle line will not be too long.
2. The second division of labor is similar to that we usually do individual study. A programmer is only

responsible for all datasets under its own project, and this programmer needs to be responsible for identifying the particularity of its own project, and then add this part of special content to the centralize program, and be responsible for all subsequent updates and modifications.

DOCUMENT MANAGEMENT

Centralized metadata

Dataset Metadata needs to be integrated in a file in the centralization process, and different columns can be used to assign values to different studies, such as Table 1, for the variable of analysis AESI flag, if the derivation is different, you need to use a separate one row to display, and then the derivation rule is updated. Each study only selects the row that matches its own study derivation rule. All modifications are placed in a compound level folder. Then, when running the metadata, copy from the overall folder to the respective studies, and run the metadata program at the study level.

What we need to pay special attention to here is that at a specific time, for a specific study, such as metadata, or even a program, can be separated from the centralization process. For example, for Phase III projects, when preparing a submission package, you can maintain metadata under the study level for use when running define. Therefore, centralization is generally limited to conventional milestones and activities. Once the submission is completed stage, perhaps centralization may not be very applicable anymore.

Dataset Name	Relative Order	Variable Name	Variable Label	Type	Derivation	ANALYSIS_STUDY A	ANALYSIS_STUDY B	ANALYSIS_STUDY C	ANALYSIS_STUDY D	ANALYSIS_STUDY E	ANALYSIS_STUDY F	ANALYSIS_STUDY G	ANALYSIS_STUDY H
ADAE	2880	AAESIFL	Analysis AE of Special Interest Flag	text	[Y if SUPPAE.AESIIN = "Y" or AE.AECAT in ("PREGNANCY DATA") or (AE.AECAT in("OVERDOSE DATA") and AEOVSYMP="Y" and (AEOVNIMP="Y" or AEOVIMP="Y")]	Y			Y	Y		Y	Y
ADAE	2880	AAESIFL	Analysis AE of Special Interest Flag	text	[Y if SUPPAE.AESIIN = "Y" or AE.AECAT in ("PREGNANCY DATA") or (AE.AECAT in("OVERDOSE DATA") and AEOVSYMP="Y" and (AEOVNIMP="Y" or AEOVIMP="Y" or AEOVIMP2="Y")]		Y	Y			Y		

Table 1. Example of centralized metadata

Centralized master shell

Master Shell is also integrated in one file, and is distinguished by different colors. For example, in the following Display 1, orange represents indication of Skin, blue represents indication of Thoracic, green represents indication of Gastric, and gray represents indication of Head and neck , red represents indication of Lymphoma, and so on. In the master shell, after each content to be displayed, abbreviations and different colors are used to indicate which content needs to be statistically analyzed in which indication.

Conventions for indication/cohort specificities

If the table is applicable only for specific indications, it will be denoted by a letter highlighted in yellow in the title table. If an item into a table is only for one of the indications (e.g disease characteristics specificities), the text itself will be in the corresponding color:

1. S: Skin
2. T: Thoracic
3. G: Gastric
4. H: Head & neck
5. L: Lymphoma
6. P1

Prior anti-cancer treatments – Exposed population

Intent of prior anti-cancer therapy [a] [n(%)] STHGL
- No prior anti-cancer therapy
- Advanced
- Neoadjuvant
- Adjuvant
- Curative
- Radio-sensitizer THL
- Induction H
- Conditioning L
- Other
- Unknown

Time from last relapse/progression to first IMP administration (months) STHGL
- Number

Type of organ(s) involved [n(%)] STGH
- XXX
- XXX

FDG-avid histology [n(%)] L
- Number
- Yes
- No

FDG Bone Marrow Uptake [n(%)] L
- Number
- Yes
- No

Display 1. Example of centralized master shell

REPOSITORY STRUCTURE

For the structure of repositories, an overall compound level repository should be created, and all the centralized programs should be saved in this repository, so you can find all the centralized SDTM/ ADaM/ TLF programs under this repository, as well as all the macro programs also need to be created under this overall repository. At each individual study level repository, all the programs will be called from overall compound level repository, however the data will be produced at study level repository, like below **Figure 1**:

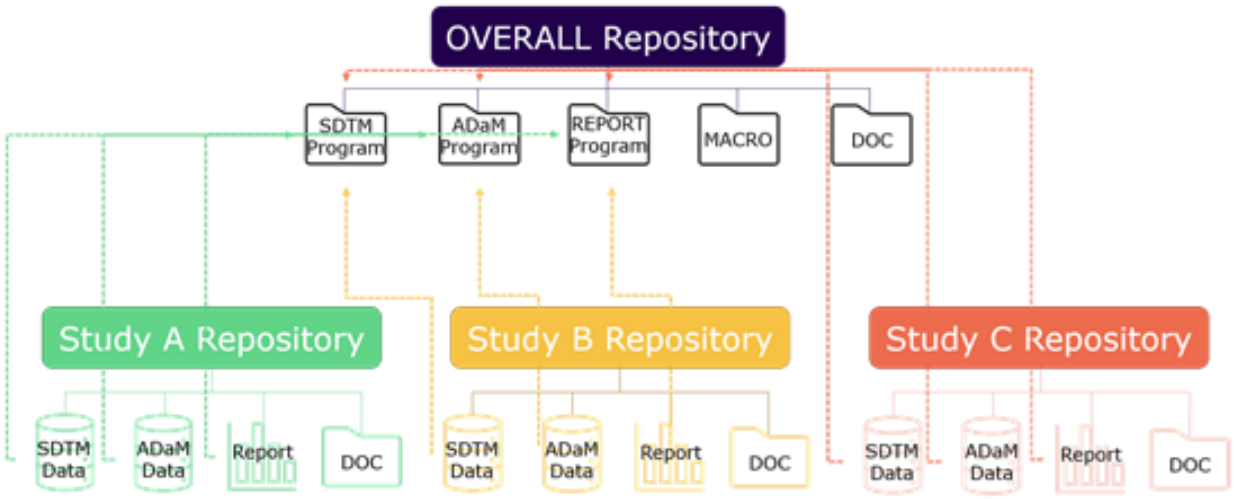

Figure 1. Repository structure workflow

PROGRAM EXECUTION STRATEGY

As this paper mentioned above, all the programs including macro programs, as well as the metadata or other related documents (If you have TLF metadata or some other similar file) should be stored under the overall compound level repository. While at the study level repository, we only need to use two programs, one is related to run SDTM/ ADaM (*run_sdtm.sas* or *run_adam.sas*) and the other is related to run reports (*run_tlf.sas*).

When we execute the run_XXX.sas program at the study level, it will call a macro under the overall repository, which is named "*run_sas_pgm.sas*", and then this macro will automatically execute all the

SDTM or ADaM or TLF programs, and will output all the datasets, logs, and reports under study level repository:

Figure 2. Program execution workflow

```
%macro run_sas_pgm(folder_name=, pgm_name=,study_log_folder=);
    options notes;

    proc printto log="../&study_log_folder./&pgm_name..log" NEW;
    run;
    ...
    %include &folder_name.(&pgm_name..sas);
    ...
    proc printto;
    run;

%mend run_sas_pgm;
```

STUDY SPECIFICITIES MANAGEMENT

This paper will introduce 3 levels on how to manage the study specificities:

1. Firstly if there is no specificities among multiple studies, that means the program will be the same to all the studies, there is no need to include any specific part in those programs. The centralized programs in this case will not include any specific part.
2. Second level is managing few specificities. This level is a very common. At this level, the specific part should be included in macro statement, that means we need to manage the study specific into a macro by using “IF” statement, like “IF the study equal to XXX then DO the specific conditions, otherwise using ‘ELSE DO’”.

```
%if &w_study. eq STUDY_A %then %do;
  if visitnum not in (110 210) then dosreduc = "XXXXX" ;
%end;
%else %if &w_study. eq STUDY_B or &w_study. eq STUDY_C %then %do;
  if upcase(extrt)='XXXXX' then do;
    if dosatg1n = . then dosreduc = "" ;
    else if visitnum not in (110 210) then dosreduc = "XXXXX" ;
  end;
  else if upcase(extrt)='XXXXX' then do;
    if dosatg1n = . then dosreduc = "" ;
    else if visitnum not in (110) then dosreduc = "XXXXX" ;
  end;
%end;
%else %do;
  if dosatg1n < prev_dose and prev_dose ne . then dosreduc = "XXXXX" ;
%end;
```

3. Third level is that everything is specific, that means we find it quite difficult to be centralized, this level is normally performed on some efficacy part, for example, on ADTTE dataset, due to the different efficacy endpoints across different studies, we put all the derivations in a new and separated macro program for each individual study, and we use a simple program to call those corresponding macros.

```
%if %upcase(&w_study.) eq STUDY_A %then %do;
  %m_adtte_studyA;
%end;
%else %if %upcase(&w_study.) eq STUDY_B %then %do;
  %m_adtte_studyB;
%end;
%else %if %upcase(&w_study.) eq STUDY_C %then %do;
  %m_adtte_studyC;
%end;
%else %if %upcase(&w_study.) eq STUDY_D %then %do;
  %m_adtte_studyD;
%end;
%else %if %upcase(&w_study.) eq STUDY_E %then %do;
  %m_adtte_studyE;
%end;
```

PROCESS MANAGEMENT

BASIC RULE ON MAINTENANCE

1. A standard document should be provided to all programmers that use centralization process, in which the good practices for analysis, reporting and data manipulation of clinical data should be agreed and followed at compound level.
2. It is recommended to clean and organize the centralization programs or documents at least once per year, because we need to keep the centralized programs more lisible and easy to understand. For example, if some codes appear to be applicable for all study, then we should delete the study specific condition (like "*%IF study equal to STUDY_A %THEN %DO*"), or we should delete some redundant or repetitive comments. Each study programmer should be responsible for this work to keep the centralized programs looking organized.

3. If any update (such as programs or metadata file) has been done under study level repository, it is very important to not forget to copy any updates from study level repository to overall compound level repository.

COMMUNICATION AND COLLABORATION

Communication is quite important in the centralization process, as there might be some conflicts when multiple programmers working in a same program, or some updates needed for some studies. Thus, to have a clear and in-time communication is very crucial.

It would be best to find a way to do better on posting message, sharing documents or even group chat. For example, creating a share folder or channel, which is dedicated in all the topics for the centralization activities, whatever questions, information updated, or any document saving could be issued and discussed in this platform.

When you are going to work on a specific program in overall compound level repository, it would be particularly important to inform other lead programmers firstly by grouping chat or other effective ways to tell them on which program you will work on as well as tell them once you have completed the working.

When you are going to modify or update a specific program in overall compound level repository, normally, you should just perform the modifications or updates only for your own study, it should not have any impact on other studies.

1. Firstly, you need to inform all other lead programmers on which centralized program you need to work on. This is to prevent just in case someone else is using it, which will cause the program to be opened in read-only mode and cause us to lose all the updates.
2. Then you could perform your works. You should only do the modifications for your own individual study, such as putting some macro conditions in the program. You need to make sure that you will not impact on all other studies.
3. Final step is communication with all other statistical programmers.
    - If you are not sure about whether the updates are applicable or not for other studies, it is highly recommended to inform all other study programmers and tell them what you have modified in which centralized program in order to let them consider whether this modification is needed or applicable for their study.
    - If you could confirm the update is just for your own study due to it is the study specific or related to the study design, then it is not necessary to emails, for example you could mention what you have done during the compound level project meeting.

The Centralization process requires team members whether within programmer team or among different function team like data management team, to cooperate with each other, coordinate their work, and achieve common goals. This kind of communication is not only conducive to problem solving, but also can improve mutual understanding and team cohesion. During the programming process, the members constantly communicate and discuss with others to achieve a unified goal. Such experience can also be extended to other fields to help team members communicate and coordinate better.

BENEFIT AND RISK

BENEFIT

The benefits are indeed obvious for centralization process. Except for what mentioned above that within the same compound, although the studies are different, it will be lots of similarities on the derivation rules,

on the statistical methods, and even from the beginning of the CRF design as well as the SAP writing, so centralization could help to keep the well consistency between multiple studies.

The second benefit is, because we share a set of central programs or documents, so when several studies have to change the same rule, only once modification would be done. Furthermore, by this way it can also limit the error rate, because updating a set of central programs is easier to control than having to update by each study's programmers separately. For example, if one of the study make some updates in the centralized program, then the programmers from other studies will evaluate whether these updates are applicable to their individual studies. Meanwhile, it is indeed equivalent to having one more programmer to review the program with the updates, that means set of central programs will be reviewed by several programmers, so that the errors in the program can be greatly reduced.

In addition to above mentioned, it is also benefit for a new study, if using centralization, the programmer only needs to add the study specificities of the new study, and here is no need to rewrite the common places which are the same with other studies. It is much faster to adjust an existing program than to create a completely new program. Centralization process is very friendly to the program development of the new study and very efficient to set up the new study.

Besides, another benefit is that if compound may have the possibility or desire of pooling activity in the future, such as ISS and ISE, then centralization process may be a basis for the pooling, because centralization makes studies keep a high level of consistency, which results in the process to be efficient.

Finally, the common benefit is that centralization is a manifestation of team spirit and cohesion, because many people maintain a set of central programs, and everyone in the compound will discuss a topic together, many ideas will be contributed. This will make everyone's communication and mutual help more abundant than when doing studies alone.

RISK

Centralization process is not all about advantages and no disadvantages, as we all know, everything has two sides. Centralization process also has some limitations.

Firstly during the initialization period of centralization process, it needs programmers to have comprehensive review of all study-related documents of multiple studies when you select the way like one programmer is responsible for multiple datasets, they need to be familiar with all the studies that are included in centralization.

Secondly, as we mentioned in above, in centralized programs, several programmers maintain a same set of programs together, or share one metadata, which is also inconvenient, because we cannot modify the same program at the same time, which is likely to overwrite the place that others are modifying. So, you need to coordinate with each other. If others are using it, you may have to wait for others to use it before modifying your place.

Next, regarding to centralized program and metadata, they are always in a dynamic status, that is, they cannot be completely fixed, because in the process of study ongoing, the program may change when Statistical Analysis Plan or table shell is changed, so not only one study is being changed, but other studies may also be changed at the same time, which causes our centralized program to be in a state of being modified for a long time, and the milestone time of each study is inconsistent, so it is difficult for us to have the opportunity to say, for example at a fixed time, we will release a version, because it is always changing. Moreover, sometimes, in the centralized program, because the programmers of the respective projects only modify the changes of their own studies, but one programmer can only check whether there is any problem with the data running under his/her own study, but for other studies, he/she cannot run other programmer's study data, so the programmer can only judge by looking at the program manually.

In addition, the readability of few centralized programs is not very high, because with the increase of centralization studies, there are more and more places to deal with the particularity of study, so this will lead to, for example, ADSL, this program will become more and more long.

Finally, with our use of centralization, it may have some limitations on our independent thinking, that is, everyone may rely on the centralized program and lack to think about why this variable should be derived in this way, especially for those same variables.

CONCLUSION

By following the guidelines of this paper, we are aimed to improve efficiency in programming and to provide a reference of programming design for the biotech industry. To sum up, readers can consider whether the centralization process is applicable to your compound in these ways:

1. First, we can see if there is any plan to do pooling at the compound level in the future. Also, it depends on the duration of your study and the date of first-patient-in of different studies. Because the front line of some studies like oncology is relatively long, but if it is like other therapy area, if the duration of the study is not long, or if the first-patient-in time of several studies are far away, then it is not suitable to use centralization process.
2. Another point is to look at the same compound, whether the trial designs of different studies are similar, which may determine whether the design of CRF can learn from each other, as well as how many studies are similar. For example, if finally, only two or three studies are similar, then in this case, if you spend time and energy on centralization, you may have to evaluate whether it is worth it.
3. Also, if it is centralized, please think about is it possible to maintain the whole process at your compound level? For example, are the milestones of different studies uneven, or are there many additional ad-hoc activities for each individual study? Maybe these ad-hocs are done only once at study level, while the program will not be used in the future and will not be integrated into the centralization. In this case it is also not suitable for the centralization process.
4. The last one is to consider the location of the team member. The more important reason for a compound that centralization can be used relatively smoothly, is depending on the location of programmers that using centralization process, if most of programmers are located in the same site, then it would be very convenient for everyone to communicate, for example for multinational companies, there would be no time difference if team members are from same site. If there is any misunderstanding, everyone can discuss in a quick manner. This is quite critical because for centralization, the more important point is communication. Any new information can be quickly and efficiently shared with the other programmers.

ACKNOWLEDGMENTS

We are sincerely acknowledged Gawain Gao and Sufang Huang for providing constructive comments with helpful discussion on this topic. We would also like to thank all the oncology colleagues who are developing or using the centralization process in Sanofi.

CONTACT INFORMATION

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

Shuang Zhang
Sanofi
6/F, HP Building, No.112, Jianguo Road, Chaoyang District, Beijing, 100022
Shuang2.Zhang@sanofi.com

Yaoyu Zhu
Sanofi
Chengdu Yintai Centre Tower 3, Tianfu Dadao Beiduan 1199, Wuhou District
Chengdu, Sichuan 610041
Yaoyu.Zhu@sanofi.com