Paper SA05

Navigating first Data Submission: Pierre Fabre's Journey and Insights into Data Submission Challenges with the NMPA

Benoît Sansas, Laboratoires Pierre Fabre, Toulouse, France
Mélanie Groc, Laboratoires Pierre Fabre, Toulouse, France
Romain Raymond, Laboratoires Pierre Fabre, Toulouse, France

ABSTRACT

Pierre Fabre conducted clinical trials in mainland China to support submission dossiers in 2024. These were the company's first National Medical Products Administration (NMPA) data submissions. The first data submission to the NMPA is not an easy and straightforward journey as guidance is only available in Chinese, and finding how to best fit to NMPA requirements can be challenging as it requires: getting the database translated into Chinese, submitting executable programming codes, and choosing the best submission strategy approach. Communication with the sponsor Chinese team and the authorities, and support from Contract Research Organization (CRO) local teams, notably for the regulatory and programming aspects, were very important in preparing this data submission. This paper outlines challenges met during the data submission preparation, the options we considered, and shares feedback on the solution we ended up choosing.

INTRODUCTION

Pierre Fabre conducted worldwide pivotal trials for oncology products in several indications, but these trials were not conducted in mainland China and the proportion of Asian participants was very low. These trials supported regulatory approvals from the FDA and EMA. In 2021, Pierre Fabre initiated specific clinical trials in mainland China to access the Chinese market, leading to the first Pierre Fabre data submission to the NMPA in August 2024. At the beginning, data submission to the NMPA was not anticipated to be a substantial challenge, but several questions were raised and the need to connect with several other teams quickly became critical. This paper summarizes the journey and insights into data submission challenges with the NMPA.

CLINICAL DATA PACKAGE

KEY ELEMENTS OF CLINICAL DATA PACKAGE

The *Guideline on the Submission of Clinical Trial Data* is issued by the NMPA and is only available in Chinese on their website. An English translation was provided by the Pierre Fabre Affiliate. There are six key elements of a clinical data package as illustrated in Figure 1:

- Raw Database in XPT format (version 5 or above) should generally contain source data collected directly from case report forms (CRFs) as well as external sources. It may contain few derived variables but missing data in raw database should not be imputed. If the Sponsor submits clinical trial data according to CDISC standards, Study Data Tabulation Model (SDTM) are expected.
- Analysis Database in XPT format (version 5 or above), is a database derived for statistical analyses and is used to produce and support statistical analysis results in Clinical Study Report (CSR). If the Sponsor submits clinical trial data according to CDISC standards, Analysis Data Model (ADaM) are expected.
- For both Raw and Analysis databases, a Data Definition file in XML or PDF format should be submitted. The Data Definition file is used to describe the submitted data, and should at least contain the name, label and basic structure of each dataset in the submitted database. The name, label, type of each variable and derivation process of each derived variable in each dataset is also expected. Define.xml file covers all requirements from the guideline.
- For both Raw and Analysis databases, a data reviewer's guide in PDF format should be submitted. Data reviewer's guide provides information in addition to what are presented in Data Definition file, including but not limited to: instructions on the use of the submitted data, relationships between the study results and the data, instructions on the use of submitted programming codes and, encoding used for the datasets. Data reviewer's guides are not intended to replace Data Definition file, but to help reviewers more accurately and

efficiently understand and use the submitted database. Both Study Data Reviewers Guides (SDRG) and Analysis Data Reviewer's Guide (ADRG) files cover all requirements from the guideline.

- Annotated Case Report Form (aCRF) in pdf format is expected to be provided to support raw database. aCRF from all studies included in the submission were included.
- Programming codes in TXT format are expected for derived variables of analysis datasets (ADaM SAS® programs) and for the analysis results of the efficacy endpoints. Programming codes submitted in a submission package should be readable (including comments), understandable, executable, and do not include external program calls.

It must be noted that even if the NMPA doesn't require to submit datasets under CDISC standards, it seems highly recommended to adhere to CDISC standards.

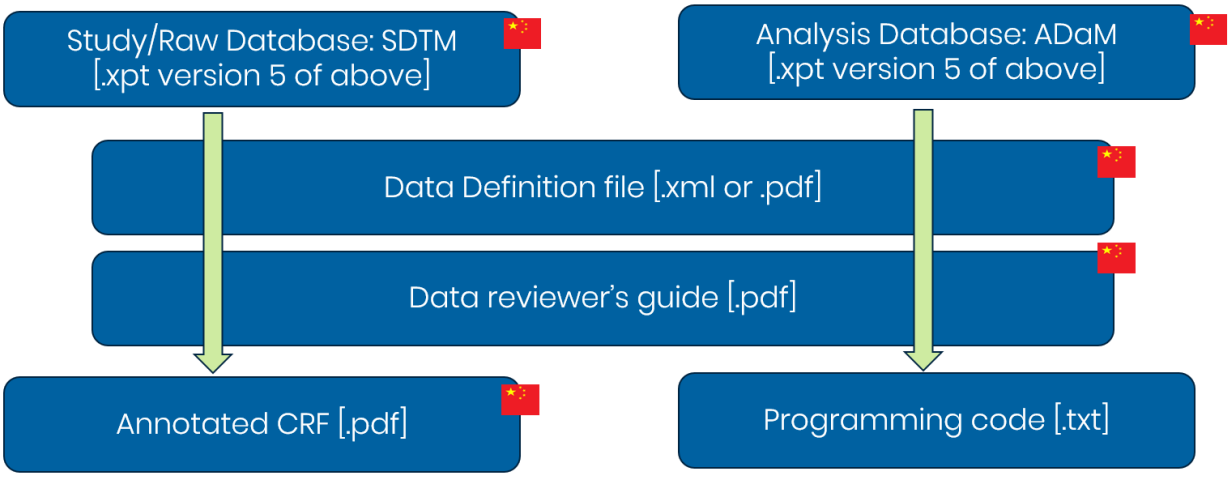

Figure 1: Six key elements of clinical data package
ADaM: Analysis Data Model; CRF; Case Report Form; SDTM: Study Data Tabulation Model

Some other elements of the clinical data package were important to consider. For instance, several rules included in the Guideline are the same as the ones requested by the FDA: e.g. maximum length of a variable name is 8 bytes, dataset and variable labels should not exceed 40 bytes in length, variable length should be set not to exceed 200 bytes.
If any submitted dataset exceeds 4 Gigabytes, then dataset must be split. Sponsor can submit the split dataset, detailed rules of splitting and detailed steps of merging it back should be specified in data reviewer's guide to ensure that the reviewer can accurately recreate the original dataset prior to splitting.
Finaly, since the Guideline does not provide the Sponsors with explicit recommendations about the efficacy endpoint programs, Pierre Fabre decided to submit all ADaM datasets programs, along with programs for the primary efficacy endpoint and key secondary efficacy endpoints.

TRANSLATION OF CLINICAL DATA PACKAGE

Among the elements of the clinical data package, a translation in Chinese is required for all elements except for the programming code. The programming team from our Global CRO in China was responsible of the translation of the entire clinical data package.

For SDTM and ADaM datasets, the following content should be in Chinese at a minimum:

- Datasets labels and variables labels (Figure 2)
- Adverse events terms and related coding (Figure 3)
- Generic name of concomitant medications
- Medical history reported in CSR and other documents

The process used to translate the verbatim terms was to first create an Excel file with all the unique verbatims collected, then a translation in Chinese was proposed by Global CRO vendor. The translation was reviewed by Pierre Fabre Affiliate to ensure consistency with other study documents (e.g. CSR). The validated list was uploaded and was used in replacement of English terms.

研究中心标识符	年龄	年龄单位	年龄组1	年龄组1(N)	性别	种族	预筛选集人群标识	筛选集人群标识
15601	68	YEARS	65 - 74	2	M	ASIAN	N	Y
15601	55	YEARS	<65	1	M	ASIAN	N	Y
15601	67	YEARS	65 - 74	2	F	ASIAN	Y	Y
15601	71	YEARS	65 - 74	2	F	ASIAN	Y	Y
15601	30	YEARS	<65	1	M	ASIAN	N	Y
15601	55	YEARS	<65	1	M	ASIAN	N	Y
15601	67	YEARS	65 - 74	2	M	ASIAN	N	Y
15601	27	YEARS	<65	1	F	ASIAN	N	Y
15601	59	YEARS	<65	1	M	ASIAN	N	Y

Figure 2: Example of ADSL translation with variable label as default view

AETERM	AEDECOD	AEBODSYS	AEBDSYCD	AELLT	AELLTCD
丙氨酸氨基转移酶升高	丙氨酸氨基转移酶升高	各类检查	10022891	丙氨酸氨基转移酶升高	10001551
关节痛	关节痛	各种肌肉骨骼及结缔组织疾病	10028395	关节痛	10003239
痤疮样皮疹	痤疮样皮炎	皮肤及皮下组织类疾病	10040785	痤疮样皮疹	10037847
腹泻	腹泻	胃肠系统疾病	10017947	腹泻	10012727
腹泻	腹泻	胃肠系统疾病	10017947	腹泻	10012727
肝痛	肝痛	肝胆系统疾病	10019805	肝痛	10019705
不适	难受	全身性疾病及给药部位各种反应	10018065	难受	10025482
色素痣	黑色素痣	良性、恶性及性质不明的肿瘤(包括囊状和息肉状)	10029104	色素痣	10062799
丙氨酸氨基转移酶升高	丙氨酸氨基转移酶升高	各类检查	10022891	丙氨酸氨基转移酶升高	10001551
贫血	贫血	血液及淋巴系统疾病	10005329	贫血	10002272

Figure 3: Example of ADAE translation with verbatim and coded terms in Chinese

For SDTM and ADaM Define.xml, the following content should be translated in Chinese at a minimum:

- Description/label and specification of each dataset in the database (Figure 4)
- Description/label and derivation of variables in dataset (Figure 5)
- Values or codelists of efficacy indicators (Figure 6).

SDTM and ADaM specification in Excel files were first translated in Chinese, and then Define.xml was recreated. As per programming code, only the primary endpoint and key secondary endpoints related values/codelists were translated.

数据集	描述	类	结构	目的	按键	文献资料	位置
ADSL	受试者水平分析数据集	SUBJECT LEVEL ANALYSIS DATASET	每名受试者一条记录	Analysis	USUBJID		adsl.xpt
ADAE	不良事件分析数据集	OCCURRENCE DATA STRUCTURE	每名受试者每次事件发生一条记录	Analysis	USUBJID, AETERM, AEDECOD, ASTDT, AENDT	仅纳入 SAFFL= "Y" 的患者	adae.xpt
ADAESI	特别关注的不良事件分析集	OCCURRENCE DATA STRUCTURE	每名受试者每次事件发生一条记录	Analysis	USUBJID, AETERM, AEDECOD, ASTDT, AENDT	仅纳入 SAFFL= "Y" 的患者	adaesi.xpt
ADCAN	抗癌全身治疗分析数据集	OCCURRENCE DATA STRUCTURE	每名受试者每次事件一条记录	Analysis	USUBJID, CMTRT, CMDECOD, ASTDT, AENDT, CMINDC, CMROUTE, CMCLAS	仅纳入 SAFFL= "Y" 的患者	adcan.xpt
ADCM	合并用药分析数据集	OCCURRENCE DATA STRUCTURE	每名受试者每次事件一条记录	Analysis	USUBJID, CMTRT, CMDECOD, ASTDT, AENDT, CMINDC, CMROUTE, CMCLAS, CMDOSE, CMDOSU	仅纳入 SAFFL= "Y" 的患者	adcm.xpt
ADDA	药物暴露分析数据集	ADAM OTHER	每名受试者每个给药记录一条记录	Analysis	USUBJID, EXTRT, ASTDT, EXTPT	仅纳入 SAFFL= "Y" 的患者	adda.xpt

Figure 4: Example of ADaM Define.xml with description, structure, and documentation translated

AVALC VLM		分析值(C)	text	59		Derived 在 VLM 中描述每个参数的来源/推导 见VLM
	PARAMCD = "BRESPC" (经确认最佳缓解)	经确认最佳缓解	text	59	经确认最佳缓解的分析值(C) [6 术语]	Derived 1）设置为 BRCNFL = "Y" 时的 ADTRL.AVALC (AVAL 数值)。 如果患者在 ADTRL 中没有 BRCNFL = "Y" 的记录，或者 BRCNFL = "Y" 的记录为 AVAL= -1，则考虑 2）如果患者有 i）AVAL = 4 或 5 的记录，并且 ii）ADY >=43，并且 iii）PRCANFL = "Y"，并且 iv）之前没有 AVAL = -1 的记录，则设置 AVALC = '疾病稳定' (AVAL = 3)。 否则 如果患者在 ADTRL 中没有符合上述标准 (1) (2) 的记录，但患者在 ADTRL 中，则设置 AVALC = "不可评估" (AVAL = -99)。 如果患者在 ADTRL 中没有记录，然后设置为 "不可评估" (AVAL = -99)。

Figure 5: Example of Best Overall Response derivation with the label, codelist and description translated

Analysis Value (C) for Best Response Confirmed

Permitted Value (Code)
COMPLETE RESPONSE
NON-CR/NON-PD
NOT EVALUABLE
PARTIAL RESPONSE
PROGRESSIVE DISEASE
STABLE DISEASE

经确认最佳缓解的分析值(C)

允许值 (代码)
完全缓解
非CR/非PD
不可评估
部分缓解
疾病进展
疾病稳定

Figure 6: Example of Best Overall Response codelist translated

Both the SDRG and ADRG must be fully translated into Chinese. When available, the Complex Algorithm document, used in complement to the data reviewer's guide to detail the complex derivations, was also translated into Chinese. For the aCRF, the description of questions designed to collect data and values or codelists of efficacy indicators should be translated at a minimum. The Blank CRF in word version was first translated and exported in PDF, then SDTM annotations in Chinese were added. Figure 7 shows the translation of all aCRF questions, of some SDTM mapping information and datasets labels.

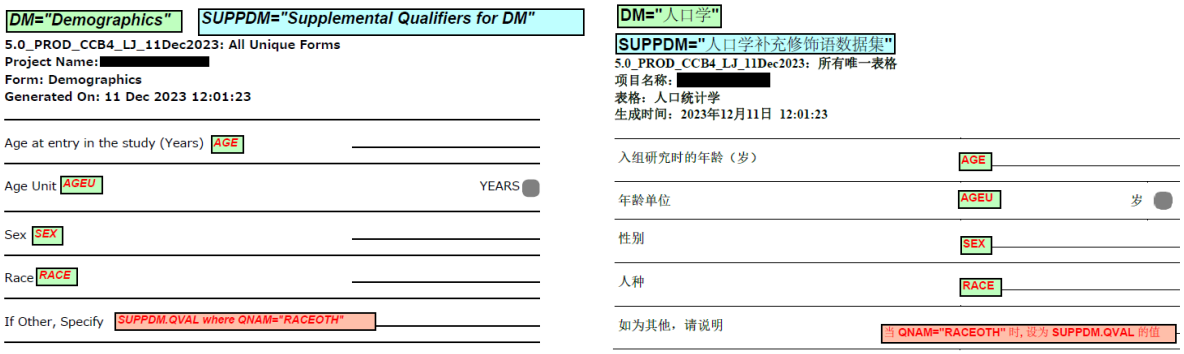

Figure 7: Example of Demographic eCRF page translated in Chinese

CENTER FOR DRUG EVALUATION (CDE) OF NMPA CONSULTATION

It is strongly recommended to consult the CDE during a pre-Investigation New Drug Application (pre-IND) meeting to ensure the clinical data package matches their expectations. Table 1 presents the data package strategy proposed and endorsed by the CDE for one of our submissions. There are no general rules or guidelines, therefore the same proposal could be rejected in a different context, highlighting the importance of a pre-IND meeting to discuss this topic.

The data package strategy was developed in very close collaboration with Pierre Fabre team in China and with Regulatory teams. The local safety lead-in trial was not included in Chinese Version as that study only enrolled 3 participants at a dosing schedule different from the pivotal trial. A translated version of the CSR was then considered sufficient. The Integrated Safety Pool was not considered for translation as the safety data came from the global worldwide trial and local pivotal trial already was translated in Chinese.

Database	English Version	Chinese Version
Global worldwide trial	Yes	Yes
Local safety lead-in trial (different dosage than Global and Local trials)	Yes	No
Local pivotal trial in China	Yes	Yes
Integrated Safety Pool	Yes	No

Table 1: Data package strategy for translation

MAIN CHALLENGES

Several challenges were faced during the preparation of data submission packages, the main ones were:

- The majority of NMPA guidance's are available only in Chinese, close work with the Affiliate team was needed for translated versions.
- Reencoding of study datasets using UTF-8 was necessary for some studies to accommodate Chinese characters, which sometimes require more than one byte. A SAS® PROC COMPARE was done after re-encoding to ensure there were no differences in the new datasets. SAS® code used for transcoding in UTF-8 is presented below and should be executed from a UTF-8 encoding SAS® session.

```
* Conversion from WLATIN to UTF-8;
libname &lib_u8 "&path_lib_u8.";
data &lib_u8..&dataset. (label="&DatasetLabel" encoding='utf-8');
       set &lib_wlt1..&dataset.;
run;
```

- Due to reencoding from wlatin1 to utf8, the length of some variables increased to more than 200 characters (e.g.: Protocol Deviation term in DV/ADDV datasets). Pierre Fabre decided to use XPT version 8 as allowed per guidance to accommodate to longer lengths. Two other options were considered to keep XPT version 5, but these options changed the original datasets and were not chosen:
    - The first one was to shorten the deviation terms to be under 200 characters by changing the special characters to normal ones and removing some spaces.
    - The second one was to use a SAS® function like tranwrd or a regular expression to convert special characters in all observations into normal characters and to remove additional spaces.
- The translation in Chinese can require some adjustments:
    - After translation, some variable labels exceeded 40 bytes in length and were required to be reworded.
    - The aCRF or Statistical Analysis Plan (SAP) pagination could change during translation, meaning the page number reference in Define.xml should be adjusted accordingly.
- Translation consistency across all documents was needed, including databases. For example, an AE term is expected to be translated in the same way as in CSR, Submission module 2.7.4, and AE/ADAE datasets. As different vendors were used for study documentation translation and databases translation, additional consistency checks were conducted by CRO programming team and Pierre Fabre affiliate.
- Programming codes submitted should be readable, understandable, and executable. They should not include external program calls and avoid using external macro programs. It was needed to rework programming codes and to include all external macro programs used at the beginning of each programming code for the local study and integrated safety pool. These updated programming codes were then re-executed to ensure consistency with the original results.

INTEGRATION IN MODULE 5

Another question was the inclusion of Chinese translated databases in module 5. Two options were considered:

- Create subfolders for the Chinese translated databases in the 'datasets' folder (Folder Level 6 in module 5) and in the 'sdtm' folder. This option is proposed by the PMDA (Pharmaceuticals and Medical Devices Agency) in a recent guidance issued in April 2024.
- Create a new « Study » folder for each Chinese translated databases. This option was endorsed by the Pierre Fabre Team (Figure 8) before the release of the PMDA guidance.

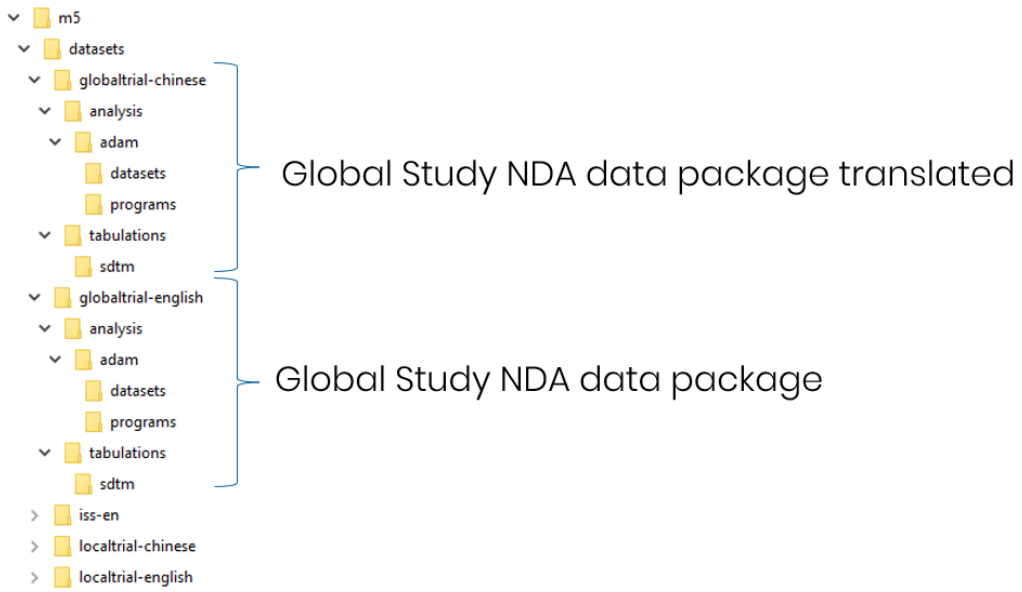

Figure 8: Pierre Fabre decision to create a new « Study » folder for Chinese translated databases (NDA: New Drug Application)

TEAMWORK

Teamwork between all the different stakeholders (Biometry, Regulatory, Clinical), from both Pierre Fabre and the Global CRO, as well as the Headquarters and Affiliate was key in the success of the data submission. Figure 9 illustrates the workflow and different stakeholders' major contribution:

- The first step was to agree on the submission strategy about the need to create an integrated efficacy pool and/or integrated safety pool. Once the strategy was defined, the determination of datasets to be translated in Chinese was discussed. The pre-NDA meeting was very important to ensure CDE alignment on the global strategy and translation. Regulatory teams from both Pierre Fabre affiliate and CRO were key contributors in these discussions because they have knowledge of CDE expectations.
- Before starting the translation activities, the data preparation was done quite in advanced to ensure that all datasets were in UTF-8 encoding, to gather all documentations needed, and to provide a translated version of other study documents: translated protocol was needed to ensure a consistency with translated aCRF, and translated SAP and CSR were needed for consistency with SDTM and ADAM.
- During the translation process, an additional step was performed by Pierre Fabre affiliate medical and CRO programming team to ensure overall consistency of translated datasets with other study documents.
- Once translated databases were delivered, both Biometry and Regulatory teams from Pierre Fabre collaborated to prepare module 5 datasets folder.

In addition, effective planning and anticipation of activities were essential to ensure a quick submission after the completion of the local study that supported the submission. It was also crucial to ensure a smooth collaboration across stakeholders.

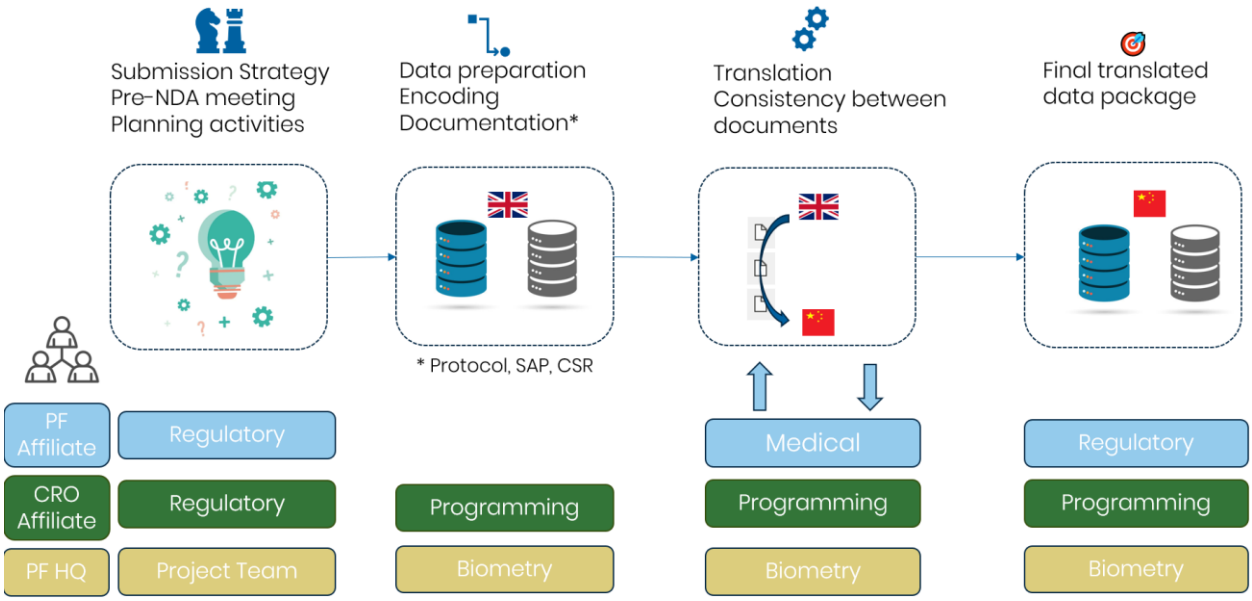

Figure 9: Team significant contribution in the process of data package preparation (PF: Pierre Fabre, HQ: Head Quarters)

CONCLUSION

Navigating the first data submission to the NMPA was a challenging journey for Pierre Fabre. This paper provides a comprehensive overview of Pierre Fabre's journey in submitting clinical data to the NMPA, highlighting the crucial elements, challenges, and strategies employed to ensure a successful submission. Effective communication with authorities, Sponsor and CRO local teams, and strong cross-functional teamwork were essential to this success. The insights gained from this experience will be invaluable for future submissions and can serve as a guide for other organizations facing similar challenges.

REFERENCES

1 Revision of Technical Conformance Guide on Electronic Study Data Submissions, PMDA, 000267935.pdf (pmda.go.jp)

CONTACT INFORMATION

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

Benoît Sansas
Laboratoires Pierre Fabre
3 avenue Hubert Curien – 31035 Toulouse
Email: benoit.sansas@pierre-fabre.com