PharmaSUG China 2025 - Paper DS-163

Bridging the Gap: How to Prepare Analyses and Electronic Submissions for Companion Diagnostics (CDx) Bridging Studies in China

Ben Gao, MSD China, Shanghai
Wei Wang, MSD China, Beijing
Nicole Zhang, MSD China, Shanghai

ABSTRACT

In the evolving landscape of clinical diagnostics, bridging studies play a pivotal role in transitioning from Clinical Trial Assays (CTA) to Companion Diagnostics (CDx). Bridging studies, which retest preserved samples with CDx to evaluate the agreement between CDx and CTA, are designed to bridge the clinical data from CTA to CDx and to evaluate the drug efficacy in CDx intent-to-test population. Primary analyses include concordance analysis (Positive Percent Agreement, Negative Percent Agreement, and Overall Percent Agreement) and efficacy analysis in CDx intent-to-test population.

This paper focuses on the analysis dataset structure essential for supporting concordance analyses and efficacy analyses in CDx bridging studies. Key variables include CTA and CDx test results, population flags and efficacy outcomes, all critical for robust statistical evaluation. We highlight regulatory requirements for source and analysis datasets, data definition files, and file formats from the China Center for Medical Device Evaluation's (CMDE) guidelines for submitting clinical trial data for CDx, i.e., Guideline on Clinical Trial Data Submission Requirements for In-vitro Diagnostic Reagents Registration and Review (2021). The types of data to submit and strategies for organizing and formatting various source data as required by CMDE are summarized. Additionally, we share our approach to create data definition files. Given the diverse data sources involved, we also suggest validation steps to ensure data quality and consistency between analysis and source datasets. Meanwhile, we present other required deliverables according to CMDE guidance (Guideline on Registration Review of Clinical Trials of Original Companion Diagnostic Reagents Co-development with Anti-Tumor Drugs, 2022), such as line data, emphasizing the key data elements that must be included in the analysis datasets to fulfill CMDE's requirements. This paper offers insights into the preparation and submission processes for China Premarket Application (PMA), which is an important element for precision medicine with the increasing demand of the drug-diagnostic co-development model.

INTRODUCTION

The area of precision medicine is experiencing rapid growth driven by technological advancements, expanded knowledge of genetics, and demanded shift towards personalized healthcare. The goal of precision medicine is to provide more effective and targeted therapies, which is particularly prominent in fields such as oncology where treatments are required to be personalized based on the biomarker or genetic profiling. Companion diagnostics (CDx) are in vitro diagnostic (IVD) devices or tests that provide essential information for both safe and effective use of specific therapies. They are designed to identify patients who are most likely to benefit from the targeted therapy. For example, the use of the HER2 test in breast cancer helps to determine whether a patient is most likely to benefit from trastuzumab. CDx are typically developed alongside the therapeutic product and must receive regulatory approval before they can be used in conjunction with the therapy.

Ideally, clinical trials use final market ready CDx for patient enrollment, and the safety and effectiveness assessment of CDx can be carried out simultaneously with the therapeutic product, which leads to more reliable data to demonstrate the drug's safety and effectiveness and streamline the regulatory process. However, the development of CDx may lag behind drug development, so CDx may not be available for patient enrollment at the time of the clinical trial. In such scenarios, a clinical trial assay (CTA) is used for patient enrollment, and specimen will be preserved for retesting by future market ready CDx to prepare regulatory registration of CDx. Clinical trials using CTA for enrollment cannot directly demonstrate the safety and effectiveness of the CDx in conjunction with the drug. Consequently, a bridging study is utilized to bridge clinical data of CTA to CDx. In the bridging study, preserved samples from subjects enrolled or screened by CTA are retested by CDx. The purpose of the bridging study is to evaluate the

consistency between CDx and CTA, thereby assessing the treatment effect in CDx intent-to-test (ITT) population. The CDx ITT population for the bridging study may be defined as subjects from one or several clinical trials of the associated therapeutic product. In some instances, there may be not enough negative samples from the original drug clinical trials, then samples from non-drug clinical trials may be included.

The Center for Medical Device Evaluation (CMDE), affiliated to National Medical Products Administration (NMPA), is responsible for the evaluation and approval of medical devices including CDx in China. The CMDE has released comprehensive guidelines that outline the requirements for the development, evaluation, and approval of CDx and ensure the safety and effectiveness of CDx. The "*Guideline on Clinical Trial Data Submission Requirements for In-vitro Diagnostic Reagents Registration and Review*" (2021) outlines the specific requirements for submitting clinical trial data for CDx. Additionally, the "*Guideline on Registration Review of Clinical Trials of Original Companion Diagnostic Reagents Co-development with Anti-Tumor Drugs*" (2022) provides further directives on the co-development of CDx with anti-tumor drugs. These regulations emphasize the importance of robust statistical evaluation, data quality, and consistency between analysis and source datasets.

This paper aims to provide insights into the preparation and submission process for CDx Premarket Application (PMA) in China. It focuses on preparing and submitting data packages for CDx bridging studies where the efficacy endpoint is binary. It covers the streamlined dataset structures required for concordance and efficacy analyses.

CDX CONCORDANCE AND EFFICACY ANALYSES

In bridging studies, concordance analysis is conducted to evaluate the contingency between CTA and CDx. CTA positive samples are retested to estimate positive percent agreement (PPA), and CTA negative samples are retested to estimate negative percent agreement (NPA). A two-way contingency table are usually reported to provide clear information for estimating PPA and NPA. Suppose that *a* represents the number of subjects with positive results for both CTA and CDx, *b* represents the number of CTA positive but CDx negative subjects, *c* represents the number of CTA negative but CDx positive subjects, and *d* represents the number of subjects with negative results for both CTA and CDx, the following two-way contingency table can be provided, see Table 1.

Table 1 CTA vs CDx results

	CTA results			
		Positive	Negative	Total
CDx results	Positive	a	c	a+c
	Negative	b	d	b+d
	Total	a+b	c+d	a+b+c+d

The PPA and NPA are estimated as

$$PPA = \frac{a}{a+b}$$

$$NPA = \frac{d}{c+d}$$

The overall percent agreement (OPA) is estimated as

$$OPA = \frac{a+d}{a+b+c+d}$$

An ideal assessment of PPA, NPA and OPA occurs when samples from all subjects enrolled in clinical trials can be retested by CDx. Good agreement demonstrated by estimates of PPA, NPA and OPA ensures the similarity of clinical efficacy observed in subjects enrolled by CTA to that by CDx. However,

challenges in sample retainment are commonly observed in bridging studies. Low sample retainment rate and poor preserved sample quality render high rate of missing data in CDx results, and consequently lead to uncertainty in performance evaluation of CDx. In such instances, a sensitivity analysis can be carried out to evaluate the impact of missing data by multiple imputations.

Besides concordance analysis in bridging studies, efficacy analysis is the other critical aspect which assesses the efficacy in targeted population identified by CDx. Based on concordance between CTA and CDx and the observed efficacy data based on CTA in clinical trials, the clinical efficacy of CDx positive subjects, denoted by δ_{CDx+}, is a weighted average,

$$\delta_{CDx+} = \Pr(\text{CTA} + |\text{CDx} +) * \delta_{CDx+\&CTA+} + (1 - \Pr(\text{CTA} + |\text{CDx} +)) * \delta_{CDx+\&CTA-}$$

where $\delta_{CDx+\&CTA+}$ denotes the clinical efficacy for both CDx and CTA positive subjects, $\delta_{CDx+\&CTA-}$ denotes the clinical efficacy for CDx positive and CTA negative subjects. The weight $\Pr(\text{CTA} + |\text{CDx} +)$ can be estimated by the concordance between CTA and CDx. $\delta_{CDx+\&CTA+}$ and $\delta_{CDx+\&CTA-}$ are estimated based on the efficacy observed in the clinical trial. In enrichment trials, $\delta_{CDx+\&CTA-}$ cannot be estimated directly based on observed data. A sensitivity analysis is conducted to estimate the range of the efficacy of CDx positive and CTA negative subjects by assuming $\delta_{CDx+\&CTA-} = c * \delta_{CDx+\&CTA+}$, where c ranges from 0 to 1 with 0 denoting the worst scenario of no efficacy and 1 denoting the best scenario of the same efficacy as that of CDx and CTA double positive subjects.

CDX ANALYSIS DATASET

To support concordance and efficacy analyses in a CDx bridging study, a comprehensive analysis dataset should be built, incorporating essential variables such as CTA and CDx test results, as well as primary efficacy outcomes like the best overall response (BOR). The structure of the dataset should be both flexible and traceable, ensuring it can accommodate the requirements of analysis and reporting effectively.

CDx bridging studies often incorporate clinical data from multiple studies to establish the clinical validity of CDx. Consequently, the data structure is essentially a pooled dataset that **integrates** information from various data sources. For oncology clinical trials with binary endpoints, the primary objectives of the bridging study are usually to estimate the overall response rate (ORR) in the efficacy population and to assess the agreement between the CTA and CDx results in the concordance population. In some instances, analysis datasets from individual clinical studies are already available, and a dedicated statistical programming team is responsible for retrieving, combining, and reshaping these datasets for CDx concordance and efficacy analyses, as summarized below.

Table 2 Analysis Dataset Needed from Individual Study

Dataset	Structure	Variables Retrieved
ADSL	One record per subject	Identifier, demographics (age, sex, race, ethnic, country), cohort, population flags, treatment group, treatment dates, tumor type, baseline characteristics (ECOG, prior lines of therapy, cancer stage etc).
ADCDX	One record per subject	CTA test results, CDx test results
ADRS	Multiple records per subject	Best overall response per independent central imaging

Although the ADRS dataset contains multiple records per subject, only the best overall response (BOR) per central imaging is needed to create the analysis dataset so only a single record per subject is retrieved from ADRS. Consequently, the resulting dataset, **ADPMA**, is structured as one record per subject. According to the CDISC ADaM Implementation Guide (ADaM-IG), this dataset can be classified as **ADaM OTHER**. This classification is appropriate because ADPMA adheres to the fundamental

principles and conventions of ADaM, even though it does not conform to one of the three defined structures (ADSL, BDS, OCCDS).

To facilitate analysis population selection, new population flags are derived in the ADPMA dataset. These include the concordance analysis population flag (CONCRDFL), efficacy analysis population flag (PRMEFFL), and the CDx evaluable population flag (CDXEVLFL). The CTA results (CTASTUDY) and CDx results (CDXSTUDY) will be utilized to calculate pair-wise metrics such as PPA, NPA, and OPA for assessing concordance between CTA and CDx. The responder variable (RESPN) is used to denote best overall response status, with a value of 1 indicating a responder and 0 indicating a non-responder. Subsequently, the overall response rate (ORR) is calculated to assess the response rate within CDx-positive patients.

Table 3 Key Variables from ADPMA

USUBJID	TRT01A	CTASTUDY	CDXSTUDY	BOR	RESPN	CONCRDFL	PRMEFFL	CDXEVLFL
ABC-001	TRT1	Positive	Positive	CR	1	Y	Y	Y
ABC-002	TRT1	Positive	Negative	PR	1	Y	Y	Y
DEF-001	TRT2	Negative	Missing	SD	0		N	N

CMDE DATA SUBMISSION GUIDANCE

Premarket Application (PMA) is a regulatory process through which the safety and effectiveness of CDx are evaluated before they can be marketed. Compliance with PMA requirements is essential for gaining regulatory approval and market access.

To guide the standardized submission of clinical trial data and related materials for IVD reagents, and to better facilitate the review of clinical trial materials, CMDE released the "***Guideline on Clinical Trial Data Submission Requirements for In-vitro Diagnostic Reagents Registration and Review***" in 2021. This guidance comprises three main sections for CDx clinical trial data submission:

Requirements for Clinical Trial Database Content

1. The data should include **subject-related information** such as clinical diagnosis, sample type, demographics, and more. **Clinical diagnosis** background information should encompass clinical diagnosis results, related symptoms and signs, and treatment information. If **subgroup** analysis is required for the submitted product, relevant subgroup classification information should also be included.
2. The data should contain the **test results** of the IVD reagent under investigation as well as the test results of the comparative method.
3. For interventional study, the dataset should also include information on the clinical trial treatment groups and clinical **evaluation endpoints**.

Requirements for Submission Format

The submission package typically includes the **original database**, **analysis database**, **data definition File**, and program code (if any). Submitting data based on CDISC standards is encouraged. When providing Chinese translations of foreign materials, it is necessary to translate at least the **dataset, variable labels**, and descriptive text in both the original and analysis database. The general requirements for the four components are similar to data standards for drug NDA submissions but offer more flexibility.

1. The guidance recommends delivering the data definition files in Excel format. These files should list the datasets, variables, variable types (e.g., character, numeric), labels, and values in the original and analysis databases.

The original database, analysis database, data definition file, and program code (if any) should be placed in four separate folders. The **database** can be submitted in **Excel** format. **Data definition files** can be submitted in PDF, Word, or **Excel** format.

CASE SHARING ON SUBMISSION PACKAGE PREPARATION

To adhere to CMDE's guiding principles for data submission, the data package should be carefully prepared to ensure data authenticity and traceability between the original and analysis database. This section will share our experiences in preparing the respective submission components.

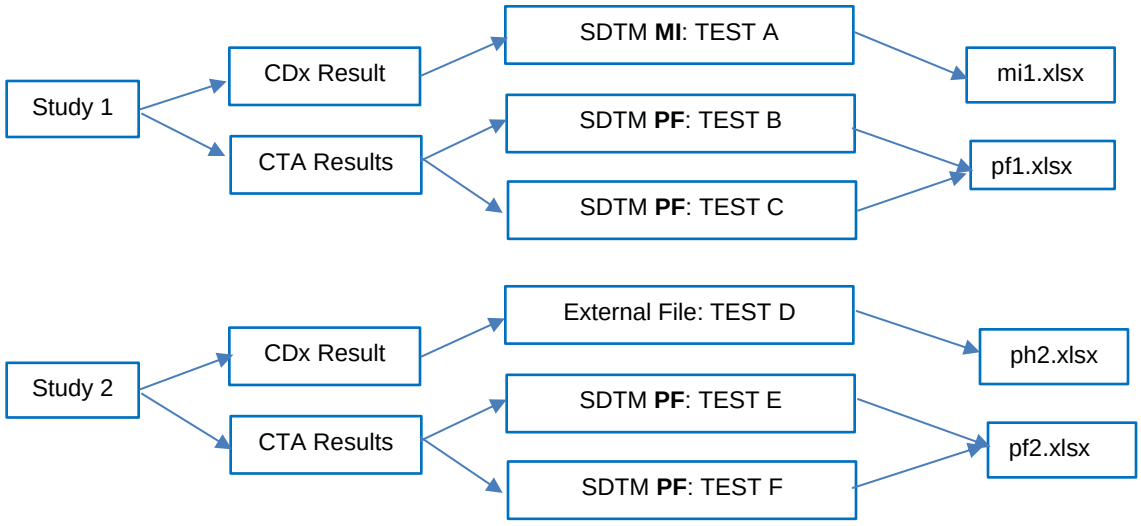
Analysis Database

In our case, the analysis dataset was already prepared and submitted for other regulatory request, so the process of preparing analysis dataset for CMDE submission is quite straightforward. Since CMDE accepts data in Excel format, we converted the ADPMA SAS dataset into Excel (.xlsx) format, which eliminated the need for converting SAS to XPT format. As we did not submit the XPT format, we did not include the Chinese translation of the variable labels within the Excel spreadsheet. Instead, these translations were provided in the data definition file discussed below. The data spreadsheet contains one sheet displaying the tabular data.

Original Database

Preparing the original database package was the most time-consuming part due to the integrated nature of the analysis dataset, which includes data from multiple studies. Consequently, we had to examine the derivation rules for each study and establish the data flow rules from the source dataset to the analysis datasets.

The first step was to determine the type of data to include in the submission package. In our case, team decided to submit the data containing CDx and CTA test results. The next step involved tracing the CTA and CDx result variables back to their source data variables. Since the original database primarily consisted of SDTM data, we chose to group the data by SDTM domain within each individual study to enhance clarity for the reviewers. For example, there were three tests in study 1(one CDx test from the MI domain and two CTA tests from the PF domain), so we prepared two source data files for this study, naming them by concatenating the SDTM domain name and the study ID. Similar to the analysis data package, the original databases were converted to Excel (.xlsx) format. The CDx and CTA test results are typically stored in the SDTM Findings domain. Therefore, only the TESTCDs that contributed to the derivation of CTA and CDx variables in the analysis dataset were included in the final original database. Some source data may be transferred from external vendors and contain non-standard SDTM variables. However, we categorized these data based on the mapping column they contained, such as the PH domain.



Source Data Assembling Example

Data Definition File

The data definition file describes the metadata of the submitted electronic datasets and is considered an important part of the electronic dataset submission for regulatory review. For typical drug NDA submissions, a define.xml file is required to clearly document the metadata structures, including datasets, variables, possible values of variables, and controlled terminologies and codes. Since it is clearly stated in the guidance that Excel file format is acceptable, we streamlined the process by providing **Excel xlsx** data definition files for both databases.

For Analysis Database

The analysis define Excel file contains 2 sheets, named Content and ADPMA respectively. All the metadata are translated into Chinese.

1. The **Content** sheet includes the dataset metadata, such as dataset name, label, class, structure and key variables.
2. The **ADPMA** sheet documents the variable-level metadata, consisting of three columns: **Variable Name**, **Variable Label**, and Description. Given that the analysis dataset is integrated, the logic descriptions for each variable were separately documented for each individual study.

数据集	描述	类	结构	关键变量	注释
ADPMA	CDx分析数据集	其他分析数据结构	每个受试者一条记录	USUBJID	有可用测量值的所有随机化受试者。

Figure 1 Sample Analysis Define – Content

For Original Database

A similar approach was used to create the define Excel for the original database. We prepared a single Excel (.xlsx) file with a Content sheet and multiple sheets for each compiled original dataset. In the example above, there would be four sheets named MI1, PF1, PH2, and PF2. This setup allows reviewers to quickly reference the metadata for each individual study.

Validation

An essential quality control procedure validates that the source data accurately maps to the corresponding variables in the analysis dataset. For example, using MI1 and PF2 it recreates the CDXSTUDY variable, while using PF1 and PF2 it recreates CTASTUDY and other CTA testing results.

Study Data Folder Structure

When submitting study datasets and their supportive files in the eCTD format, they should be organized into a specific file directory structure. As mentioned earlier, the original database includes some datasets that do not conform to the standard SDTM structure, also known as legacy data. Consequently, we placed these datasets and their corresponding define files in the **legacy** folders, referencing the eCTD folder structure guidelines outlined in Appendix III of the *Guidelines for Submission of Clinical Trial Data* from the Center for Drug Evaluation (CDE).

1. **/analysis/legacy**: ADPMA and define.
2. **/tabulations/legacy**: source data files and define.

OTHER REQUESTS FOR CDX SUBMISSION

In 2022, CMDE released the guidance "*Guideline on Registration Review of Clinical Trials of Original Companion Diagnostic Reagents Co-development with Anti-Tumor Drugs*", which aims to guide applicants in preparing and writing registration application materials for clinical trials of original companion diagnostic reagents developed concurrently with anti-tumor drugs. In the section of *Clinical trial documentation requirements*, it recommends the applicant to submit the **clinical trial data summary table**, also called line data. This table is used to summarize the results of companion diagnostic reagent tests in a tabular format. It should include the following information: Subject ID, age, gender, sample type, clinical background information, CDx test results, treatment group, and anti-tumor drug efficacy. The

background information of the case should clearly specify the patient's tumor type, stage, and previous treatment regimen.

The sample table can be found from the appendix of the guidance and is listed below for reference. To fulfill this line data request, statistical programmers should work closely with statisticians, the biomarker team and the regulatory team to align on the required data to put on this table and ensure that the analysis dataset contains the variables as required from the line data. Below are some points to consider:

- Should CTA test results be added to facilitate side-by-side comparison,
- Confirm the variables to represent Sample Type and Clinical Diagnosis and their value formatting,
- Determine the variable to use for anti-tumor efficacy, e.g., best overall response.

Table 4 Summary Table of CDx Test Results

Subject ID	Age	Sex	Sample Type	Clinical Diagnosis	CDx Results	Treatment Group	Anti-Tumor Efficacy
ABC-001	53	M	Tumor	HNSCC	Positive	Trt 1	PR

CONCLUSION

This paper has provided a high-level overview of the methodologies for concordance and efficacy analyses in CDx bridging studies. We have detailed the structure of the analysis datasets needed to support these analyses, along with the CMDE guidelines and requirements for data submission. Additionally, we have outlined the steps for preparing the analysis data package, source data package, and data definition files, as well as organizing the folder structures. The shared submission preparation experiences are based on our current understanding of the guidelines and are intended to serve as a reference. Ensuring thorough documentation and adherence to data standards is crucial for facilitating regulatory review. In the era of precision medicine with the increasing demand of the drug-diagnostic co-development model, statistical programmers will have more chance to engage in CDx PMA work, and can further explore and refine the submission process based on insights gained from their respective study experiences.

REFERENCES

Center for Medical Device Evaluation, NMPA. "Guideline on Clinical Trial Data Submission Requirements for In-vitro Diagnostic Reagents Registration and Review." CMDE. 2021. https://www.cmde.org.cn/flfg/zdyz/zdyzwbk/20211126085958548.html

U.S. Food and Drug Administration. "STUDY DATA TECHNICAL CONFORMANCE GUIDE." FDA. 2025. https://www.fda.gov/media/153632/download.

Hu, Danxia, Gao Mandy and Tong, Jianyong. 2024. "Implementing Statistical Methods in Companion Diagnostic: A Case Study in NSCLC", *PharmaSUG China 2024*, Paper SA-045.

Song, Changhong, et al. "Statistical considerations for some issues in clinical bridging studies evaluating companion diagnostic devices." *Journal of Biopharmaceutical Statistics* 34.3 (2024): 441-452.

ACKNOWLEDGMENTS

The authors would like to thank Frédéric Coppin, Helen Wu and Amy Gillespie for reviewing this paper.

CONTACT INFORMATION

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

Ben Gao
Statistical Programming
MSD China, Shanghai
peng.gao4@merck.com

Wei Wang
Biostatistics
MSD China, Beijing
wei.wang42@merck.com

Nicole Zhang
Statistical Programming
MSD China, Shanghai
jing.zhang18@merck.com