Paper AS21

Advancing Precision Medicine and Regulatory Readiness: Data Harmonization and Statistical Methods for Companion Diagnostics Bridging Studies

Ummul Vara Qurratul Ain, Bayer, Whippany, USA

ABSTRACT

Companion Diagnostics (CDx) play a pivotal role in precision medicine by linking drug efficacy to diagnostic device performance and ensuring targeted therapies reach the right patients. Bridging studies are essential for demonstrating concordance between Clinical Trial Assays (CTAs) and candidate CDx assays providing evidence to support regulatory decision-making.

This paper focuses on preparation of harmonized analysis datasets according to CDISC standards, role of supplemental negative and surrogate data in enhancing specificity and evaluating CDx performance relative to CTAs. Additionally, the framework describes statistical methods with a focus on clinical efficacy, concordance, and sensitivity analyses incorporating advanced approaches to address missing data in CDx studies and strengthen robustness of the results. The contents and preparation of a submission package (Master Analysis File) are also highlighted, emphasizing the need for reproducible workflows to support regulatory readiness, thereby showcasing the evolving role of statistical programmers in shaping CDx validation and precision medicine strategies.

INTRODUCTION

Precision medicine has transformed oncology by enabling therapies tailored to a patient's molecular profile [1]. A key enabler of this is the Companion Diagnostic (CDx), an in vitro diagnostic device developed to identify patients most likely to benefit from a specific treatment [2, 3]. Regulatory guidance on the co-development of diagnostics and therapeutics underscores this linkage, establishing that a CDx is essential for the safe and effective use of its corresponding drug [6].

In many clinical trials, patient enrollment is based on a Clinical Trial Assay (CTA), often a laboratory-developed test (LDT), rather than the final, market-ready CDx. This practical choice avoids potential treatment delays but necessitates subsequent bridging study [6, 8]. Analytical validation must be completed prior to using an investigational CDx. Bridging studies are vital to demonstrate that the candidate CDx possesses performance characteristics akin to those of the CTA used in the clinical trial. This comprehensive approach facilitates the accurate identification of the target patient population and provides critical evidence of analytical and clinical validity required for regulatory approval [4-6].

From a statistical programming perspective, CDx bridging studies present unique and complex challenges. The first is data harmonization. Data originates from disparate sources, including the clinical trial database and proprietary diagnostics vendor files, each with its own structure and terminology. Genomic data requires intensive standardization to reconcile inconsistent variant nomenclature, assay technologies, and quality control (QC) failure reasons, such as "insufficient tissue" or “incomplete reads” [8-10].

The second challenge lies in the statistical analysis itself, which must extend beyond simple concordance to include advanced sensitivity analyses that address these real-world data issues. This paper presents a practical, end-to-end programming and statistical framework for CDx bridging studies, showcasing the modern statistical programmer's pivotal role in solving complex data challenges, implementing advanced statistical methods, and preparing reproducible, submission-ready packages.

The following case study illustrates this framework in a retrospective CTA-CDx bridging setting, including harmonized analysis dataset development and sensitivity analysis to evaluate robustness under missing CDx results.

CASE STUDY OVERVIEW

A retrospective CDx bridging case study was designed to provide evidence intended to support regulatory decision-making for a candidate companion diagnostic (CDx) assay.

This was achieved through two co-primary objectives:

(1) to establish the clinical validity of the CDx by evaluating its ability to predict patient response, and

(2) to assess the concordance between the candidate CDx and the Clinical Trial Assay (CTA) used in the parent clinical trial.

To measure these objectives, specific endpoints were defined. The primary endpoint for assessing clinical validity was Objective Response Rate (ORR), as determined by a Blinded Independent Central Review (BICR) using RECIST v1.1 criteria. This endpoint was chosen for consistency with the primary endpoint used in the parent clinical trial. The endpoints for assessing concordance were Positive Percent Agreement (PPA) and Negative Percent Agreement (NPA).

METHODS

STUDY DESIGN AND ANALYSIS POPULATIONS

This illustrative retrospective CDx bridging case study was constructed to reflect a Formalin-Fixed Paraffin-Embedded (FFPE) tissue workflow using specimens from two primary sources, which formed the basis for all analysis populations. A Clinical Trial Cohort of approximately 130 biomarker-positive patients from the parent clinical trial was included to provide data for both the efficacy and concordance analyses. A Supplemental Cohort of approximately 120 biomarker-negative patients was included to ensure an adequate number of negative cases for a robust concordance analysis. This was necessary because CTA testing in the parent clinical trial was primarily performed in biomarker-positive patients, and CTA-negative samples may be limited (e.g., not routinely collected and/or not available for retesting with the candidate CDx due to sample availability or consent constraints). All samples from both sources were tested with the candidate CDx assay. The primary analyses were designed to establish both clinical validity and concordance. To assess clinical validity, the efficacy of the targeted therapy, as measured by Objective Response Rate (ORR), was estimated and compared across subgroups defined by their CTA and CDx status.

A key methodological consideration is that the supplemental samples were procured from patients not treated with the investigational drug; therefore, efficacy data was not available for this subgroup. Concordance analysis required negative samples and therefore included the supplemental cohort. In addition, concordance-derived predictive values (Positive Predictive Value (PPV) and Negative Predictive Value (NPV)) were evaluated in a sensitivity analysis to support ORR estimation for the overall CDx-positive population; however, these results are beyond the scope of this paper. Concordance analysis between CTA and the CDx was conducted using the combined set of samples from the clinical trial and the supplemental negative cohort.

SAMPLE SIZE DETERMINATION

The sample size for the bridging study was determined based on separate considerations for efficacy and concordance objectives to ensure adequate precision for the planned evaluations.

- Biomarker-positive samples (efficacy-driven): The number of biomarker-positive samples was informed by prior clinical trial experience and aimed to provide sufficient precision around Objective Response Rate (ORR) estimates relative to a clinically meaningful threshold. This ensured that confidence intervals around ORR were interpretable for clinical validation purposes.

- Biomarker-negative sample (concordance-driven): The number of supplemental biomarker-negative samples followed published methodological guidance by Song et al. [12] to ensure stable estimation of Negative Percent Agreement (NPA) and reliable precision of the corresponding confidence intervals.

ANALYSIS SET DEFINITIONS

From these cohorts, two key analysis sets were defined and implemented within ADaM datasets. A sample was considered evaluable only if it yielded a valid, interpretable result from the candidate CDx assay.

- Efficacy Analysis Set (EES): This set included all patients from the Clinical Trial Cohort who had a valid CDx result, making them evaluable for the ORR endpoint.

- Concordance Analysis Set (CES): This set included all samples from both the Clinical Trial Cohort and the Supplemental Cohort that had a valid result on both CTA and the candidate CDx assay.

DATA HARMONIZATION AND ANALYSIS FRAMEWORK

The foundation of this bridging case study was a robust data harmonization framework designed to integrate disparate data sources into submission-ready ADaM datasets. This was essential for managing the complexity of the biomarker data and enabling advanced statistical analyses.

BIOMARKER DATA HARMONIZATION FRAMEWORK (PRE-AXCDX)

The biomarker data originated from multiple heterogeneous sources and required substantial upstream harmonization prior to creation of analysis datasets. These data sources varied in structure, nomenclature, and level of biological detail, reflecting differences in vendor-specific conventions, assay platforms, and data transfer practices. Addressing this variability was necessary to ensure consistency, interpretability and traceability across downstream analyses.

A significant programming challenge was the systematic standardization to CDISC convention, guided by pre-defined mapping rules across these sources. Data from different vendors, each with unique variable names and structures, was mapped to a consistent format aligned with the Genomic Findings (GF) domain. This included harmonizing inconsistent terminology, such as mapping dozens of vendor-specific text strings to a controlled vocabulary suitable for analysis and reporting.

The raw genomic data also included multiple classes of genetic alterations. During upstream data preparation, records were filtered to retain only clinically relevant variants. All records for synonymous variants, changes in the DNA sequence that do not alter the final protein's amino acid sequence, were removed. This step efficiently reduced the data volume to only those variants (such as missense mutations) with the potential to impact protein function and, therefore, drug efficacy.

Following this, an additional annotation step was performed to add biological context. To execute this efficiently, all unique variants were first extracted from the dataset. This complex annotation process, typically performed by bioinformatics team, provides information such as a variant's known pathogenicity and is determined from public databases (e.g., COSMIC). These annotations were then mapped back to the harmonized biomarker data by the statistical programming team, preserving traceability while avoiding redundant annotation of duplicate records.

While this upstream process preserves variant-level findings, it does not by itself provide the sample-level CDx endpoint, assay/QC context, and evaluability flags needed to define analysis populations and support efficacy and concordance endpoints. Therefore, the sample-level companion diagnostic dataset (AXCDX) was created.

CREATION OF SAMPLE-LEVEL COMPANION DIAGNOSTIC DATASET(AXCDX)

AXCDX served as the central sample-level analysis dataset, integrating biomarker results generated by next-generation sequencing (NGS) from the central CDx vendor with subject-level clinical data, samples not sent for CDx testing, and data from supplemental negative and surrogate cohorts. For each specimen, key attributes including specimen type (e.g., plasma DNA, tumor tissue, TNA), assay platform and version, method of biomarker testing, and assay quality control status were captured. Quality control records were retained to explicitly identify samples that failed assay performance criteria, enabling transparent classification of CDx evaluability and supporting downstream missing data handling.

At the variant level, AXCDX incorporated harmonized genomic information, including gene symbol, variant type, nucleotide and amino acid changes, genomic position, transcript identifier, read depth, and percent genetic change. Although these attributes do not directly contribute to CDx endpoint definitions, they provide critical biological and technical context for interpreting assay concordance and investigating discordant CTA/CDx results. Annotation variables, such as COSMIC classification and somatic status, were included to support traceability to public genomic databases and to justify variant inclusion or exclusion decisions.

Following integration of all data sources, a set of derived variables were created to support downstream analyses. These include population flags to identify samples originating from the CTA-positive and CTA-negative source cohorts, as well as CDx evaluability indicators to differentiate valid CDx results from invalid or missing results. In addition, a concordance outcome was derived by comparing the final CDx result with the corresponding CTA result, providing basis for concordance assessment. The resulting AXCDX dataset contains one or more records per subject, detailing every aspect of the biomarker assessment, and is fully prepared for subsequent analysis and reporting.

CREATION OF THE SUBJECT-LEVEL DATASET (AXSL)

The final subject-level analysis dataset (AXSL) served as a primary source for efficacy and concordance analyses. This dataset, containing one record per subject, was constructed by merging the subject-level biomarker outcomes from AXCDX with several key datasets from the clinical database. The final AXSL dataset integrated several key components for each patient including

- Harmonized CDx and CTA assay information and results
- Core demographic and baseline disease characteristics
- Prior treatments/ prior therapies
- CDx evaluability and missing/invalid result indicators
- Concordance classification derived from CTA/CDx result agreement
- Key clinical efficacy endpoints including ORR
- Variant-level genomic information (e.g., gene symbol, variant annotation).

This consolidation resulted in a complete, analysis-ready dataset that linked each patient's clinical outcome directly to their biomarker status as determined by both CTA and the candidate CDx assay.

Table 1 provides an illustrative example of the AXSL dataset. It showcases how the harmonized biomarker results from both CTA and the candidate CDx assay are integrated at the subject level with patient cohort information and clinical outcomes. For clarity, this example focuses on CDx evaluability and the derived analysis flags used for concordance classification and ORR endpoint assessment. Additional variables have been omitted for brevity.

Table 1: Illustrative Example of the Subject-Level Analysis Dataset (AXSL)

Subject Identifier (USUBJID)	Cohort (COHORT)	Treatment Group (TRT01AN)	CTA Result (RESLCTAC)	CDx Result (RESCDXC)	CDx Evaluable Flag (CDX01AG)	CDx Missing Reason (XBRESND)	Concordance Status (CONCORDC)	Objective Response (ORRRESPC)
A75-1001	Clinical	ARM A	POSITIVE	POSITIVE	Y		Concordant	Y
A75-1002	Clinical	ARM A	POSITIVE	POSITIVE	Y		Concordant	Y
A75-1003	Clinical	ARM A	POSITIVE	.	N	Sample not sent	Not Evaluable	N
A75-1004	Clinical	ARM B	POSITIVE	POSITIVE	Y		Concordant	N
A75-1005	Clinical	ARM B	POSITIVE	NEGATIVE	Y		Discordant	N
A75-1006	Clinical	ARM B	POSITIVE	INVALID	N	Insufficient Quality	Invalid	Y
A75-1007	Clinical	ARM A	POSITIVE	POSITIVE	Y		Concordant	Y
A75-1008	Clinical	ARM A	POSITIVE	.	N	Sample not available	Not Evaluable	N
B92-5001	Supplemental	N/A	NEGATIVE	NEGATIVE	Y		Concordant	N/A
B92-5002	Supplemental	N/A	NEGATIVE	NEGATIVE	Y		Concordant	N/A
B92-5003	Supplemental	N/A	NEGATIVE	POSITIVE	Y		Discordant	N/A
B92-5004	Supplemental	N/A	NEGATIVE	.	N	Sample not available	Not Evaluable	N/A

Legend: ". "= Missing/Not Performed; N/A = Not Applicable (Supplemental cohort not treated; ORR not assessed).

PROGRAMMING TRACEABILITY AND REPRODUCIBILITY

To support reproducibility and regulatory traceability, all derived analysis variables and flags were created using metadata-driven programming and documented mapping rules from source biomarker datasets into Pre-AXCDX, AXCDX, and AXSL. Key derivations (e.g., CDx evaluability, cohort flags, concordance classification, and analysis population inclusion) were validated through independent QC checks, frequency-based reconciliation, and cross-dataset consistency checks. This ensured that every reported concordance and efficacy endpoint could be traced back to the originating assay records and specimen-level results.

STATISTICAL ANALYSIS METHODS

The statistical framework was designed to assess the primary endpoints of concordance and clinical efficacy, and to ensure the robustness of the conclusions through a series of pre-specified sensitivity analyses.

PRIMARY EFFICACY AND CONCORDANCE ANALYSES

The primary analysis for clinical efficacy compared the Objective Response Rate (ORR) between the CDx-positive subgroup and the overall CTA-positive population. Concordance was evaluated by quantifying agreement between the candidate CDx assay and the CTA using Positive and Negative Percent Agreement (PPA/NPA). Confidence intervals for all proportions were calculated using appropriate statistical methods.

FRAMEWORK FOR HANDLING MISSING DATA AND SENSITIVITY ANALYSES

A central feature of this study was a structured framework to address potential bias arising from patients with missing CDx results (e.g., samples not sent or not available for re-testing). This was implemented using a multi-step approach followed by a final estimation and combination step.

- Covariate Preparation and Screening: A wide range of baseline covariates were harmonized and re-categorized to support statistical modeling. A data-driven screening process was then executed, using

statistical tests, to identify a final, refined set of covariates, most predictive of both CDx evaluability and the final clinical outcome.

- Multiple Imputation (MI): To quantitatively assess the impact of missing data, a multiple imputation (MI) procedure was performed. A logistic regression model was built using the key predictive covariates identified in the screening step (such as baseline demographics, prior therapies, clinical response, and other significant baseline factors). This model was then used to generate 60 complete datasets by simulating plausible values for each missing CDx result based on the relationships observed in the complete data.

- Illustrative code for Proc MI (subset of covariates shown):

```
PROC MI DATA=AXSL_FOR_IMPUTE NIMPUTE=60 SEED=2024 OUT=IMPUTED_DATASETS;
CLASS RESCDXC AGEGR SEX RACE ETHNIC PRIORTRT;
MONOTONE LOGISTIC (RESCDXC = AGEGR SEX RACE ETHNIC PRIORTRT/ details
likelihood=augment);
VAR RESCDXC AGEGR SEX RACE ETHNIC PRIORTRT;
RUN;
```

The output is a single dataset containing 60 imputed versions of the analysis data. **Table 2** illustrates how missing CDx results for two subjects A75-1003 and A75-1008 are "filled in" differently across the first few imputed datasets, reflecting statistical uncertainty. The presentation approach for illustrative imputation and bootstrap dataset structures (Tables 2-3) was informed by Hu (2024).

Table 2: Illustrative structure of imputed datasets (IMPUTED_DATASETS) across multiple imputations (only variables relevant to illustrating variability across imputations are shown)

Imputation (IMPNUM)	Subject Identifier (USUBJID)	CDx Missing Reason (XBRESND)	CDx Result (RESCDXC)	CTA Result (RESLCTAC)	CDx Evaluable Flag (CDX01AG)	Concordance Status (XBRESND)	Objective Response (ORRRESPC)
1	A75-1001		POSITIVE	POSITIVE	Y	Concordant	Y
1	A75-1002		POSITIVE	POSITIVE	Y	Concordant	Y
1	A75-1003	Sample not sent	**POSITIVE**	POSITIVE	Y	Concordant	N
1	A75-1008	Sample not available	**NEGATIVE**	POSITIVE	Y	Discordant	N
…	…	…	…	…	…	…	…
2	A75-1001		POSITIVE	POSITIVE	Y	Concordant	Y
2	A75-1002		POSITIVE	POSITIVE	Y	Concordant	Y
2	A75-1003	Sample not sent	**NEGATIVE**	POSITIVE	Y	Discordant	N
2	A75-1008	Sample not available	**POSITIVE**	POSITIVE	Y	Concordant	N
(and so on for all 60 imputations)							

- MI-Boot Nested Balanced Bootstrap Resampling: Standard methods for combining multiple imputation results can sometimes underestimate total variability when resampling uncertainty is not fully accounted for. To address this, the MI-Boot method was implemented. This technique involves performing a nested bootstrap resampling procedure for each of the 60 imputed datasets. This step was executed within a macro loop, iterating from 1 to 60. Within each iteration, PROC SURVEYSELECT was used to generate 500 bootstrap samples from the corresponding imputed dataset with the balanced bootstrap (method=balbootstrap) option. This approach ensures that each original subject is represented an equal number of times across the full set of bootstrap samples, thereby improving the stability of the final estimates. This step is crucial for estimating the "within-imputation" variance, reflecting uncertainty arising from random patient sampling. This procedure resulted in 30,000 datasets which are used to generate the final endpoint estimates.

- Illustrative code for balanced bootstrap resampling step (PROC SURVEYSELECT)

```
/* --- This code is within a macro loop from i=1 to 60 --- */
/* Generate 500 balanced bootstrap resamples for the i-th imputed dataset */
PROC SURVEYSELECT DATA=IMPUTED_DATASET_&i OUT=BOOTSTRAP_SAMPLE_&i
METHOD=BALBOOTSTRAP NOPRINT
REP=500 SEED=2025;
RUN;
```

Table 3: Example illustrating the structure of the BOOTSTRAP_SAMPLES Dataset (60 imputations x 500 balanced bootstrap replicates; subset shown)

Imputation (IMPNUM)	Replicate (REPNUM)	Subject Identifier (USUBJID)	CTA Result (RESLCTAC)	CDx Result (RESCDXC)	CDx Missing Reason (XBRESND)
1	1	A75-1001	POSITIVE	POSITIVE	
1	1	A75-1005	POSITIVE	NEGATIVE	
1	1	A75-1003	POSITIVE	**POSITIVE**	Sample not sent
…... (and so on for all subjects in the first bootstrap sample)					
1	2	A75-1004	POSITIVE	POSITIVE	
1	2	A75-1007	POSITIVE	POSITIVE	
1	2	A75-1008	POSITIVE	**NEGATIVE**	Sample not available
... (and so on for all subjects in the second bootstrap sample)					

*The **Rep** variable identifies each of the 500 unique bootstrap samples created within each of the 60 imputations. For example, all rows where Imputation=1 and Rep=1 represent the first complete bootstrap dataset. The imputed values for subjects (like A75-1003 and A75-1008) are carried through from their source imputation.*

- MI-Boot Combination of Estimates and Confidence Intervals: The final step combined the results from all 30,000 bootstrap samples into a single, robust point estimate and corresponding 95% confidence interval. This was achieved using the MI-Boot procedure proposed by Schomaker and Heumann [11] which integrates uncertainty arising from both multiple imputation and resampling. Specifically, two distinct components of variance were estimated: (1) within-imputation variance, derived non-parametrically from the 500 bootstrap samples within each imputed dataset, reflecting uncertainty due to random patient sampling and (2) between-imputation variance reflecting variability across the endpoint estimates obtained from the 60 imputed datasets. These two variances were then combined to calculate a total standard error and a final, robust 95% confidence interval. This process was applied to ORR, PPA, and NPA, providing a comprehensive and robust assessment of the study's primary endpoints.

SENSITIVITY ANALYSIS FOR UNOBSERVED EFFICACY IN THE INTENDED-USE POPULATION

A key methodological challenge was estimating the ORR for the full CDx-positive population, as this group includes a subgroup of CTA-/CDx+ patients for whom no efficacy data was collected. To address this, a modeling-based sensitivity analysis following the approach proposed by Li [7] was implemented.

The purpose of this analysis was to evaluate whether the overall ORR for the CDx-positive population remained clinically meaningful across a range of plausible "what-if" scenarios for the unknown efficacy of this CTA-/CDx+ subgroup. The analysis was repeated under systematically varied assumptions, as described below.

- Assumption on Unobserved Efficacy: The ORR of the unobserved CTA-/CDx+ group varied across a pre-defined range, spanning from a worst-case scenario (ORR = 0%) to a best-case scenario in which ORR was assumed equal to that observed in the CTA+/CDx+ subgroup.

- Assumption on Concordance Uncertainty: To reflect uncertainty in assay concordance, the Negative Percent Agreement (NPA) was allowed to vary within its estimated 95% confidence interval. For each scenario, confidence intervals for the resulting ORR estimates were calculated using the delta method.

  The primary objective was to assess whether the estimated ORR for the CDx-positive population remained clinically meaningful across a broad range of pessimistic and optimistic assumptions. To keep the paper concise, detailed scenarios of ORR estimates are not presented; instead, this sensitivity framework was designed to evaluate the robustness of the study conclusions to uncertainty arising from unobserved efficacy.

RESULTS

ANALYSIS POPULATIONS, DEMOGRAPHICS AND BASELINE CHARACTERISTICS

The analysis included a Clinical Trial Cohort of approximately 130 biomarker-positive patients and a Supplemental Cohort of approximately 120 biomarker-negative patients. Of the 250 total samples available for CDx testing, 190 (76%) yielded a valid CDx result and were considered CDx-evaluable. The remaining 60 samples (24%) were CDx-non-evaluable (including missing/invalid results), with non-evaluability observed in both clinical cohort (50/130) and the supplemental cohort (10/120). Comparison of baseline characteristics between evaluable and non-evaluable groups revealed no clinically meaningful differences, suggesting a limited risk of bias due to CDx non-evaluability (summary statistics not shown).

CONCORDANCE ANALYSIS

The concordance analysis was performed on the 190 CDx-evaluable samples; 60 CDx non-evaluable samples were excluded. The results are presented in Table 4.

Table 4: Concordance of Candidate CDx vs. Clinical Trial Assay (CTA) (N=190)

	CTA Positive	CTA Negative	Total
CDx Positive	68	4	72
CDx Negative	12	106	118
Total	80	110	190

The key agreement metrics were a Positive Percent Agreement (PPA) of 85.0% (68/80) and a Negative Percent Agreement (NPA) of 96.4% (106/110) demonstrating a high level of agreement between the candidate CDx and CTA.

CLINICAL EFFICACY ANALYSIS

The clinical validity of the CDx was assessed by comparing the ORR in the CDx-positive population to the CTA-Positive Population. A key success criterion was to demonstrate that the lower bound of the ORR's 95% CI would remain above a pre-specified 30% threshold. The results, summarized in Table 5, demonstrate that this criterion was met. All ORR estimates were based on treated clinical trial cohort patients with evaluable clinical response data and therefore, may differ from the concordance sample size.

Table 5: Objective Response Rate (ORR) in Efficacy Subgroups

Subgroup	N	ORR (%)	95% CI (%)
CTA positive	86	58.1%	47.0%, 68.7%
CDx positive	74	60.8%	48.8%, 72.0%

The ORR in the CDx-Positive Population (60.8%; 45/74) was comparable to the CTA-Positive Population (58.1%; 50/86). Importantly, the lower bound of the 95% CI for the CDx-positive ORR (48.8%) remained above the 30% success threshold.

SUMMARY OF SENSITIVITY ANALYSES

The robustness of the primary findings was confirmed via the pre-specified sensitivity analyses. Tables 6A and 6B summarize sensitivity analyses comparing complete-case results versus results obtained after imputing missing CDx outcomes using the MI-Boot method. ORR was evaluated in the treated clinical trial cohort and summarized by CTA-positive and CDx-positive subgroups. Concordance endpoints (PPA/NPA) were calculated in the CDx-evaluable population.

Table 6A: Sensitivity Analysis Comparing Complete-Case vs. MI-Boot Imputed Analysis Results for Efficacy

Analysis Type	Endpoint	Complete-Case Analysis (95% CI)	Analysis with Imputed Data (95% CI)
Efficacy	ORR (%)	60.8% (48.8%, 72.0%)	59.2% (47.5%, 70.6%)

Table 6B: Sensitivity Analysis Comparing Complete-Case vs. MI-Boot Imputed Analysis Results for Concordance

Analysis Type	Endpoint	Complete-Case Analysis (95% CI)	Analysis with Imputed Data (95% CI)
Concordance	PPA (%)	85.0% (75.5%, 92.2%)	84.1% (74.9%, 91.6%)
	NPA (%)	96.4% (91.1%, 98.9%)	96.0% (90.8%, 98.7%)

Results from the MI-Boot imputed analyses were consistent with the complete-case analyses, with only minor differences in ORR and concordance estimates. Overall, these sensitivity analyses support the robustness of the primary conclusions regarding clinical validity and concordance after accounting for missing CDx results using the MI-Boot method.

DISCUSSION

The results of this illustrative bridging study provide evidence supporting both the clinical validity of the candidate CDx assay for its intended use and strong concordance with the original CTA. This was achieved by providing robust evidence for the two primary pillars of the bridging argument: comparable clinical efficacy and high assay concordance.

Clinical validity was demonstrated by successfully bridging the efficacy observed in the CTA-positive population to the corresponding CDx-positive population. The ORR in the CDx-positive population (60.8%) was comparable to the CTA-positive population (58.1%), with the lower bound of its 95% CI (48.8%) remaining above the pre-specified 30% success threshold.

Concordance was demonstrated through a high degree of agreement between the candidate CDx assay and the original CTA. The analysis, which utilized 130 samples from the clinical trial cohort and 120 supplemental negative samples, yielded a PPA of 85.0% and an NPA of 96.4% in the primary analysis of 190 CDx-evaluable samples. Importantly, investigation into the discordant results (CTA+/CDx- cases) suggested that these may be attributable to variants outside the defined analytical range of the new CDx assay.

Robustness of these findings was supported by comprehensive sensitivity analyses. Results for both efficacy and concordance remained stable after applying MI-Boot imputation to account for missing CDx data. Additional sensitivity modeling analyses (not shown) provided consistent results under conservative assumptions for unobserved outcomes.

Together, these findings support the candidate CDx as a tool to identify patients with the target biomarker who may benefit from the targeted therapy.

Despite these favorable results, CDx bridging studies present important methodological and operational challenges that must be addressed to ensure interpretability and regulatory readiness. Biomarker data are often derived from multiple vendors and platforms, each employing different internal standards, result representations, and levels of biological annotation. In addition, genomic biomarker analyses often require cross-functional collaboration across statistical programming, bioinformatics, and clinical teams to translate complex biological outputs into analysis-ready datasets. These challenges, together with the evolving nature of industry standards for genomic data representation within SDTM and ADaM, underscore the importance of robust upstream data harmonization and clearly defined, reproducible analysis frameworks.

THE REGULATORY SUBMISSION PACKAGE

A critical deliverable for the case bridging study was the creation of a comprehensive and standardized submission package, often referred to internally as the Master Analysis File (MAF), designed to ensure full reproducibility and facilitate efficient regulatory review.

The package was structured to include a main PDF document containing key metadata and several ZIP files containing the underlying data and programs. The core components included:

- Analysis Datasets: A ZIP file containing all final, locked ADaM datasets in a regulatory-compliant format (e.g., SAS® XPT v5).
- Analysis Programs: A ZIP file containing all well-annotated SAS® programs used to generate the analysis results, typically provided as plain text files to ensure accessibility.
- Listings: A ZIP file containing patient-level listings provided in spreadsheet format (e.g., *.xlsx)
- Documentation and Metadata: A central PDF document containing key submission documentation (e.g., cover letter, table of contents, SAP, and TLFs), along with essential metadata documents such as the Analysis Data Specifications (ADS), which describe the structure and content of the ADaM datasets, and the Analysis Result Metadata (ARM), which provides traceability between datasets, programs and outputs.

This structured approach, governed by an internal review and approval process within a controlled document management system, ensures that the submission is consistent, complete, and ready for regulatory scrutiny, highlighting the programmer's role in ensuring data integrity and reproducibility.

KEY LESSONS LEARNED

This case study demonstrated that technical challenges in CDx data integration are inseparable from broader organizational and knowledge-based challenges. While statistical programmers are experts in clinical data standards, analysis datasets, and regulatory reporting, CDx studies require close interaction with bioinformatics pipelines that often operate outside traditional clinical programming workflows.

A key lesson learned is the need for expanded genomic and bioinformatics literacy within statistical programming teams, particularly in understanding variant-level data structures, assay quality control metrics, and annotation workflows. Conversely, effective CDx delivery also requires bioinformatics teams to understand regulatory expectations, CDISC standards, and the downstream implications of data transformation on concordance and clinical efficacy analyses [9-10].

Many challenges encountered during dataset development were resolved not through new statistical methods, but through iterative cross-functional collaboration, shared data reviews, and joint interpretation of assay behavior and biological context. These experiences underscore that CDx development is fundamentally a multidisciplinary effort, requiring statistical programmers to act as integrators across clinical, bioinformatics, and regulatory domains.

Early cross-functional engagement, alignment with vendors through consistent data transfer specifications, shared documentation, and continuous knowledge exchange were critical to maintaining traceability, ensuring analytical consistency, and delivering submission-ready outputs under compressed timelines.

CONCLUSION

This paper presents an end-to-end framework for the analysis and validation of a CDx bridging study, from harmonization of disparate data sources to the implementation of advanced statistical methods. This case study illustrates how the role of the modern statistical programmer has evolved beyond output generation to become a critical partner in design and execution of complex validation strategies.

Techniques such as multiple imputation for missing data, MI-Boot-based confidence interval estimation, and sensitivity modeling for unobserved outcomes support analytical rigor and enable reproducible, submission-ready packages. Collectively, these capabilities help inform regulatory decision-making and ultimately support precision medicine.

DISCLOSURE

Editorial support was provided using ChatGPT (OpenAI) [13] to refine language and improve clarity. All technical content was developed and verified by the author.

REFERENCES

[1] Maliepaard M, et al. Evaluation of Companion Diagnostics in Scientific Advice and Marketing Authorisation Procedures in Europe. *Frontiers in Medicine*. 2022;9.

[2] Kang SL, Kwon JY, Kim SM. Insights into Post-Marketing Clinical Validation of Companion Diagnostics with Reference to the FDA, EMA, PMDA, and MFDS. *Frontiers in Pharmacology*. 2023; 14:1191216.

[3] La Rocca C, et al. Companion Diagnostics: State of the Art and New Regulations. *Diagnostics (Basel)*. 2021;11(10):1864.

[4] Menu R. Navigating Precision Medicine: The Role of Companion Diagnostics. *Sophia GENETICS Resource*. 2022.

[5] Hu D, Gao M, Tong J. Implementing Statistical Methods in Companion Diagnostic: A Case Study in NSCLC. *Pharma SUG China 2024*. Paper SA-045. 2024.

[6] U.S. Food and Drug Administration (FDA). Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product: Draft Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD. July 2016.

[7] Li M. Statistical consideration and challenges in bridging study of personalized medicine. *Journal of Biopharmaceutical Statistics*. 2015;25(3):397–407.

[8] U.S. Food and Drug Administration (FDA). Use of Standards in Next-Generation Sequencing (NGS) for Molecular Diagnostics: Guidance for Industry. 2018.

[9] Yun L, Zheng H. Challenges of Genomic Data in Clinical Trials. *PHUSE US Connect*. Paper DS12. 2025.

[10] Zittlau L, Ständler S. Biomarkers Unleashed: A Programming Pipeline from Chaos to Clarity in Genomic Data. *PHUSE EU Connect*. 2024.

[11] Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Statistical Methods in Medical Research*. 2018;37(14):2252–2266.

[12] Song S, Xiong X, Kim S, Xu Z, Xu D, Biswas B. Statistical considerations for some issues in clinical bridging studies evaluating companion diagnostic devices. *Journal of Biopharmaceutical Statistics.* 2023;33(6):1067-1077.

[13] OpenAI. ChatGPT (Jan 2026 version) [Large language model]. https://chat.openai.com

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Lingsong Yun and Satish Patil for their careful review and valuable guidance, which significantly enhanced clarity and strength of this manuscript.

I am also grateful to Anupama Sheoran for encouraging me to submit this work and for her consistent support throughout its development.

CONTACT INFORMATION

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

Author Name: Ummul Vara Qurratul Ain
Company: Bayer
Address: 100 Bayer Boulevard, Whippany, NJ 07981
Email: fnu.ummulvaraqurratulain@bayer.com