Paper DH11

Unlocking the Secrets of Health: A Unified Approach to Biomarker and Clinical Data Integration

Pavan Kumar Tatikonda, Takeda, Cambridge, MA, USA
Xiaoyi Zhu, PPD, Santa Clara, CA, USA
Sunithra Kadambi, Phastar, Aldie, VA, USA

ABSTRACT

In pharmaceutical research, biomarker data analyses are often exploratory and used to better understand the study drug and inform future trials. Integrating biomarker data and clinical data is necessary to perform meaningful analyses. Biomarker data such as genetic variations, protein levels, and metabolic profiles can be high-dimensional and complex. This data is often not CDISC compliant. This paper explains a standardized integration process utilizing ADaM datasets and raw biomarker data. We produced a comprehensive, integrated dataset by creating a unified template and applying rigorous validation, enabling robust exploratory analyses. Our findings demonstrate the feasibility and potential of this approach to enhance clinical insights and therapeutic strategies. This work paves the way for more efficient and accurate biomarker and clinical data integration.

INTRODUCTION

Biomarker-driven clinical trials are widely adopted, with nearly half of the oncology studies and 16.5% of all clinical trials[4]. They produce vast amounts of data from various biospecimens through multiple assay modalities, often leading to data silos. The rapid generation of large data volumes presents significant challenges to linking data across modalities and centralizing reports. These are needed for timely access to critical information for decision-making during trials. Establishing a centralized, searchable biomarker database at the commencement of a trial that integrates clinical, biomarker, and sample data allows study teams to collaborate effectively with access to a cohesive dataset. This setup facilitates quality control (QC), data transformation, and the identification of trends across subjects, time points, treatment groups, and response statuses.

This paper explores key challenges and strategies in managing high-dimensional biomarker data in exploratory analyses, where biomarkers are increasingly prevalent. While not required to submit to regulatory agencies or meet CDISC standards, these biomarker datasets can still benefit from a CDISC-like structure to improve analytical efficiency and traceability. We address gaps in current practices, such as relying on email and SharePoint for clinical data requests and storage, performing limited QC through sporadic spot checks, and struggling with the non-standard, high-dimensional data format. This paper highlights the critical need for a robust and standardized integration process to ensure the integrity and reliability of biomarker data in clinical research.

DATA INTEGRATION KEY ELEMENTS

An Exploratory Analysis Plan (EAP) details the key questions and hypotheses about biomarkers to be used and their intended functions in exploratory analyses. It describes the analysis approach of the exploratory data and is aligned with the exploratory objectives and endpoints outlined in a clinical trial protocol.

Biomarker data could encompass cell surface markers, cytokine/chemokine levels, gene signatures, immune cell subsets, and activation markers. They can be characterized using gate definitions, source sample types, and reagents. The combination of these characteristics can yield several thousand unique combinations. In our proof-of-concept, we leveraged Cytometry by Time-of-Flight (CyTOF) data, though this approach applies equally to other biomarker datasets. CyTOF is a recently introduced technology that uses mass spectrometry to measure the abundance of metal isotope labels on antibodies and other tags, such as peptide-MHC tetramers, for labeling specific T cells on single cells. This advanced method provides unparalleled depth and breadth in cellular analysis, uniquely positioned to explore immune responses, particularly in oncology.

We also reference ‘clinical data’ collected in the clinical database as part of the clinical trial conduct. A comprehensive dataset was constructed with key demographics, baseline characteristics, safety, and efficacy variables.

The statistical open-source software R provides a robust platform for processing the huge volume of data generated during biomarker analysis.

Amazon Simple Storage Service (Amazon S3) is a scalable, high-performance, secure cloud object storage service offered by AWS that allows users to store and retrieve data from anywhere. Convenient R packages are available to retrieve and upload data to and from Amazon S3 buckets.

SHAPING YOUR DATA

Now, given a bunch of biomarker data and related clinical data available, how would you integrate them into a robust dataset for exploratory analysis? Would you opt for a long or wide format? Would you split the datasets by a meaningful parameter? Would you combine multiple variables into one? We sought answers to these questions based on the analysis outlined in the EAP, aiming to produce a standardized integrated biomarker dataset.

Understanding dataset orientations and effectively transforming data into an appropriate format is essential for effective data manipulation, analysis, and visualization. In a wide format, results for various biomarkers across visits are presented as columns, making them visually interpretable. Whereas the long format organizes data over multiple rows, enabling easier aggregated analysis. Given the high-dimensional nature of biomarker data, creating additional analysis variables can dramatically increase data points. A wide format could result in datasets with thousands of variables, making the long format more practical and appropriate. However, can we also leverage the advantages of the wide format? Let us explore together.

Another aspect of data integration is managing multiple analysis parameters. Biomarker characteristics, such as gate definitions, marker strings, source sample types, and reagents, can form distinct data subsets or combine to define analysis parameters like mean intensity or percent cells. The decision to retain all data points in a single dataset or split them into separate datasets depends on convenience, storage considerations, and analysis requirements. For ongoing analyses where data is consistently updated as the study progresses, the integrated dataset must also be refreshed subsequently. In such scenarios, retaining all data points in a single dataset is often the best approach for minimal downstream impact.

We thus adopted an approach inspired by the CDISC ADaM BDS structure, which organizes data in a long format. The PARAM and PARAMCD variables facilitate easy filtering of analysis parameters. In contrast, variables like BASE and CHG and result variables enable seamless analysis and comparisons across visits. In addition to these standard CDISC variables, we included the result, baseline, change from baseline, and fold change over baseline variables for two specific biomarker parameters: event count and percent of leukocytes. This slight deviation from CDISC allowed us to combine the analytical advantages of the ADaM structure with the visual benefits of a wide format.

PROCESS FLOW

Generating an integrated biomarker dataset starts with securing access to work environments and analysis documents. The programming team creates the Clinical Baseline Dataset specification and dataset guided by the statistician based on the exploratory analysis plan. Once the biomarker data is received from the vendor, the primary programmer combines it with the Clinical Baseline Dataset to create the Integrated Biomarker Dataset, following the specifications. A QC programmer independently validates this dataset and generates a comparison report. A clean report finalizes the Integrated Biomarker Dataset (Figure 1).

DATA FLOW

The data flow begins with creating a Clinical Baseline Dataset using SAS from SDTM/ADaM data. The Clinical Baseline Dataset and biomarker vendor datasets are uploaded to a specified, secure location, such as a refined Amazon S3 bucket. Both datasets are then pulled from the refined bucket, processed, and integrated in a validated R programming environment (i.e., Posit Workbench). This produces the final biomarker analysis dataset, which is then uploaded back to the designated location on the refined bucket for further use (Figure 2).

Figure 1: Biomarker Data Integration Process Flow Steps

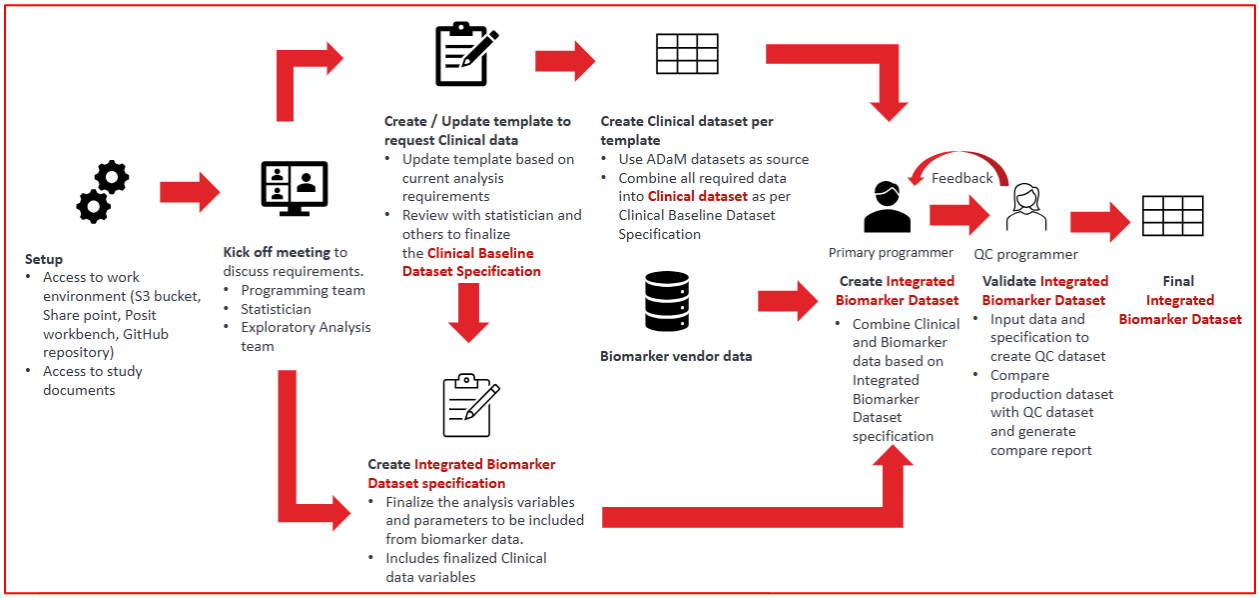
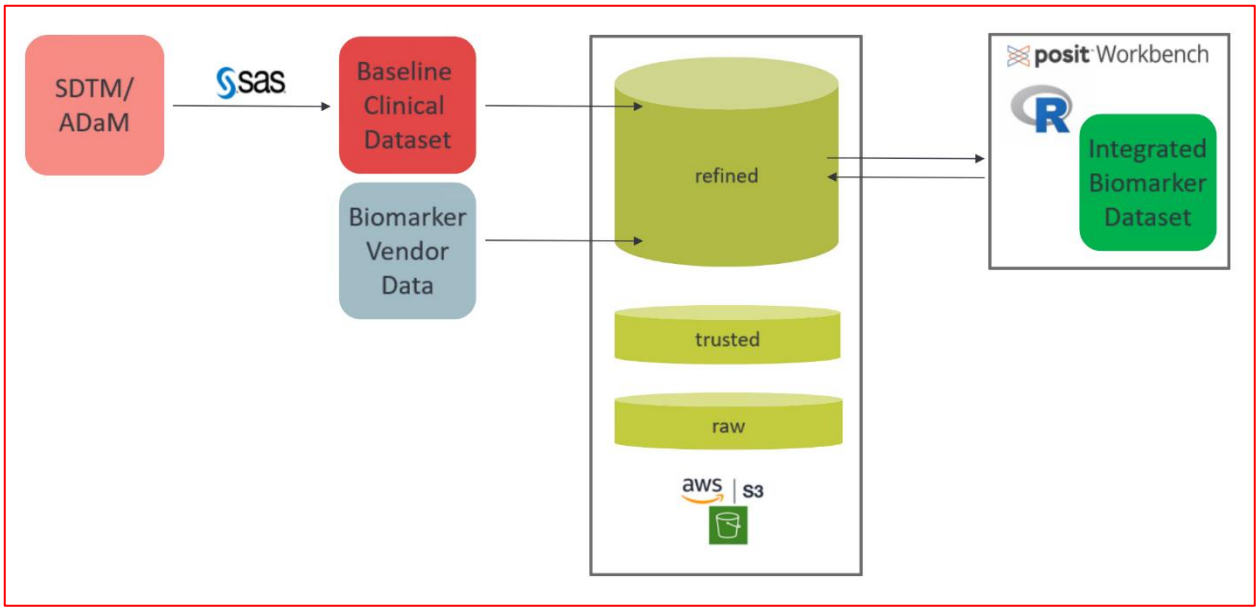


Figure 2: The Biomarker Data Integration POC Data Flow



CLINICAL BASELINE DATASET

Clinical Baseline Dataset specification (Table 1) is a template for requesting clinical data, functioning similarly to the data transfer specification or agreement (DTA) used during the integration of EDC data with external vendor data. The document is maintained in Microsoft Excel and consists of columns that aid in generating the Data Definition files used in submissions. A list of variables is created, and their metadata and derivation methods are detailed. The programming team develops and finalizes the specification in line with the EAP and informs the development of the Clinical Baseline Dataset.

The Clinical Baseline Dataset serves as a repository for all subject-level attributes and clinical measurements and adheres to a one-record-per-subject structure. It is a one-stop-shop for all the data that aids in meaningful biomarker data analysis. Data is sourced from several SDTM/ ADaM domains, including ADSL, ADEX, ADTTE, ADCM, GF, etc. Baseline parameters, such as those for vital signs, questionnaires, laboratory parameters, immune response, etc., can be used to stratify the population and identify biomarkers specific to subgroups. Demographic variables such as age, sex, race, country, and ethnicity provide the essential context for the study population and potential subgroup analyses. Further, exposure-related variables such as those to study drugs and other relevant medications are

essential data points for exploratory research. Efficacy data points that serve as study endpoints are also included in this dataset. Since SDTM/ADaM datasets are used as sources, traceability is improved. Structural metadata is maintained to ensure consistency with source datasets. However, the length allowed for variable names and labels is set to 15 characters and 100 characters length to accommodate exploratory analysis. Controlled terminology is applied to ensure uniformity and enable pattern identification. Leveraging standardized datasets to construct the Clinical Baseline Dataset optimizes efficiency and advances the workflow.

Table 1: Example of Clinical Baseline Dataset Specification

Dataset	Order	Domain	Variable Name	Variable Label	Type	Expected Length	Takeda Core	Controlled Terms	Format	Origin	Predecessor Name	Derivation	Assigned Value	Comments
CLINDSBL	1	ADSL	USUBJID	Unique Subject Identifier	Char	50	Req			Predecessor	ADSL.USUBJID			
CLINDSBL	3	ADSL	TRT01A	Actual Treatment for Period 01	Char	100	Req			Predecessor	ADSL.TRT01A			
CLINDSBL	4	ADSL	AGE	Age	Num	8	Req			Predecessor	ADSL.AGE			
CLINDSBL	6	ADSL	SEX	Sex	Char	10	Req	(Female, Male)		Predecessor	ADSL.SEX			
CLINDSBL	7	ADSL	RACE	Race	Char	50	Req	(ASIAN, BLACK OR AFRICAN AMERICAN, WHITE, NOT REPORTED)		Predecessor	ADSL.RACE			
CLINDSBL	8	ADSL	ECOGCAT	ECOG Category	Char	10	Perm	(0-1, 2, Missing)		Derived				If ADSL.ECOGBL in (0,1) then set to "0-1" else if ADSL.ECOGBL eq 2 then set to "2" else if ADSL.ECOGBL is missing then set to "Missing"
CLINDSBL	14	ADEX	CD38E90D	Exp to Anti-CD38 w/i 90 Days Prior Trt	Char	1	Perm	(NY)		Derived				If ADEX.EXDOSE is not missing and ADEX.EXTRT is "Treatment" and ADEX.ANL01FL eq "Y" then set to "Y", else set to "N"
CLINDSBL	39	ADTTE	DORMINV	Dur of Response per INV (months)	Num	8	Req			Derived				Set to ADTTE.AVAL where ADTTE.PARAMCD eq "DORM"
CLINDSBL	51	ADVS	WEIGHT	Baseline Weight (kg)	Num	8	Perm			Derived				Set to ADVS.AVAL where ADVS.PARAMCD = "WEIGHT" and ABLFL eq "Y"

THE DATA WRANGLING JOURNEY

The integration process involved examining the study's clinical database and vendor-transferred biomarker raw datasets to propose an effective integrated biomarker dataset structure. The integrated datasets were produced on the validated R environment by automating the importation of various source files, performing data cleaning and manipulations, and uploading the integrated data back to the central repository.

SETTING UP THE WORK ENVIRONMENT

The initial step involves connecting Posit Workbench and an Amazon S3 bucket for secure and organized data storage, management, and transfer. This requires configuring access by setting up AWS credentials, including AWS Access Key ID, Secret Access Key, and Session Token. These credentials can be configured within the Posit Workbench terminal using commands such as `aws configure`. For data transfer between Posit and the Amazon S3 bucket, R users can utilize the `aws.s3` package. For example, the command `s3readusing(FUN=read.csv, object='path/to/s3object', bucket='bucketname')` supports seamless data import into R, and while `putobject(file='path/to/output', bucket='bucketname', object = 'path/to/s3bucket')` facilitates the upload of output files back to the S3 bucket.

Integrating GitHub with Posit enhances collaboration and version control. This integration enables efficient code collaboration and tracking changes through `git` commands. These configurations collectively ensure streamlined data management, reproducibility, and teamwork within a robust statistical programming environment. Study teams can adapt and refine these steps to align with their specific requirements and workflows.

BIOMARKER RAW DATA

Our current dataset was produced by analyzing specimens such as RBC-lysed whole Blood and Bone Marrow Assay (BMA) using gate definitions and markers. This produced results such as Mean, Percent of parent Population results, event counts, and Leukocyte Counts.

The raw biomarker dataset consists of the Subject ID (SUBJID) variable for subject identification and variables such as Gate Definition (LBGATDEF), Marker String (LBMRKSTR), Test Detail (LBTSTDTL), etc. as the biomarker identifiers. Specimen and reagent-related variables are also present. The dataset further includes result variables relevant to the tests conducted. The structure of the dataset is one record per subject per biomarker per test per timepoint. Multiple result variables are present for this combination (Table 2, 3).

Gate Definition refers to the process of identifying and isolating specific cell populations from a complex sample (like blood or tissue) based on their unique characteristics, such as size, granularity, and the presence or absence of surface markers[1,2].

A marker string is a notation used to describe the specific combination of cell surface markers expressed (positive) or not expressed (negative) by a particular population of cells. It is commonly used in immunophenotyping, flow cytometry, or CyTOF analyses to identify and classify cell populations. A marker string acts like a "fingerprint" to identify specific cell types [3].

Reagents are antibodies conjugated to fluorescent dyes or metal isotopes used to detect specific cell surface markers. Each reagent is designed to bind to a unique marker on the cell surface, allowing us to identify cell types, measure their activation states or functionality[5].

Table 2: Variables of the raw biomarker data

Variable Name	**Description**
SUBJID	Subject ID
LBREFID	Sample ID
LBGATDEF	Gate Definition
LBMRKSTR	Marker String
LBTSTDTL	Test Detail (e.g., “mean” or “percent”)
REAGMMI	Metallic Based Mean Intensity Reagent
LBSPEC	Specimen Type
VISIT	Visit
LBORRES	**Original Result**
PCTLEUK	**Percent of Leukocytes**
EVNTCT	**Event Count**

Table 3: Raw biomarker data with an example of one subject

LBSBMRKS	LBGATDE	LBMRKSTR	LBTSTDTL	REAGMMI	LBORRES	LBSPEC	LBMETHOD	PCTLEUK	EVNTCT	VISIT
mDC	mDC	CD14-CD16-CD56-CD3-CD19-	mean	162Dy_CD86	16.46166151	RBC Lysed Whole Blood	CyTOF	0.116477285	274	V21 C5D1
mDC	mDC	CD14-CD16-CD56-CD3-CD19-	mean	172Yb_CD83	0.435098903	RBC Lysed Whole Blood	CyTOF	0.116477285	274	V21 C5D1
pDC	pDC	CD14-CD16-CD56-CD3-CD19-	mean	162Dy_CD86	5.920483597	RBC Lysed Whole Blood	CyTOF	0.017003983	40	V21 C5D1
pDC	pDC	CD14-CD16-CD56-CD3-CD19-	mean	172Yb_CD83	0.28795932	RBC Lysed Whole Blood	CyTOF	0.017003983	40	V21 C5D1
mDC	mDC	CD14-CD16-CD56-CD3-CD19-	mean	162Dy_CD86	16.47284876	RBC Lysed Whole Blood	CyTOF	0.138468238	283	V2 Baseline
mDC	mDC	CD14-CD16-CD56-CD3-CD19-	mean	172Yb_CD83	0.83951769	RBC Lysed Whole Blood	CyTOF	0.138468238	283	V2 Baseline
pDC	pDC	CD14-CD16-CD56-CD3-CD19-	mean	162Dy_CD86	6.12556888	RBC Lysed Whole Blood	CyTOF	0.036696529	75	V2 Baseline
pDC	pDC	CD14-CD16-CD56-CD3-CD19-	mean	172Yb_CD83	0.810534398	RBC Lysed Whole Blood	CyTOF	0.036696529	75	V2 Baseline

For illustration purposes, we will focus on handling Mean Metal Intensity (MMI) data (where LBTSTDTL = “mean”) only.

PROCESSED BIOMARKER DATA

An integrated biomarker dataset specification was developed to guide the creation of the final dataset. The specification details a list of variables from the Clinical Baseline Dataset, ADaM BDS variables like PARAM, PARAMCD, BASE, and CHG, and other time point and biomarker-identifying variables. Metadata and the corresponding derivation methodologies for each variable are also specified. This document assists in the streamlined creation of an Integrated Biomarker dataset.

As part of data preparation, a few variables from the biomarker data were re-mapped to new variable names to ensure clarity, and variables such as visit, and study-specific controlled terminology were applied to them for data standardization. For example, the TIMEPOINT variable was derived to distinguish baseline records from cycle visits and to include dosing references such as “PREDOSE”. For variables retained names from the source data, such as LBGATDEF, LBMRKSTR, and REAGMMI, labels describing the intended purpose were provided to ensure interpretability to end-users. Additional identifiers derived, such as the TESTDETAIL variable, were created to aid in efficient data screening. This variable was developed using source variables, including LBGATDEF, LBMRKSTR, and LBTSTDTL. Further, descriptive identifiers for each test performed are provided through the PARAM and PARAMCD variables. The primary result corresponding to a parameter was captured under the ORIGRES variable. The variable name deviates from standard ADaM to be able to differentiate from other result variables pertinent to a parameter. In this study, Event Count and Percent of Leukocytes were additional results maintained in the wide format. The result variables corresponding to these results were named EVNTCTORIGRES and PCTLEUKORIGRES. Usage of baseline (BASE), change from baseline (CHG), and fold change from baseline (FCHG) variables increase efficiency

in handling data across time points and test types. The structured dataset design optimizes compatibility with statistical and visualization tools, allowing for more effective cross-functional analyses in a regulated clinical research environment (Table 4, 5).

The first key step in data processing is to derive baseline results by filtering the records at the initial time point.

```
dat_long_baseline <- dat_long %>%
  filter(timepoint=="C1D1_0 PREDOSE") %>%
  select(subject_id, specimen_type, LBGATDEF, LBMRKSTR, LBTSTDTL, TESTDETAIL,
         REAGMMI, PARAMCD, PARAM, MMI, EVNTCT, PCTLEUK) %>%
  rename(BASE=MMI, EVNTCTBASE=EVNTCT, PCTLEUKBASE=PCTLEUK)
```

This is the foundation for calculating change from baseline and fold change over baseline as in the code below. Baseline data was merged with the main dataset using key variables and then calculated the change from baseline (CHG) as {original result – baseline result}, and fold change over baseline (FCHG) as {original result / baseline result}. These statistics were calculated for all three result variables - Original Results, Event Count, and Percent of Leukocytes.

```
dat_long_merged <-
  left_join(dat_long, dat_long_baseline,
            by=c("subject_id","specimen_type","LBGATDEF","LBMRKSTR",
                 "LBTSTDTL","TESTDETAIL","REAGMMI","PARAMCD","PARAM")) %>%
  rename(ORIGRES=MMI, EVNTCTORIGRES=EVNTCT, PCTLEUKORIGRES=PCTLEUK) %>%
  mutate(CHG=ORIGRES-BASE,
         EVNTCTCHG=EVNTCTORIGRES-EVNTCTBASE,
         PCTLEUKCHG=PCTLEUKORIGRES-PCTLEUKBASE) %>%
  mutate(FCHG=ifelse(!is.na(BASE) & !is.na(ORIGRES) & BASE != 0,ORIGRES/BASE,NA),
         EVNTCTFCHG=ifelse(!is.na(EVNTCTBASE) & !is.na(EVNTCTORIGRES) & EVNTCTBASE !=
0,EVNTCTORIGRES/EVNTCTBASE,NA),
         PCTLEUKFCHG=ifelse(!is.na(PCTLEUKBASE) & !is.na(PCTLEUKORIGRES) & PCTLEUKBASE !=
0,PCTLEUKORIGRES/PCTLEUKBASE,NA))
```

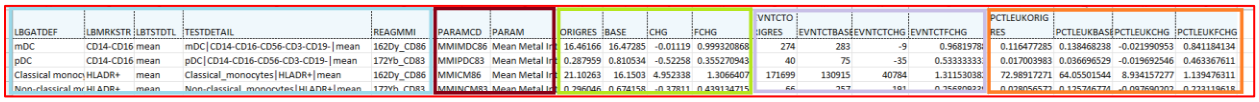
Table 4: Processed dataset variables

Variable Name	Description
SUBJID	Subject ID
SAMPLEID	**Sample ID**
LBGATDEF	Gate Definition
LBMRKSTR	Marker String
LBTSTDTL	Test Detail (e.g., “mean” or “percent”)
REAGMMI	Metallic Based Mean Intensity Reagent
TESTDETAIL	**Combined Value of Gate Definition Marker String Test Detail**
SPECIMENTYP	**Specimen Type**
TIMEPOINT	**Timepoint**
PARAM	**Parameter**
PARAMCD	**Parameter Code**
ORIGRES	**Original Result**
BASE	**Baseline Result**
CHG	**Change from Baseline**
FCHG	**Fold Change over Baseline**
EVNTCTORIGRES	**Event Count Original Result**
EVNTCTBASE	**Event Count Baseline Result**
EVNTCTCHG	**Event Count Change from Baseline**
EVNTCTFCHG	**Event Count Fold Change over Baseline**

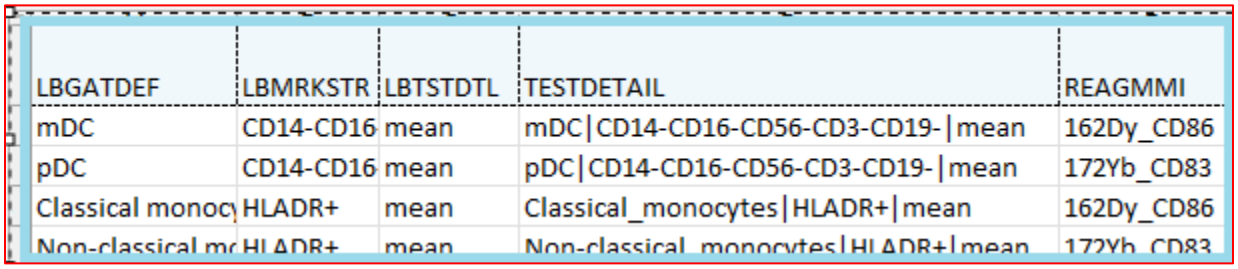
PCTLEUKORIGRES	**Percent of Leukocytes Original Result**
PCTLEUKBASE	**Percent of Leukocytes Baseline Result**
PCTLEUKCHG	**Percent of Leukocytes Change from Baseline**
PCTLEUKFCHG	**Percent of Leukocytes Fold Change over Baseline**

The original result, baseline, change from baseline, and fold change over baseline are derived respectively for intensity, event count, and percent of leukocytes. Unlike the traditional ADaM dataset structure, which follows a long-format approach where all test information is grouped into parameters, we retained key details in separate columns for this dataset. This design enhances clarity, streamlines analysis, and simplifies reporting using tools like R Shiny. Furthermore, we used ORIGRES instead of AVAL to preserve the original result values, ensuring a more illustrative and meaningful data representation for the study team. The tailored approach maintains key ADaM principles while addressing specific analysis and reporting needs. Below are snippets from the integrated dataset along with a magnified view (Table 5).

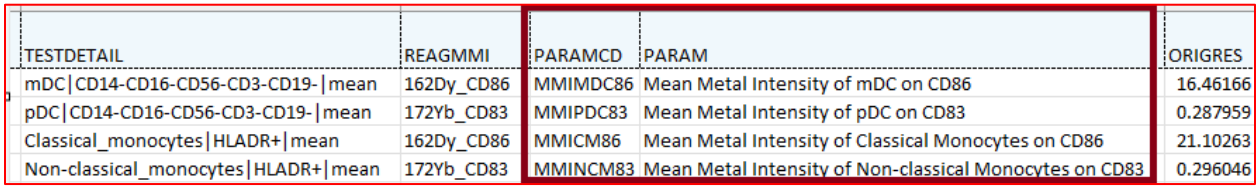
Table 5: Integrated Biomarker dataset of the same subject data for one parameter



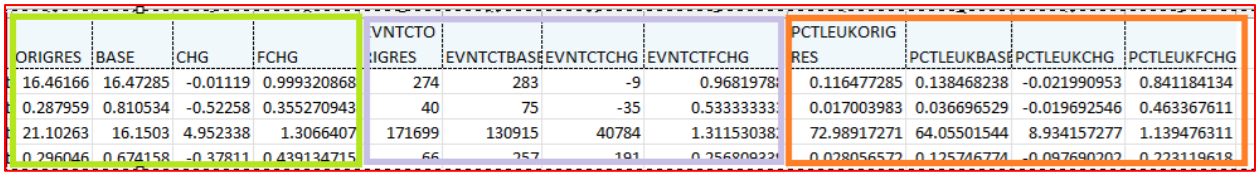
Below are snippets in a magnified view



LBGATDEF	LBMRKSTR	LBTSTDTL	TESTDETAIL	REAGMMI
mDC	CD14-CD16	mean	mDC|CD14-CD16-CD56-CD3-CD19-|mean	162Dy_CD86
pDC	CD14-CD16	mean	pDC|CD14-CD16-CD56-CD3-CD19-|mean	172Yb_CD83
Classical monocy	HLADR+	mean	Classical_monocytes|HLADR+|mean	162Dy_CD86
Non-classical mc	HLADR+	mean	Non-classical_monocytes|HLADR+|mean	172Yb_CD83



TESTDETAIL	REAGMMI	PARAMCD	PARAM	ORIGRES
mDC|CD14-CD16-CD56-CD3-CD19-|mean	162Dy_CD86	MMIMDC86	Mean Metal Intensity of mDC on CD86	16.46166
pDC|CD14-CD16-CD56-CD3-CD19-|mean	172Yb_CD83	MMIPDC83	Mean Metal Intensity of pDC on CD83	0.287959
Classical_monocytes|HLADR+|mean	162Dy_CD86	MMICM86	Mean Metal Intensity of Classical Monocytes on CD86	21.10263
Non-classical_monocytes|HLADR+|mean	172Yb_CD83	MMINCM83	Mean Metal Intensity of Non-classical Monocytes on CD83	0.296046



ORIGRES	BASE	CHG	FCHG	VNTCTOIGRES	EVNTCTBASE	EVNTCTCHG	EVNTCTFCHG	PCTLEUKORIGRES	PCTLEUKBASE	PCTLEUKCHG	PCTLEUKFCHG
16.46166	16.47285	-0.01119	0.999320868	274	283	-9	0.9681978	0.116477285	0.138468238	-0.021990953	0.841184134
0.287959	0.810534	-0.52258	0.355270943	40	75	-35	0.5333333	0.017003983	0.036696529	-0.019692546	0.463367611
21.10263	16.1503	4.952338	1.3066407	171699	130915	40784	1.3115303	72.98917271	64.05501544	8.934157277	1.139476311
0.296046	0.674158	-0.37811	0.439134715	66	257	191	0.2568093	0.028056572	0.125746774	-0.097690202	0.223119618

DATA VALIDATION

The integrated biomarker dataset is validated through a parallel programming QC. A QC programmer utilizes the source datasets and Integrated Biomarker dataset specification to recreate the dataset in a validated R programming environment independently. Adherence to the specification, consistency, and completeness of data and their metadata are corroborated, and discrepancies are resolved. A comparison report is generated from the R environment, and a clean report finalizes the Integrated Biomarker Dataset generation. This validation process minimizes the risk of critical errors while strengthening the reliability and accuracy of the dataset.

CONCLUSION

In conclusion, establishing a centralized biomarker database is paramount for effective data analysis and exploratory research. The extensive, high-dimensional data produced in biomarker-driven trials requires a streamlined quality control and standardization process. The integrated biomarker dataset, combining clinical and biomarker data specifications, facilitates robust data management, ensuring the integrity and reliability of analyses. Advanced methods like CyTOF provide in-depth cellular analysis, which is crucial for understanding immune responses in oncology. Implementing standardized processes in clinical research is essential for producing cohesive datasets supporting decision-making and future exploratory analyses. Centralized databases and advanced analytical tools promote improved collaboration and data consistency, ensuring the success and effectiveness of biomarker-driven clinical trials.

REFERENCES

1. *Flow Cytometry gating for beginners*. (2023, September 27). Antibodies | ELISA kits | Proteins | Proteintech Group. https://www.ptglab.com/news/blog/flow-cytometry-gating-for-beginners/
2. *Gating: First steps*. (2021, June 26). LearnHaem | Haematology Made Simple. https://www.learnhaem.com/courses/flow-cytometry/lessons/gating-first-steps/
3. *A guide to flow Cytometry markers*. (2023, August 30). BD Biosciences | Flow Cytometry Instruments and Reagents. https://www.bdbiosciences.com/en-sg/learn/science-thought-leadership/blogs/flow-cytometry-markers
4. *How biomarkers impact clinical trial study start-up*. (2024, December). https://www.precisionformedicine.com/blog/how-biomarkers-impact-clinical-trial-study-start-up/
5. *Part 2. CyTOF assays*. (n.d.). Institute for Immunity,Transplantation and Infection. https://iti.stanford.edu/himc/immune-monitoring-virtual-course-2020/part-2.html

ACKNOWLEDGMENTS

We thank our cross-functional colleagues for their guidance and support during the conception and implementation of the biomarker data integration process. We also thank Maria Dalton, Venky Chakravarthy, Samanthi Perera and Kevin Galinsky for their valuable feedback on this paper.

CONTACT INFORMATION

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

Author Name: Pavan Kumar Tatikonda
Company: Takeda Development Center Americas Inc.,
Address: 650 E Kendall St, Cambridge, MA, 02142
Email: pavan.tatikonda@takeda.com
Website: https://www.takeda.com/

Author Name: Xiaoyi Zhu
Company: PPD, part of Thermo Fisher Scientific
Address: 929 North Front Street, 6th Floor, Wilmington, NC 28401-3331, United States
Email: sherry.zhu1@thermofisher.com
Website: https://www.ppd.com/

Author Name: Sunithra Kadambi
Company: Phastar Inc.
Address: 2D Bollo Ln, Chiswick, London W4 5LE, United Kingdom
Email: sunithra.kadambi@phastar.com
Website: https://phastar.com/

Brand and product names are trademarks of their respective companies.