

Homogenizing Unique and Complex data into Standard SDTM Domains with TAUGs

Sowmya Srinivasa Mukundan and Charumathy Sreeraman, Efficacy Lifescience Analytics

ABSTRACT

Clinical research supports discovery of new and better ways to detect, diagnose, treat and prevent disease. Furthermore, the core focus of each therapeutic area (TA) is on research and development of treatments, together with prevention of specific diseases. It must be envisaged that each TAs demands diverse way of collecting, measuring and analyzing data based on the focus of the research. SDTM is one of the pioneer CDISC foundational standards. It defines and underpins the strategy for submitting data tabulations to regulatory authorities. The SDTMIG organizes and formats data to support streamlined data collection and analysis across different TAs. TAUGs are extended Foundational Standards to represent data that pertains to specific disease areas. It supports pharmaceutical / biotech companies with implementation of these CDISC standards for a specific disease and facilitate resolutions for mapping additional or unique data points needed to support any given TA for their analysis. In this paper we will explore the TAUG (focusing on two different Therapeutic Areas) and will be elaborately discussing how to map the unique and custom data to the standard SDTM domains with the help of TAUG.

INTRODUCTION

Therapeutic Area (TA) in a simple way it is defined as a Clinical discipline. It is a ground that focus on the treatment and prevention of diseases which undesirably impact the health of living being. Data collection and analysis differs when there is a change in the therapeutic areas.

TAUG provides models of CDISC implementation of TA specific data that cannot be found in existing CDISC documentation. TAUG development identifies gaps in CDISC models which are then taken back to respective team for consideration and incorporation. There are many user guidelines for different therapeutic areas published by CFAST Team. The TA Standards include disease-specific metadata, examples and guidance on implementing CDISC standards for a variety of uses, including global regulatory submissions. Already there are many TA user guidelines available.

In this paper, we will focus on the Breast Cancer and Vaccines TA. The TAUG-BrCa is a standard guideline to use CDISC standards to represent data pertaining to breast cancer studies. It consists of metadata file, CRF examples, SDTM dataset mapping and ADaM dataset mapping examples. TAUG-Vax is a standard guideline relevant to Vaccine studies which consists of metadata file, SDTM dataset mapping.

TAUG

TAUG is the acronym for Therapeutic Area User Guide. TAUGs are developed under the Coalition for Accelerating Standards and Therapies (CFAST) initiative. CFAST, a joint initiative of the Clinical Data Interchange Standards Consortium (CDISC) and the Critical Path Institute (C-Path), was launched to accelerate clinical research and medical product development by facilitating the establishment and maintenance of data standards, tools, and methods for conducting research in therapeutic areas.

As standards work takes place across the TA teams, there are regularly new collection and submission domains and/or variables identified that may be needed to support any given therapeutic area. These new articles are evaluated by the SDS teams for inclusion in the next publication of the standard. In the meantime, any new domain or variable in a TAUG would be designated as "Provisional" until such time as they appear in a succeeding publication of the standard.

The design and mapping of specialized data to current standards provides many opportunities to implement new and different submission strategies, while remaining compliant with the published standard. The target

of the CFAST initiative is to identify a core set of clinical therapeutic area biomedical concepts and endpoints for targeted therapeutic areas and translate them into CDISC standards to improve semantic understanding, support data sharing, and facilitate global regulatory submission.

TAUG – BRCA

OVER VIEW OF BREAST CANCER

Breast cancer is a disease in which malignant (cancer) cells form in the tissues of the breast. It is the most common cancer among women, although it exists in men in rare cases. Most breast cancers, almost 95% are cancer tumors that develop in the milk ducts.

Breast cancers can be caused by a variety of factors such as, the heredity factor, Menstruation and age, the diet relationship. Traditionally, cancers have been treated with a variety or a combination of methods. The most typical therapies are surgery, radiation therapy, chemotherapy, hormone therapy, radioactive substances, or immunotherapy. With breast cancers, almost all cases can be treated, but only on the condition that they are discovered at an early stage, mainly through physical inspection or through mammography. The most common method of dealing with a breast cancer is mastectomy where the cancer cells are removed through surgery.

TAUG-BRCA

TAUG-BrCa v1.0 describes common kinds of data relevant for breast cancer studies, so that those handling the data (e.g., data managers, statisticians, programmers) understand the data and can apply standards appropriately.

It represents Non-Standard Variables (NSVs) appended to the end of the parent domain and enable use of consistent terminology for aggregation, comparison of data across studies and drug programs. Followed by sample value-level metadata for the NSVs.

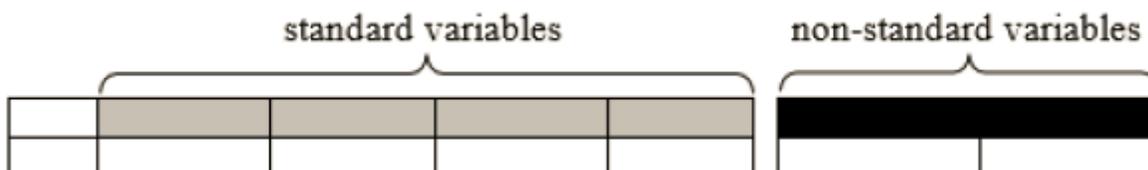


Figure 1. Figure Presentation of Non-standard Variables

The SDTM domains specific to Breast cancer studies is broadly classified in the below table.

	Domains
Disease Characteristics	MI, CM, PR, AE, MH
Disease managements and Assessments	TU, TR, RS
Associated Person data collection	APDM, APMH

Table 1. SDTM Domain Classification

Disease Characteristics

The Disease characteristics is broadly classified as shown in the below figure.

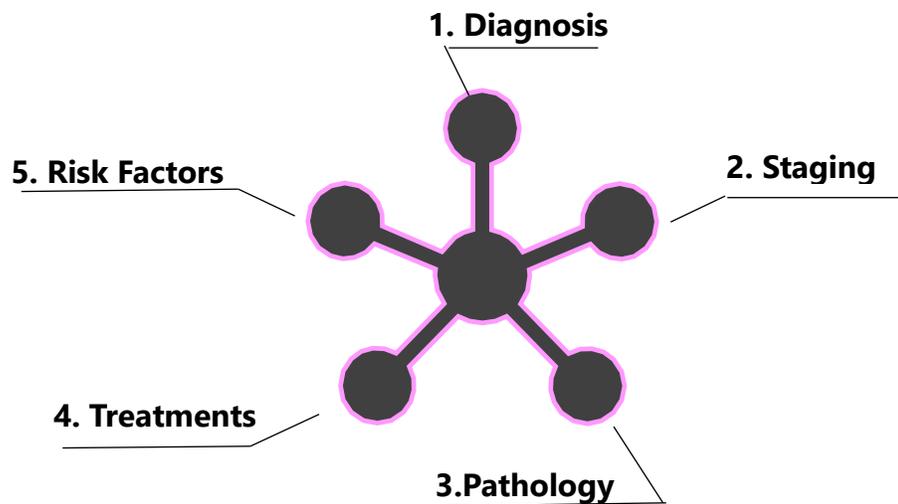


Figure 2. Disease Characteristics

Diagnosis

The Diagnosis of a disease is categorized based on the Presence of the Breast Cancer, Type of Breast Cancer and Severity or the Extent of the Breast Cancer. Determination on the severity/extent of the BrCa the diagnosis part is usually captured in Finding domains.

The following tests and procedures may be used to diagnose:

- Physical exam **and** history
- Clinical breast exam (CBE)
- Mammogram
- Ultrasound exam
- MRI (magnetic resonance imaging)
- Blood chemistry studies
- Biopsy
- Fine-needle aspiration (FNA) biopsy
- Core biopsy
- Excisional biopsy

Staging

Staging is defined as the Extent to which the malignancy has spread in the body. The staging system of a breast cancer is based on TNM classification. T, N, and M stand for primary **tumor**, regional lymph **node**, and distant **metastasis**.

T (TumorSize)	T1 (Tumor Size <2cm)	T2 (Tumor Size 2-5 cm)	T3 (Tumor Size >5 cm)	T4 (Tumor extent to other parts)
N (Lymph Nodes)	N0 (No Lymph node metastasis)	N1 (Metastasis to ipsilateral, movable, axillary LNs)	N2 (Metastasis to ipsilateral fixed, axillary, or IM, LNs)	N3 (Metastasis to infraclavicular IN, or to axillary and IM LNs)
M (Metastasis)	M0 (No distant metastasis)	M1 (Distant metastasis)		

Table 2. TNM Classification

Pathology

Pathologic assessments are both macroscopic and microscopic and provide further detail as to the specific type and aggressiveness of the cancer. Pathologic assessment of breast cancer includes immunohistochemistry (IHC) assessment for estrogen receptor (ER) and progesterone receptor (PR) as well as IHC or *in situ* hybridization (ISH) determination of human epidermal growth factor receptor 2 (HER2) status. Mostly, this information is collected in MI or TU dataset.

Prior Treatments

Any prior treatments or newly diagnosed with breast cancer will be capture either in CM or PR domains. Using TAUG, the non-standard variables TRTINT(Treatment Intent), TRTSTT(Treatment Setting) and RSDISC(Reason for Discontinuation) will be added to CM domain.

Risk Factors

Any major diseases other than Breast cancer considered as a Major Comorbid Conditions and history of family members will be captured in AP domain.

Disease Managements and Assessments

Treatment of Breast Cancer

Local treatment affects cancer cells in the tumor and the area near it .The local treatments are,

- Surgery
- Radiation therapy

Systemic treatment travels through the bloodstream reaching cancer cells all over the body. The Systemic treatments are,

- Chemotherapy
- Hormone therapy
- Bone Directed
- Targeted therapy

Tumor Identification, Assessment and Disease Response

Tumor Identification (TU), Tumor Results (TR), and Disease Response (RS) are already present in SDTMIG in finding Domains. Despite that, a few variables are identified in TAUG metadata is captured in the below table 3.

TU	TR	RS
TUYN	TRDIAM	RSYN
TULOCDTL	TRDIAMU	OVRLRESP
TUCHANGE	TRTOOSM	OVRLDAT
TUDAT	TRINEVAL	TRGRES
	TUMSTATE	

Table 3. TAUG Metadata Variables

The below sample CRF annotated to show mapping for Tumor Identification/Results of Target Lesions, Disease Response and mapping of New Lesion. SDTM variables are annotated in Red. If CDASH variable differs from SDTM, the CDASH variables are in Blue.

Annotated CRF: Tumor Identification/Results Target Lesions

This CRF is only an example and is not meant to imply that any particular layout is preferable over another.

CRF annotated to show mapping. SDTM variables are in **Red**. If CDASH variable differs from SDTM the CDASH variable is in **Blue**. * : new variable request submitted. Refer to the corresponding CDASH Metadata table for more information on Sponsor-related Implementation decisions and TA specific usage rules.

Target **TU** **TU**TESTCD=TUMIDENT and **TU**ORRES=TARGET

Response Criteria: RSCAT		RECIST 1.1	
<i>Pre-specified</i>			
Were tumors identified? TUYN*	<input type="checkbox"/> Yes <input type="checkbox"/> No	Tumor ID: TULNKID	(A10 or Sponsor-Defined CT)
Location: TULOC	<Select appropriate values from LOC CT>	Location Text: SUPPTULOCXT	TULOCTXT* (A200)
Laterality: TULAT	<input type="checkbox"/> Left <input type="checkbox"/> Right <input type="checkbox"/> Bilateral <Select appropriate values from LAT CT>	Directionality: TUDIR	<input type="checkbox"/> Distal <input type="checkbox"/> Intermediate <input type="checkbox"/> Proximal <input type="checkbox"/> Inner <input type="checkbox"/> Outer <Select appropriate values from DIR CT>
Changes to Tumor Identified: TUCHANGE*	Split TU TESTCD=TUSPLIT Merged TU TESTCD=TUMERGE		
Method of Evaluation: TUMETHOD	<input type="checkbox"/> Clinical Exam <input type="checkbox"/> Endoscopy <input type="checkbox"/> PET Scan <input type="checkbox"/> Physical Examination <input type="checkbox"/> CT Scan <input type="checkbox"/> Mammography <input type="checkbox"/> PET/CT Scan <input type="checkbox"/> Scintigraphy <input type="checkbox"/> Ductography <input type="checkbox"/> MRI <input type="checkbox"/> PET/MRI Scan <input type="checkbox"/> Ultrasound <input type="checkbox"/> DXA Scan <input type="checkbox"/> MUGA <input type="checkbox"/> Photography <input type="checkbox"/> X-Ray <input type="checkbox"/> Echocardiogram <Select appropriate values from METHOD CT>		
Date of Evaluation: (DD-MMM-YYYY) TUDTC	__/__/----	Diameter: TR TESTCD=DIAMETER TORRES	
Evaluator TUEVAL	<input type="checkbox"/> Investigator <input type="checkbox"/> Independent Assessor	Diameter Unit: TORRESU	mm TRDIAMU*
Evaluator Identifier: TUEVALID	<input type="checkbox"/> Radiologist 1 <input type="checkbox"/> Radiologist 3 <input type="checkbox"/> Radiologist 2 ...	Diameter Too Small to Measure: TORRES	<input type="checkbox"/> Yes TRTOOSM*
Tumor Inevaluable? TRINEVAL*	<input type="checkbox"/> Inevaluable TR TESTCD=TUMSTATE TRSTAT=NOT DONE	Lymph Node State: TR TESTCD=LNSTATE TORRES	<input type="checkbox"/> Pathological <input type="checkbox"/> Non-Pathological
If Tumor is Inevaluable, Reason Not Done: TRREASND	<input type="checkbox"/> Cavitation <input type="checkbox"/> Necrosis <input type="checkbox"/> Insufficient Images/Anatomy <input type="checkbox"/> Site Error <input type="checkbox"/> Fibrosis <input type="checkbox"/> Poor Scan Quality <input type="checkbox"/> Inconsistent Modality <input type="checkbox"/> Other		

Figure 3. Sample aCRF with CDASH and SDTM Mapping – TU & TR Domains

Annotated CRF: Disease Response

This CRF is only an example and is not meant to imply that any particular layout is preferable over another.

CRF annotated to show mapping. SDTM variables are in **Red**. If CDASH variable differs from SDTM, the CDASH variable is in **Blue**. *new variable request submitted. Refer to the corresponding CDASH Metadata table for more information on Sponsor-related Implementation decisions and TA specific usage rules.

Response Criteria: <i>Pre-specified</i> RSCAT	RECIST 1.1
Was the response assessment performed? RSYN	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No RSSTAT=NOT DONE where RSTESTCD=OVRLRESP
Reason Response Assessment Not Performed: RSREASND* RSREASND where RSTESTCD=OVRLRESP	<input type="checkbox"/> Not Imaged <input type="checkbox"/> Patient Refusal <input type="checkbox"/> Site Error <input type="checkbox"/> Other
Evaluator RSEVAL	<input type="checkbox"/> Investigator <input type="checkbox"/> Independent Assessor
Evaluator Identifier: RSEVALID	<input type="checkbox"/> Radiologist 1 <input type="checkbox"/> Radiologist 2 <input type="checkbox"/> Radiologist 3
Overall Response: OVRLRESP* RSTESTCD=OVRLRESP RSTEST=Overall Response	RSORRES (RSSTRESC) <input type="checkbox"/> Complete Response (CR) <input type="checkbox"/> Partial Response (PR) <input type="checkbox"/> Stable Disease (SD) <input type="checkbox"/> Non Complete Response/Non Progressive Disease (NON-CR/NON-PD) <input type="checkbox"/> Progressive Disease (PD) <input type="checkbox"/> Not Evaluable (NE)
Date of Procedure for Overall Response (e.g., scan date): (DD-MMM-YYYY) OVRLDAT* RSBTC	___/___/___

Figure 4. Sample aCRF with CDASH and SDTM mapping - RS Domain

Annotated CRF: Tumor Identification/Results New Lesions

This CRF is only an example and is not meant to imply that any particular layout is preferable over another.

CRF annotated to show mapping. SDTM variables are in **Red**. If CDASH variable differs from SDTM, the CDASH variable is in **Blue**. *new variable request submitted. Refer to the corresponding CDASH Metadata table for more information on Sponsor-related Implementation decisions and TA specific usage rules.

New **TUTESTCD=TUMIDENT** and **TUORRES=NEW**

Response Criteria: <i>Pre-specified</i>	RECIST 1.1
Were tumors identified? TUYN*	<input type="checkbox"/> Yes <input type="checkbox"/> No
Tumor ID: TULNKID	(A10 or Sponsor-Defined CT)
Location: TULOC	<Select appropriate values from LOC CT>
Laterality: TULAT	<input type="checkbox"/> Left <input type="checkbox"/> Right <input type="checkbox"/> Bilateral <Select appropriate values from LAT CT>
Directionality: TUDIR	<input type="checkbox"/> Distal <input type="checkbox"/> Intermediate <input type="checkbox"/> Proximal <input type="checkbox"/> Inner <input type="checkbox"/> Outer <Select appropriate values from DIR CT>

Figure 5. Sample aCRF with CDASH and SDTM Mapping - NEW Lesion

The below screenshots show the data collected in CE, FACE and VS domains on the reactogenicity events.

CELNKID	CELNKGRP	CETERM	CELLT	CELLTCD	CEDECOD	CEPTCD	CEHLT
DELTOID MUSCLE-LEFT-FATIGUE/TIREDNE..	Unplanned Visit..	FATIGUE/TIREDNE..	Fatigue	10016256	Fatigue	10016256	Asthenic condit..
DELTOID MUSCLE-LEFT-HEADACHE	Unplanned Visit..	HEADACHE	Headache	10019211	Headache	10019211	Headaches NEC
DELTOID MUSCLE-LEFT-NEW OR WORSEN..	Unplanned Visit..	NEW OR WORSEN..					
DELTOID MUSCLE-LEFT-NEW OR WORSEN..	Unplanned Visit..	NEW OR WORSEN..					
DELTOID MUSCLE-LEFT-PAIN AT INJECTIO..	Unplanned Visit..	PAIN AT INJECTIO..	Injection site p..	10022086	Injection site p..	10022086	Injection site re..
DELTOID MUSCLE-LEFT-REDNESS	Unplanned Visit..	REDNESS	Redness	10038198	Erythema	10015150	Erythemas
DELTOID MUSCLE-LEFT-SWELLING	Unplanned Visit..	SWELLING	Swelling	10042674	Swelling	10042674	General signs..
DELTOID MUSCLE-LEFT-VOMITING	Unplanned Visit..	VOMITING	Vomiting	10047700	Vomiting	10047700	Nausea and vo..

Figure 7. Screenshot of CE Domain

FALNKID	FALNKGRP	FATESTCD	FATEST	FAOBJ	FACAT	FASCAT	FAORRES	FAORRESU	FASTRESC	FASTRESN
DELTOID MUSCLE..	Unplanned Visit..	G4CRIMET	Grade 4 Criteri..	Redness	UNPLANNED..	ADMINISTRAT..	N		N	
DELTOID MUSCLE..	Unplanned Visit..	G4CRIMET	Grade 4 Criteri..	Swelling	UNPLANNED..	ADMINISTRAT..	N		N	
DELTOID MUSCLE..	Unplanned Visit..	MAXDIAM	Maximum Dia..	Redness	UNPLANNED..	ADMINISTRAT..	8	cm	8	8
DELTOID MUSCLE..	Unplanned Visit..	MAXDIAM	Maximum Dia..	Swelling	UNPLANNED..	ADMINISTRAT..	5	cm	5	5
DELTOID MUSCLE..	Unplanned Visit..	MINDIAM	Minimum Diam..	Redness	UNPLANNED..	ADMINISTRAT..	7	cm	7	7
DELTOID MUSCLE..	Unplanned Visit..	MINDIAM	Minimum Diam..	Swelling	UNPLANNED..	ADMINISTRAT..	3	cm	3	3

Figure 8. Screenshot of FACE Domain

From the above screenshot, we can relate that the assessments about the reactogenicity events collected in CE Domain like Severity, measurements are captured in FACE Domain. They are also connected in RELREC to show the exact relationship.

RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
CE		CELNKGRP		MANY	CJ1
FA		FALNKGRP		MANY	CJ1

Figure 9. Screenshot of RELREC

The occurrence of the SYSTEMIC event FEVER is captured in the FA domain and the corresponding temperature is captured in VS. the below screenshots showing the VS domain with the temperature values and RELREC establishing relationship between the occurrence of FEVER and the temperature.

VSLNKID	VSLNKGRP	VSTESTCD	VSTEST	VSCAT	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.8	F	36		36 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.3	F	35.72		35.72 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.9	F	36.06		36.06 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	97.2	F	36.22		36.22 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.6	F	35.89		35.89 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.8	F	36		36 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.8	F	36		36 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	97.2	F	36.22		36.22 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	97.2	F	36.22		36.22 C
VACCINATION..	VACCINATION..	TEMPMAX	Temperature(M..	REACTOGENI..	96.6	F	35.89		35.89 C

Figure 10. Screenshot of VS Domain

⚠ IDVAR	⚠ RDOMAIN	⚠ IDVARVAL	⚠ RELTYPE	⚠ RELID
FALNKID	FA	VACCINATION 1-Fever-DAY 1		VD1
VSLNKID	VS	VACCINATION 1-Fever-DAY 1		VD1
FALNKID	FA	VACCINATION 1-Fever-DAY 2		VD2
VSLNKID	VS	VACCINATION 1-Fever-DAY 2		VD2
FALNKID	FA	VACCINATION 1-Fever-DAY 3		VD3
VSLNKID	VS	VACCINATION 1-Fever-DAY 3		VD3

Figure 11. RELREC between FA and VS for SYSTEMIC event FEVER

CONCLUSION

The understanding of data and standardizing is always of great interest in the clinical research industry and with the advantage of TA knowledge it will be a quicker process. The availability of provisional standardization leading to CDISC compatibility saves lots of resources. The TAUGs will continue to be a mechanism by which sponsors can get targeted clinical trials up and running in a faster and more efficient manner by collecting and representing the trial data in a standard way.

REFERENCES

Fred Wood (2016), Considerations and Conventions within the Therapeutic Area User Guides (TAUGs) PharmaSUG 2016 - Paper DS07.
Available at: <https://www.lexjansen.com/pharmasug/2016/DS/PharmaSUG-2016-DS07.pdf>

Anita Umesh, Breast Cancer TAUG Overview and Implementation Meeting 151209 - Breast Cancer TAUG - Anita Umesh.pdf
Available at:
http://cdiscportal.digitalinfuzion.com/CDISC%20User%20Networks/North%20America/Bay%20Area/_layouts/mobile/dispform.aspx?List=98c6dea6-cbe8-4301-98d6-f66d6281af62&View=a55a0fa5-cfeb-447c-a9c2-e290ebc135c2&ID=52

ACKNOWLEDGMENTS

We would like to express my sincere gratitude to the management of my organization – “Ephicity” for the encouragement and support in helping us with all the necessary facilities to write this paper.

We would also like to thank members of the Ephicity family for their encouragement, insightful comments and hard questions.

Our sincere thanks to our Director Mr. Tyagrajan Swaminathan and our India center head Mr. Siva Ramamoorthy for their unwavered support.

RECOMMENDED READING

Study Data Tabulation Model Implementation Guide: Human Clinical Trials, Version 3.2, CDISC Submission Data Standards Team (November 26, 2013).

Study Data Tabulation Model, Version 1.4, CDISC Submission Data Standards Team

(November 26, 2013)

TAUG – BrCa V1.0

TAUG-VAX V1.0

CONTACT INFORMATION

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

Name: Sowmya Srinivasa Mukundan

Enterprise: Efficacy Lifescience Analytics Pvt. Ltd., India Address: No.6, 2nd Floor, 2nd Main Rd, Arekere, Off. Bannerghatta Road, City, State ZIP: Bangalore, Karnataka 560076

E-mail: sowmya.mukundan@efficacy.in

Web: www.efficacy.com

Name: Charumathy Sreeraman

Enterprise: Efficacy Lifescience Analytics Pvt. Ltd., India Address: No.6, 2nd Floor, 2nd Main Rd, Arekere, Off. Bannerghatta Road, City, State ZIP: Bangalore, Karnataka 560076

E-mail: charumathy.sreeraman@efficacy.in

Web: www.efficacy.com

Any brand and product names are trademarks of their respective companies.