PharmaSUG 2023 - Paper SS-304

As Simple as Falling Off a Log?: An Unusual Case Study of Mapping Data into the SDTM DA (Drug Accountability) Domain

Susan Mutter, PROMETRIKA, LLC

ABSTRACT

The SDTM DA (Drug Accountability) domain tabulates the amount of treatment units dispensed to a subject and the amount returned to gauge dosing compliance for each treated subject in a study and is often collected in a log form format. It seems like a fairly straightforward domain, so mapping subject data to it should be as easy as falling off a log. Or is it? Add together a sponsor and two different CRO's (Contract Research Organizations) with an evolving protocol, separate databases for the double-blind and open label extension portions of the study, and a creative data entry approach and you have a recipe for complexity. This paper will present a case study of mapping drug accountability data that was anything but simple.

INTRODUCTION

While the CDISC SDTM (Standard Data Tabulation Model) as a whole continues to evolve, the DA (Drug Accountability) domain has changed very little since its inception. This is a findings domain that tabulates the accountability of study products, such as information on their receipt, dispensing, return, and packaging, or, in other words, how much treatment was provided to the subject and how much of it was returned. This information can be used to determine dosing compliance. By including the actual treatment associated with the dispensed treatment, the domain can be used to verify that the treatment dispensed was consistent with the study randomization. While this information seems fairly straightforward, it is crucial to the study analysis to know whether the study subjects received the treatment they were assigned to and were compliant with the study dosing strategy. The importance of this domain should not be underrated.

In addition, when you work for a CRO (Contract Research Organization), you often need to map data for a sponsor organization from a study database that your CRO did not build which can be very frustrating. You may not have insight into how the study build evolved or access to view the database instance, so you only see what the data looks like once you receive it. You may not be able to control or give input into how the database is updated or how new components are added. If the data management portion is not handled by your organization, it can be difficult to get clarity into how the data is being entered and effect data corrections. You often have to deal with whatever you get and make it work, which also makes it difficult to try to automate the annotation and mapping process. The following is a case study of a challenging mapping example from an innocuous looking Drug Accountability log.

MAPPING THE DATA

Prior to starting the mapping process, the sponsor provided the study protocol, while the study eCRF (electronic Case Report Form), CRF (Case Report Form) completion guidelines and datasets came from the CRO that built the database. The protocol confirmed that the project was a fairly standard double-blind, parallel, active-control study. The Drug Accountability log CRF form contained nine variables: Visit, Bottle Number Assigned, Lot Number Dispensed, Date Dispensed, Number of Capsules Dispensed, Bottle Returned Y/N, Date Returned, and Number of Capsules Returned were all automatically populated within the database from the IWRS (interactive web response system). The only thing the site needed to enter was the Bottle Number Dispensed. The datasets didn't have much data in them yet, but it was enough to get started, so the form was annotated as shown in Figure 1 and the resulting mapping specifications are shown in Table 1. As illustrated, the form and specifications are very straightforward and look like a good candidate for automated mapping or to give to a less experienced programmer to get practice with mapping.

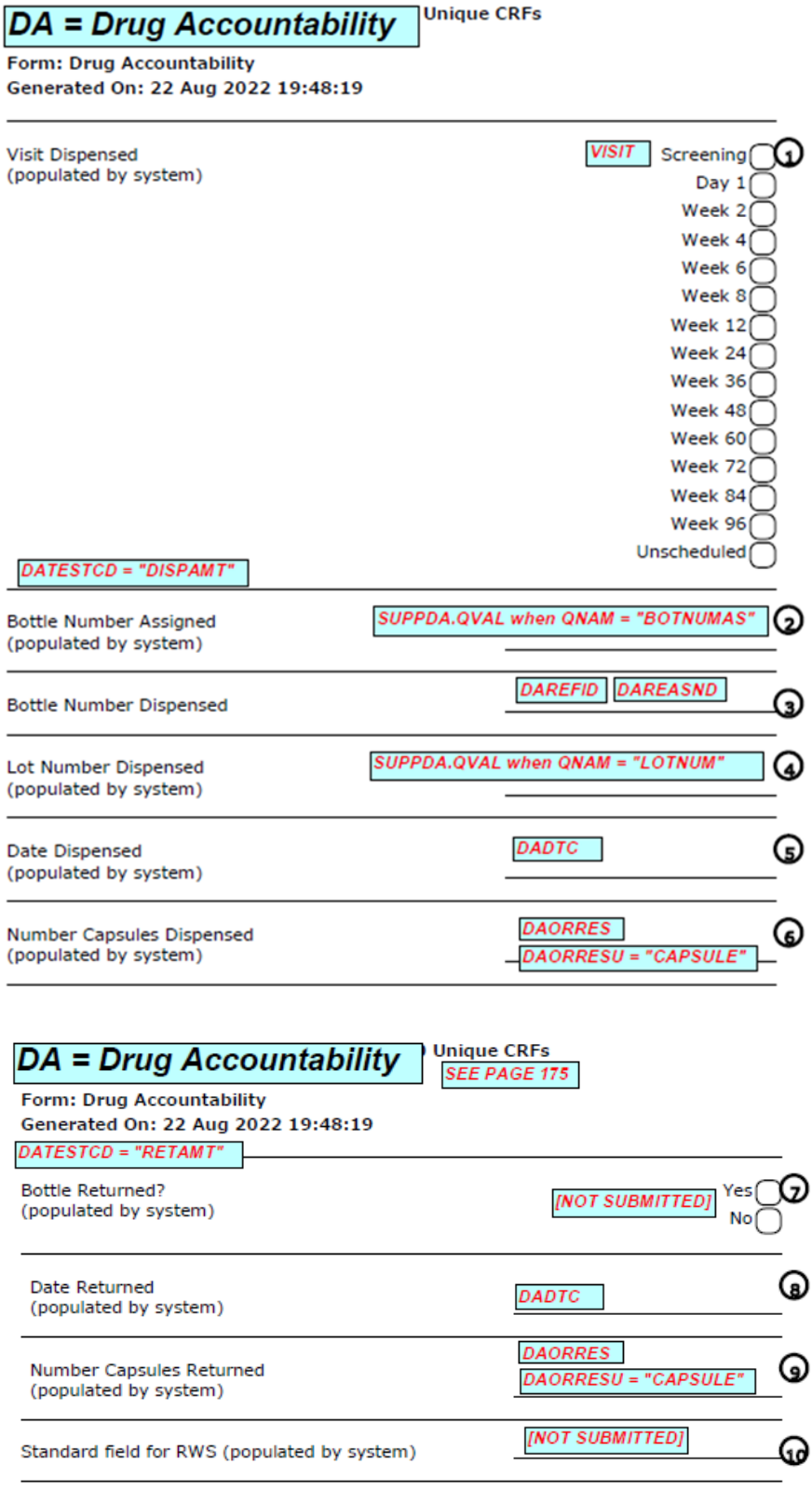

Figure 1. DA aCRF Annotations

	Variable	Mapping Specification	Mapping Specification
1	STUDYID	"OAK LOG"	"OAK LOG"
2	DOMAIN	"DA"	"DA"
3	USUBJID	Subject ID	Subject ID
4	DASEQ	Incremental sequence number	Incremental sequence number
5	DAREFID	Bottle Number Dispensed	Bottle Number Returned
6	DATESTCD	"DISPAMT"	"RETAMT"
7	DATEST	"Dispensed Amount"	"Returned Amount"
8	DACAT	"LEAVES" or "ACORNS"	"LEAVES" or "ACORNS"
9	DAORRES	Number of Capsules Dispensed	Number of Capsules Returned
10	DAORRESU	"CAPSULE"	"CAPSULE"
11	DASTRESC	same as DAORRES	same as DAORRES
12	DASTRESN	numeric of DASTRESC	numeric of DASTRESC
13	DASTRESU	same as DAORRESU	same as DAORRESU
14	DASTAT	"NOT DONE" if not dispensed	"NOT DONE" if Bottle Returned is "N"
15	DAREASND	Reason not done	Reason not done
16	VISITNUM	Derived	Derived
17	VISIT	value of VISIT	value of VISIT
18	EPOCH	Derived	Derived
19	DADTC	Date Dispensed	Date Returned
20	DADY	Derived	Derived

Table 1. DA Mapping Specification

As the study progressed, there were several amendments, and then an OLE (Open Label Extension) was added. The OLE visits were set up in a new, entirely separate database. Some information from the initial screening in the double-blind database was mirrored in the new system and some ongoing records such as Adverse Events and Concomitant Medications were also electronically pulled in. Information from the double-blind database Drug Accountability log was not pulled in, and the form had the same fields. The new datasets and visits were simply added to the mapping specifications and programming.

DATA ISSUES

As more data came in and analysis work started, it became clear that there were some anomalies in the data which caused the results to not make sense. After several discussions internally and with the sponsor, and some time spent looking at the data values, a number of major issues were identified and confirmed:

- If drug wasn't available during a planned visit, the site was dispensing the drug later during an unplanned visit but was sometimes entering the information under the original planned visit
- The Bottle Number Dispensed wasn't always being entered
- Values other than Bottle Number were being entered in Bottle Number Dispensed, such as "NOT DISPENSED" or "NO" or other comments
- Some dispensed records from the double-blind database were being re-entered into the OLE database, along with the returned information
- Date Dispensed was the date that the bottle number was assigned in IWRS
- The Date Returned was the date that the site returned the drug to the warehouse, not the date that the subject returned the drug to the site

- Some bottles that were not dispensed and then were returned had fewer capsules reported in the Number of Capsules Returned field than when they were not dispensed

Table 2 shows examples of raw data issues for bottles indicated as not returned with a returned date, and not dispensed with a returned date. Table 3 shows examples of raw data issues for the same bottle being dispensed at different visits on the same date, as well as a record with a bottle number dispensed but no other information provided.

eCRF page name	Number Tablets Dispensed	Number Tablets Returned	Visit Dispensed	Bottle Number	Date Dispensed(Character)	Bottle Returned?	Date Returned(Character)	MEDID_D	DALOT
Drug Accountability	68	40	Day 1	131580	10-Feb-2020	No	24-Sep-2020	131580	23071010
Drug Accountability	68	32	Week 2	144104	24-Feb-2020	No	24-Sep-2020	144104	23071010
Drug Accountability	68	0	Week 2	147460	24-Feb-2020	No	24-Sep-2020	147460	23071010
Drug Accountability	68	68	Week 6	124416	23-Mar-2020	No	24-Sep-2020	NOT DISPENSED	23071010
Drug Accountability	68	68	Week 6	190035	23-Mar-2020	No	24-Sep-2020	NOT DISPENSED	23071010

Table 2. Raw Data Examples of Bottles Not Dispensed and Returned

eCRF page name	Number Tablets Dispensed	Number Tablets Returned	Visit Dispensed	Bottle Number	Date Dispensed(Character)	Bottle Returned?	Date Returned(Character)	MEDID_D	DALOT
Drug Accountability	68		Week 6	196488	2-Jan-2019	No		196488	23071010
Drug Accountability	68		Unscheduled	196488	2-Jan-2019	No		196488	23071010
Drug Accountability	68	23	Unscheduled	200002	27-Jul-2020	Yes		200002	4171884
Drug Accountability	68							200002	
Drug Accountability	68	23	Week 112	200002	27-Jul-2020	Yes	5-Aug-2021	200002	4171884

Table 3. Raw Data Example of Bottles Dispensed at Different Visits on the Same Date

MAPPING ISSUES

The data anomalies had a detrimental effect on the mapping results in the SDTM DA domain:

- Visit values were inaccurate
- The actual bottle dispensed value was sometimes missing
- It wasn't always clear which bottles were not dispensed
- Double-blind records added to the OLE database were creating duplicate records
- The Date Returned could occur long after the subject finished the study

Table 4 shows examples from the SDTM DA domain of bottles that were not dispensed but then were returned, while Table 5 shows duplicate bottle returned records in the SDTM DA domain.

Sequence Number	Group ID	Reference ID	Sponsor-Defined Identifier	Short Name of Accountability Assessment	Name of Accountability Assessment	Category of Assessment	Assessment Result in Original Units	Original Units
1			1229971	DISPAMT	Dispensed Amount	ACORNS		
2			1229172	DISPAMT	Dispensed Amount	ACORNS		
3		NOT DISPENSED	1229971	RETAMT	Returned Amount	ACORNS	68	CAPSULE
4		NOT DISPENSED	1229172	RETAMT	Returned Amount	ACORNS	68	CAPSULE
5		131580	1157352	DISPAMT	Dispensed Amount	ACORNS	68	CAPSULE
6		147460	1176997	DISPAMT	Dispensed Amount	ACORNS	68	CAPSULE
7		131580	1157352	RETAMT	Returned Amount	ACORNS	40	CAPSULE
8		147460	1176997	RETAMT	Returned Amount	ACORNS	0	CAPSULE
9		144104	1176996	DISPAMT	Dispensed Amount	LEAVES	68	CAPSULE
10		144104	1176996	RETAMT	Returned Amount	LEAVES	32	CAPSULE

Table 4. Bottles Not Dispensed and Returned Records in SDTM

Sequence Number	Group ID	Reference ID	Sponsor-Defined Identifier	Short Name of Accountability Assessment	Name of Accountability Assessment	Category of Assessment	Assessment Result in Original Units	Original Units
163		151079	475659	DISPAMT	Dispensed Amount	ACORNS	68	CAPSULE
201		151079	475659	RETAMT	Returned Amount	ACORNS	20	CAPSULE
202		151079	4489104	RETAMT	Returned Amount	ACORNS	20	CAPSULE

Table 5. Duplicate Bottles Returned Records in SDTM

The resulting SDTM DA domain dataset made trying to calculate dosing compliance in the analysis extremely challenging. Many subjects appeared to have returned more capsules than they were dispensed, an example of which is displayed in Table 6. Visit windows and DM.RFPENDTC (Date/Time of End of Participation) were also thrown off by the Bottle Returned dates. The analysis programmers tried to work with the domain but the results were less than ideal. A revised mapping approach was needed.

Parameter Code	Parameter	Analysis Value
NUMDISP	Number of Dispensed Capsules	3844
NUMEXP	Number of Capsules Expected to be Taken	2882
NUMRET	Number of Returned Capsules	4212
TRTCOMP	Treatment Compliance Rate (%)	-9.299097849

Table 6. Resulting Negative Treatment Compliance Rate in ADEX

NEW MAPPING SOLUTION

The new mapping approach focused only on drug that was dispensed to the subject and what, if any, was returned, while removing any duplicate information contained in both the double-blind and OLE databases. Figure 2 shows the updated aCRF annotations and Table 7 shows the updated mapping specifications. VISIT changed to a derived variable based off of the Date Dispensed, which solved the issue of the existing visit values being inaccurate. Date Returned changed to Not Submitted to reflect the fact that the information was not subject related. Number of Capsules Returned was retained but only applied to bottles that were actually dispensed. DASTAT and DAREASND were entirely removed from the specifications.

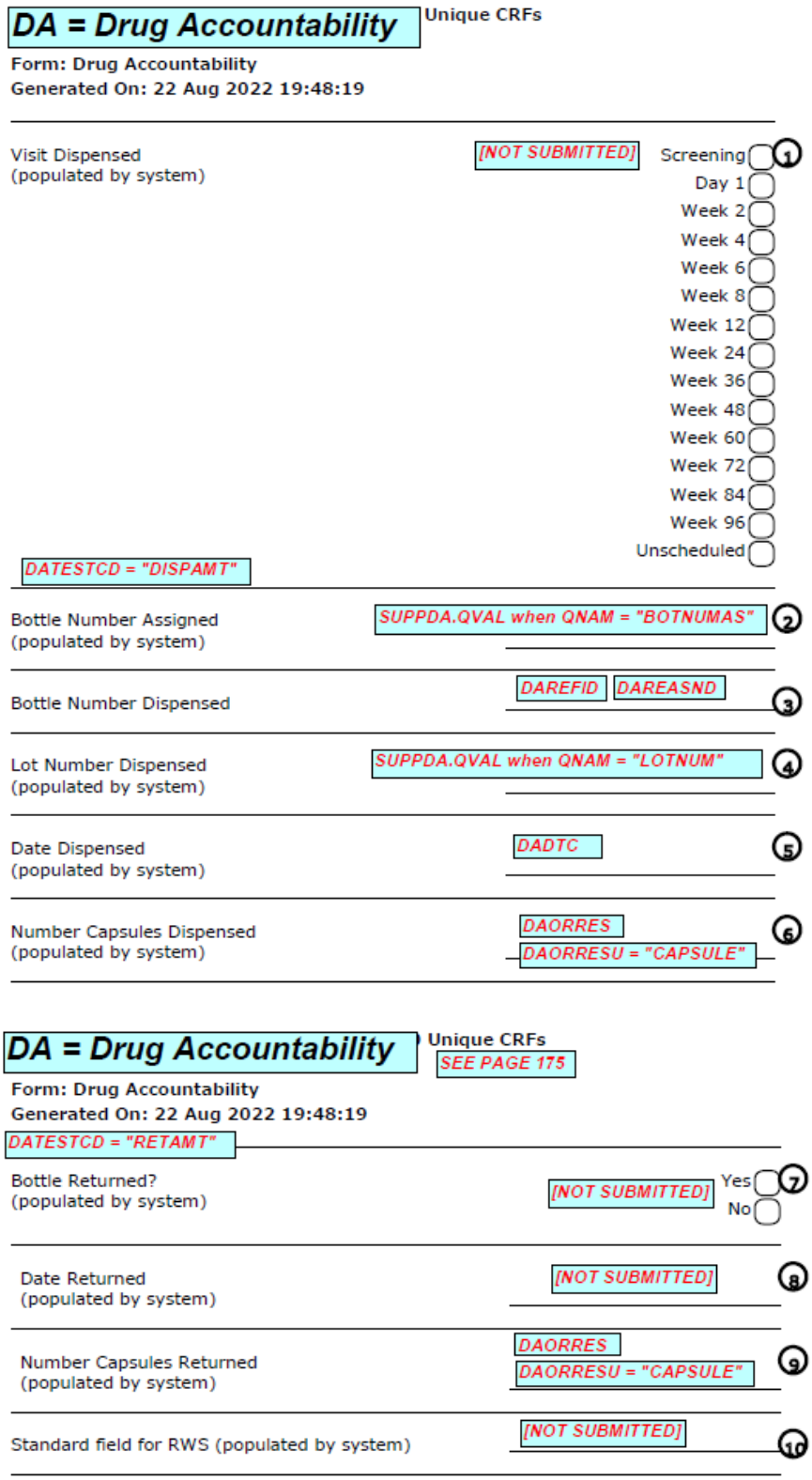

Figure 2. New DA CRF Annotation

	Variable	Mapping Specification	Mapping Specification
1	STUDYID	"OAK LOG"	"OAK LOG"
2	DOMAIN	"DA"	"DA"
3	USUBJID	Subject ID	Subject ID
4	DASEQ	Incremental sequence number	Incremental sequence number
5	DAREFID	Bottle Number Dispensed	Bottle Number Returned
6	DATESTCD	"DISPAMT"	"RETAMT"
7	DATEST	"Dispensed Amount"	"Returned Amount"
8	DACAT	"LEAVES" or "ACORNS"	"LEAVES" or "ACORNS"
9	DAORRES	Number of Capsules Dispensed	Number of Capsules Returned
10	DAORRESU	"CAPSULE"	"CAPSULE"
11	DASTRESC	same as DAORRES	same as DAORRES
12	DASTRESN	numeric of DASTRESC	numeric of DASTRESC
13	DASTRESU	same as DAORRESU	same as DAORRESU
14	VISITNUM	Derived	
15	VISIT	Derived from DADTC	
16	EPOCH	Derived	
17	DADTC	Date Dispensed	
18	DADY	Derived	

Table 7. Updated DA Mapping Specification

While the updates appear to be fairly simple, the tricky part was ensuring that records for bottles dispensed and their matching return information were accurately selected and that duplicate records were eliminated. This was accomplished using the following algorithms:

- In the combined double-blind and OLE raw datasets, if Number of Capsules Dispensed is populated, select a valid number value from Bottle Number Dispensed or if Bottle Number Dispensed is blank, select the value in Bottle Number Assigned to create DRUG DISPENSED records, using Date Dispensed to derive the matching Visit value. Then collapse the resulting records on Bottle Number, Number of Capsules Dispensed and Date Dispensed to remove duplicates.
- In the combined double-blind and OLE raw datasets, if Number of Capsules Returned is populated, select a valid number value from Bottle Number Dispensed or if Bottle Number Dispensed is blank, select the value in Bottle Number Assigned to create DRUG RETURNED records and do not populate any timing variables. Then collapse the resulting records on Bottle Number and Number of Capsules Returned to remove duplicates.

To facilitate the process, requests were funneled to data management to address the following issues: complete missing Bottle Number Dispensed values, verify Date Dispensed, clarify which bottles were not dispensed and the corresponding values in Number of Capsules Returned, and remove duplicate information from the double-blind database residing in the OLE database. Data management confirmed that the Date Returned could not be changed to the date that the subject returned their drug to the site because that information was not available in IWRS.

The new SDTM DA dataset was cleaner. Table 8 displays the same subject from Table 4, and shows a much clearer picture of how much drug was dispensed and returned, with all of the non-subject data

points and duplicate records removed. Table 9 displays the updated records within ADEX (Exposure Analysis Dataset), which shows a clear picture that the subject was 100% compliant with treatment.

Sequence Number	Group ID	Reference ID	Sponsor-Defined Identifier	Short Name of Accountability Assessment	Name of Accountability Assessment	Category of Assessment	Assessment Result in Original Units	Original Units
1		131580	1157352	DISPAMT	Dispensed Amount	ACORNS	68	CAPSULE
2		147460	1176997	DISPAMT	Dispensed Amount	ACORNS	68	CAPSULE
3		131580	1157352	RETAMT	Returned Amount	ACORNS	40	CAPSULE
4		147460	1176997	RETAMT	Returned Amount	ACORNS	0	CAPSULE
5		144104	1176996	DISPAMT	Dispensed Amount	LEAVES	68	CAPSULE
6		144104	1176996	RETAMT	Returned Amount	LEAVES	32	CAPSULE

Table 8. Revised SDTM DA Domain Records

Parameter Code	Parameter	Analysis Value
NUMDISP	Number of Dispensed Capsules	204
NUMEXP	Number of Capsules Expected to be Taken	132
NUMRET	Number of Returned Capsules	72
TRTCOMP	Treatment Compliance Rate (%)	100

Table 9. Revised Records in ADEX

CONCLUSION

When mapping data into SDTM, even if it is not a newer domain, and the CRF form looks straightforward, it is very important to check the data carefully at the beginning of the process and as a study progresses to ensure that the mapping algorithm properly represents the data being collected. For automatically populated data points, a description of the source and the purpose of the data should be provided to the team at the outset to ensure transparent data traceability. Full disclosure of data sources helps to ensure that the data being mapped into subject level domains is in fact subject data. When using automated mapping tools or if a more junior programmer is writing the specifications, check the results carefully. When working with multiple organizations, it is important to maintain clear communication across the whole team, so everyone is aware of the data collection strategy and any adjustments that are made during the study. If, after all that, the data still is not what you were expecting, don't be afraid to adjust the mapping to tabulate the data in a way that helps the analysis. Keep in mind that every new study seems to have a different twist in it somewhere, so even something as straight forward as Drug Accountability may not be as simple as you think.

ACKNOWLEDGMENTS

Thank you to my colleagues Andrew Park, Elena Prosekova and Assir Abushouk for their assistance in reviewing this paper.

DISCLAIMER: The contents of this paper are the work of the author and do not necessarily represent the opinions, recommendations, or practices of PROMETRIKA, LLC.

RECOMMENDED READING

- *Study Data Tabulation Model Version 1.7*

- *Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.3*

CONTACT INFORMATION

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

Susan Mutter, Director, Statistical Programming
PROMETRIKA, LLC
Email: smutter@prometrika.com
Web: https://www.prometrika.com/