ORIGINAL PAPER

A use-case analysis of Clinical Data Interchange Standards Consortium/Study Data Tabulation Model in academia in an investigator-initiated clinical trial

Shizuko Takahara[1,2], Toshiki I. Saito[3], Yasuhito Imai[1], Takahiro Kawakami[4] and Toshinori Murayama[1,2]

[1]*Innovative Clinical Research Center, Kanazawa University, Kanazawa, Japan*
[2]*Clinical Development, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan*
[3]*Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan*
[4]*G-link System Consulting K.K., Tokyo, Japan*

ABSTRACT

Submitting data compliant with the Clinical Data Interchange Standards Consortium (CDISC) standards is mandatory for new drug applications (NDAs). The standards set by CDISC are widely adopted in the pharmaceutical business world. Introduction of CDISC standards in academia can be necessary to reduce labor, resolve the shortage of data managers in academia, and gain new knowledge through standardized data accumulation. However, the introduction of CDISC standards has not progressed in communities within the academia that do not apply for NDAs. Therefore, herein, we created study data tabulation model (SDTM)-compliant datasets within the academia, without outsourcing, to reduce costs associated with investigator-initiated clinical trials. First, we input data from paper case report forms (CRFs) into an electronic data capture system with minimal function for paper CRFs, "Ptosh," which is compatible with SDTM. Then, we developed a generic program to convert data exported from Ptosh into fully SDTM-compliant datasets. The consistency was then verified with an SDTM validator, Pinnacle21 Community V3.0.1 (P21C). This resulted in generation of SDTM datasets, resolving all "Rejects" in P21C, thereby achieving the required quality level. Although Ptosh directly exports data in SDTM format, manual mapping of items on CRFs to SDTM variables prepared in Ptosh is necessary. SDTM mapping requires extensive knowledge and skills, and it was assumed that mapping is challenging for the staff without in-depth knowledge of CDISC standards and datasets. Therefore, for CDISC dissemination in academia, it is crucial to secure the staff, time, and funding to acquire the knowledge.

Keywords: CDISC, SDTM, investigator-initiated clinical trial, clinical data management, academia

Abbreviations:
ALL-RET trial: a phase I/II, open-label, single-arm study of CH5424802 for patients with advanced non-small-cell lung cancer harboring a RET fusion gene
CDISC: Clinical Data Interchange Standards Consortium
CDMS: clinical data management system
CRF: case report form
EDC: electronic data capture
NDA: new drug application

P21C: Pinnacle21 Community V3.0.1
PMDA: Pharmaceuticals and Medical Devices Agency
RECIST: response evaluation criteria in solid tumors
SDTM: Study Data Tabulation Model

INTRODUCTION

In Japan, the Ministry of Health, Labour and Welfare, Pharmaceutical and Food Safety Bureau Director published Notification No. 0620-6 "Basic Principles on Electronic Submission of Study Data for New Drug Applications" in 2014. From October 2016, submission of electronic data that conforms to the Clinical Data Interchange Standards Consortium (CDISC) standards, which are the global standards for clinical trial data, is required for new drug applications (NDAs).[1-3] Currently, the CDISC/Study Data Tabulation Model (SDTM), which is the raw data for each study, and CDISC/Analysis Data Model, which is the dataset processed for analysis but not mentioned in this work, often tend to be submitted to the Pharmaceuticals and Medical Devices Agency (PMDA) for NDAs.[2] The CDISC/SDTM defines a standard structure for clinical trial data tabulations. Each of these tables is called a domain. For example, data on subject demographics are assigned to the domain coded as "DM." Domains are defined for subject demographics, laboratory data, vital signs, adverse events and so on, according to the attributes of the dataset. Usually, clinical trial data are acquired and mapped to the SDTM variables as part of the data management process. Each domain code, variable name, variable order, and display format is defined in detail in the SDTM Implementation Guide.[4] When receiving CDISC standard data, the PMDA uses software called Pinnacle21 Enterprise (P21C) to validate whether the data were created according to the rules of the CDISC standard.[5]

Overseas, in the United States of America, the Food and Drug Administration requires the submission of CDISC data[6]; in China, CDISC standards are now the preferred standards for electronic data submission.[6] European and South Korean authorities are considering adopting these standards.[7] Thus, the submission of CDISC data for NDAs is becoming a standard. CDISC standards have spread rapidly in the Japanese pharmaceutical industry. As of October 31, 2019, 302 consultations have been submitted for electronic data submission by the PMDA, and electronic application data have already been submitted for 94 product applications.[8]

Originally, CDISC standards were not meant for NDAs. CDISC states that accessibility, interoperability, and reusability of data enable more meaningful and efficient research that has a greater impact on global health.[9] Barrie Nelson, vice president of CDISC, quotes "The benefits of implementing CDISC standards in research studies are numerous -- fostered efficiency, enhanced innovation, increased predictability, complete traceability, improved data quality, reduced costs, and streamlined processes -- all ensuring the integrity of your data from end to end".[10] The CDISC standards will promote the standardization of operations, leading to laborsaving and high-quality trial operations. The standards will also enable meta-analysis of accumulated, standardized data, leading to advancements in medicine. PMDA not only requires CDISC data for regulatory review but also envisions the use of accumulated, standardized data.[11] These benefits are not only for pharmaceutical companies but also for academia. Further, the Japan Agency for Medical Research and Development (AMED) states that investigator-initiated clinical trials will need to comply with CDISC standards from the planning and implementation stages, and they expect that the clinical trials contracted with them will be performed according to CDISC standards.[12]

In academia, unlike in pharmaceutical companies, clinical trial protocols are created by investigators. In a pharmaceutical company, a description of items to be captured is usually standardized. For instance, the number of white blood cells will be denoted as "WBC" even if companies have different protocols. However, because each clinical trial protocol within a single university is created by a different principal investigator, an item will be denoted in different ways in each protocol, such as "White Blood Cell," "WBC," or "Leukocytes." This leads to different nomenclatures and database structures for clinical trial data. This will increase data management labor, and meta-analysis using multiple datasets from different clinical trials will be more difficult. Therefore, the introduction of CDISC standards will help resolve these problems.[13]

Although there are good reasons to introduce CDISC standards in academia, the standardization is not widespread, at least in Japanese academia. One of the reasons being insufficient research funding,[14] not only for hiring staff but also to cover the expenses of implementing the required information technology (IT) system and the associated staff education. Therefore, human resources, such as data managers, are limited, and the daily workload is overwhelming, which implies that the scope to exceed the necessary minimum work is limited. Furthermore, conditions of employment within academia makes it difficult to hire people with expertise.[15] In addition, because most of the data management staff are employed by hospitals, priority is given to those with medical licenses, in contrast to those with IT knowledge. The use of CDISC standards is not mandatory in investigator-initiated trials that are not used for NDAs. These trials are considered to be the reason for the delayed introduction of CDISC standards in academia.

Therefore, we assessed whether fully SDTM-compliant datasets can be generated within academia without outsourcing the task to contract research organizations in an investigator-initiated clinical trial in the field of oncology and clarified adequate methods and problems.

METHODS

The investigator-initiated clinical trial "A phase I/II, open-label, single-arm study of CH5424802 for patients with advanced non-small-cell lung cancer harboring a RET fusion gene" (University Hospital Medical Information Network (UMIN) ID: UMIN000020628) (hereinafter referred to as the "ALL-RET trial")[16] was the target of this study (hereinafter referred to as "CDISC-study"). The CDISC-study is neither an activity through understanding the cause of diseases and their pathology, nor is it an activity through improving measures to prevent injury and disease as well as diagnostic and treatment measures in medical care or through verifying those measures' validity. Therefore, we think that it is outside the scope of the Ethical Guidelines for Medical and Health Research Involving Human Subjects.[17]

The ALL-RET trial had already started data collection with paper case report forms (CRFs) at the planning stage of the CDISC-study. We used "Ptosh" as the clinical data management system (CDMS). "Ptosh" was jointly developed by the National Hospital Organization Nagoya Medical Center and the Nonprofit Organization for Supporting Clinical Research.[18] It is an electronic data capture (EDC) system with minimal functions for paper CRFs, and it can export data in SDTM format. An EDC is a system that creates electronic CRFs, and the data entered in the database becomes the original CRF. However, as data collection was already progressing with the paper CRFs in the ALL-RET trial, we decided not to use Ptosh as the EDC system, but rather as a CDMS for paper CRFs. In contrast to the EDC system, which is a web-based application where the data are directly input at trial sites, the CDMS is a system for data management staff to input and manage CRFs at the datacenter and is based on the paper-based CRF data.

A general CDMS or an EDC system exports files for each input screen, and the mapping to
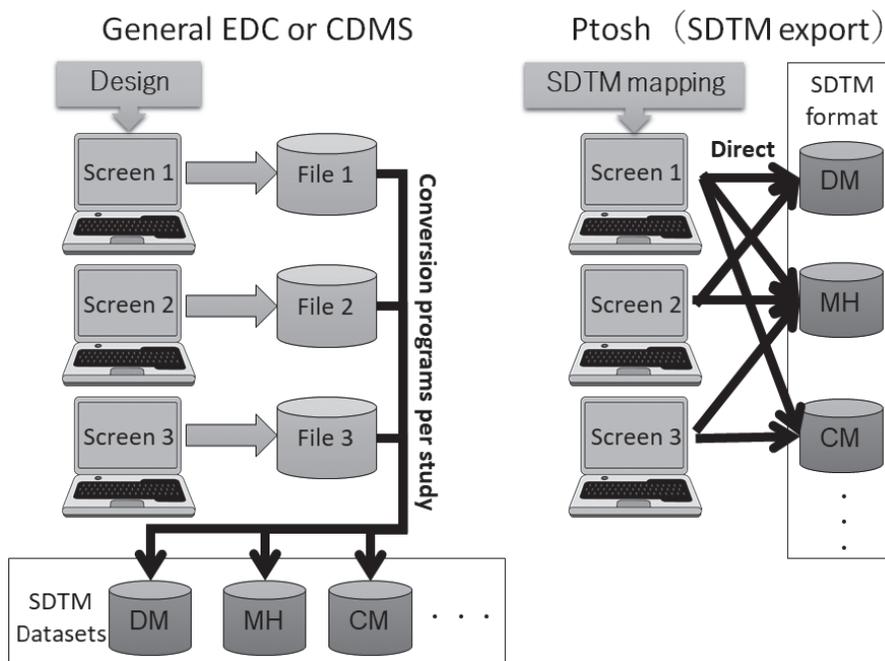
Fig. 1 Contrast for concept of creating SDTMs between general EDC or CDMS, and Ptosh
Concept of creating SDTMs using general EDC or CDMS is shown on the left side, and Ptosh on the right side. Ptosh can output data in SDTM format directly, whereas general EDC and CDMS create files for each input screen and later create SDTM datasets using conversion programs.

SDTM will be performed using those files. However, in Ptosh, regardless of the input screen and structure, variables can be consistently mapped to SDTM (Fig. 1). Therefore, datasets in SDTM-format can be obtained without any programming; however, the datasets are not complete SDTM datasets for NDA. Ptosh can output integrated data in SDTM format from one input screen with multiple domains, or vice versa, and can aggregate data from multiple input screens for one domain.

We mapped CDISC/SDTM variables to CRF items for the ALL-RET trial, built a database and input screens on Ptosh (Fig. 2(A)), and entered CRF data into the database using input screens (Fig. 2(B)).

SDTM datasets exported by Ptosh are incomplete; thus, post-processing is required to make them fully SDTM-compliant. Therefore, we developed a program using Base SAS, that is a software provided by SAS Institute Inc., to process these datasets. (Fig. 2(C))

In addition to CRF data, SDTM standards also require a description of a protocol outline called a trial design dataset. We manually created the trial design datasets using Microsoft Excel to avoid rejection by P21C. (Fig. 2(D))

The processed SDTM domain file and the minimum trial design datasets were verified by P21C. (Fig. 2(E))

Finally, as a part of the data management work of the ALL-RET trial, we used the complete SDTM datasets to create a list of patients, adverse events, and response evaluation criteria in solid tumors (RECIST) transitions for safety monitoring (Fig. 2(F)). Although RECIST is used for efficacy evaluation, the transition was monitored so that investigational drug administration
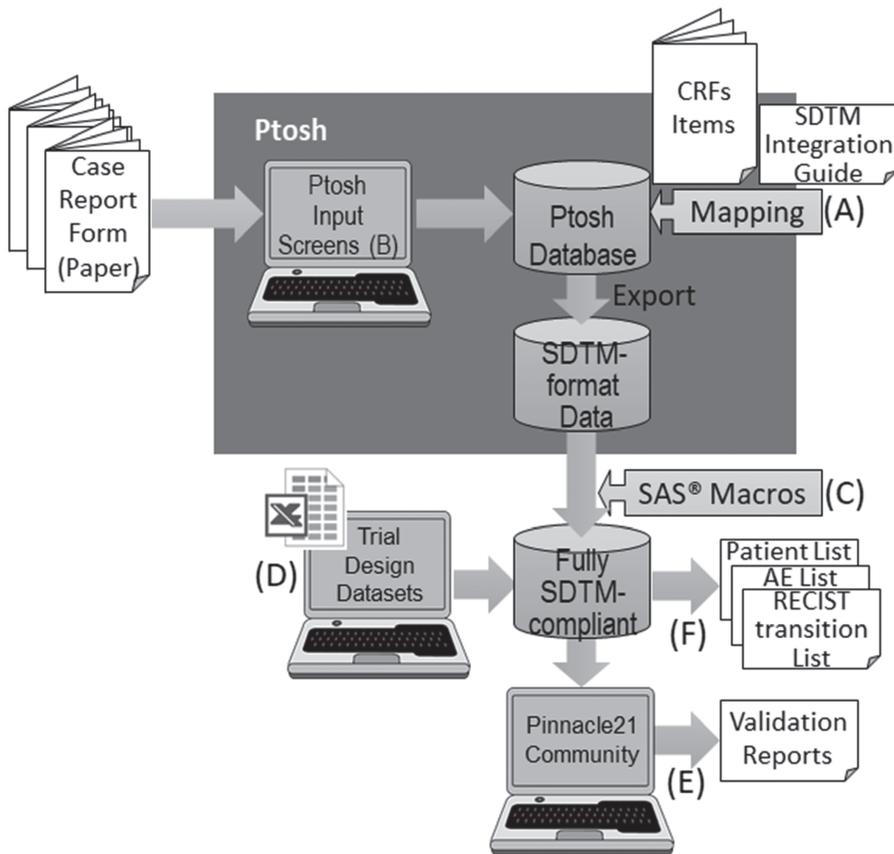
Fig. 2 Outline of the CDISC-study

First, (A) we mapped CDISC/SDTM variables to CRF items for the ALL-RET trial and built a database and input screens on Ptosh, (B) entered the CRF data of the ALL-RET trial into the Ptosh database, (C) processed the exported SDTM-format data from Ptosh with SAS Macros, (D) manually created trial design datasets, and then (E) used P21C to validate SDTM datasets. (F) We created three lists using the SDTM datasets.
AE: adverse events

would not be continued even though it deserves progressive disease (PD).

In addition to the data management system, the first author, who is in charge of data management, self-learned SDTM using materials published on the web, attended a half-day seminar conducted by UMIN, the official SDTM training provided by CDISC, and participated in a CDISC Japan User Group/SDTM team.

RESULTS

We, the academic staff for data management, mapped the CRF items to SDTM variables and built a database in Ptosh. The contents of paper CRFs were input into the database, and data were exported in SDTM format. The exported data required minimal processing to avoid problems in P21 validation.

In the CDISC study, we obtained 22 domains directly from Ptosh (Fig. 3). In Ptosh, multiple

AE	2019/08/16 11:35	Microsoft Excel CSV ファイル	96 KB
CE	2019/08/16 11:35	Microsoft Excel CSV ファイル	307 KB
CM	2019/08/16 11:35	Microsoft Excel CSV ファイル	113 KB
CO	2019/08/16 11:35	Microsoft Excel CSV ファイル	114 KB
DA	2019/08/16 11:35	Microsoft Excel CSV ファイル	94 KB
DD	2019/08/16 11:35	Microsoft Excel CSV ファイル	4 KB
DM	2019/08/16 11:35	Microsoft Excel CSV ファイル	11 KB
DS	2019/08/16 11:35	Microsoft Excel CSV ファイル	13 KB
DV	2019/08/16 11:35	Microsoft Excel CSV ファイル	4 KB
EX	2019/08/16 11:35	Microsoft Excel CSV ファイル	8 KB
LB	2019/08/16 11:35	Microsoft Excel CSV ファイル	180 KB
MH	2019/08/16 11:35	Microsoft Excel CSV ファイル	93 KB
MI	2019/08/16 11:35	Microsoft Excel CSV ファイル	41 KB
PE	2019/08/16 11:35	Microsoft Excel CSV ファイル	178 KB
PR	2019/08/16 11:35	Microsoft Excel CSV ファイル	19 KB
QS	2019/08/16 11:35	Microsoft Excel CSV ファイル	25 KB
RS	2019/08/16 11:35	Microsoft Excel CSV ファイル	124 KB
SS	2019/08/16 11:35	Microsoft Excel CSV ファイル	8 KB
SU	2019/08/16 11:35	Microsoft Excel CSV ファイル	4 KB
SUPPQUAL	2019/08/16 11:35	Microsoft Excel CSV ファイル	242 KB
TR	2019/08/16 11:35	Microsoft Excel CSV ファイル	292 KB
TU	2019/08/16 11:35	Microsoft Excel CSV ファイル	58 KB

Fig. 3 Exported files from Ptosh

State where 22 domain files for the ALL-RET trial were exported from Ptosh (unprocessed). Each domain file for CDISC was created directly from Ptosh.

domains can be mapped on the same input screen, and variables from one domain can be distributed on multiple input screens, exporting data in SDTM format in an integrated way.

An adverse event entered into Ptosh in Japanese (Fig. 4a) was exported directly in SDTM format in English as the second row of AE (Adverse Events) domain data (Fig. 4b). With adequate license, a medical dictionary for regulatory activities (MedDRA) can be automatically coded by selecting and inputting the symptom name of the adverse event on the input screen. Even if the input screen was in Japanese, all Ptosh outputs were created using Controlled Terminology in English; the terminologies were specified by CDISC.

RECIST data is divided into three domains, TU (Tumor Identification), TR (Tumor Results), and RS (Disease Response) in SDTM. TU comprises the names of target and non-target lesions. Each TU record represents one lesion. TR includes the size of target lesions or presence of non-target lesions. A TR record represents each lesion for each visit. RS includes the assessments for target lesions, non-target lesions, new lesions, or overall responses. An RS record represents an assessment on a visit. It is necessary to link these TU and TR records using Link ID, and this link can be generated by Link ID on pre-mapped CRFs in Ptosh (Fig. 5).

The data exported from Ptosh are not completely SDTM-compliant. Most variables for prefixed data or input data will be exported from Ptosh. However, most variables for derived data will be

missing or exported as blank. For example, an exported DM (Demographics) domain does not contain any data derived from other domains (eg, study drug administration end date, death status, date of death, etc). We needed to post-process the exported data to fill these derived variables.

Fig. 4a Screen for adverse events in Ptosh (test data)

All input screens are shown in Japanese. It also shows that MedDRA coding can be performed directly in the system.

Fig. 4b Unprocessed AE domain data output from Ptosh

Part of the unprocessed AE domain data output from Ptosh. As shown in Fig. 4a, the data was entered by selecting the Japanese options, but all the data were stored in English. MedDRA coding stores not only the code but also English terms. AEACN, AEREL, and AEOUT are all stored in the Controlled Terminology terms in English characters. The dataset item names are SDTM-compliant.

Additionally, we needed to delete extra records with blank data, which will not be exported from the current version of Ptosh. For example, a patient has 3 target and 3 non-target lesions in the ALL-RET trial (Fig. 5). However, because 5 and 7 input fields were prepared on the input screen for target and non-target lesions, 5 and 7 records are exported with 2 extra-records for target lesions and 4 extra for non-target lesions (Fig. 5). We developed a processing program to complete SDTM datasets using Base SAS. This program was created as a general-purpose macro so that it can be used in other studies.

In the validation of P21C, we reduced "Reject" and "Error" items by modifying the mapping in Ptosh for the ALL-RET trial. Finally, 0 rejects and 8 errors were left. All 8 errors were explainable. They were caused by a lack of data on the cases that dropped out before the start of treatment and the inconsistency in the dates of cases that were discontinued without treatment. PMDA will not accept any data with "Reject," but it will accept data with "Error" that include explanations regarding the causes of the violations and the reasons why such errors cannot be corrected.[19] Therefore, if the data included no "Reject" by the Pinnacle 21 Community, a free version of Pinnacle 21 Enterprise,[20] then it implies that our data is suitable for submission to the PMDA. There was no "Reject" from the P21C validator; therefore, we were able to obtain SDTM-compliant datasets, which is sufficient for NDA.

STUDYID	DOMAIN		TULNKID	TUTESTCD	TUTEST	TUORRES	TUSTRESC	TULOC	TULAT	TUPORTOT	
ALL-RET	TU		NT01	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET	LYMPH NODE		MULTIPLE	
ALL-RET	TU		NT02	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET	BRAIN		MULTIPLE	
ALL-RET	TU		NT03	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET	PLEURAL FLUID	RIGHT		
ALL-RET	TU		NT04	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET				
ALL-RET	TU		NT05	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET				
ALL-RET	TU	···	NT06	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET				···
ALL-RET	TU		NT07	TUMIDENT	Tumor Ident	NON-TAR	NON-TARGET				
ALL-RET	TU		T01	TUMIDENT	Tumor Ident	TARGET	TARGET	LUNG	LEFT		
ALL-RET	TU		T02	TUMIDENT	Tumor Ident	TARGET	TARGET	LYMPH NODE HILUM	LEFT		
ALL-RET	TU		T03	TUMIDENT	Tumor Ident	TARGET	TARGET	BONE		MULTIPLE	
ALL-RET	TU		T04	TUMIDENT	Tumor Ident	TARGET	TARGET				
ALL-RET	TU		T05	TUMIDENT	Tumor Ident	TARGET	TARGET				

STUDYID	DOMAIN	USUBJID	TRSEQ	TRSPID	TRLNKID	TRTESTCD	TRTEST	TRORRES	TRORRESU	
ALL-RET	TR	ALL-RET-00	25	53_RECIST	T01	DIAMETER	Diameter	30	mm	
ALL-RET	TR	ALL-RET-00	26	53_RECIST	T02	DIAMETER	Diameter	11	mm	
ALL-RET	TR	ALL-RET-00	27	53_RECIST	T03	DIAMETER	Diameter	21	mm	
ALL-RET	TR	ALL-RET-00	28	53_RECIST	T04	DIAMETER	Diameter		mm	
ALL-RET	TR	ALL-RET-00	29	53_RECIST	T05	DIAMETER	Diameter		mm	
ALL-RET	TR	ALL-RET-00	30	53_RECIST	NT01	TUMSTATE	Tumor State	PRESENT		···
ALL-RET	TR	ALL-RET-00	31	53_RECIST	NT02	TUMSTATE	Tumor State	PRESENT		
ALL-RET	TR	ALL-RET-00	32	53_RECIST	NT03	TUMSTATE	Tumor State	PRESENT		
ALL-RET	TR	ALL-RET-00	33	53_RECIST	NT04	TUMSTATE	Tumor State			
ALL-RET	TR	ALL-RET-00	34	53_RECIST	NT05	TUMSTATE	Tumor State			
ALL-RET	TR	ALL-RET-00	35	53_RECIST	NT06	TUMSTATE	Tumor State			
ALL-RET	TR	ALL-RET-00	36	53_RECIST	NT07	TUMSTATE	Tumor State			

STUDYID	DOMAIN	USUBJID	RSSEQ	RSSPID	RSTESTCD	RSTEST	RSCAT	RSORRES	
ALL-RET	RS	ALL-RET-00	14	92_FollowUp1	BESTRESP	Best Overall Response	RECIST 1.1	SD	
ALL-RET	RS	ALL-RET-00	15	92_FollowUp2	BESTRESP	Best Overall Response	RECIST 1.1		
ALL-RET	RS	ALL-RET-00	16	53_RECIST_Resu	NEWLIND	New Lesion Indicator	RECIST 1.1		
ALL-RET	RS	ALL-RET-00	17	53_RECIST_Resu	NEWLIND	New Lesion Indicator	RECIST 1.1	ABSENT	
ALL-RET	RS	ALL-RET-00	18	53_RECIST_Resu	NEWLIND	New Lesion Indicator	RECIST 1.1	ABSENT	
ALL-RET	RS	ALL-RET-00	19	53_RECIST_Resu	NTRGRESP	Non-target Response	RECIST 1.1		···
ALL-RET	RS	ALL-RET-00	20	53_RECIST_Resu	NTRGRESP	Non-target Response	RECIST 1.1	Non-CR/Non-PD	
ALL-RET	RS	ALL-RET-00	21	53_RECIST_Resu	NTRGRESP	Non-target Response	RECIST 1.1	NE	
ALL-RET	RS	ALL-RET-00	22	53_RECIST_Resu	OVRLRESP	Overall Response	RECIST 1.1		
ALL-RET	RS	ALL-RET-00	23	53_RECIST_Resu	OVRLRESP	Overall Response	RECIST 1.1	CR	
ALL-RET	RS	ALL-RET-00	24	53_RECIST_Resu	OVRLRESP	Overall Response	RECIST 1.1	SD	

Fig. 5 Unprocessed output data from Ptosh for the TU, TR, and RS domains
Data output from Ptosh for the TU, TR, and RS domains for the RECIST evaluation. TU and TR data are appropriately linked by TULNKID and TRLNKID. In this sample data, there are 3 target lesions and 3 non-target lesions, but the input screen allows for 5 target lesions and 7 non-target lesions to be entered. Therefore, the lesions that have not been input are output as blanks, and these records require processing to be deleted later.

DISCUSSIONS

This study makes the following contributions:

1. We developed a method with which the data management staff of academia can output fully SDTM-compliant datasets, validated by P21C, in investigator-initiated trials without outsourcing.

2. The software, which exported data in SDTM format, was shown to be useful and an adequate method in academia.

Effects of Standardization and Laborsaving using SDTM Datasets

We developed a program to create three lists from the fully SDTM-compliant datasets (Fig. 2(F)). The program is not specific to the ALL-RET trial, and it can be applied without modification for SDTM datasets from any clinical trial regardless of subspecialties. Likewise, other general programs can be developed for other lists, aggregations, graphs, statistics, etc, if SDTM datasets are available. Standardizing SDTM datasets may lead to reduced labor, which is a significant advantage even for clinical studies without NDA. In addition to the CRF data management work, it seems possible to build a common central monitoring system within academia with SDTM datasets.

How Did We Choose Ptosh?

We usually use DEMAND, a CDMS provided by Densuke Systems Co., Ltd. (currently transferred to G-Link System Consulting K.K. by business transfer), as the CDMS for paper CRFs in Kanazawa University. DEMAND is built using the same design concept as the "general EDC or CDMS" shown in Fig. 1. Therefore, DEMAND has templates corresponding to the SDTM domains and Controlled Terminology, which is a CDISC standard dictionary. For domains with a simple structure, such as an adverse event that consists of one event and one record, a template with the AE domain structure can be easily created. However, clinical trial data are not limited to such simple structures. In particular, data belonging to the "Findings" domain class in SDTM have a single record per item; therefore, the data structure is completely different from the "general EDC and CDMS" referred to in Fig. 1 (Fig. 6 and 7). Also, in SDTM, data must be written in English character strings, and not in codes, such as numerical values (Fig. 7); however, it is cumbersome to manually enter all data in English character strings. If there is no function to convert input code to terminology or select input on the input screens, it cannot be put to practical use. There are other limitations with the data. For instance, the TU, TR, and RS domains must be linked to each other.

We concluded that using DEMAND alone may not be sufficient to output the data of all CRFs from the ALL-RET trial to SDTM and decided to use Ptosh for SDTM output. However, Ptosh has the following limitations. To use Ptosh, it is necessary that staff who are familiar with SDTM perform SDTM mapping to build the input screens. They must carefully design the input screens. In a general system, shown in Fig. 1, SDTM knowledge is not required for the construction of input screens; this knowledge is required only during the development stage of the conversion program. Also, for long-term clinical studies, Ptosh design will be based on the SDTM version prior to the start of operations, and new SDTM versions or therapeutic areas may be newly released at the time of data fixing. This may require minor modifications of exported data from Ptosh in the future.

Data image of general EDC or CDMS

SUBJ-ID	Fever	Chills	Vomiting	Diarrhea	Constipation	Arthralgia
01-001	N	N	Y	Y	N	Y
01-002	N	N	N	N	Y	N

CRF Example

SYMPTOM

Fever	☑ N	☐ Y
Chills	☑ N	☐ Y
Vomiting	☐ N	☑ Y
Diarrhea	☐ N	☑ Y
Constipation	☑ N	☐ Y
Arthralgia	☐ N	☑ Y

Data image of CDISC/SDTM

SUBJ-ID	ITEM	RESULT
01-001	Fever	N
01-001	Chills	N
01-001	Vomiting	Y
01-001	Diarrhea	Y
01-001	Constipation	N
01-001	Arthralgia	Y
01-002	Fever	N
01-002	Chills	N
01-002	Vomiting	N
01-002	Diarrhea	N
01-002	Constipation	Y
01-002	Arthralgia	N

Fig. 6 Contrast for image of output data between general EDC or CDMS, and Ptosh
General CRF for the presence of symptoms and an image of the output data. In the output data from general EDC or CDMS, one record represents the contents of one CRF. In SDTM, however, one record represents each item. Thus, SDTM can be used universally irrespective of the items prepared for each clinical trial.

Data image of general EDC or CDMS

SUBJ-ID	Event Name	Level	Relation
01-001	Constipation	2	2

Code List
1:Mild
2:Moderate
3:Severe

Code List
1:Not Related
2:Related

(Input a numeric code on the screen)

CRF Example

Adverse Events

Event Name	*Constipation*
Level	☐¹ Mild ☑² Moderate ☐³ Severe
Relation	☐¹ Not Related ☑² Related

Data image of CDISC/SDTM

SUBJ-ID	Event Name	Level	Relation
01-001	Constipation	MODERATE	RELATED

Select the word in
SDTM Controlled
Terminology

Fig. 7 Contrast for image of output data between general EDC or CDMS, and Ptosh
CRF that describes an adverse event and images of output data. In a general EDC or CDMS, choices are stored as code. However, SDTMs must be created as English strings specified by Controlled Terminology. In the case of a system that can be entered only with a code, it will be necessary to create another program later to convert every data point entered by code.

Use of Ptosh, an EDC, as a CDMS and Role Sharing Between It and a General CDMS
　　In the CDISC-study, because we planned to output SDTM datasets with paper CRFs and

CDMS, and the ALL-RET trial had started with paper CRFs, we decided to use Ptosh as the CDMS. However, because Ptosh is an EDC system with minimal functionalities as a CDMS, it is designed to create electronic CRFs. Usually, only investigators, sub-investigators, or research coordinators at trial sites are allowed to input data directly into the electronic CRFs in EDC. We used the Ptosh version available as of January 2019 because it had a minimum CDMS function called "substitute input" that allowed data center staff to input data from paper CRFs filed by trial site members. However, this is not recommended as the function is now obsolete and could be abolished in the future.

Ptosh is an EDC system without the general data management functions present in many CDMSs. For example, Ptosh does not have the double-entry-and-compare function (the method of two people entering the same CRFs separately and comparing the data to find and correct input errors). Therefore, we performed collating by reading out all items in all cases. However, collating by reading out is more error-prone than double-entry.[21] It is more efficient to consider using a CDMS with double-entry function at least for the numerical data in the LB (Laboratory Test Results) and VS (Vital Signs) domains of SDTM.

Issues of SDTM Output by Academia

Regardless of the disease area, many items are easily standardized. These include the patient's basic background items (such as date of birth and sex), adverse events, concomitant medications, clinical tests and vital signs, RECIST evaluation (oncology area only), and dropouts. However, current disease history, study intervention, and special tests for efficacy evaluation depend largely on the disease and the characteristics of the clinical study, and complete standardization is difficult.[13] The items that are difficult to standardize are also the items where SDTM mapping is difficult. Advanced knowledge of CDISC standards is required to perform these mappings. Therefore, advanced education on CDISC standards is essential for the staff.

We saved time and costs for the CDISC-study by obtaining scientific research grants. As we had prior knowledge of CDISC standards, we could map items on CRFs to SDTM variables when we started to work on Ptosh. However, it is conceivable that sparing the required time and costs is difficult for most academic institutions, especially in Japan.

The first author is a clinical data manager and has previously worked as a system engineer, having acquired knowledge of programming and relational databases. However, it may be difficult for staff with only user-level knowledge of the system to be ready to work immediately after CDISC training. Therefore, it is reasonable to consider hiring those who have a good working knowledge and experience with databases and IT systems.

It is difficult to establish an employment quota for such human resources. In academia, there are some organizations in which medical staff simultaneously serve as data managers because these organizations belong to a hospital; they may not have an employment quota for the data management department unless they are medical staff. Also, there are many organizations that have employees on short-term contracts, eg, for up to five years. It is futile for the department to educate such employees, given that they will leave the organization after the contract expires, and by that time they may have become full-fledged CDISC experts.

We hope that academia as a whole will take a top-down approach in introducing CDISC standards from the perspective of anticipating future standardization and labor savings, and in securing research funding and staff, as well as staff training. In addition to Ptosh, other SDTM output tools are becoming available for SDTM dataset output in academia. For example, REDCap is an EDC system developed at Vanderbilt University in the US and widely used in academia worldwide. It does not have the capability to output SDTM by itself, but REDCap2SDTM, an SDTM dataset output add-on program, has been developed by Osaka University (now transferred

to Osaka City University).[22] In this way, an environment for CDISC compliance in academia is being arranged.

To spread the use of CDISC standards in academia, it is important to ensure that they are fully used, sufficient education regarding CDISC is provided, and costs are secured to build and use the system.

REFERENCES

1   Pharmaceutical and Medical Device Agency. Holding of the briefing session for the operation based on the "Basic Principles on Electronic Submission of Study Data for New Drug Applications [in Japanese]." Pharmaceutical and Medical Device Agency. https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/p-drugs/0026.html. Accessed May 10, 2021.

2   Head of Pharmaceutical and Food Safety Bureau, Ministry of Health, Labour and Welfare. Basic Principles on Electronic Submission of Study Data for New Drug Applications. Pharmaceutical and Medical Device Agency. https://www.pmda.go.jp/files/000153708.pdf [in English], https://www.pmda.go.jp/files/000159962.pdf [in Japanese]. Accessed May 10, 2021.

3   Head of Pharmaceutical Evaluation Division, Pharmaceutical Safety and Environmental Health Bureau, Ministry of Health, Labour and Welfare. Revision of Basic Principles on Electronic Submission of Study Data for New Drug Applications. Pharmaceutical and Medical Device Agency. https://www.pmda.go.jp/files/000234977.pdf [in English], https://www.pmda.go.jp/files/000234491.pdf [in Japanese]. Accessed May 10, 2021.

4   Study Data Tabulation Model Version 1.7 (Final). CDISC. https://www.cdisc.org/standards/foundational/sdtm/sdtm-v1-7/html. Accessed May 10, 2021.

5   Technical Information Regarding Submission of Application Electronic Data (FAQ, Data Standard Catalog etc.)[in Japanese]. Pharmaceutical and Medical Device Agency. https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/p-drugs/0028.html. Accessed May 10, 2021.

6   Global Regulatory Requirements. Clinical Data Interchange Standards Consortium. https://www.cdisc.org/resources/global-regulatory-requirements. Accessed May 10, 2021.

7   CDISC Vision and Mission. CDISC Wiki [in Japanese]. https://wiki.cdisc.org/pages/viewpage.action?pageId=32812989. Last Updated January 17, 2017. Accessed May 10, 2021.

8   Business track record in 2019 and future approaches. Pharmaceutical and Medical Device Agency [in Japanese]. https://www.pmda.go.jp/files/000233224.pdf. Published December 2019. Accessed May 10, 2021.

9   CDISC. Clear data. Clear impact. Clinical Data Interchange Standards Consortium. https://www.cdisc.org/. Accessed May 10, 2021.

10   Barrie N. FDA Binding guidance: a pivotal milestone for CDISC Standards. *Appl Clin Trials*. 2016;25(12):42–42.

11   New Drug Review with Electronic Data. Pharmaceutical and Medical Device Agency. https://www.pmda.go.jp/english/review-services/reviews/0002.html. Accessed May 10, 2021.

12   Joining the Clinical Data Interchange Standards Consortium (CDISC) and Activity Status. Japan Agency for Medical and Research Development. https://www.amed.go.jp/en/aboutus/collaboration/cdisc.html. Last Updated April 7, 2017. Accessed May 10, 2021.

13   Takahara S: CDISC Public Symposium Proceedings – Significance and issues regarding CDISC deployment in academia. Japan Agency for Medical and Research Development [in Japanese]. https://www.amed.go.jp/content/000003726.pdf#page=135. Published March 24, 2017. Accessed May 10, 2021.

14   Sawanobori K, Sakushima K, Arato T, Shichinohe H, Sato N, Houkin K. Challenges Identified in the Project for the Development of Human Cell Therapy Products in Hokkaido University: Mid-term Review. *Regul Sci Med Prod*. 2017;7(2):91–97. doi:10.14982/rsmp.7.91.

15   Todaka K. Clinical Study and Medical Product Development: Role of the Academic Research Organization [in Japanese]. *J Jpn Surg Assoc*. 2016;77(5):1292–1294. doi:10.3919/jjsa.77.1292.

16   Browsing of UMIN-CTR Clinical Trial Registration Information. University Hospital Medical Information Network. https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr.cgi?recptno=R000023806. Accessed May 10, 2021.

17   Ethical Guidelines for Medical and Health Research Involving Human Subjects. Ministry of Health, Labour and Welfare. https://www.mhlw.go.jp/content/10600000/000757250.pdf. Accessed May 10, 2021.

18   Saito IT, Saito MA, Kondo S, Nagai K, Nishioka E, Horibe K. Computerization of Clinical Trial Data Management at the Clinical Research Core Hospital in Japan [in Japanese]. *Regul Sci Med Prod*. 2015;5(1):61–71. doi:10.14982/rsmp.5.61.

19   FAQs on Electronic Study Data Submission. Pharmaceutical and Medical Device Agency [in Japanese]. https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/p-drugs/0029.html#Q116. Accessed April 5, 2021.

20   Is there a timeline for when Pinnacle21 Community begins to support new PMDA validation rule? Pinnacle21. https://www.pinnacle21.com/forum/there-timeline-when-pinnacle21-community-begins-support-new-pmda-validation-rule. Accessed May 10, 2021.

21   Ohashi Y, Tsuji A. *Clinical Trial Data Management [in Japanese]*. Tokyo: Igaku-Shoin Ltd; 2004.

22   Yamamoto K, Ota K, Akiya I, Shintani A. A pragmatic method for transforming clinical research data from the research electronic data capture "REDCap" to Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM): Development and evaluation of REDCap2SDTM. *J Biomed Inform*. 2017;70:65–76. doi:10.1016/j.jbi.2017.05.003