TCP Transl Clin Pharmacol
2017;25(1):34-42
http://dx.doi.org/10.12793/tcp.2017.25.1.34

Development of an automated appendix generation system (ARGUS) for clinical study reports

Dooyeon Jang[1,2], Seunghoon Han[1,2] and Dong-Seok Yim[1,2]*

[1]Department of Clinical Pharmacology and Therapeutics, Seoul St. Mary's Hospital, [2]PIPET (Pharmacometrics Institute for Practical Education and Training), College of Medicine, The Catholic University of Korea, Seoul 06591, Korea

*Correspondence: D. S. Yim; Tel: +82-2-2258-7327, E-mail: yimds@catholic.ac.kr

Received 20 Feb 2017
Revised 7 Mar 2017
Accepted 7 Mar 2017

Keywords
ARGUS,
CDISC,
CSR,
appendix

pISSN: 2289-0882
eISSN: 2383-5427

ORIGINAL ARTICLE

Data handling and tabulation are a time-consuming job when writing appendices for clinical study reports. The authors have developed an automated appendix generation system (ARGUS) conforming to the CDISC/SDTM standard using SAS (version 9.3) and R (version 3.3.1: for PK plot generation). It consists of the one main program and three subprograms. The program runs to convert a database file into an appendix document with about 100 tables and plots in MS Word format within one min after pressing the submit button under common desktop environments. We found that tasks of constructing appendices for a typical 2×2 crossover design study that have taken our team about 8 days were completed within 6 or 7 hours using the ARGUS system.

Introduction

A clinical study report (CSR) is official documentation of the results of a clinical trial.[1] To ensure the credibility of the CSR, appropriate management of clinical trial data is necessary.[2] When a CSR is written, many kinds of tables are produced from clinical trial data. Converting diverse data into tables, especially those for a CSR appendix, without error, has been a huge burden to scientific writers.[3] Typically, tables for CSRs and their appendices have been created using copy and paste from the database file (delivered as an Excel file type). Such conventions are prone to errors, especially when the workload is increased.[4] Thus, automated and reproducible processes are needed to improve the reliability and accuracy of reports.[5]

Here we introduce a system named "automated report generation and update code script (ARGUS)" that allows us to make complicated tables in CSRs and their appendices automatically. The system was designed to be compatible with the Clinical Data Interchange Standards Consortium (CDISC) standard and only database files created according to variable name rule of Study Data Tabulation Model (SDTM) can be processed. There is also a module that can process non-standard data to work with plasma concentration data.

Methods

System components

The ARGUS system was written using SAS (version 9.3; the dynamic data exchange (DDE) programming was only compatible with SAS for Windows, version 9.3 or higher) except for the parts for PK plot generation (R, version 3.3.1 was used). The report document is created by sequential runs of SAS code scripts. Linux audit daemon was used for monitoring (file tracking) of CRF database files and code script files. The system of code script was designed for the appendix of the CSR of a 2 × 2 crossover design (bioequivalence test or drug–drug interaction study). The ARGUS system components are shown in Figure 1.

The structure of the SAS code script

The table form of the system was developed based on a CSR appendix of a 2 × 2 crossover design we made earlier. The SAS code script consists of a main program, subprograms, and modules. The main program executes each of the subprograms, and the subprograms are classified according to input, manipulation procedure, and output. Each subprogram contains two or three modules (Fig. 2). The module script can be revised according to users' needs.

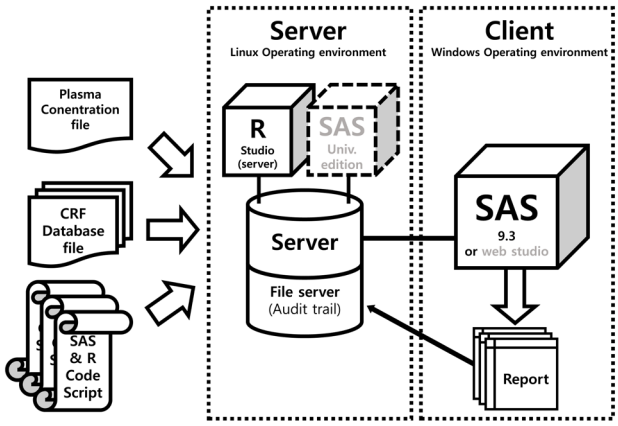

Figure 1. System components of the ARGUS.

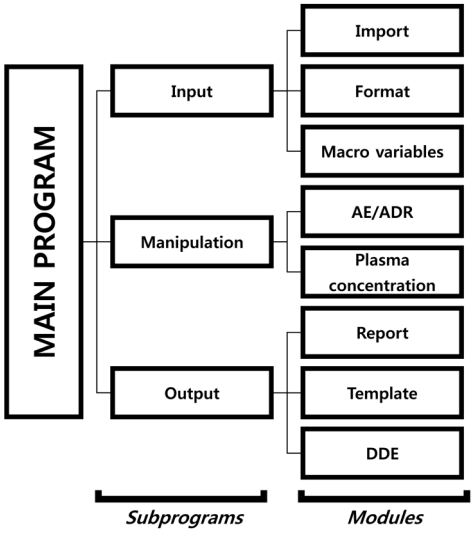

Figure 2. Code structure of the ARGUS.

Input subprogram: The input subprogram loads a database file and clinical trial information. The import module imports database files and the macro variable module contains information regarding the clinical trial, which is used as the argument for the system. The format module uses the SAS format procedure. It changes numeric or acronym data to certain characters. For example, if number '2' is defined as "Nonsmoker", then the module outputs "Nonsmoking" automatically.

Manipulation subprogram: The manipulation subprogram transforms the database file into a form that can be presented in the report. It consists of the AE/ADR module and plasma concentration module.

Output subprogram: The output subprogram consists of a report, template, and DDE modules. The report module outputs the reportable processed data in an rtf. The template module constructs headers and borderlines (line width or style) of tables. The DDE module is a post-processing module that is used to add text or modify fonts. It may also save the document in MS Word format. In SAS, the file is saved through the output subprogram and the DDE module is not activated.[6]

System operation

For the ARGUS to automate the process, it has the following prerequisites before running the main program.

✓ The database file path, the script file path, and a macro variable should be entered.

✓ The R code for individual PK plots should be run before executing the main program.

After executing the main program, a CSR Appendix report is completed. In the main program, when the user presses the Submit button, the system is activated and the process is started. When the process is completed, the output file is saved and MS Word is automatically launched to generate the report file (in the Windows version of SAS). If the program is modified in accordance with clinical trials, we recommend recording the modifications in the main program. The operation flow is as follows:

✓ Input and modification (input path and macro variables, format, module add/fix, etc.) -> execute (submit) -> Output

System constraints

✓ The system was created according to the CDISC/SDTM standard, thus nonstandard CRF database files are not compatible.

✓ Hangul (Korean characters) cannot be processed by SAS.

✓ When the hierarchy structure of the code is changed, the system does not run properly.

✓ The current version is designed for the 2 × 2 crossover design (bioequivalence test or drug–drug interaction study), code modification is necessary for other study designs.

Results

The run took about 50–60 sec to produce 45 tables and 56 plots (Table 1) after pressing the submit button in our desktop environment (x86-64 Intel Pentium Processor 4.6 GHz, 8 GB RAM).

The operation of the system is described as follows.

1. ARGUS makes new directories for the data file and code script file at the Linux server below the ARGUS and project name directories, respectively. Then, tracking is initiated.
2. ARGUS uploads the NCA dataset, individual PK parameter table (results from NCA analysis), original CRF database file (compatible file type: SAS7bdat, CSV), and code script file to each subdirectory under the project directory through the authorized FTP account.
3. Before running the main program, the R plot code should

Table 1. List of tables and plots generated by the ARGUS.

Types	Category	Name of tables	Number of items
Tables			
	Subject data listings		
		Discontinued Subjects	1
		Protocol Deviations	1
		Subjects Excluded from the Pharmacokinetic Analysis	1
		Demographic Characteristics	1
		Screening Check	1
		Drug Administration	1
		Visit Dates	1
		Inclusion/Exclusion Criteria	1
		Physical Examination	1
		Lead Electrocardiography	1
		Concurrent Medications	1
	Vital Signs		
		SBP/DBP/Pulse rate/Body Temperature	4
		Medical History	2
		Study Closure	1
	Pharmacokinetics		
		Pharmacokinetic Parameters	4
		Plasma concentration	4
		Sampling time	2
	Laboratory Measurements		
		Hematology	3
		Chemisty	5
		Urinalysis	3
		Coagulation/serology	2
	Adverse Events		
		Individual Adverse Events	1
		Adverse Events	1
		Adverse Drug Reactions	1
		Adverse Event Which is not Adverse Drug Reaction	1
	Total		**45**
Plots			
		Individual Plasma Concentration-Time Profiles	**56**

be run. The user should connect to the R studio server to draw individual PK plots and store them using r plot code.
4. The user should connect to the SAS Studio (or run SAS, version 9.3), and then load the main program. The code script file and macro variables may be modified for new reports.
5. If all files are ready, the submit button can be pressed. When the process is finished, the system opens MS Word and the Word file is stored on the server.

The original CRF database files of clinical laboratory test results are huge. Tabulation of clinical laboratory tests has been time-consuming work. Using the ARGUS system, two variables (LBTESTCD and LBORRES) with a SAS format procedure generate a cross table using very simple code. Part of a clinical laboratory (hematology) table produced by the ARGUS is shown in Figure 3. Similarly, vital sign data are also processed by the manipulation and report modules. Figure 4 shows part of a systolic blood pressure (SBP) table. The bottom line of the table header of all figures are marked with double lines using the template module.

AE (Fig. 5) and ADR (Fig. 6) tabulation require complicated procedures. All kinds of system organ classes (SOC) of MedDRA were stored in the SAS format module. The AE/ADR

	WBC			RBC			Hemoglobin			Hematocrit			Platelet			seg.Neutrophils		
Enrollment No.	Scr	Day 1	DAY 8	Scr	Day 1	DAY 8	Scr	Day 1	DAY 8	Scr	Day 1	DAY 8	Scr	Day 1	DAY 8	Scr	Day 1	DAY 8
A010	4.67	3.69	3.64	5.18	4.75	4.78	16.5	15.5	15.1	47.7	43.8	45	264	210	211	61.3	57.2	65.4
A020	5.24	.	.	5.16	.	.	15.4	.	.	44.6	.	.	182	.	.	58.9	.	.
A021	5.07	4.46	6.07	4.62	4.5	4.52	13.8	13.5	13.6	42.2	40.3	41.1	296	301	293	47.9	43.3	66.4
A030	8.16	5.17	5.38	4.52	4.14	4	14.8	13.8	13.3	45.9	42.9	40.7	294	258	217	68.6	56.8	59.5
A040	5.64	5.6	6.16	5.41	5.3	5.53	15.7	15.2	15.5	45.8	45.6	47.9	199	176	194	56.5	61.4	56.2
A050	3.98	5.13	4.98	5.32	4.68	4.96	16.3	14.5	15.1	48	41.9	44.7	249	211	218	46	40.1	44.6
A060	3.98	4.2	5.38	5.3	5.38	4.95	15.2	15.6	13.9	45.6	45.8	42.5	218	232	231	51.2	57.2	58.5
A070	5.93	4.86	5.5	4.9	4.6	4.63	14.9	13.8	13.8	42.5	41.3	41.3	324	308	332	56.2	61.7	58.8
A080	6.21	5.38	5.13	5.55	5.39	5.26	16.1	15.7	15	47	46.4	45.4	314	300	298	47.6	52.8	52.7
...																		
B130	12.89	8.35	6.4	4.97	4.46	4.42	15.5	14	13.5	45.4	40.5	41.1	269	323	302	68.9	50	51.8
B140	8.08	5.13	4.88	5.45	5.14	4.96	16.6	16	15.1	49.2	46.7	46	198	227	202	69.5	53.3	58.3
N	29	28	27	29	28	27	29	28	27	29	28	27	29	28	27	29	28	27
MEDIAN	5.77	4.83	4.98	5.16	4.76	4.77	15.4	14.6	14.1	45.8	43.35	42.8	249	240.5	217	56.2	54.65	54.5
MIN	3.92	3.69	3.64	4.52	4.14	4	13.8	13.1	12.6	41.7	40.1	38.2	164	115	121	31.9	24.2	28.3
MAX	12.89	8.35	6.62	5.56	5.39	5.53	16.9	16.1	15.5	50	46.8	47.9	329	323	332	83	66.5	67
ARITHMETICMEAN	6.23	5.02	5.04	5.11	4.78	4.73	15.45	14.58	14.19	45.79	43.15	42.91	246.28	236.32	223.67	56.77	52.88	54.79
SD	1.99	0.96	0.81	0.28	0.35	0.35	0.8	0.83	0.78	2.09	2.15	2.28	48.51	48.67	48.56	10.87	9.21	8.62
CV	31.89	19.18	16.11	5.51	7.37	7.35	5.16	5.72	5.48	4.57	4.98	5.32	19.7	20.59	21.71	19.15	17.41	15.73

Figure 3. Clinical laboratory test (hematology) table produced by the ARGUS.

SBP(mmhg)																		
	SCR	Period 1								Period 2								
	SCR	D-1	Day 0				Day 1	Day 2	Day 3	Day 6	Day 7				Day 8	Day 9	Day 10	PSV
A010	130	136	125	119	126	135	121	136	135	130	124	122	122	124	134	131	121	128
A020	115	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A021	108	119	116	107	94.0	93.0	107	114	124	106	109	110	93.0	100	125	115	126	.
A030	123	124	113	112	101	104	110	110	112	111	107	115	113	109	107	108	108	115
A040	125	137	122	118	110	105	117	128	121	131	112	120	107	116	127	127	119	124
A050	108	110	114	116	97.0	125	114	124	125	122	120	127	122	126	126	122	127	107
A060	129	122	107	109	120	99.0	98.0	107	112	130	114	108	102	120	113	114	.	.
...																		
B130	122	124	113	109	95.0	111	113	110	129	118	136	128	108	108	116	110	115	144
B140	113	123	110	110	100	100	103	113	118	105	102	104	107	106	102	115	121	120
N	29.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0	27.0	27.0	27.0	27.0	27.0	27.0	27.0	26.0	25.0
MEDIAN	124	123	115	110	109	108	116	124	121	122	113	115	111	109	114	120	121	124
MIN	97.0	99.0	92.0	90.0	90.0	93.0	93.0	101	103	99.0	102	99.0	91.0	92.0	93.0	101	98.0	105
MAX	137	139	138	139	141	139	140	140	141	143	140	139	130	139	134	139	139	144
ARITHMETICMEAN	122	123	116	112	109	111	115	121	122	122	115	116	111	112	115	119	120	123
SD	10.0	9.3	10.8	11.0	13.0	11.5	11.2	11.6	10.9	11.6	9.0	10.0	9.9	10.5	10.1	9.5	9.5	10.7
CV	8.2	7.6	9.3	9.8	11.9	10.4	9.7	9.5	8.9	9.5	7.9	8.6	8.9	9.4	8.7	7.9	7.9	8.7

Figure 4. SBP Systolic blood pressure (SBP) table produced by the ARGUS.

A

	Severity			Event		Causal Relationship	
	Mild	Moderate	Severe	n*	%	Possibly related (n*)	Unlikely (n*)
Nervous system disorder							
Headache	1	0	0	1(1)	2.78	0	1(1)
Dizziness	3	0	0	2(3)	5.55	2(3)	0
Musculoskeletal and connective tissue disorders							
Myalgia	1	0	0	1(1)	2.78	0	1(1)
Respiratory, thoracic and mediastinal disorders							
Rhinorrhoea	1	0	0	1(1)	2.78	0	1(1)
General disorders and administration site conditions							
Pyrexia	1	0	0	1(1)	2.78	0	1(1)

B

	Severity			Event		Causal Relationship	
	Mild	Moderate	Severe	n*	%	Possible	Unlikely
Nervous system disorders							
Dizziness	3	0	0	2(3)	5.56	2(3)	
Headache	1	0	0	1(1)	2.78		1(1)
Musculoskeletal and connective tissue disorders							
Myalgia	1	0	0	1(1)	2.78		1(1)
General disorders and administration site conditions							
Pyrexia	1	0	0	1(1)	2.78		1(1)
Respiratory, thoracic and mediastinal disorders							
Rhinorrhoea	1	0	0	1(1)	2.78		1(1)

C

	Severity			Event		Causal Relationship	
	Mild	Moderate	Severe	n*	%	Possible	Unlikely
GASTROINTESTINAL DISORDERS							
ABDOMINAL PAIN LOWER	2	0	0	2(2)	7.14	2(2)	
DIARRHOEA	3	0	0	3(3)	10.7	3(3)	
INVESTIGATIONS							
ANION GAP INCREASED	1	0	0	1(1)	3.57		1(1)
ASPARTATE AMINOTRANSFERASE INCREASED	1	0	0	1(1)	3.57		1(1)
BLOOD CREATINE PHOSPHOKINASE INCREASED	2	0	0	2(2)	7.14		2(2)
BLOOD PH DECREASED	1	0	0	1(1)	3.57		1(1)
BLOOD URINE PRESENT	1	0	0	1(1)	3.57		1(1)
WHITE BLOOD CELLS URINE POSITIVE	1	0	0	1(1)	3.57		1(1)
NERVOUS SYSTEM DISORDERS							
DIZZINESS	1	0	0	1(1)	3.57		1(1)
HEADACHE	0	1	0	1(1)	3.57	1(1)	
RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS							
EPISTAXIS	2	0	0	2(2)	7.14		2(2)
NASAL CONGESTION	1	0	0	1(1)	3.57		1(1)
PRODUCTIVE COUGH	2	0	0	2(2)	7.14		2(2)
INFECTIONS AND INFESTATIONS							
NASOPHARYNGITIS	2	0	0	2(2)	7.14		2(2)
PHARYNGITIS	1	0	0	1(1)	3.57		1(1)

Figure 5. Adverse events (AE) tables. A) Existing appendix table (made manually) for AE, B) Table reproduced by the ARGUS: AE example 1, C) Table reproduced by the ARGUS: AE example 2.

A

	Test* N=36		Reference+ N=36		Total N=36	
	n§	%	n§	%	n§	%
Nervous system disorders						
Dizziness	2(3)	5.55	-	-	2(3)	5.55

B

	Test N=36		Reference N=36		Total N=36	
	n*	%	n*	%	n*	%
Nervous system disorders						
Dizziness	2(3)	5.56	.	.	2(3)	5.56

C

	Test N=28		Reference N=27		Total N=28	
	n*	%	n*	%	n*	%
GASTROINTESTINAL DISORDERS						
ABDOMINAL PAIN LOWER		.	2(2)	7.41	2(2)	7.14
DIARRHOEA	1(1)	3.57	2(2)	7.41	3(3)	10.7
NERVOUS SYSTEM DISORDERS						
HEADACHE	1(1)	3.57		.	1(1)	3.57

Figure 6. Adverse drug reaction (ADR) tables. A) Existing appendix table (made manually) for ADR, B) Table reproduced by the ARGUS: ADR example 1, C) Table reproduced by the ARGUS: ADR example 2.

A

	Test* N=36		Reference+ N=36		Total N=36	
	n§	%	n§	%	n§	%
Nervous system disorder						
Headache	1(1)	2.78	-	-	1(1)	2.78
Musculoskeletal and connective tissue disorders						
Myalgia	1(1)	2.78	-	-	1(1)	2.78
Rhinorrhoea	-	-	1(1)	2.78	1(1)	2.78
General disorders and administration site conditions						
Pyrexia	1(1)	2.78	-	-	1(1)	2.78

B

	Test N=36		Reference N=36		Total N=36	
	n*	%	n*	%	n*	%
Nervous system disorders						
Headache	1(1)	2.78		.	1(1)	2.78
Musculoskeletal and connective tissue disorders						
Myalgia	1(1)	2.78		.	1(1)	2.78
General disorders and administration site conditions						
Pyrexia	1(1)	2.78		.	1(1)	2.78
Respiratory, thoracic and mediastinal disorders						
Rhinorrhoea		.	1(1)	2.78	1(1)	2.78

Figure 7. AE not ADR tables. A) Existing appendix table (made manually) for AE not ADR, B) Table reproduced by the ARGUS: AE NOT ADR example 1, C) Table reproduced by the ARGUS: AE NOT ADR example 2.

C

	Test N=28		Reference N=27		Total N=28	
	n*	%	n*	%	n*	%
INVESTIGATIONS						
ANION GAP INCREASED	1(1)	3.57		.	1(1)	3.57
ASPARTATE AMINOTRANSFERASE INCREASED	1(1)	3.57		.	1(1)	3.57
BLOOD CREATINE PHOSPHOKINASE INCREASED	1(1)	3.57	1(1)	3.70	2(2)	7.14
BLOOD PH DECREASED	1(1)	3.57		.	1(1)	3.57
BLOOD URINE PRESENT	1(1)	3.57		.	1(1)	3.57
WHITE BLOOD CELLS URINE POSITIVE	1(1)	3.57		.	1(1)	3.57
NERVOUS SYSTEM DISORDERS						
DIZZINESS	1(1)	3.57		.	1(1)	3.57
RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS						
EPISTAXIS	1(1)	3.57	1(1)	3.70	2(2)	7.14
NASAL CONGESTION		.	1(1)	3.70	1(1)	3.57
PRODUCTIVE COUGH	1(1)	3.57	1(1)	3.70	2(2)	7.14
INFECTIONS AND INFESTATIONS						
NASOPHARYNGITIS	2(2)	7.14		.	2(2)	7.14
PHARYNGITIS		.	1(1)	3.70	1(1)	3.57

Figure 7. Continued.

A

	Admission & Discharge (YYYY-MM-DD)				PSV		RD*
Enrollment No.	Period 1		Period 2		PSV	UV	RD*
SUBJID	Admission	Discharge	Admission	Discharge	PSV	UV	RD*
A010	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A020	.	.	.	.	.	.	WC
A021	2016-04-12	2016-04-13	2016-04-19	2016-04-20	.	.	WC
A030	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A040	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A050	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-29	.	
A060	2016-04-12	2016-04-13	2016-04-19	2016-04-20	.	.	WC
A070	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A080	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B100	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B110	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-28	.	
B120	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B130	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B140	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-26	.	

Figure 8. Study closure table with or without postprocessing (DDE). A) Before postprocessing, B) After postprocessing.

B

Table 00.00.12 Study Closure

Enrollment No.	Admission & Discharge (YYYY-MM-DD)				PSV	UV	RD*
	Period 1		Period 2				
SUBJID	Admission	Discharge	Admission	Discharge			
A010	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A020	.	.	.	.	.	.	WC
A021	2016-04-12	2016-04-13	2016-04-19	2016-04-20	.	.	WC
A030	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A040	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A050	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-29	.	
A060	2016-04-12	2016-04-13	2016-04-19	2016-04-20	.	.	WC
A070	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
A080	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B090	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B100	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B110	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-28	.	
B120	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B130	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-27	.	
B140	2016-04-12	2016-04-13	2016-04-19	2016-04-20	2016-04-26	.	

*RD(Reason for Discontinuation)

Figure 8. Continued.

module counts the numbers of subjects with AE and events, and calculates the percentage from the number of all subjects administered the drug. Counted and calculated numbers are combined for tabulation according to the SOC.

The ADR table is divided into two treatments in the 2 × 2 design. Treatment programming was needed for AE/ADR tabulation. The treatment program was constructed using the administration date (EXSTDTC), the AE (ADR), start date (AESTDTC) in the CRF database file and clinical trials information input by macro variables. The adverse events (which are not adverse drug reactions, Fig. 7) program is identical to that for the ADR table except for the relationship of AE variable (AEREL) criteria.

The DDE module uses Word Basic command code. The DDE module ensures SAS controls MS Word in Windows environments. Fig. 8 shows table changes before and after the postprocessing by the DDE module.

Discussion

The appendix jobs that have taken our team about 8 days to complete, were completed within 6 or 7 hours using the ARGUS system. The current version requires minor code modifications for each CRF database file and appendix item selection by the user. The user may complete the input of variables and execution within 30 minutes, but it will take time to fix the code after execution. Most tables are created properly, but some may require code modifications. Because plasma concentration data are delivered in formats varying by assaying institutions, the plasma concentration module needs to be modified for each clinical trial. If it is standardized, this modification step may be omitted in pharmacokinetic analysis and its tabulation.[7]

The AE table has a column to describe the causality relationship to treatments of clinical studies (causality). The variable AEREL was originally classified into five (not related, unlikely related, possibly related, probably related, definitely related) categories in the CRF database file.[8] After the manipulation process, the number of columns in the AE table is adjusted according to the extent of causality. The empty columns are automatically omitted.[9] This algorithm will be applied to the test date of the clinical laboratory table.

There still remains some issues in ARGUS, such as compatibility with Hangul where most of concomitant medication data are written. Postprocessing (DDE module) of tracking the document modification is also needed to assure the completeness of the document. The DDE module generally needs a huge amount of coding work. For example, about 10–20 lines of

code are necessary for a simple change of cell width in a table. Moreover, it needs modification to be used for another project because the data size, trial design, and subject number may differ.

The system described in this report is in its prototype stage, but it is meaningful first step in building a knowledge-based system. This system requires verification by beta testing in actual clinical trials. Through tuning of the prototype version, training materials and manuals are to be published.

Acknowledgements

This research was supported by the EDISON※ Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (grant number: 2016M3C1A6936614).

※ EDucation-research Integration through Simulation On the Net

Conflict of interest

The authors declared no conflict of interest.

References

1. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline - Structure and Content of Clinical Study Reports (E3). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf Accessed October 2 2016
2. Krishnankutty B, Bellary S, Kumar NB, Moodahadu LS. Data management in clinical research: An overview. Indian J Pharmacol 2012;44:168-172. doi:10.4103/0253-7613.93842.
3. Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. BMJ 2012; 344:d8141. doi:10.1136/bmj.d8141.
4. Hong MK, Yao HH, Pedersen JS, Peters JS, Costello AJ, Murphy DG, et al. Error rates in a clinical data repository: lessons from the transition to electronic data transfer--a descriptive study. BMJ Open 2013;3:pii:e002406. doi:10.1136/bmjopen-2012-002406.
5. Peng RD. Reproducible research and Biostatistics. Biostatistics 2009;10: 405-408. doi:10.1093/biostatistics/kxp014.
6. Gilmore J. Using Dynamic Data Exchange with Microsoft Word http://www2.sas.com/proceedings/sugi22/SYSARCH/PAPER308.PDF Accessed August 1 2016
7. Wood F, Schaefer P, Carolina N, Lewis R. Considerations in the Submission of Pharmacokinetics (PK) Data in an SDTM-Compliant Format. http://www.pharmasug.org/proceedings/2012/DS/PharmaSUG-2012-DS10.pdf Accessed July 15 2017
8. Analysis Data Model (ADaM) Data Structure for Adverse Event Analysis. https://www.cdisc.org/sites/default/files/.../adam/adam_ae_final_v1.pdf Accessed December 1 2016
9. Babcock G, York N. Dropping variables from a large SAS data set when all their values are missing. http://www.lexjansen.com/nesug/nesug13/90_Final_Paper.pdf Accessed December 1 2016