Paper PP03

Challenges involved with using legacy clinical trial study data during integration

Adebola Opasanya, Vitals Healthcare Ltd, London, UK
Adrienne M Bonwick, GSK, Stevenage, UK

ABSTRACT

Effective data integration is critical to the Integrated Summary of Safety (ISS) analysis and the integrity of the submission of a new compound to Health Authorities. It is in a pharmaceutical company's best interest and also an FDA requirement to review the safety of drugs in clinical development in order to perform meta-analyses on multiple studies.

In this paper, we review the challenges in pooling 27 very different Phase I studies to produce an Integrated Summary of Safety. Variations in the historic nature of the studies, the design and study indications all contributed to the resource intensive nature of producing integrated data and summaries. The author reflects on the recent experience of creating such summaries and how some of these challenges were overcome.

INTRODUCTION

This paper provides context for the integration of analysis datasets (ADaM) from 27 phase I studies. The compound was acquired by GSK from another company. GSK received Raw data and completed Clinical Study Reports (CSR). For some of the studies, statistical analysis plans were received but no programming was provided. The studies spanned from 2005 to 2016 with 2183 patients (n=2183) and consisted of 24 healthy population studies, 2 mixed populations (healthy subjects (HS) and non-healthy subjects (NHS)) and 1 opioids population.

The 27 studies have various designs; some are single arm, single period, some are parallel arm studies, whereas others are cross-over studies. Some studies treat subjects with a single dose of the study drug, and others are repeated doses of the study drug.

The primary source for the ADaM datasets were the integrated SDTM datasets. The SDTMs were created using the locked databases of the 27-individual phase 1 studies and appended onto each other to create a single set of SDTM. There was no data cut-off applied to the SDTM datasets.

The primary aims of this integration were 3-fold:

- Create ADaM datasets to support the analysis defined in the ISS Summary Document Analysis Plan (SDAP)
- Support potential re-production of selected CSR outputs which were previously created for individual studies
- Deliver a quality Integrated Summary of Safety (ISS)

PLANNING

The process of data integration from multiple clinical trials is often labour intensive and can be prone to errors, it was very important to define a clear strategy with clear goals as early as possible in the process.

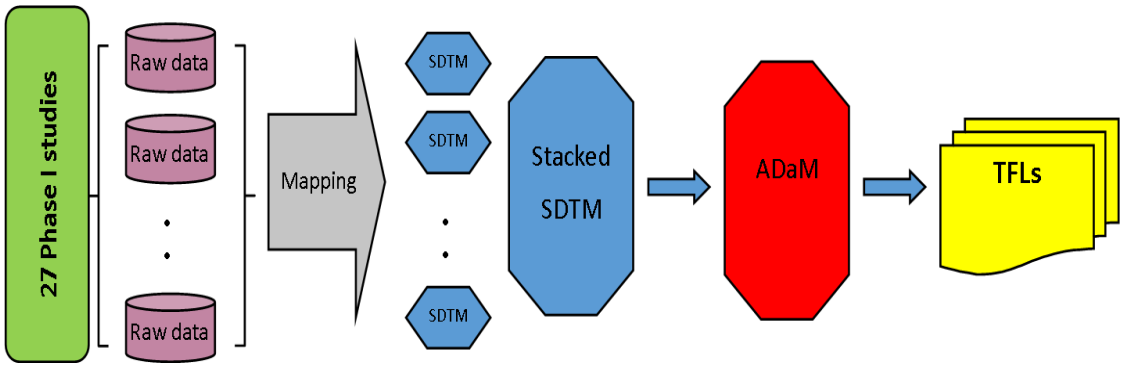
There are several milestones in integrated analysis which require input from a wide scope of team members and great planning. The initial phase of this integration therefore involved the coming together of several teams to build the Integration team. Statistics and programming team were the centre of the core team. Several additional key teams such as Regulatory Affairs, Clinical, Drug Safety, amongst others were involved. During the planning phase the following steps were put in place;

- A strong statistics and programming team to handle the integration was built
- The Analysis plan was developed
- Stats and programming team drew up timelines for the integration
- Resource was planned considering the number of ADaM datasets and numbers of outputs needed and the timeline required
- Validation processes was defined for the integration. Double programming (main and qc programmer) and Statistical review, following the GSK QC SOP.
- The outputs from the original CSRs were specified for reproduction, these were primarily laboratory, vital signs and pharmacokinetic analysis.
- Standards for integration, as defined in the Study Data Standardization Plan (SDSP) was agreed during the planning stage as follows

Standard or Dictionary	Versions Used
SDTM	SDTM v1.3 SDTM IG v3.1.3
ADaM	ADaM v2.1 ADaM IG v1.0 ADaM Structure for Occurrence Data (OCCDS) v1.0
Controlled Terminology	CDISC Controlled Terminology dated 29 September 2017
Data Definitions	define.xml v2.0
Medications Dictionary	GSKDRUG v1.4
Medical Events Dictionary	MedDRA v19.1
DAIDs Grading	Division of AIDS (DAIDS) Table for Grading the Severity of Adult and Paediatric Adverse Events Version 2.0 November 2014

PROCESS

The process for the creation of the integrated data was that individual studies had SDTMs created from the raw data, these were then stacked to create an SDTM for all studies. These were then transformed into integrated ADaM data from which the TFLs were created and also used for the comparison against the original CSRs.



CHALLENGES

Varying study designs

The 27 studies have various designs, some are single arm, single period, some are parallel arm studies, whereas others are cross-over studies. Some studies treat subjects with a single dose of the study drug, and others are repeated doses of the study drug.

The 27 studies being integrated were categorized into being single period, washout or non-washout studies, as well as if a subject on the study received either a single study drug/Placebo dose or repeated Study drug/Placebo doses as per the table below.

Study	Display Populations	Single dose/ Repeated Dose	Washout	Study Drug alone	Study Drug + Other	Placebo
001	Healthy Subjects	Single dose	X	X	X	X
002	Healthy Subjects	Single dose	X	X		
003	Healthy Subjects	Single dose	X	X	X	
004	Healthy Subjects	Repeated	X	X	X	X
005	Healthy Subjects	Single dose	X	X	X	
006	Healthy Subjects	Repeated		X	X	
007	Healthy Subjects	Repeated			X	
008	Healthy Subjects	Single dose	X	X		
009	Healthy Subjects	Repeated		X	X	
010	Healthy Subjects	Single dose	X	X		
011	Healthy Subjects	Repeated	X	X		X
012	Healthy Subjects	Single dose	X *	X	X	
013	Healthy Subjects	Repeated			X	
014	Healthy Subjects	Repeated	X *	X	X	
015	Healthy Subjects	Repeated		X	X	
016	Healthy Subjects	Repeated	X	X		
017	Healthy Subjects	Repeated		X	X	
018	Healthy Subjects	Repeated	X *	X	X	
019	Healthy Subjects	Repeated	X *	X	X	
020	Healthy Subjects/ Hepatic Impairment Patients	Single dose	X	X		
021	Healthy Subjects	Single dose	X	X		
022	Opioid Subjects	Repeated			X	
023	Healthy Subjects / Renal Impairment Patients	Single dose	X	X		
024	Healthy Subjects	Single dose	X	X	X	
025	Healthy Subjects	Single dose		X		
026	Healthy Subjects	Single dose	X	X		
027	Healthy Subjects	Single dose	X	X		X
* Washout period after treatment period 1 followed by non-washout periods						

Various treatment groups in different studies, including washout periods.
Each of the 27 studies has a Protocol Design which dictates the treatment periods and when they start and stop. The treatment had different doses in several studies, some with single dose period or multiple dose, some studies were healthy subjects only and some with non-healthy subjects. Some of these studies also included wash out periods.

For this integration, consideration was given to the CSR treatment periods (for CSR reproduction) and the ISS treatment period (for integration outputs). It was important that the required mappings are well documented.

Mapping the study level data to pooled data whilst enabling recreation of CSR outputs from pooled data
Developing integration specification to define how data from the individual study should be mapped to integrated dataset was a complex exercise. The studies had different data standards. Mapping began with a review of the analysis plans developed for each study but not all analysis plans for the 27 studies were received and no programming codes were received.

Dealing with missing values from the different studies
The integrated SDTM were just stacked individual SDTMs with no review of consistency between the studies. Therefore, in cases where time is missing and the date of the event or assessment falls onto a day when the subject crossed over between 2 ISS periods, then it is not possible to know which of the 2 ISS periods the event or assessment occurred in. In these situations, the later of the 2 ISS periods was associated with the event or assessment.

Windowing or imputation rules
Creating a fit for purpose Windowing and imputation rule for 27 studies was challenging, therefore kept as simple as possible.

- Use of Visit Windowing, Unscheduled Visits, and Record Selection

Was windowing used in one or more analysis datasets?
No
Were unscheduled visits used in any analyses?
Yes, for the ISS there were no outputs summarized “by visit”. However, when selecting records deemed as “Worst Case” results from unscheduled visits were included (i.e. ECGs).
The variables AVISIT/AVISITN were simply taken from the SDTM variables VISIT/VISITNUM. These were not harmonized across the 27 studies. No visit windowing was applied.

- Imputation/Derivation Methods

If date imputation was performed, were there rules used in multiple analysis datasets?
Yes, date imputation was applied if a partial date was collected. If start day was missing then 01 was imputed, if start month was missing then January was imputed. If end day was missing, then the last day of the month was imputed, if the end month was missing, then December was imputed. This is applicable to the Adverse Events (ADAE), Concomitant Medications (ADCM) and Medical History (ADMH) datasets.

Consistency of code lists, formats and units across studies.
In this integration, as it is often found, individual studies were completed using different versions of MedDRA and WHO drug dictionaries. The FDA requires that all coded terms need to be recoded to the latest available dictionary version prior to submission and this was completed in the integrated SDTM. It was important to also maintain the original coding for traceability back to the original study terms. This was challenging when integrating 27 studies as all terms need to be checked and compared back to the individual studies.

Summary statistics for some lab, PK and vital signs to match original CSRs
The ISS contained outputs for

- Populations – 4 outputs
- Disposition – 4 outputs
- Demographics – 2 outputs
- Medical History – 2 outputs
- Concomitant medication – 2 outputs
- Exposure – 4 outputs
- Adverse Events – 30 outputs
- Laboratory data – 11 outputs
- ECG – 14 outputs

Additional data was added into the ADaM datasets to enable the reproduction additional adverse event and laboratory data and new data for Vital signs and pharmacokinetic/pharmacodynamic data.

There were several data points where the original raw data, and hence the mapped SDTM data, did not match the original CSR. These were documented in file notes and the ADRG.

SOLUTION

Creation of Additional CSR and ISS variables
CSR vs ISS variable. The ADaM datasets were setup to support both standard CSR reporting as well as ISS specific analysis. Therefore, when the CSR reporting and ISS reporting can share a variable (such as Serious AE definition), there is just one variable (i.e. AESER), but when the CSR and ISS differ in approach, there might be two or more similarly named variables. Typically, variables added which are related specifically to the ISS are prefixed with "I", "ISxxxxxx" or "ISSxxxxx".

For example:

CSR Variable(s)	ISS Variable(s)
APERIOD/APERIODC	IPERIOD/IPERIODC
APxxSDT/APxxSDTM	ISxxSDT/ISxxSDTM
TRTxxA/TRTxxAN	ISSxxA/ISSxxAN
TRTA/TRTAN	ISSA/ISSAN
<none>	ISSUBTYP*

* Additionally, there are variables needed for the ISS which had no CDISC basis, such as ADSL.ISSUBTYP which gives the high-level type for each subject i.e. Healthy Subject, Non-Healthy Subjects etc.

Simplified treatment group to capture all patients needed
The treatment groups were simplified into treatment periods as follows

HS TRT Alone	Healthy subjects with GSK Treatment only
TRT + Other	GSK Treatment period with treatment co-administered with at least one other drug
TRT NHS	Non-Healthy Subjects who received at least one dose of GSK Treatment within the treatment period
Any TRT	Treatment period with treatment given alone or with others, i.e. any treatment given within this treatment period
Placebo	Treatment period with Placebo only
Total	Total treatment or Placebo. (If other drug in treatment period alone, this is not included)

Harmonisation at the ADaM level instead of the integrated SDTM level
The ADaM datasets set up to support both standard CSR reporting as well as ISS specific analysis
The mapping challenges was dealt with by using one mapping definition for all studies for windowing, code lists, units and missing values. Reports was developed to compare:

- The variables required in integrated dataset vs those available in the study level datasets
- The attributes of variables required in the integrated dataset vs those available in the study level datasets e.g. label, length, type (numeric value, character, date or time, numeric or character code lists) etc. This was done thoroughly and issues such as truncation, miscoding of categorical fields etc. was avoided.

Reproduction of CSR outputs was performed from the pooled data and compared to individual study CSR outputs. For the CSR Treatment Periods**,** each of the 27 studies has a Protocol Design which dictates the treatment periods and when treatment was started and stopped.

The variables TRxxSDT, TRxxSDTM, TRxxSTM, TRxxEDT, TRxxEDTM and TRxxETM provided the first and last dose dates (and times) within each treatment period. The variables TRTxxP/TRTxxA provided the planned and actual treatment groups for each period.

The variables APxxSDT, APxxSDTM, APxxSTM, APxxEDT, APxxEDTM and APxxETM defined the ranges of time (without gaps) between the treatments. i.e. the end of period 1 will be one second before the start of period 2. For example:

ADSL:

SUBJID	TRT01P	TRT01A	AP01SDTM	AP01EDTM	TRT02P	TRT02A	AP02SDTM	AP02EDTM
001	Drug A	Drug A	01JAN2018 09:00	06JAN2018 09:09	Drug B	Drug B	06JAN2018 09:10	10JAN2018 13:30
002	Drug B	Drug B	02JAN2018 09:30	07JAN2018 09:59	Drug B	Drug A	07JAN2018 10:00	08JAN2018 16:00

Creation of new parameters to match original reports and traceability was one of the key elements in this integration. The original study content was kept as much as possible so the regulatory authorities can trace back to the original individual studies.

Documentation of all derivations in the metadata and a well-versed Analysis Data Reviewers Guide (ADRG) was created as part of the production process. It meant a detailed ADRG was produced (107 pages). Any record excluded from the source study, was documented in ADRG with explanations. An additional document was also created where discrepancies between source study and the integrated data as a part of our quality checks process was documented to make sure no subject was excluded, and all subjects were reported as per the expectations.

CONCLUSION

As pharmaceutical companies continue to expand its pipeline through acquisitions and purchase of compounds from other pharmaceutical companies, the consolidation of multiple clinical studies for ISS / ISE submissions continues to pose real challenges.

Planning early and in detail, implementing solutions at the ADaM level, using data standards such as CDISC, making sure there is consistency between integrated data and individual study data and cross-checking outputs from ISS data to original individual study reports are some of the solutions to a good integration.

The proposed solution demonstrates reliable methods by which data integration issues can be easily discerned. However, the concept of integrating or pooling data is neither new nor defined by a pre-determined set of methodologies or rules. Even with a well-defined data standard in place, a certain amount of variability is to be expected. This is compounded if the integration includes studies of special interest.

This integration was done in two parts; the production of the ISS data and related outputs were completed first to enable the writing of the summary document. Subsequently, the addition of data to support the CSR was performed off the critical path. The ISS outputs were rerun on the new data to ensure that they matched.

REFERENCES

John R. Gerlach and John C. Bowen. 'Facilitating Data Integration for Regulatory Submissions', SCSUG 2010 Available at: **http://www.scsug.org/SCSUGProceedings/2010/Gerlach 2/Data_Integration.pdf**

Abraham Yeh, Xiaohong (Grace) Zhang and Shin-Ru Wang. 'Pooling strategy of clinical data' Available at: https://www.lexjansen.com/nesug/nesug13/21_Final_Paper.pdf

P Majcher 'Tools to Facilitate the Creation of Pooled Clinical Trials Databases' PharmaSUG. Available at: https://www.lexjansen.com/pharmasug/2004/ApplicationsDevelopment/AD10.pdf

Ryan DeCosta. 'Challenges Involved with the Integration of Clinical Study Data'. PHUSE 2006. Available at: https://www.lexjansen.com/phuse/2006/po/PO13.pdf

ACKNOWLEDGEMENTS

The authors greatly appreciate the Clinical Programming team at GSK for their participation in the development of this paper.

CONTACT INFORMATION

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

Adebola Opasanya
Email: adebolaopasanya@gmail.com

Adrienne M Bonwick
GlaxoSmithKline
Gunnels Wood Road,
Stevenage,
Hertfordshire,
SG1 2NY
Email: Adrienne.M.Bonwick@gsk.com