Paper SM10

The Forgotten Output - Data for ClinicalTrials.gov and EudraCT

Erik Brun, H. Lundbeck A/S, Valby, Denmark

ABSTRACT

As programmers we are generally familiar with study outputs like CDSIC® data packages and TLFs. But besides the study report, it is also a requirement for many studies that the results must be published at ClinicalTrials.gov and/or EudraCT (insert link). This is sometimes referred to as the plain language summary. Despite being called plain language summary, a considerable amount of this reporting are Adverse Event tabulations. These results can be uploaded as an xml-file. With a programmer's point of view, this paper will give a short introduction to these result outputs and some of the challenges the author has encountered in creating them. One of the quirks with these outputs that can lead to some frustration is that the results aren't necessarily going to match the study report. I will give examples in the paper to explain why there can be a miss-match. Another challenge is that both ClinicalTrials.gov and EudraCT are only designed to handle very simple study designs, as in *n* parallel groups with only screening, treatment and follow-up phases, so whenever you have just a slightly complex study design, say having an open label run-in phase, you will have to be creative.

INTRODUCTION

It is commonly assumed that once the outputs for the study report are complete, but for most clinical trials it also required to make information about the trial public available for all interested on ClinicalTrials.gov (https://clinicaltrials.gov/) and/or EudraCT (https://www.clinicaltrialsregister.eu/ctr-search/search/). The purpose of this paper is to give a short introduction to the data required for these sites.
While other disclosure sites exist, this paper will only cover the two major disclosure sites as I only have experience in delivering data to those.

The purpose of these disclosure sites is to provide information about clinical research studies to all interested.

- These amongst other
  - The sponsor of the trial
  - The trial design
  - Disease(s) studied
  - Subject disposition
  - Baseline characteristics
  - Endpoints, including results
  - Tables of SAEs and frequent non-serious AEs. Which is the main focus of this paper.

Generally, the disclosure of results must be published no later than 12 months after last participant last visit, but special rules apply for some studies for example pediatric studies, where the deadline is six months after.

WHY THE FORGOTTEN OUTPUT?

Well, I don't know, but I have my hypotheses about what could contribute to it being forgotten.
The combination of that is not a part of the study report per se or a submission package, and it is a relatively small task makes it easy to overlook. On top of that the deadline is typically a while after the trial report makes it easy to postpone. Such when it is time to deliver the disclosure outputs those, who worked on the trial, have been allocated to new projects.
When it at the same time as the output isn't defined as a CDISC or similar standard, it is less like likely to find a spot in the standards libraries.
And from the authors experience, this is not the only way these outputs become the forgotten outputs. Even when a study is outsourced, it can be forgotten. Sometimes it's not either in the workorder with the CRO or sometimes the CROs are forgetting to deliver it.

OBSERVATIONS

XML_FILES

When publishing trial information there are two methods for providing information to the disclosure sites. Either manual data entry or upload of a xml-file. Historically we have taken a mixed approach, where the AE results was uploaded as xml and everything else have been entered manually. Currently we are using a CRO to whom we provide one xml-file with the AE results that they can use for the publishing for both sites.

It can be an uphill struggle to get the xml-files right the first time. The documentation is not easy to approach unless you are rather xml savvy. And when you upload an invalid file to the disclosure sites, the error messages from both sites are often not very informative about what went wrong.

WHY THE RESULTS RARELY MATCH THE TRIAL REPORT AND OTHER FRUSTRATING QUIRKS OF THE DISCLOSURE SITES

1. The simple descriptive statistics are for disclosure sites based on randomized patients in the trial report are typically based on treated patients. A rather common cause for a difference is participants withdrawing their consent after randomization but before dosing.
2. The cut-off percentage for most frequent AEs must be an integer. Typically, this is 5%, but for some trials where the tested drug appears to have a very good safety profile this number may be lowered, otherwise one might end up with a table only consisting of ailments like common cold and COVID-19. When the study team then decides that halving the limit to 2.5% the result cannot be replicated at the disclosure sites.
3. At disclosure sites only *non-serious* treatment emergent AEs contribute to the count of frequents AEs. SAEs are reported in another section of the display for a study. A consequence of this could be a frequent AE term which both appear as serious and non-serious in a trial might get under the cut-off for display at the disclosure sites. Another consequence is that the counts don't match the clinical study report, there can be fewer counts in the disclosure output. A relatively frequent example of such AE is asthma. These differences compared to the trial report can be a trigger for QC-findings, when the QC'er is not aware of this or just plainly forgot about it.
4. EudraCT for some reason applies a 12-digit code to the 8-digit MedDRA codes to the MedDRA terms. Not only does this add another layer of confusion but these EuCTid codes can be a bit tricky to come by.
5. Sometimes treatment groups are pooled at the disclosure sites compared to the study report. This mostly happens in phase 1 and early phase 2.
6. Both sites can only handle very simple study designs, luckily they do it in the same way. When dealing with "complex" studies one must transform the data into a structure that resembles a simple study that the sites can handle. Complex has been put in quotation marks because a study with two treatment phases is to be handled as a "complex" study. Figure 1 illustrates a study where the participants start in a double-blind randomized phase with three treatments (Two active and placebo). This phase is followed by an open label phase where the participants on active continues their dose and the placebo group is allocated to either of the two active groups. At the disclosure sites, this will be presented as no less than seven different "parallel" treatment groups.

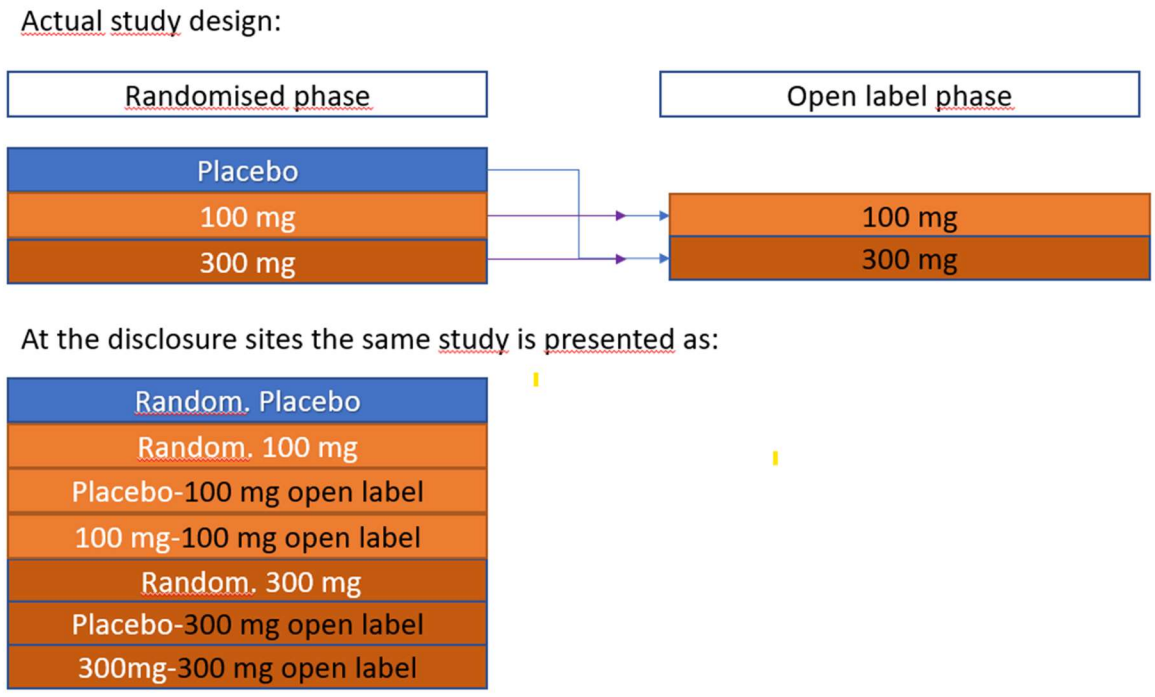

Figure 1 - Example how a study is designed and how it is presented at the disclosure sites

OUR SOLUTION FOR CREATING XML-FILES IN 6 SIMPLE STEPS

Step 1 & 2 are only needed when dealing with a "complex" study.

1. "Clone" the participants by making a representation of each participant for each of the phases the participant was in. To "clone" is just to assign a separate identifier (USUBJID) for each phase of the study, such it

appears that the study only has one treatment phase. See fig. 2

ADAM.ADSL

USUBJID	TRT01P	TRT02P
12345A-XX0001-1234	300 mg	300 mg
12345A-XX0001-5678	Placebo	100 mg

USUBJID	TRTP
12345A-XX0001-1234	Random. 300 mg
12345X-XX0001-1234	300mg-300 mg open label
12345A-XX0001-5678	Random. Placebo
12345X-XX0001-5678	Placebo-300 mg open label

Figure 2 - "Cloning " the participants

2. AEs are assigned to the new subject identifiers to the phase in accordance with the phase the AEs started.

ADAM.ADAE

USUBJID	TRT01P	TRT02P	APHASE	AEDECOD, AREL, AOUT…
12345A-XX0001-1234	300 mg	300 mg	1	
12345A-XX0001-1234	300 mg	300 mg	2	
12345A-XX0001-1234	300 mg	300 mg	2	
12345A-XX0001-5678	Placebo	100 mg	1	
12345A-XX0001-5678	Placebo	100 mg	1	
12345A-XX0001-5678	Placebo	100 mg	2	
12345A-XX0001-5678	Placebo	100 mg	2	

USUBJID	TRTP	AEDECOD, AREL, AOUT…
12345A-XX0001-1234	Random. 300 mg	
12345X-XX0001-1234	300mg-300 mg open label	
12345X-XX0001-1234	300mg-300 mg open label	
12345A-XX0001-5678	Random. Placebo	
12345A-XX0001-5678	Random. Placebo	
12345X-XX0001-5678	Placebo-300 mg open label	
12345X-XX0001-5678	Placebo-300 mg open label	

Figure 3 - Assigning AEs to a "clone"

3. Using the engine from one of our standard TLFs for counting AEs, the AEs are counted 5 times with different selection criteria:
   - Non-serious treatment emergent AEs
   - Treatment emerergent SAEs
   - Treatment emerergent related SAEs
   - Treatment emerergent SAEs with fatal outcome
   - Treatment emerergent related SAEs with fatal outcome (Used in EUDRACT only)

   Each run generates a SAS® data set and a pdf report. The pdf is used to ease the work for the QC person, such they don't have to go and read the xml-files. Unfortunately, we haven't been able to get our hands on appropriate style sheets.

4. The data sets are then combined as described below using a simple data step, such that all the subset counts of the SAEs are added to the full list of SAEs.

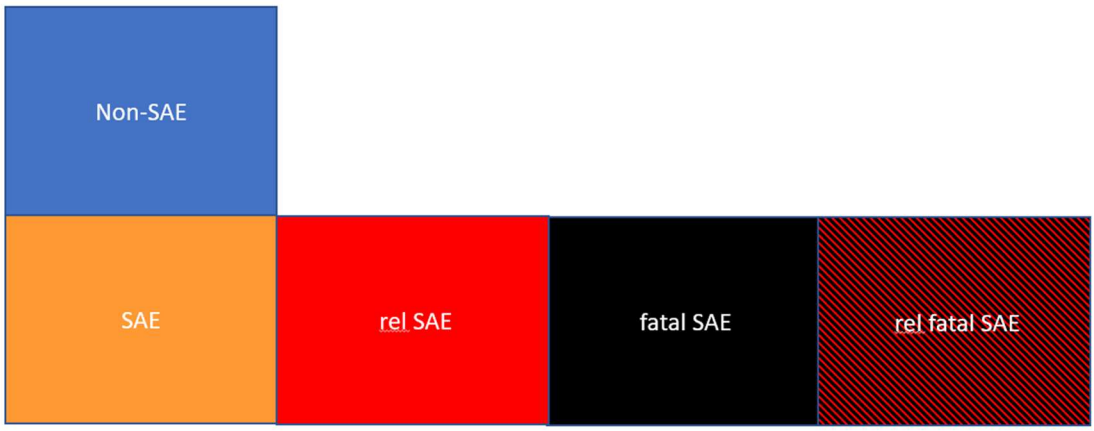

Figure 4 - Stacking the AE count data sets

5. Create the xml-file(s) using a SAS data step with many put statements (among some SAS dinosaurs this approach might also be known as proc put). Be careful when making this part of the code, the disclosure sites are not forgiving, and their error messages are far from plenty and informative.
6. Create the descriptive statistics in a separate program. Until now we haven't been asked to wrap that in xml. Note that these outputs must include all randomized/enrolled participants and not only the treated as it is common for the study report. These tables include:
    - Number of participants by gender and race for each treatment
    - Mean age by treatment
    - Participants per country and the EudraCT defined age groups

CONCLUSION

Typically, using the results from the study report do not work, as they are slightly different definitions, like SAEs are not considered a subset of AEs. They are apparently considered a separate category. That said, the study report can be used as a sanity check of your results. The numbers often differ, but not a lot. Another factor that can make the result for disclosure outputs differ from the trial report is that some counts are at the disclosure sites are based on enrolled patients rather than treated patients. Remember to brief your QC'er that these differences most likely occur, as that will save the both of you from panic.

Due to the quite limited possibilities when presenting results at the disclosure sites sometimes it requires a little creativity to present "complex" trials. But therein lies the fun challenge for the programmer.

We have made a solution by putting parts from other systems together in a not entirely seamless fashion. But it works and it is rather easy to use. And sometimes that is all what you need. Don't overdo stuff.

Communication, especially with medical documentation, is important, especially whether they plan to pool treatment groups or when the study design is complex.

The results will only match the study report when the study design is simple and there are no SAEs

No, it is not a big task, but ensure that time and resources are allocated. Make this output a part of the standards package, so it does not become the forgotten output, that someone must scramble to make before deadline.

Don't be afraid to plunge into the unknown, you will learn something and if no-one else around you know about that subject you are suddenly the expert. That was the authors experience being thrown into making these output.

REFERENCES

The disclosure sites…

ACKNOWLEDGMENTS

Thanks to Kerstin Dalström, Director, Transparency & Outsourcing, H. Lundbeck A/S for providing input to this paper.

CONTACT INFORMATION

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

Erik Brun
H. Lundbeck A/S
Ottiliavej 9

2500 Valby
Email: erik@lundbeck.com
Web: www.lundbeck.com
LinkedIn: https://www.linkedin.com/in/erik-brun-149a884/