

PhUSE 2017

Paper DH02

The Secure Data Office Concept

Annemie Van Dijck, SGS, Mechelen, Belgium

ABSTRACT

Pharmacokinetic (PK) and pharmacodynamic (PD) data can be unblinding. This implies that for double blind studies the PK/PD data are often only included in the clinical database after database lock. The PK/PD data cannot be cleaned and the programming of input files for analysis will only start after the lock. This can bring your PK/PD analysis on the critical path for submission. Here comes the introduction of an unblinded party to the rescue. The Secure Data Office will do the data management activities and programming of PK/PD input files already before the lock. It is critical that this is an independent party, not involved in the interpretation of the data. The quality of the data will improve and the PK/PD analysis can start significantly earlier. An additional benefit is that this dedicated team will build up experience in handling these specific types of data.

INTRODUCTION

In a double blind study, it is critical that the subjects as well as investigators and everybody involved in interpretation and analysis of the trial outcome measurements is unaware of the subject's treatment. This means that all randomisation information and unblinding data (e.g. lab data), are only accessible for a very limited group of people. For example, pharmacists, who need to prepare the dosing, and people of the lab analyzing blood samples are unblinded during study conduct. On the other hand, other groups of employees are absolutely forbidden to have access to any unblinding information. Data managers, who have the opportunity to manipulate data, could misuse treatment information to make a certain cohort/treatment group look better. Statisticians, PK analysts and modelers are only allowed access to study data after it is locked (data manipulation is no longer possible) and the analysis plan is final.

The downside is that this unblinding data does not get cleaned. This can cause delays in database lock as mismatches between e.g. lab samples and Case Report Form (CRF) information captured on those samples are only detected too late. In some cases the unblinding lab data is not even included in the clinical database lock, but is added later on. Every PK analyst will be able to tell you that this missed start will result in crunched timelines, because the PK results still need to be in the clinical study report, together with all other analysis results. This goes even further than only PK data, as the same applies for biomarkers, immunogenicity data and all other data that might be unblinding.

So how can you make your PK/PD input files available at database lock? This is where the secure data office (SD office) comes into play.

SD office is a dedicated team that is physically separated from all other departments and works on its own secure servers. Access to all folders on the server and all databases is limited to SD office employees only. No one from the SD office is involved in any analysis of the data. They are not allowed to make any changes to the data content wise. Having such a dedicated separate group gives you the benefit of being able to handle any type of unblinding data in a secure way. This process is also fully supported by procedures (documented in SOPs and WIs).

Lab data from PK samples or samples for any other purpose that might be unblinding can be transferred from the lab to the SD office. The SD Office sends sample identifier transfers to the data management department for the purpose of reconciliation between CRF and lab vendor data (and central lab data if applicable). At this time the trial is still in conduct. Any inconsistencies, such as missing samples, can still be queried. This leads to an improvement of the data quality and integrity. When sample results are ready, they are sent from the lab to the SD office. The data management department will send the CRF data to the SD office. The SD Office will make sure all the received data from the data management department and the lab will be merged and converted into the required Study Data Tabulation Model (SDTM) dataset structure. They create the domains for pharmacokinetic concentrations (PC), (part of) laboratory data (LB), immunogenicity data (IS) and other datasets as needed before the lock. During the lock process, they deliver the datasets to the data management department. As such, the PC and other unblinding data can be included immediately in the clinical database lock. If needed, datasets can also be delivered to safety committees or unblinded statistical support groups before the data management team is unblinded. It is also possible to deliver blinded test files or files with dummy data to statistical programmers or other parties to prepare scripts before database lock.

PhUSE 2017

For each transfer between an external party (e.g. a lab) and SD Office, or between SD Office and the data management department there is always a data transfer agreement (DTA) needed. This document describes the expected transfers: content, format, file type, frequency. In case of specific blinding requirements, it is also described in detail which parties can receive which variables. In case SD Office needs to clear the content of specific variables to avoid unblinding, this is also clearly indicated in the DTA.

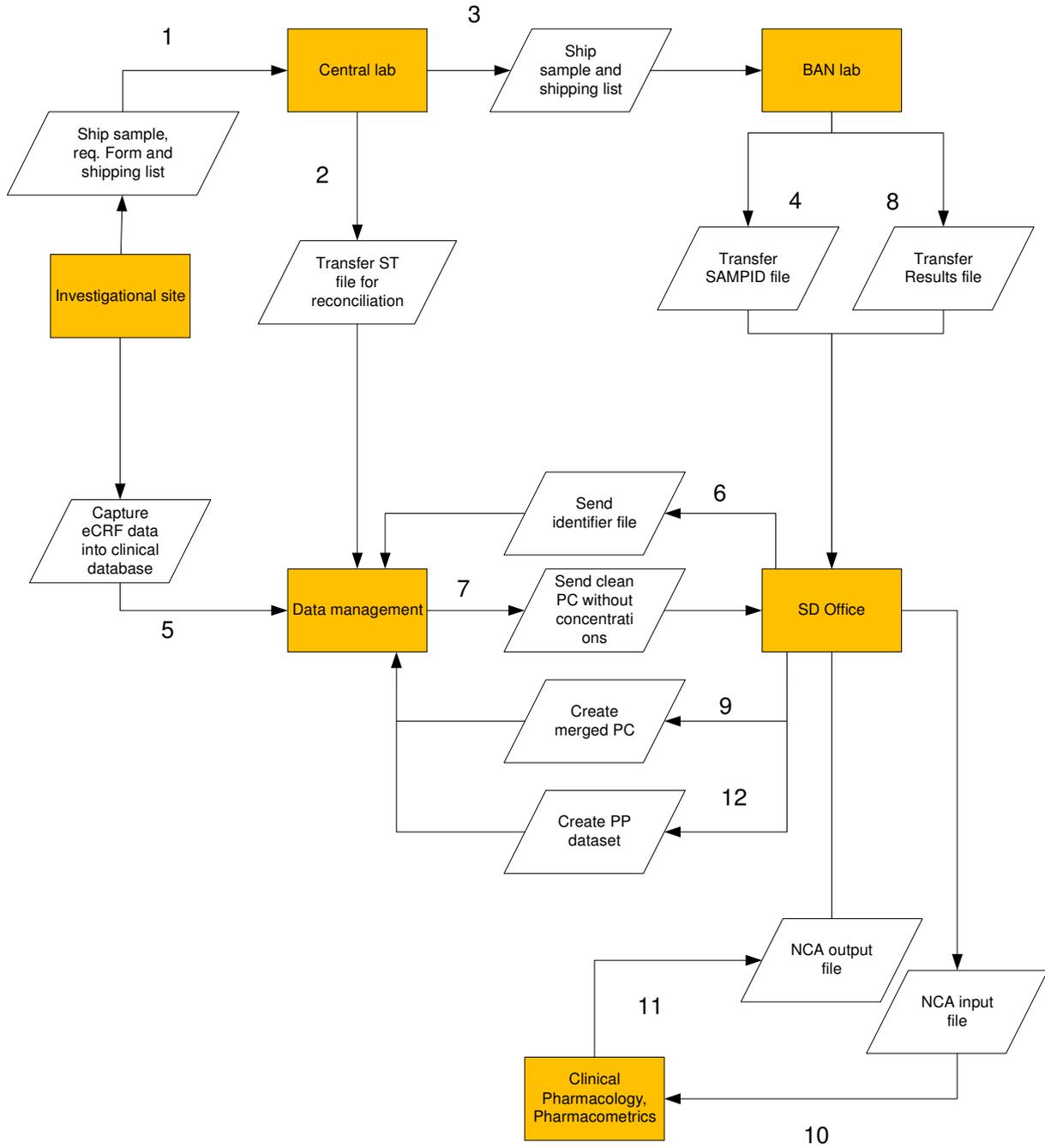
RANDOMISATION DATA

SD office can create randomisation lists. This makes sense, as they are unblinded throughout the entire study and not involved in any analysis. For simple studies, the randomisation list can be created in SAS ® and distributed on paper. Code breaking envelopes are provided in case emergency unblinding is needed. For more complex studies (multiple sites, stratification factors, larger number of subjects, ...) an Interactive Web Response System (IWRS) module can be set up. For blinded studies, SD Office will provide support and handle the unblinding data in the IWRS system, such as uploading the final medication kit list and randomisation list in the IWRS module or staged release of randomisation numbers in the system. For example, in some studies sentinel subjects are used. Sponsors might want to wait with dosing of the next subject a fixed number of hours to make sure no serious adverse events are occurring in this period. In that case, SD Office can manually release the next randomisation number at the expected moment.

Moreover, SD Office can create the SDTM datasets containing the randomisation information for double blind studies. There are two possible scenarios: an external IWRS vendor is involved in the study, or SD officers created the randomisation list themselves. In case an external IWRS vendor is involved, the random assignment of subjects to a certain treatment group happens automatically in the system according to predefined rules. An unblinded transfer of the randomisation data from the IWRS vendor to SD Office is needed. At the IWRS vendor, they have both the treatment information and the real subject IDs in the database. Both of these need to be transferred to SD Office. In case the randomisation list was created by SD Office, the link between randomisation number and treatment information is already available. The data management department should provide the link between the real subject ID and the randomisation number to SD Office. SD Office will then convert all this information into one SDTM dataset. If there is other information related to the randomisation (such as stratification factors), this is also included in the same SDTM domain. The advantage of having SD Office creating the randomisation datasets is that these can already be prepared before the lock. Possible issues can be detected sooner. SD Office can release this information to other unblinded parties, such as statistical support groups, independent data monitoring committees, independent review committees ... The randomisation information is also necessary for internal use in preparing the input files for pharmacokinetic or pharmacodynamic analysis before database lock.

PhUSE 2017

THE PK DATA FLOW



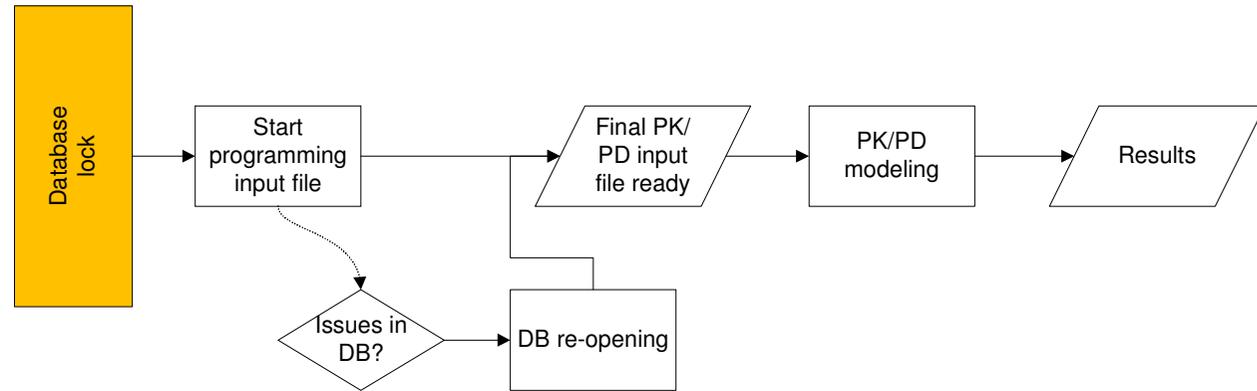
PhUSE 2017

1. The central lab collects all samples and sample shipping lists from the investigational sites. In addition, sample tracking data (subject number, collection date and time, sample reference number) from the lab requisition form will be stored in the central lab's database.
2. After entry of the requisition form data in the central lab database, these data are provided to the data management department and the bioanalytical (BAN) lab for storage in their database. This data is referred to as the "sample tracking file". The ST file can be used for data reconciliation by the data management department. After reconciliation (CRF data versus ST file), the data management department can query the site and/or central lab and request updates to the central lab database. New sample tracking files will be requested until all issues are resolved.
3. The samples with incremental shipping list(s) are sent to the BAN lab. The BAN laboratory will receive PK samples from investigational sites for analysis of drug concentrations via the central laboratory.
4. The BAN lab verifies if samples match with the sample shipping list. The BAN lab will store the sample identifier data in their database and will provide SD Office with the BAN sample identifier file(s) containing the unique sample reference number (REFID) and subject ID.
5. The data management department collects CRF data (including all information related to the PK samples, such as collection date and time, REFID) into the clinical database.
6. To ensure that there are no inconsistencies between the central lab database and the BAN lab database, reconciliation between the central lab sample tracking data, the BAN sample identifier file and the CRF data should be performed. Therefore, SD Office will provide the data management department with the SAS dataset containing the post processed BAN sample identifier file (converted to the same format as the SDTM PC dataset) for cleaning. Post processing might include e.g., derivation of subject ID if BAN delivers the randomisation number or incomplete subject ID; addition of extra fields (VISITNUM). The data management CRO can query the site and/or central lab. After reconciliation, the data management department can request updates to the BAN lab database. Only updates to subject ID and unique sample reference number will be performed by the BAN lab. To avoid unblinding, concentrations will not be included in this PC sample identifier file. Also other variables might be blinded as specified in the DTA. A new BAN sample identifier file should be sent to SD Office in case of updates to the data.
7. After reconciliation of the PK sample related CRF data and (Central) Lab data, the data management department provides SD Office with the clean PC domain without concentrations. This PC domain from data management contains:
 - o Merged (e)CRF data with Central Lab data Sample taken/not taken, scheduled times, actual dates and times, unique sample reference numbers from CRF and/or Central Lab
 - o All applicable test names (PCTEST, PCTESTCD) per sample
 - o The data management department will also send other SDTM datasets to SD Office: AE, CM, DS, CO, EX, DM, LB, VS, etc. to be used for creation of the NCA Input Files or NONMEM Input files
8. After analysis of the samples: the BAN laboratory provides the concentration results in the BAN sample result file(s) to SD Office. BAN is responsible for the accuracy of the PK results; this should not be cleaned by the data management department nor SD Office.
9. The clean PC file received from data management will be merged by SD Office with the PK results. For central lab studies, merging is done on subject ID and REFID. The merged PC file is sent to data management to be included in the locked database.
10. SD Office provides the NCA input files and/or NONMEM input files to the PK analysts.
11. SD Office receives the NCA output file and/or the NONMEM analysis parameters (if required) from the PK analysts.
12. SD Office provides pharmacokinetic parameters (PP) datasets and corresponding Define.xml files to the data management department.

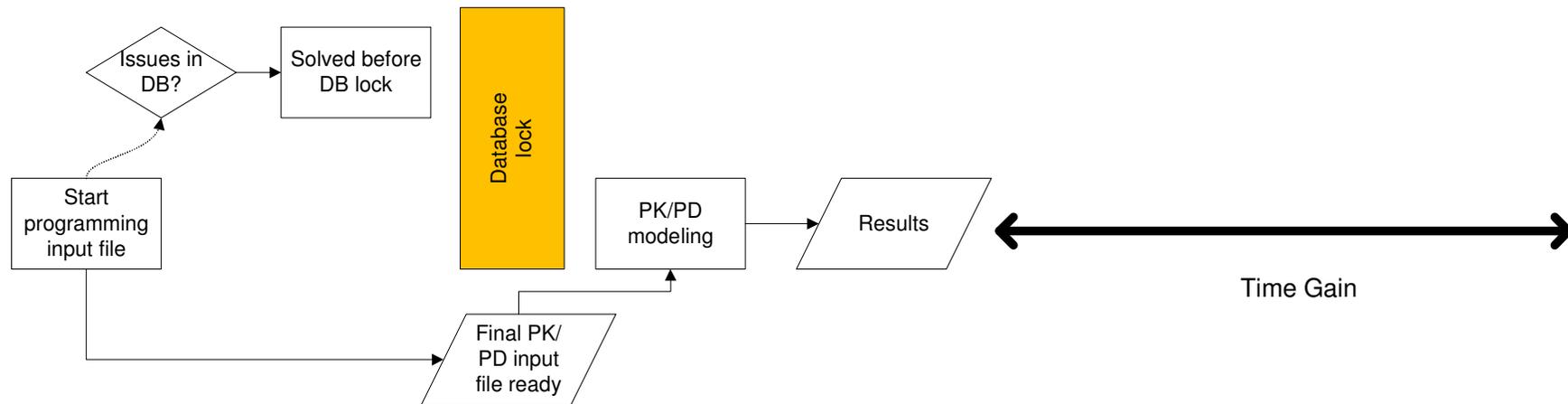
In case of a local lab study, there is no central lab involved. Similar to the central lab reconciliation, SD Office will provide the data management department with a BAN sample identifier file converted to the PC SDTM format. However, in this process, the reconciliation will be done between the CRF sample tracking data and the BAN sample identifier file. There is no sample tracking file from the central lab. The data management department can query the site or the BAN lab. After reconciliation and if applicable, the data management department can request updates to the BAN lab database. Only updates to subject ID, nominal time points and other identifier variables (clearly detailed in the DTA) will be performed by the BAN lab. The data management department will provide SD Office with a clean PC file. For local lab studies, there is no REFID. Merging is done on variables that uniquely identify each record: subject ID, visit, planned time point, specimen, analyte...

CREATION OF THE PK/PD INPUT DATASETS

Without SD Office



With SD Office



PhUSE 2017

At SD Office, input files for PK/PD analysis are created by SAS programmers. These input files are usually files for Non Compartmental Analysis (NCA) or Non-linear Mixed Effects Modeling (NONMEM). The SAS programmers will use the SDTM datasets, including the unblinding datasets available at SD Office, as source for the creation of the input datasets for PK/PD analysis. Sometimes also the SAS analysis datasets containing derived variables already calculated for analysis purposes will be used (ADaM datasets). This will ensure consistency between the PK/PD analysis and the statistical analysis. The ADaM datasets need to be provided by the statistical programmer to SD Office. Any variable derivations not required for statistical analysis will be derived by SD Office.

The requirements of the input files are described in a data specifications document. To create this document, the SAS programmer is in close contact with the PK/PD analyst, data management and the statistical programmers (if applicable). The document describes all sources to be used and contains the definitions of clear derivation rules. Once the specifications are in good shape, the programmers start working with the data from the clinical database before the database lock. All SDTM datasets need to be shared between the data management department and the SD office. If a stable draft of the data specifications document is reached and draft SDTM data is available, the programming of the PK and PD input files can start before the database lock. In this way, the programmers at SD office can start working with the clinical data, such as demographic data, vital sign and meal information. In addition, they have access to the randomisation information and the unblinding data. This can be PK data, but also outcome measurements of the study such as survival, time-to-event, questionnaires, ECG, EEG and viral load data. Any strange and unexpected values can be reported to the data management team. This will lead to an improved data quality. If the source data is confirmed to be correct, the PK/PD analyst will be contacted to discuss how to handle this unexpected data. SAS scripts will be written by the programmers to deal with this. At the time of the database lock, only a rerun of the prepared scripts with the final data (SDTM/ADaM) needs to be done. This results in a significant time gain. In case of very short timelines between database lock and study report writing, even more is possible. A blinded PK/PD input file can be delivered before unblinding. By using dummy subject IDs and removing all treatment information, the analyst can already have a look at the data before the study is unblinded. They can already start to build their models. After database lock, a compare of values before and after can be done, or small modifications can be made to the model.

OTHER UNBLINDING DATA STREAMS

The flow for other unblinding lab data streams is very similar to the PK data flow. Both the central lab and the local lab scenario are possible. Only the unblinding analytes should come via SD Office. All other analytes can be handled by the data management department. Depending on the nature of the lab data, the data can end up in different SDTM domains: for some sponsors PD is used as domain for pharmacodynamic measurements, IS for immunogenicity data, and LB for most other lab data.

When there is a need for a release of unblinding lab data before data management is unblinded, the interference of SD Office is required to create the SDTM dataset. SD Office will do the integration of the unblinding analytes into the dataset of data management. Examples of these releases are: release to statistical support group, independent review committee and independent data monitoring committee. In some cases the unblinded part is provided to the SD Office SAS programmers to be used for the preparation of the NONMEM files.

Not only lab data can be unblinding, in some trials also imaging data, tumor sizes, and answers on questionnaires can also be unblinding. To decide if the involvement of SD Office is useful, several questions should be considered:

- Will the identifier data be already available before the unblinding of the study? If not, there is no gain in involving an unblinded party.
- Are blinded transfers to data management sufficient to perform reconciliation? If yes, it is preferred that the reconciliation is done by data management. Can the data delivering party provide blinded cleaning transfers directly to the data management department? If not, the blinding should be done by SD Office.
- Does the unblinded content need to be cleaned: checks written by, and output reviewed by SD Office? In case it is decided that only an unblinded party can clean the data, who will resolve the issues detected by SD Office? Is there an unblinded party at the site that can answer to queries, such as an unblinded pharmacist, independent drug monitor, unblinded CRA? In case the answer to the last question is yes, the involvement of SD Office will result in a significant improvement of data quality.

DETECTION OF PROTOCOL DEVIATIONS

In some cases SD office is involved in the identification of unblinding protocol deviations (usually related to medication kit errors). For some studies the medication kit numbers can be unblinding, e.g. placebo and active kits starting with a different number. In that case, data management cannot receive the medication kit numbers at all and SD Office will do the medication kit reconciliation. If the medication kit numbers are not unblinding, it depends on

PhUSE 2017

how a protocol deviation is defined if SD Office should be involved or not. If any medication kit misallocation is considered a deviation, the detection can be done by data management and SD Office is not involved. If the misallocation to another treatment group is considered a major deviation, SD Office, who has access to the unblinded medication kit list (containing the medication kit content information), should verify if the medication kit misallocations led to a different treatment.

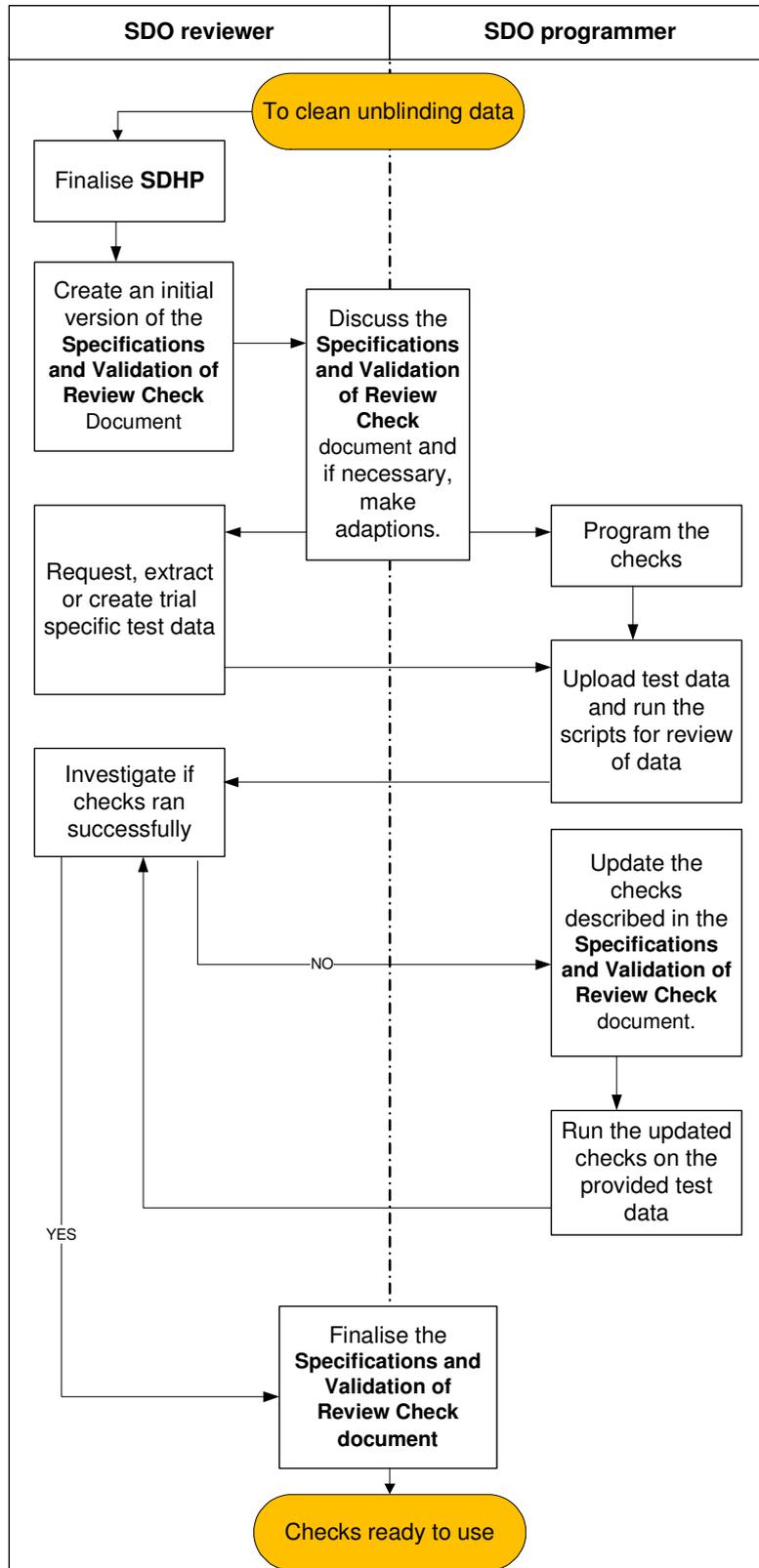
		Protocol deviation defined as	
		Any medication kit misallocation	Only treatment misallocations
Medication kit number unblinding?	Y	SD Office involved	SD Office involved
	N	Detection of medication kit misallocations by data management	SD Office involved

Medication kit numbers are assigned to the subject by the IWRS system at every medication kit dispensation. The assignment is based on the randomisation number and in line with the treatment group to which the subject was randomised. The assigned medication kit numbers represent the planned data. This data can be made available to SD Office in several ways, either via direct access of the SD Office employee to the eCRF (if this information is captured in the eCRF), an external transfer of medication kit data (from the independent drug monitor manager, combined information from the pharmacists of all sites) or via a report containing extracted medication kit data from the IWRS database. Dispensed medication kits are the ones that are physically delivered to the subject and represent the actual data. The dispensed medication kit numbers are usually captured in the eCRF.

Planned medication kit numbers must be reconciled with actual medication kit numbers. A distinction needs to be made between errors in medication kit reporting (=typing error) or real medication kit misallocations. The dispensed medication kit numbers must be mapped into the clinical datasets (EXREFID/DAREFID). The REFIDs are required to derive the corresponding batch lot numbers. In case the medication kit numbers are unblinding, this task cannot be performed by data management. The protocol deviations are categorized by SD Office into minor or major. Usually, a medication kit error is considered major when a wrong drug or dose level was taken by/administered to the subject. SD Office creates a part of the DV domain, containing only the unblinded protocol deviations. This contains records for the medication kit misallocations and the protocol deviations detected by the independent drug monitor manager, e.g. expired medication kit given. This unblinded part of the DV domain is delivered to the data management department at database lock and merged with the DV domain created by data management.

PhUSE 2017

CLEANING OF UNBLINDING DATA



PhUSE 2017

In case reconciliation can be done by data management, but the vendor cannot provide blinded transfers for identifier cleaning, SD Office is involved. SD Office will request transfers of unblinded data from external vendors (e.g. lab data, drug preparation data ...). The data is always stored in a protected environment. SD Office checks if the transfer is consistent with the specifications in the DTA. If not, they will request updates from the external vendor. SD Office sends blinded transfers (remove content for certain variables) and provides them to the data management department. In case the reconciliation can be done by data management, this is the preferred situation.

If cleaning of the secure data itself is required, SD office will perform the cleaning activities. The specific cleaning requirements are documented in the secure data handling plan (SDHP). This document is reviewed and approved by the sponsor. When the SDHP is finalized, the expected review checks will be defined in the Specifications and validations of review check document. The reviewer and the programmer will discuss the technical details and include them in the same document. The programmer will program the checks as described. The programmer will run the checks on the trial specific test data. The reviewer will verify that the checks ran successfully and that the expected predefined discrepancies are identified. This process is repeated until all checks are proven valid.

At the moment production data is received, the checks will be executed on the production data. The programmer will provide the reviewer with the outcome of the checks. Feedback of the review is returned to the vendor or queries are written in the eCRF by SD Office. If no blinded eCRF pages are available, all queries need to be written in a blinded fashion. A better option is that blinded pages are foreseen. Settings should allow only unblinded trial personnel to access these pages. In these blinded pages it is possible to send unblinding queries to the correct people. In this case, SD office needs access to the relevant forms in the eCRF system.

An example of cleaning often performed by SD Office is the cleaning of drug preparation data. For some trials, the drug preparation is done by an unblinded pharmacist in advance of drug administration. The details of the drug preparation are captured in the eCRF. The unblinded pharmacist and SD Office have access to the unblinding drug preparation pages in the eCRF, SD Office reviews the unblinding data and can send queries in the eCRF system to the unblinded pharmacist.

CONCLUSION

The introduction of SD office in your data management processes can be very beneficial. SD Office can as unblinded independent party start cleaning unblinding data or send cleaning transfers to data management before the database is locked. As such they can play an important role in the PK flow. Since the PK data can be cleaned earlier and the input files are prepared upfront, the analysis of the PK and PD data can start significantly earlier. When the same programmers are always handling PK and PD data, they become very experienced on the subject. Next to PK data, SD Office can handle any type of unblinding data. They create randomisation lists and the SDTM datasets containing the randomisation data. They do cleaning on unblinding data and detection of medication kit misallocations. As they can already perform this work before database lock, the unblinding data will be of good quality and can already be included in the first lock. Unblinded transfers can be delivered to data review committees or statistical support groups. The blinded or dummy transfers can be provided to statisticians and PK analysts to perform dry runs or prepare their scripts before database lock. The introduction of SD office as a partner in your data management process will result in better quality of your data at database lock and earlier access to the analysis results.

CONTACT INFORMATION

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

Annemie Van Dijck
SGS
Generaal De Wittelaan 19A
2800 Mechelen
Work Phone: +32 488 245 699
Email: annemie.vandijck@sgs.com

Brand and product names are trademarks of their respective companies.