**Paper DH04**

# Automating your safety data reconciliation: An approach towards reliable and efficient SAE reconciliation process

Ritu Dogra, SGS Health Science, Mechelen, Belgium
Agnieszka Bel, SGS Health Science, Mechelen, Belgium

## ABSTRACT

Serious Adverse Event (SAE) reconciliation is a critical step in the management of clinical trials. It involves comparing the safety and clinical database to identify, resolve, and document any discrepancies. Poor quality of the reconciliation may result in erroneous information being reported to health authorities, potentially jeopardizing the approval of new drugs/devices. To mitigate these risks, SGS implemented an automated process that enhances the accuracy and efficiency of SAE reconciliation.

This paper will guide you through our automated SAE reconciliation process. It begins with the upload of safety database extract into our internal database. Simultaneously, the AEREFIDs from the safety database are added into the SDTM database, linking both databases along with other variables, e.g. AEDECOD and USUBJID. Afterwards, automatic checks run comparing both databases and listing all discrepancies for review by the Clinical Data Manager. The process concludes with automatic generation of the reconciliation report, listing all identified discrepancies to be reviewed by all stakeholders.

## INTRODUCTION

As per ICH GCP, a Serious Adverse Event (SAE) is any unfavorable medical occurrence that is considered serious at any dose if it results in death, is life-threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity or is a congenital anomaly/birth defect[1].

SAEs are the most important indicators of the safety of a new drug, as they provide key information about the potential risks, that are associated with an investigational drug/device. As the safety of the patient is of utmost importance in the clinical trials, various strict guidelines have been put in place to protect the patient rights and wellbeing. These guidelines provide strict guidance on the reporting of safety data[2]. As a result, Adverse Event data is required to be collected in multiple databases i.e. in the safety database which is managed by the pharmacovigilance team/safety officers and in the clinical database which is managed by the Clinical Data Managers. Since the SAE data is stored in two separate databases, it is critical to make sure that data is consistent in between both databases. To achieve this, SAE reconciliation is performed.

SAE reconciliation process is the process of comparing the safety database and the clinical database to identify if there are any discrepancies between the two databases, whether they are acceptable or not, to resolve the identified discrepancies and to document the acceptable discrepancies in a clear SAE report.

The reconciliation process needs to have short-turnaround times to make sure that safety database and the clinical database remain consistent at any timepoint. However, if the reconciliation process is of poor quality, it may result in erroneous or incomplete information being reported to health authorities. This would impact the integrity of clinical data and potentially jeopardizes the approval of new drugs and medical devices. In some cases, it may delay the submission and result in financial loss. This highlights the need for an efficient SAE reconciliation process that is reliable and minimizes the risks of manual errors. In addition to this, reconciliation outcome needs to be reviewed by pharmacovigilance and data management team to take appropriate action or document any potential discrepancies. Therefore, the SAE reconciliation outcome should be available in a clear report that can be reviewed by all stakeholders. All these goals can be achieved by automating the SAE reconciliation process as much as possible.

Automatic SAE reconciliation has multiple benefits over manual reconciliation process. Automatic processes are capable of efficiently handling large volumes of data which makes it easier to manage and reconcile SAEs in a short timeframe in multisite trials and late phase trials that create high amounts of SAEs. By using an automatic process, we can minimize the need of manual inputs, which in turn lowers the risk of human error and minimizes inconsistencies. Automated systems are also capable of generating quick reports that are easy to comprehend.

We can say that overall, SAE reconciliation automation streamlines and speeds up the process, which allows for quicker identification of discrepancies and leads to higher quality and clear safety reporting. Ultimately, this enhances patient safety and supports regulatory compliance in clinical trials.

This paper will showcase the strategies implemented at SGS to achieve an efficient automatic SAE reconciliation process.

## KEY CONSIDERATIONS FOR AUTOMATING THE SAE RECONCILIATION PROCESS

From a CRO perspective, we want to create a flexible process, that can be implemented for different trials. As we work with different clients and safety database owners, we will be dealing with different data formats of safety database extracts such as SAS file, excel file, csv file, etc. Our automated system must be capable of uploading the safety database extract, regardless of the data format. Once the extract is uploaded, the system should be able to identify discrepancies between the two databases. Although automated systems can identify the discrepancies efficiently, we cannot ignore the importance of human involvement in resolving those discrepancies. Therefore, we would prefer a system that allows Clinical Data Managers to still review the data in an efficient way with tracking of the reconciliation status ensuring high quality and effective oversight. Finally, we want to have a system capable of generating clear reports that can be used by data management and safety database owners, to follow-up on SAE discrepancies.

Along with managing data from various safety database owners, CROs must also be able to handle data coming from different EDC systems with each having a different set-up of forms. However, to implement an automated process, it is essential to first establish a stable data structure. One solution to this is to automate the process on SDTM data instead of raw EDC data, as it provides a stable structure defined by CDISC[3].

At SGS, as a standard we have SDTM database available within few weeks from the time EDC goes live (duration varies based upon trial phase and complexity) (Figure1). We perform data cleaning and all the vendor reconciliations including the safety data reconciliation on SDTM datasets, instead of raw data from EDC systems, as they can have varying structure across the studies.

*Figure 1: Availability of SDTM database and data cleaning*


## STRATEGY FOR AUTOMATING SAE RECONCILIATION

Our strategy to automate SAE reconciliation can be divided in 4 steps:

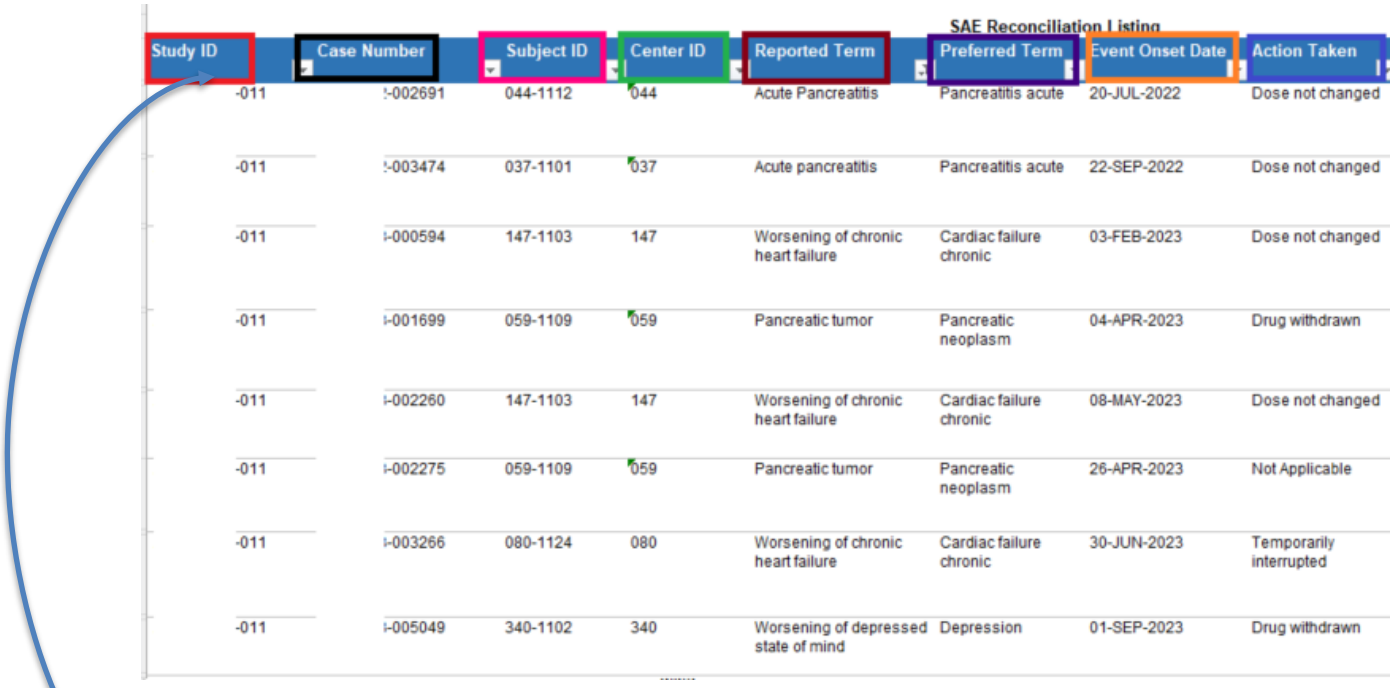
### STEP 1: RECEIVING AND UPLOADING THE SAFETY DATABASE EXTRACT.

First step is to receive a safety database extract from the safety database owner and to upload the extract into our database. As mentioned previously, safety database extracts can have different formats or file structure, for example SAS files, Excel files, csv files etc. At SGS, the safety database extract is uploaded by the Clinical Data Programmer (CDP). During the setup of the trial, the CDP needs to create a trial specific mapping script allowing the automation of the upload. This script is based on the specific structure of the safety database extract file for every trial and/or safety database owner. The script is created once at the beginning of the trial and is validated by the Clinical Data Manager to ensure the accurate upload of the safety database extract.

During the conduct phase of the trial, Clinical Data Programmer (CDP) runs the script, and the safety database extract will get automatically uploaded in a specific help table of our database. This help table has a fixed structure and therefore irrespective of different formats of safety database extract, it can be used to automate the SAE reconciliation process in any trial (Figures 3 and 5).

As the script has been programmed based on the structure of safety database extract during the setup phase, it is expected that no changes are made to the structure of the extract during conduct phase. We need to make sure that the safety database owner always provides the extract with same format, therefore, it is important to clearly discuss it at the start of the trial.

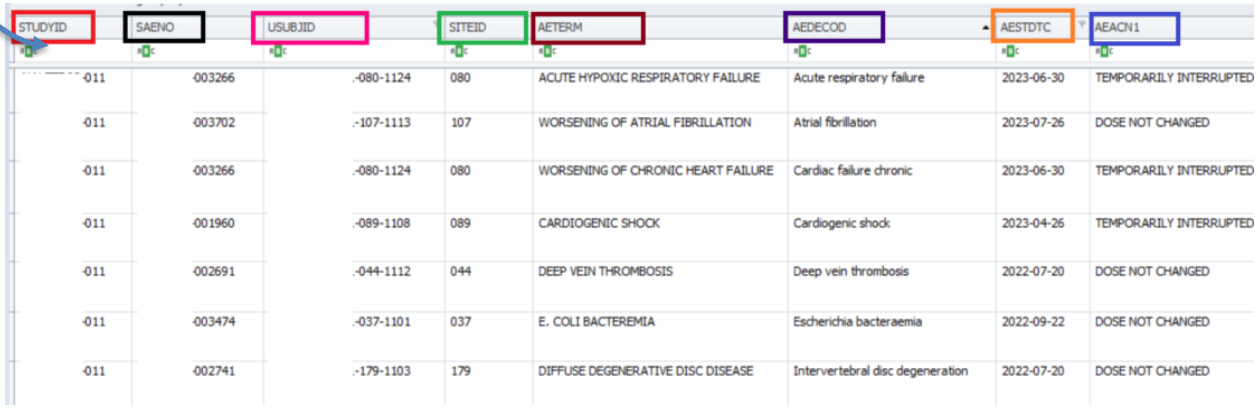
*Figure 2: An of example safety database extract (Excel format)*



SAE Reconciliation Listing

| Study ID | Case Number | Subject ID | Center ID | Reported Term | Preferred Term | Event Onset Date | Action Taken |
|---|---|---|---|---|---|---|---|
| -011 | -002691 | 044-1112 | 044 | Acute Pancreatitis | Pancreatitis acute | 20-JUL-2022 | Dose not changed |
| -011 | -003474 | 037-1101 | 037 | Acute pancreatitis | Pancreatitis acute | 22-SEP-2022 | Dose not changed |
| -011 | -000594 | 147-1103 | 147 | Worsening of chronic heart failure | Cardiac failure chronic | 03-FEB-2023 | Dose not changed |
| -011 | -001699 | 059-1109 | 059 | Pancreatic tumor | Pancreatic neoplasm | 04-APR-2023 | Drug withdrawn |
| -011 | -002260 | 147-1103 | 147 | Worsening of chronic heart failure | Cardiac failure chronic | 08-MAY-2023 | Dose not changed |
| -011 | -002275 | 059-1109 | 059 | Pancreatic tumor | Pancreatic neoplasm | 26-APR-2023 | Not Applicable |
| -011 | -003266 | 080-1124 | 080 | Worsening of chronic heart failure | Cardiac failure chronic | 30-JUN-2023 | Temporarily interrupted |
| -011 | -005049 | 340-1102 | 340 | Worsening of depressed state of mind | Depression | 01-SEP-2023 | Drug withdrawn |

*Figure 3: Help table (data from the SAE extract is mapped to specific variables in help table)*



| STUDYID | SAENO | USUBJID | SITEID | AETERM | AEDECOD | AESTDTC | AEACN1 |
|---|---|---|---|---|---|---|---|
| -011 | -003266 | -080-1124 | 080 | ACUTE HYPOXIC RESPIRATORY FAILURE | Acute respiratory failure | 2023-06-30 | TEMPORARILY INTERRUPTED |
| -011 | -003702 | -107-1113 | 107 | WORSENING OF ATRIAL FIBRILLATION | Atrial fibrillation | 2023-07-26 | DOSE NOT CHANGED |
| -011 | -003266 | -080-1124 | 080 | WORSENING OF CHRONIC HEART FAILURE | Cardiac failure chronic | 2023-06-30 | TEMPORARILY INTERRUPTED |
| -011 | -001960 | -089-1108 | 089 | CARDIOGENIC SHOCK | Cardiogenic shock | 2023-04-26 | TEMPORARILY INTERRUPTED |
| -011 | -002691 | -044-1112 | 044 | DEEP VEIN THROMBOSIS | Deep vein thrombosis | 2022-07-20 | DOSE NOT CHANGED |
| -011 | -003474 | -037-1101 | 037 | E. COLI BACTEREMIA | Escherichia bacteraemia | 2022-09-22 | DOSE NOT CHANGED |
| -011 | -002741 | -179-1103 | 179 | DIFFUSE DEGENERATIVE DISC DISEASE | Intervertebral disc degeneration | 2022-07-20 | DOSE NOT CHANGED |

*Figure 4: An example of SAS based safety database extract*

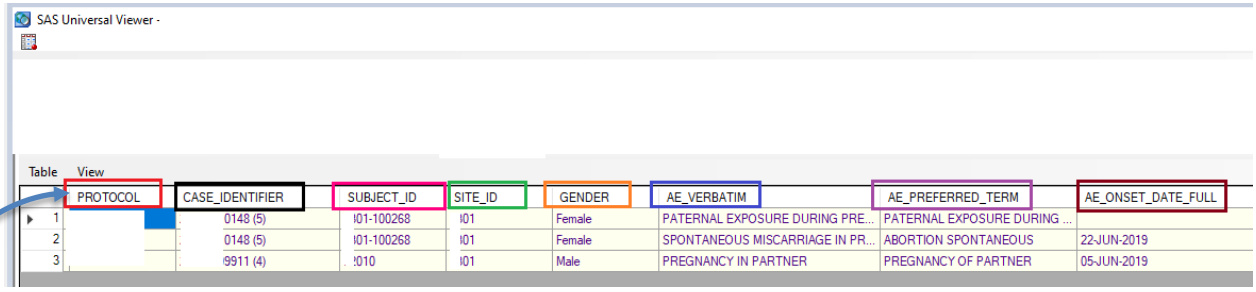
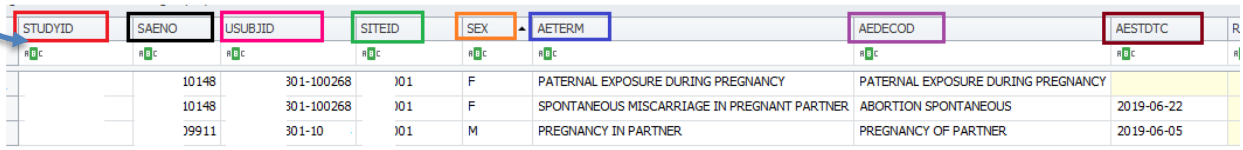



*Figure 5: Help table (data from the SAE extract is mapped to specific variables in help table)*




In Figures 2-5, it can be seen how the data from the different formats of safety database extract are mapped to specific variables in the help table (illustrated with same color boxes). Irrespective of the safety database extract formats, the structure of help table remains the same.

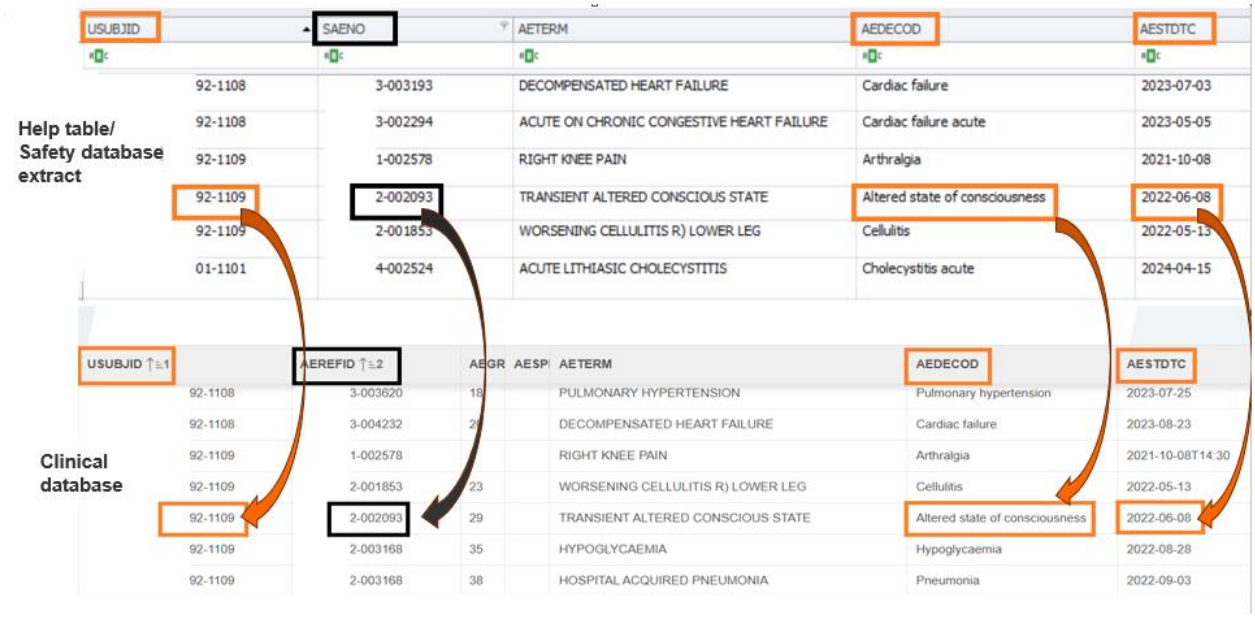
#### STEP 2: AUTOMATIC COMPLETION OF AEREFID VARIABLE IN THE CLINICAL DATABASE

Next, we proceed with the automatic completion of the AEREFID variable in the clinical database as this data is only present in the safety database extract. At SGS, we are using an in-house data entry tool that allows the automatic completion of AEREFIDs in the clinical database. Once the safety database extract is uploaded in the help table, the AEREFID variable in the clinical database is automatically populated with the SAENO data from the help table based on a linking between three key variables: USUBJID, AESTDTC and AEDECOD (Figure 6).

However, in some cases exact match between these three variables i.e. USUBJID, AESTDTC and AEDECOD may not be found in the clinical and the safety database. For instance, when AE term is coded differently in the clinical and safety database. In this scenario the linking will not work due to missing exact match and therefore the AEREFIDs will remain blank (Figure 7). To address this issue, it is important to have a contingency plan in place allowing the Clinical Data Manager to manually complete the AEREFID variable from the SAENO variable present in the safety database help table.

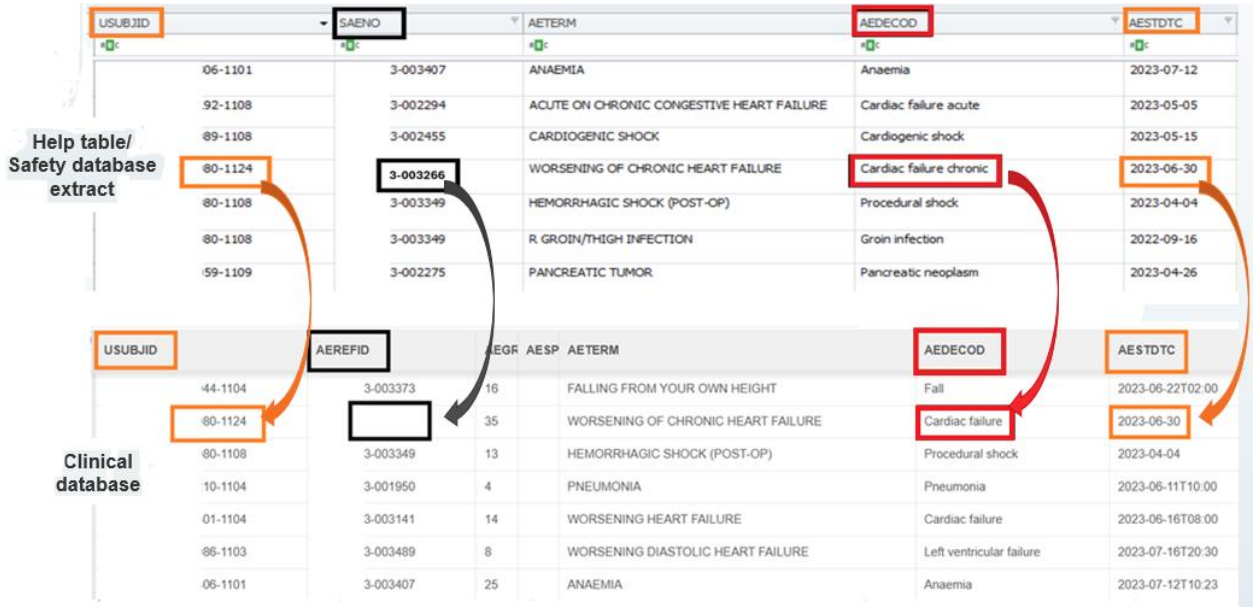
Our in-house data entry tool, which automatically completes the AEREFID variable, also allows Clinical Data Managers (CDMs) to manually complete the AEREFID variable in cases where automatic linking is not possible

*Figure 6: An example of automatically completed AEREFID in clinical database as there is an exact match between the two databases for USUBJID, AESTDTC and AEDECOD*




.

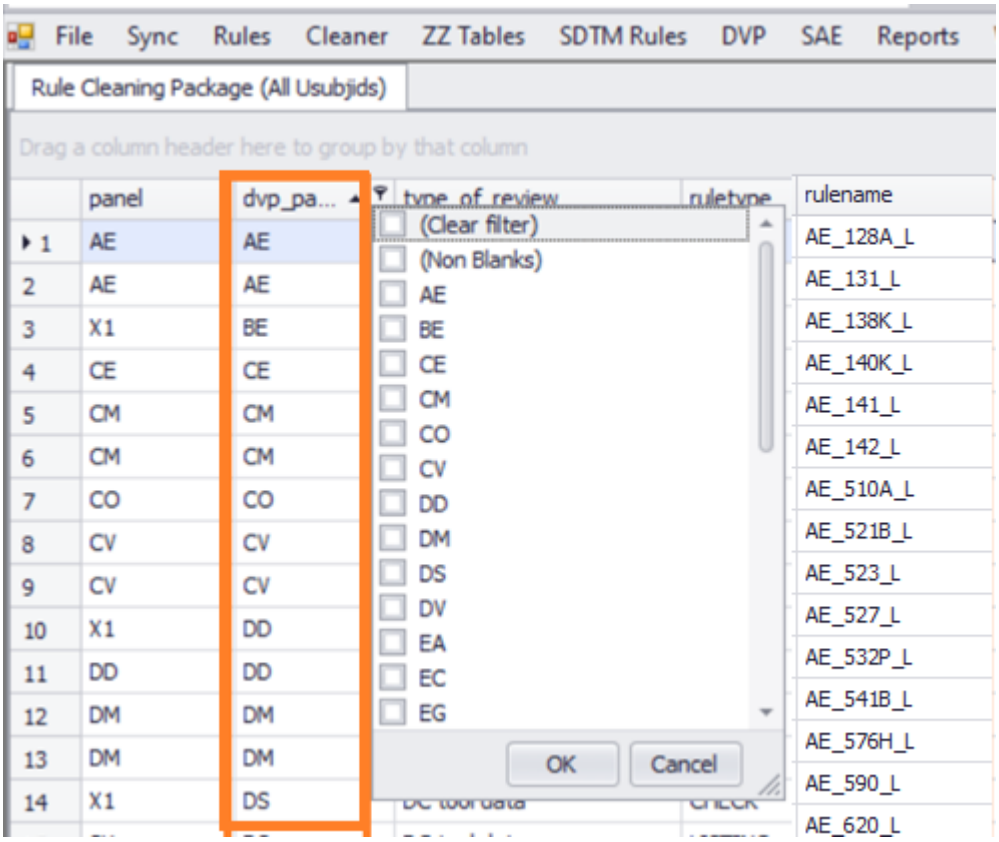
*Figure 7: An example of AEREFIDs not completed in clinical database, as coding is not matching with the coding in safety database*



### STEP 3: SAE RECONCILIATION

Next step is to have automatic rules in place, flagging discrepancies, and having the Clinical Data Manager take the appropriate actions. At SGS we have created an in-house data management cleaning tool that automates this process. This tool automatically executes the different programmed rules on the SDTM database and the help table, generating output for review by the Clinical Data Managers. For ease of the review, the output is organised per SDTM domain for example AE, BE, CM, CE etc. (Figure 8)

*Figure 8: Various domains in the SGS cleaning tool*




At SGS we maintain an in-house library of programmed rules. For each trial, either the corresponding rules are taken from inhouse library or new trial specific rules can be created by validating them in the tool itself. We can divide the SAE reconciliation rules in main four categories:

1. Safety: For each SAE in clinical database a corresponding SAE is present in safety database
2. Clinical: For each SAE in safety database a corresponding SAE is present in clinical database
3. Discrepancy: Discrepancies between clinical and safety database: Difference in AESTDTC, AETERM, AEDECOD and SAE criteria, e.g. AESDTH, AESHOSP, etc.
4. Other issues: If an AE is serious, then it has an AEREFID completed and vice versa.

To get maximum accuracy, for each category mentioned above, multiple rules can be programmed if needed. For example, the first rule (For each SAE in clinical database a matching SAE is present in safety database and vice versa) can be checked in two ways:

1. matching on AEDECOD and AESTDTC and
2. matching on reference number (AEREFID/SAENO).

All these rules are validated by the Clinical Data Managers during the set-up phase to make sure the correct output is created by each rule.

Let us understand in detail with an example of a rule that checks the differences or discrepancies between the safety

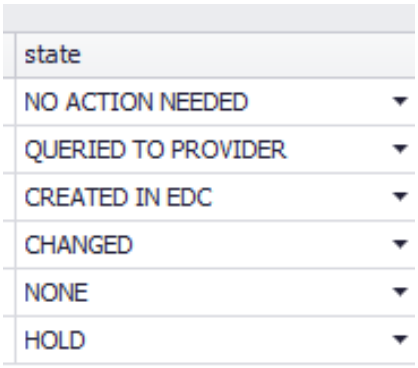
database and the clinical database. The presented rule compares various parameters/variables in both databases such as the coded AE term, start date and Serious Adverse Event criteria such as hospitalization, whether the SAE resulted in death or disability etc. and lists the differences that were found.

*Figure 9: Output from a specific programmed check comparing different variables in clinical and safety database*

| aedecod#crf | aedecod#sae | aestdtc#crf | aestdtc#sae | aesdisab#crf | aesdisab#sae | aesdth#crf | aesdth#sae | aeshosp#crf | aeshosp#sae | diff#all | state |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pneumonia | Pneumonia | 2024-12-27T09:00 | 2024-12-27 | N | N | N | N | Y | Y | AETERM - AEOUT - AESM... | QUERIED TO PROVIDER |
| Arrhythmia | Arrhythmia | 2022-12-11 | 2022-12-11 | N | N | N | N | Y | Y | AEREL | NO ACTION NEEDED |
| Confusional state | Confusional state | 2024-09-20 | 2024-09-20 | N | N | N | N | Y | Y | AEOUT | CREATED IN EDC |
| Confusional state | Confusional state | 2024-09-24 | 2024-09-24 | N | Y | N | N | N | N | AESDISAB | QUERIED TO PROVIDER |
| Urinary tract infection | Urinary tract infection | 2024-11-01 | 2024-11-01 | N | N | N | N | Y | Y | AESMIE | QUERIED TO PROVIDER |
| Bundle branch block left | Bundle branch block left | 2022-12-03 | 2022-12-03 | Y | Y | N | N | Y | Y | AETERM - AEACN | NO ACTION NEEDED |
| Septic shock | Septic shock | 2024-07-29T23:08 | 2024-07-29 | Y | Y | Y | Y | Y | Y | AEACN - AEOUT | QUERIED TO PROVIDER |
| Hypovolaemic shock | Hypovolaemic shock | 2022-03-13 | 2022-03-13 | N | N | N | Y | Y | Y | AEACN - AEOUT - AESDTH | NO ACTION NEEDED |
| Cardiac failure | Cardiac failure | 2024-12-04 | 2024-12-04 | N | N | N | N | Y | Y | AESEV | CREATED IN EDC |
| Acute kidney injury | Acute kidney injury | 2024-12-25 | 2024-12-25 | N | N | N | N | Y | Y | AEOUT | QUERIED TO PROVIDER |

In Figure 9, the data from corresponding variables between the clinical and safety database can be seen. For example, the first 2 columns highlighted in black are the coded AE terms in the clinical and safety database. In next 2 columns you can see the comparison of AE start date between the two databases. The 2nd last column 'diff#all' lists out all the differences that were present for a particular record. Clinical Data Managers review this output and take the appropriate actions (last column). The tool gives an option to choose from a list of states that can be provided to each record based on what action is needed or taken for a particular record (Figure 10).

*Figure 10: List of states that can be assigned to each record*




In Figure 10, you can see the list of states that can be assigned to the records allowing for efficient tracking of the SAE reconciliation. In case a new safety database extract is loaded, the status is retained for the records where there is no update and all the new records which need to be reviewed by Clinical Data Managers are automatically assigned to NONE. For SAE reconciliation mainly 4 states are used:

1. NO ACTION NEEDED is assigned to the records that either have no discrepancies or have already been checked by the safety database owner and agreed as acceptable discrepancies, for example, slight variations in terminology used in describing events may be of no consequence or the cases that have been agreed as unresolvable.
2. QUERIED TO PROVIDER is used for the records when an update request has been sent to the safety database owner.
3. CREATED IN EDC is assigned to the records for which a query has been raised in the EDC.
4. HOLD is used for records which are added recently in clinical database. For example, a new SAE is added in EDC shortly before we received the safety database extract. This SAE will not be present in the current safety database extract therefore it cannot be reconciled during this round and will be kept on hold till we receive the next safety database extract.

The system should also be able to identify and highlight the changes in the data. In our data cleaning tool, there is a special state named as CHANGED which is automatically assigned to the records that have been changed in comparison to previous safety database extract or previous EDC extract (clinical database). This feature is essential for tracking the changes in the data (Figure 11).

*Figure 11: Changed state with details of the variables that were changed in comparison to previous safety database extract or EDC extract*

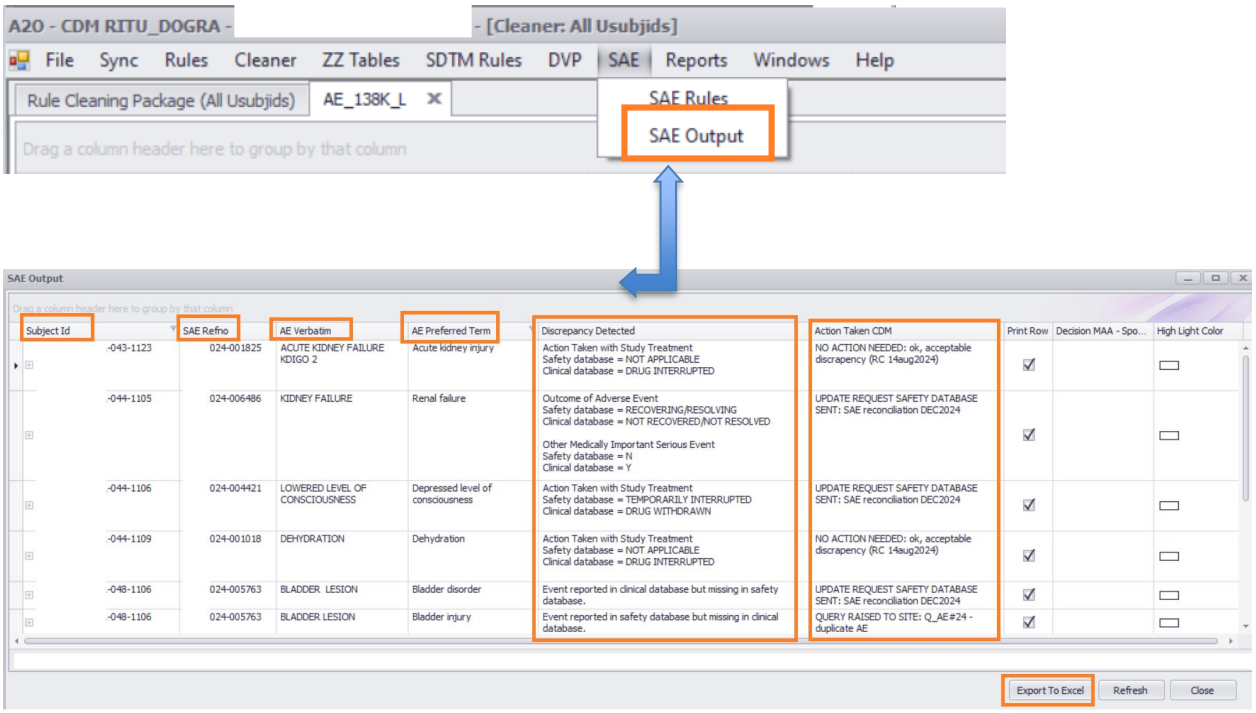
| state | comments |
|---|---|
| CHANGED | AESTDTC#CRF has changed, OLD value was 2024-03-24,DIFF#ALL has changed, OLD value wa |
| CHANGED | AESTDTC#CRF has changed, OLD value was 2024-04-10,DIFF#ALL has changed, OLD value wa |
| CHANGED | AESEV#CRF has changed, OLD value was MODERATE,AEOUT#CRF has changed, OLD value wa |
| CHANGED | AEACN#CRF has changed, OLD value was DOSE NOT CHANGED,DIFF#ALL has changed, OLD v |
| CHANGED | AEACN#CRF has changed, OLD value was DRUG INTERRUPTED,DIFF#ALL has changed, OLD va |
| CHANGED | AEACN#CRF has changed, OLD value was DRUG INTERRUPTED |
| CHANGED | AEACN#CRF has changed, OLD value was DRUG INTERRUPTED |
| CHANGED | AETERM#CRF has changed, OLD value was PROGRESSIVE ATTR AMYLOID CARDIOMYOPATHY, |
| CHANGED | AEOUT#CRF has changed, OLD value was FATAL,AESDTH#CRF has changed, OLD value was Y, |

The Clinical Data Manager reviews the output from all the programmed checks and assigns the specific states to all the records.

#### STEP 4: AUTOMATIC REPORT GENERATION

Last step is the automatic generation of the reconciliation report. At SGS, the report is automatically created in the same tool where cleaning is performed. This report is based on the output from all the programmed checks that are used for SAE reconciliation.

*Figure 12: Structure of an automatically generated SAE report*

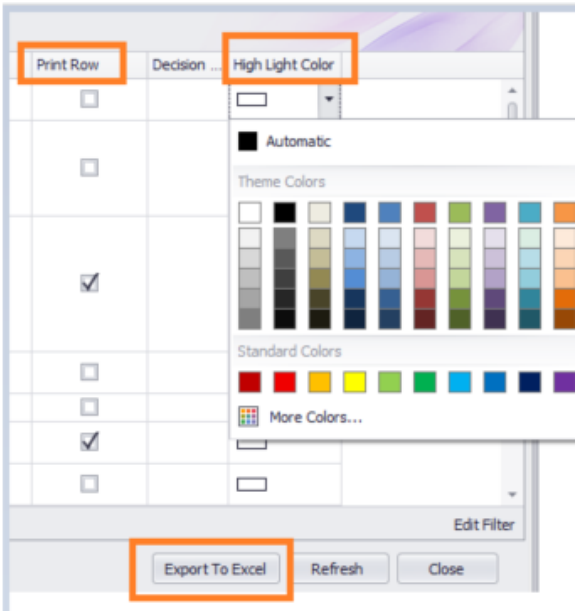



The report contains all the important variables such as Subject Id, SAE reference number, AE verbatim as well the coded term (AE preferred term) along with the list of discrepancies that were found in all the programmed rules and the action taken by the Clinical Data Manager. The states assigned by Clinical Data Managers in the cleaning tool are replaced in the report as follows for clarity:
QUERIED TO PROVIDER --> UPDATE REQUEST SAFETY DATABASE SENT
CREATED IN EDC --> QUERY RAISED TO SITE
NO ACTION NEEDED and HOLD remain the same as they are self-explanatory.

An interesting feature that we implemented in our tool is that it provides an option to modify or customize the report by using below two options and an option to export this report to Excel:

1. Print row
2. Highlight color

*Figure 13: Print row and Highlight color option to customize the report*




Print row option enables the Clinical Data Managers to select or deselect the record/rows that should be present in the downloaded report.
Similarly highlight color option can be used to make the report visually comprehensible. For example, the records which have already been checked by the safety database owner and do not need any further action can be colored green. The records which need to be reviewed can be highlighted in orange or other agreed colors (Figure 14).

*Figure 14: SAE report in excel format*

**Protocol Reference:** *Version: Draft*
**SGS Internal Reference:** *Date: 13/Dec/2024*

| Subject ID | SAE refno | Discrepancies Y/N | AE Verbatim | AE Preferred Term | Discrepancy detected | Decision Safety database owner/Sponsor | Action taken CDM |
|---|---|---|---|---|---|---|---|
| AA-BB-CC | AA-BB-CC | Y | LIGHTHEADEDNESS | Dizziness | Reported Term for the Adverse Event<br>Safety database = LIGHT HEADED DUE TO 2:1 HIGH DEGREE AV BLOCK<br>Clinical database = LIGHTHEADEDNESS<br><br>Severity/Intensity<br>Safety database = SEVERE<br>Clinical database = MILD<br><br>Serious Event<br>Safety database = Y<br>Clinical database = N | | UPDATE REQUEST SAFETY DATABASE SENT: SAE reconciliation DEC2024 |
| AA-BB-CC | AA-BB-CC | Y | ATRIAL FIBRILLATION | Atrial fibrillation | Reported Term for the Adverse Event<br>Safety database = ATRIAL FIBRILLATION WITH HEART BLOCK<br>Clinical database = ATRIAL FIBRILLATION | | UPDATE REQUEST SAFETY DATABASE SENT: SAE reconciliation DEC2024 |
| AA-BB-CC | AA-BB-CC | Y | LEFT SHOULDER PAIN | Arthralgia | Event reported in safety database but missing in clinical database. | | NO ACTION NEEDED: ok, acceptable discrepancy RC14aug2024 |
| AA-BB-CC | AA-BB-CC | Y | SEPSIC SHOCK | Septic shock | Reported Term for the Adverse Event<br>Safety database = SEPTIC SHOCK<br>Clinical database = SEPSIC SHOCK | | NO ACTION NEEDED: Ok, AETERM consistent(Wrong spelling in EDC) |

In the above Figure 14 you can see an example of downloaded report in excel format.

Once the report is ready, it is shared with the client and safety database owners and feedback is requested. Based on the feedback, the appropriate action is taken by the Clinical Data Managers to complete the review round.

## CONCLUSION

At SGS, we implemented automation at each step of SAE reconciliation process. Starting with the automatic upload of the safety database extract, the process includes the automatic completion of AEREFIDS through the linking of the safety and clinical databases. Listings for review by the Clinical Data Managers are also generated based on automatic checks that run on the SDTM database. At the end of the process, an automatic report generation completes the workflow.

In our practice, using the automated SAE reconciliation process enables efficient reconciliation of even large number of SAEs in a short timeframe while maintaining high quality. The checks that run on SDTM datasets are validated by the Clinical Data Managers which ensures accuracy. With our automated systems in place, we can easily track the records that have already been reconciled and require no further updates. These records do not need to be checked during next round of reconciliation, saving valuable time. Using SDTM datasets ensures consistency and simplifies the reconciliation process. Automated tools allow us to generate comprehensive and customized reports which would take more time and effort if done manually and reduce the risks of human errors in the report.

## REFERENCES


1. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). (2025). *Guideline for Good Clinical Practice E6(R3)*. ICH_E6(R3)
2. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). (1994). *ICH Topic E2A: Clinical Safety Data Management*.
3. Clinical Data Interchange Standards Consortium (CDISC). *Study Data Tabulation Model (SDTM)*.


## CONTACT INFORMATION

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

Author Name: Ritu Dogra, Agnieszka Bel
Company: SGS Health Science
Address: Generaal De Wittelaan 19A bus A5
2800 Mechelen - Belgium
Work Phone:
EUROPE: +32 15 27 32 45
Email: clinicalresearch@sgs.com
Website: www.sgs.com/cro