Paper DS19

# How unique is the subject identifier in a study having multiple screenings and enrollments?

Anahita Vyas, Merck & Co., Inc., Rahway, NJ, USA
Rama Kudaravalli, Merck & Co., Inc., Rahway, NJ, USA

## ABSTRACT

In clinical studies, the process of mapping Subject Identifiers (SUBJID) and Unique Subject Identifiers (USUBJID) within Study Data Tabulation Model (SDTM) datasets poses significant challenges, particularly when subjects undergo multiple screenings across one or more clinical trials. These challenges are primarily related to maintaining data integrity, ensuring accurate subject tracking, and adhering to the regulatory standards established by authorities such as the FDA. By addressing these challenges through systematic processes and robust data management practices, sponsors can significantly enhance the integrity of their clinical data submissions and improve compliance with regulatory standards. When subjects participate in multiple studies or are re-enrolled within the same study, the USUBJID is designed to uniquely identify a subject across all studies and submissions, preventing data duplication and ensuring consistency. This paper highlights the challenges in the data collection, mapping SUBJID and USUBJID uniquely in SDTM domains across multiple screenings and adhering to the conformance rules.

## INTRODUCTION

The concept of multiple screenings and enrollments was first articulated in the FDA Study Data Technical Conformance Guide (TCG) version 4.2, published in October 2018. According to this guidance, if a subject is screened more than once during a study, each unique screening or enrollment must correspond to a different SUBJID. For subjects undergoing multiple screenings, the primary screening should be recorded in the DM (Demographics) domain, while subsequent screenings should be documented in a custom domain designated as DC (Demographics as Collected). Furthermore, in studies that involve multiple screenings, the SUBJID must be added in other SDTM domains, beyond just the DM domain. Comprehensive explanations regarding these requirements can be found in the Study Data Reviewer's Guide (cSDRG).

## MULTIPLE SCREENINGS AND ENROLLMENTS IN CLINICAL STUDIES

The purpose of implementing multiple screenings and enrollments in clinical trials serves several important functions that contribute to the overall effectiveness and efficiency of the study. Firstly, it enhances the accuracy and tracking of each subject, allowing for more reliable data collection and management throughout the trial. Additionally, this approach can help save costs associated with subject recruitment, as it enables researchers to retain potential participants who may require more than one screening attempt to meet the eligibility criteria. Understanding the reasons for re-screening is also crucial; factors such as conflicts in consent timing, the need for stabilization of medication, or the resolution of acute events can all necessitate a second screening. Furthermore, robust record-keeping is facilitated through this process, as it provides detailed documentation of the number of visits for screening and the dates on which informed consent was obtained. Overall, multiple screenings and enrollments promote data transparency and clarity, ensuring compliance with regulatory standards and enhancing the integrity of the trial data.

### KEY BENEFITS OF STRUCTURED SCREENING PROCESSES

Implementing a structured screening process in clinical trials offers numerous advantages that enhance the overall efficacy and reliability of participant management. This method not only ensures accurate enrollment but also aids in meticulous record-keeping to comply with regulatory standards. Below are few key benefits described:

- **Enhanced Accuracy**: A thorough evaluation of subjects during the screening process significantly increases the likelihood of enrolling only eligible participants. By utilizing comprehensive criteria, researchers can ensure that candidates meet all necessary requirements before entering the trial.

- **Improved Data Reliability**: The introduction of unique identifiers for each screening attempt facilitates accurate data collection and tracking. This system allows researchers to monitor all screening activities effectively, thus ensuring that the data collected is trustworthy and reflective of each participant's true status.

- **Flexibility in Recruitment**: The ability to retain potential participants who may require additional screenings offers great flexibility. This approach not only helps in keeping interested subjects engaged but also improves overall recruitment efforts, ultimately leading to a more diverse participant pool.

- **Cost Efficiency**: By reducing recruitment costs, this structured process maximizes the utilization of individuals who may need multiple attempts to meet the eligibility criteria. This efficiency leads to more effective allocation of resources, allowing research teams to focus on other critical areas of the trial.

- **Documenting Eligibility Changes**: Systematic documentation of any changes in a subject's eligibility status over time is essential for understanding participant dynamics. This tracking helps researchers adapt their strategies and maintain the integrity of the trial.

- **Regulatory Transparency:** Finally, a structured method promotes compliance with regulatory standards, ensuring precise record-keeping and subject tracking. This transparency is crucial for maintaining the trust of regulatory bodies and stakeholders, which is vital for the success of any clinical trial.

By adopting these structured screening practices, clinical trials can achieve improved accuracy, reliability, and efficiency, ultimately contributing to the successful completion of research objectives.

**CHALLENGES WITH MULTIPLE ENROLLMENTS**

Managing Subject Data Tabulation Model (SDTM) datasets presents significant challenges, particularly when subjects enroll multiple times in the same clinical trial. Given the complexity of tracking and maintaining data integrity under these circumstances, several key challenges arise as below:

- **Distinct SUBJID Assignments**: Each subject must have a unique Subject Identification (SUBJID) for each screening attempt. This requirement complicates the tracking and management of identifiers, necessitating clear associations to ensure data accuracy.

- **Maintaining a Unique USUBJID**: The Unique Subject Identification (USUBJID) must remain consistent across all screening attempts. Balancing distinct SUBJIDs while preserving the integrity of the USUBJID requires careful planning to prevent data management complications.

- **Impact on Data Integrity**: Discrepancies in the assignment of SUBJID and USUBJID can lead to critical errors in data merging, undermining the overall integrity of the dataset. Inconsistencies or changes in these identifiers complicate data analysis and can adversely affect study outcomes.

- **Data Duplication**: Re-enrollment can lead to duplicated data entries, where similar or identical data points are noted for multiple screening attempts. This duplication can confuse data analysis, skew results, and complicate the interpretation of outcomes.

- **Inconsistent Data Collection**: Variability in data collection protocols across different screening attempts may result in inconsistencies in data type and quality. Changes in survey formats or assessment methods can hinder the comparability and aggregation of results.

- **Subject Withdrawal and Re-enrollment**: Tracking subjects who withdraw and later re-enroll can complicate data management. Researchers must ensure the current data reflects the subject’s status while retaining historical records, requiring a robust data management system.

- **Regulatory Compliance**: Managing multiple enrollments introduces compliance challenges regarding the documentation and reporting standards set forth by regulatory authorities, such as the FDA or EMA. Meeting these requirements can be resource intensive.

- **Increased Administrative Burden**: The complexity of managing multiple identifiers increases the administrative burden on clinical trial teams. This complexity often leads to higher operational costs and necessitates additional personnel or specialized training.

- **Complicated Adverse Event Reporting**: Reporting adverse events becomes more complex for subjects enrolled multiple times, particularly when accurately attributing incidents to the correct enrollment period. Compliance with reporting regulations becomes increasingly challenging.

- **Statistical Analysis Complexity**: The need to account for multiple enrollment instances complicates statistical analysis. Standard techniques may not apply directly, requiring analysts to adopt methods that address repeated measures and correlations among data points from the same subject.

- **Subject Retention Issues**: Increased enrollment rates can lead to higher dropout rates, particularly if subjects have negative experiences during prior attempts. Understanding the reasons behind withdrawals is essential for maintaining study power and validity.

To effectively mitigate these challenges, researchers can implement standardized protocols, provide comprehensive training, utilize advanced software solutions for data management, conduct regular audits, and maintain open communication with subjects. By addressing these challenges proactively, clinical trial teams can enhance data management practices and ensure more reliable and robust study outcomes.

**Overcoming Data Collection Challenges**
Successful data collection in clinical trials is crucial to ensure the integrity and validity of research findings. However, there are several challenges that researchers frequently encounter. Addressing these issues with targeted solutions can greatly enhance the data collection process.

- **Personnel Training**: Implement comprehensive training programs that emphasize the importance of data accuracy. These training sessions should cover best practices for data entry and the implications of errors, fostering a culture of meticulousness among staff.

- **Design Data Entry Guidelines**: Develop and enforce clear data entry guidelines that outline the required formats and procedures. This will help unify the data collection process across different personnel and sites, ensuring that all information is gathered uniformly.

- **Accurate Visit Assignments**: Use sophisticated scheduling software that allows for accurate tracking and management of participant visits. Regularly updating this system can help ensure that each visit corresponds correctly to the participant involved, thus maintaining data integrity.

- **Regulatory Adherence**: Establish robust compliance monitoring systems that include regular audits to confirm adherence to regulations. This proactive approach can help catch any discrepancies early and ensure that all aspects of the study meet legal and ethical standards.

- **Standardized Data Mapping**: Create a standardized data mapping plan before data collection begins. This plan should detail how different types of data will be categorized and formatted, ensuring that all information is compatible and can be easily analyzed.

- **Ensuring Data Integrity**: Implement rigorous data validation checks at various stages of data collection. Periodic monitoring and audits can help detect and rectify errors, preserving the reliability of the data collected.

By proactively addressing these data collection challenges through training, standardization, compliance, and integrity measures, researchers can enhance the quality of their data collection processes and ultimately contribute to more reliable clinical trial outcomes.

## DATA COLLECTION

Prior to the implementation of the FDA Study Data Technical Conformance Guide (TCG), the mapping of the subject identification variable (SUBJID) followed a different approach that was more straightforward but less nuanced. In earlier practices, the randomization number (RANDNUM) was typically mapped to the SUBJID variable within the Demographics (DM) domain (see Figure 1 and 2 below) and the randomization number collected on case report form was annotated as SUBJID (see Figure 3 below). This mapping served as a means of identifying and tracking subjects within clinical trials, providing a consistent reference point throughout the study. However, as clinical trial designs became increasingly complex and the need for improved data management became evident, the regulatory landscape evolved. The introduction of the TCG marked a significant shift in these practices, emphasizing the necessity for more precise identification methods, especially in trials involving multiple screenings and enrollments. The TCG clearly delineated the need for each unique screening to correspond to a different SUBJID, thereby enhancing the accuracy and reliability of participant tracking. This shift reflects a broader commitment to data integrity and compliance with regulatory standards, ensuring that clinical trials can effectively manage subjects and maintain comprehensive records.

**Figure 1.**

**Source.DM**

| STUDYID | FORMNM | USUBJID | RANDNUM | RFICDTC |
|---|---|---|---|---|
| 1234-001 | DM | 1234-001_000100001 | 100001 | 2024-01-29 |
| 1234-001 | DM | 1234-001_000100002 | 100601 | 2024-02-20 |

**Figure 2.**

SDTM.DM

| STUDYID | DOMAIN | USUBJID | SUBJID | RFICDTC |
|---|---|---|---|---|
| 1234-001 | DM | 1234-001_000100001 | 100001 | 2024-01-29 |
| 1234-001 | DM | 1234-001_000100002 | 100601 | 2024-02-20 |

**Figure 3.**

aCRF

| Subject Assignment | |
|---|---|
| 4. Did Subject Meet All Eligibility Criteria? | ○ No ○ Yes [NOT SUBMITTED] |
| 5. Was Treatment/Randomization Number Given? | ○ No ○ Yes [NOT SUBMITTED] **If Yes, Please Enter Randomization Number:** [ ] [NOT SUBMITTED] |
| 6. Treatment/Randomization Number | [ ] SUBJID |

**AN INNOVATIVE APPROCH**

With the implementation of the FDA Study Data Technical Conformance Guide (TCG), there was a significant enhancement in the approach to subject identification within the Study Data Tabulation Model (SDTM) domains, reflecting the need for greater precision in managing clinical trial data. To resolve the challenges mentioned in this article before and facilitate comprehensive data collection for re-screened subjects, one of the key innovations is to introduce the addition of a new form known as the "Participation Number Application (PNA)" to the Case Report Form (CRF) and an addition of a new form field named as "Participation Number" (see Figure 4 below). This new form is specifically designed to collect unique participation numbers that serve as identifiers for each subject involved in a clinical trial and each participation number will be unique and differ for every screening a subject goes through in a single trial. In addition, the new PNA form is collected as a new source data will be referenced as PNA_DC. This new data source is designed to gather comprehensive information regarding the needs and preferences of participants throughout their engagement in the study. In conjunction with this, the implementation of a systematic approach of collecting participation numbers within the PNA_DC source data is collected under a variable name referred to as DCPATNUM (see Figure 5 below). For each participant, a unique participation number will be assigned that corresponds to every screening or session they undergo. To maintain clarity and organization in the data records, the value can be appended in an alphabetical character to the base participation number for each instance of screening. For example, if a participant is assigned the initial participation number "001," subsequent screenings will be documented as "001A," "001B," "001C," and so forth. In the figure 5 below, the subject with identification number "000100001" is screened two times in a trial and therefore there are two participation numbers assigned to the subject. Therefore, to distinguish those participation numbers the base value is appended with "A" and "B" and assigning the values as "000100001A" as first screening participation number and "000100001B" as a second screening participation number. This incremental alphabetical system allows us to effectively track and identify multiple participation events without creating confusion or ambiguity.

**Figure 4.**

Blank CRF

| Participation Number Assignment | |
|---|---|
| 1. Participation Number | [ ] |

**Figure 5.**

Source.PNA_DC

| STUDYID | FORMNM | USUBJID | DCPATNUM | RFICDTC | DCDTC | VISIT | VISITNUM |
|---|---|---|---|---|---|---|---|
| 1234-001 | PNA | 1234-001_000100001 | 000100001A | 2024-01-29 | 2024-01-29 | Screening | 10100 |
| 1234-001 | PNA | 1234-001_000100001 | 000100001B | 2024-01-29 | 2024-03-05 | Re-Screening | 10200 |
| 1234-001 | PNA | 1234-001_000100002 | 000100002A | 2024-02-20 | 2024-02-20 | Screening | 10100 |

The handling of subject identification underwent significant refinement, particularly in how unique identifiers are

managed within the SDTM framework. In this approach, the participation number collected from the Participation Number Application (PNA) form and referenced in the source PNA_DC data is systematically mapped to the subject identification variable (SUBJID) in both the Demographics (DM) and Demographics as Collected (DC) SDTM domains (see Figure 6 below). In this strategy, the SDTM for Demographics as Collected (DC) will be thoroughly mapped to include all screening records associated with each individual subject. This comprehensive mapping approach ensures that every screening event is documented, providing a detailed overview of each participant's journey throughout the study. In contrast, the SDTM for Demographics (DM) will be structured to comply strictly with CDISC (Clinical Data Interchange Standards Consortium) guidelines, featuring only a single record per subject. This single entry will represent the primary demographic information for each participant, incorporating the most recent participation number linked to their primary screening. This design choice not only streamlines the demographic data but also highlights the most current and pertinent information, facilitating ease of interpretation and analysis. By differentiating between the extensive data captured in the DC and the succinct representation in the DM, we aim to maintain data integrity while enhancing clarity and usability in alignment with regulatory standards.

**Figure 6.**

SDTM.DC

Unique SUBJID for each screening visit

| STUDYID | DOMAIN | USUBJID | SUBJID | RFICDTC | DCDTC | VISIT | VISITNUM |
|---|---|---|---|---|---|---|---|
| 1234-001 | DC | 1234-001_000100001 | 000100001A | 2024-01-29 | 2024-01-29 | Screening | 10100 |
| 1234-001 | DC | 1234-001_000100001 | 000100001B | 2024-01-29 | 2024-03-05 | Re-Screening | 10200 |
| 1234-001 | DC | 1234-001_000100002 | 000100002A | 2024-02-20 | 2024-02-20 | Screening | 10100 |

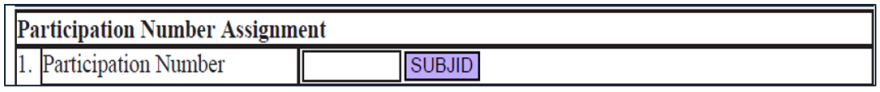
SDTM.DM

| STUDYID | DOMAIN | USUBJID | SUBJID | RFICDTC |
|---|---|---|---|---|
| 1234-001 | DM | 1234-001_000100001 | 000100001B | 2024-01-29 |
| 1234-001 | DM | 1234-001_000100002 | 000100002A | 2024-02-20 |

Moreover, in the CRF the “Participation Number” field which is expressly designed to capture a unique participation number for each screening period or enrollment will be annotated with “SUBJID” (see Figure 7 below). This enhancement is intended to bolster data integrity by precisely documenting each screening attempt in connection with the relevant subject identifier.

**Figure 7.**

aCRF




After the collection of rescreening information for the subjects in a trial, the randomization number collected in source DM can be mapped to SUPPDM domain (see Figure 8) and the randomization number collected on case report form will then be annotated as “QVAL when QNAM = ‘RANDNUM’ in SUPPDM” (see Figure 9 below). This structure will allow for a clear linkage between participant randomization and their demographic data while maintaining compliance with regulatory standards established by the TCG. The separation of these identifiers, where SUBJID reflects unique participation and RANDNUM addresses randomization specifics, plays a critical role in enhancing the integrity and traceability of data across clinical trial submissions. Overall, this refined methodology exemplifies an evolution toward more rigorous data management practices in clinical trials, ensuring that each participant's data is accurately represented and easily navigable for both regulatory review and analysis.

**Figure 8.**

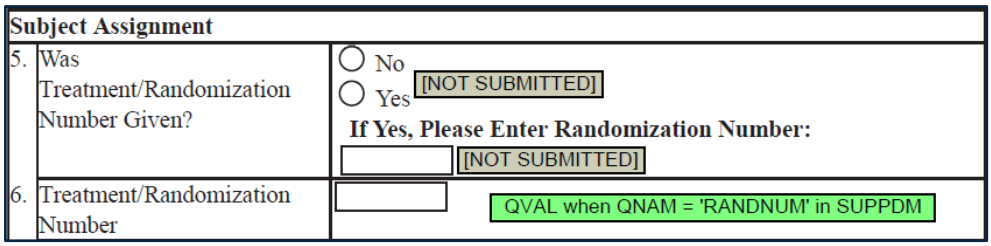
Source.DM

| STUDYID | FORMNM | USUBJID | RANDNUM | RFICDTC |
|---|---|---|---|---|
| 1234-001 | DM | 1234-001_000100001 | 100001 | 2024-01-29 |
| 1234-001 | DM | 1234-001_000100002 | 100601 | 2024-02-20 |

SDTM.SUPPDM

| STUDYID | RDOMAIN | USUBJID | QNAM | QLABEL | QVAL |
|---|---|---|---|---|---|
| 1234-001 | DM | 1234-001_000100001 | RANDNUM | Randomization number | 100001 |
| 1234-001 | DM | 1234-001_000100002 | RANDNUM | Randomization number | 100601 |

**Figure 9.**



| Subject Assignment | |
|---|---|
| 5. Was Treatment/Randomization Number Given? | ○ No<br>○ Yes [NOT SUBMITTED]<br>**If Yes, Please Enter Randomization Number:**<br>[ ] [NOT SUBMITTED] |
| 6. Treatment/Randomization Number | [ ] QVAL when QNAM = 'RANDNUM' in SUPPDM |

In addition to mapping the re-screening information to DM and DC domains, it is mapped to all other SDTM domains. We would like to show how the disposition records and other finding domains are mapped for re-screening information. Re-screening information is mapped to various SDTM domains to ensure comprehensive data capture in clinical trials. In the DM (Demographics) domain, records indicate subject's primary screening status out of the number of times a subject has been re-screened. The DC (Demographics as collected) domain captures all details about the re-screening event, including all the unique PNA numbers. Within the DS (Disposition) domain, specific records document the re-screening status, such as whether the subject is eligible or ineligible post-re-screening. Additionally, findings domains such as LB (Laboratory Tests) and VS (Vital Signs) capture relevant clinical measurements taken during re-screening, reflecting how these results might affect eligibility. Each mapping adheres to CDISC standards to maintain clarity and compliance, facilitating accurate data representation and regulatory submission.

One of the critical advancements in this context is the systematic appending of the SUBJID to the Findings domains such as Vital Signs (VS) SDTM domain (see Figure 10), a process guided by the VISIT/VISITNUM variables that is informed by both the Participation Number Application (PNA) and the visit number documented in the Vital Signs dataset (VS01). This meticulous mapping ensures that each set of vital signs collected during the study is accurately connected to the corresponding participant number and their specific visit. By linking SUBJID to each vital signs record based on the VISIT, researchers enhance the clarity of the dataset, allowing for precise tracking of participant health metrics throughout the clinical trial. Such a structured approach not only improves data integrity but also aligns with regulatory standards that mandate comprehensive documentation of participant data. As a result, this integration serves a dual purpose: it facilitates efficient data analysis and reporting while ensuring compliance with the stringent requirements set by regulatory authorities. Ultimately, the inclusion of SUBJID in the SDTM.VS domain based on VISIT underscores the importance of rigorous data management practices in clinical trials, contributing to higher data quality, improved participant tracking, and enhanced transparency throughout the research process, all of which are vital for successful regulatory submissions and the overall integrity of clinical research.

**Figure 10.**

Source.VS01_VS

| STUDYID | FORMNM | USUBJID | VSTESTCD | VSTEST | VSORRES | VSORRESU | VISIT | VISITNUM | VSDTC |
|---|---|---|---|---|---|---|---|---|---|
| 1234-001 | VS01 | 1234-001_000100001 | TEMP | Temperature | 36.5 | C | Screening | 10100 | 2024-01-29 |
| 1234-001 | VS01 | 1234-001_000100001 | TEMP | Temperature | 36.8 | C | Re-Screening | 10200 | 2024-03-05 |
| 1234-001 | VS01 | 1234-001_000100002 | TEMP | Temperature | 37 | C | Screening | 10100 | 2024-02-20 |

SDTM.VS

| STUDYID | DOMAIN | USUBJID | VSTESTCD | VSTEST | VSORRES | VSORRESU | VISIT | VISITNUM | VSDTC | SUBJID |
|---|---|---|---|---|---|---|---|---|---|---|
| 1234-001 | VS | 1234-001_000100001 | TEMP | Temperature | 36.5 | C | Screening | 10100 | 2024-01-29 | 000100001A |
| 1234-001 | VS | 1234-001_000100001 | TEMP | Temperature | 36.8 | C | Re-Screening | 10200 | 2024-03-05 | 000100001B |
| 1234-001 | VS | 1234-001_000100002 | TEMP | Temperature | 37 | C | Screening | 10100 | 2024-02-20 | 000100002A |

The incorporation of the unique participation number in SUBJID, into the SDTM Disposition (DS) domain stands as a significant advancement in the data management practices outlined by the FDA (TCG), enhancing the overall integrity and transparency of clinical trial data. By appending SUBJID to the DS domain (see Figure 11 and 12), researchers can create a direct and clear linkage between each participant and their specific treatment status as well as other critical aspects of the study related to disposition, such as reasons for discontinuation or completion, thus ensuring comprehensive tracking of participants throughout the clinical trial process. This methodical integration not only adheres to regulatory standards that mandate thorough documentation of subject participation but also facilitates a more robust analysis of the data by allowing for better insight into participant outcomes and the effects of the intervention being studied. Furthermore, the inclusion of SUBJID within the DS domain contributes to enhanced data quality and consistency, offering a structured approach to the documentation that aids in the identification of potential discrepancies or patterns in subject behavior. This meticulous record-keeping is vital for maintaining the integrity of clinical research, as it fosters transparency and accountability, essential elements that regulatory authorities expect in the evaluation and

review of clinical trial results. Ultimately, the practice of appending SUBJID to the SDTM.DS domain not only enriches the dataset by providing a systematic way to connect participants to their study status but also reinforces the commitment of clinical researchers to uphold high standards of quality and compliance in their data management practices, thereby supporting the successful submission of data for regulatory approval and advancing the field of clinical research.

**Figure 11.**

**Multiple Source Data**

| STUDYID | FORMNM | USUBJID | DSTERM | DSDECOD | DSCAT | DSSCAT | DSSTDTC |
|---|---|---|---|---|---|---|---|
| 1234-001 | DM | 1234-001_040100003 | FULL STUDY INFORMED CONSENT OBTAINED | INFORMED CONSENT OBTAINED | PROTOCOL MILESTONE | | 7/24/2024 |

| STUDYID | FORMNM | USUBJID | DSTERM | DSDECOD | DSCAT | DSSCAT | DSSTDTC |
|---|---|---|---|---|---|---|---|
| 1234-001 | FBRC | 1234-001_040100003 | GENETIC AND BIOMEDICAL CONSENT OBTAINED | INFORMED CONSENT OBTAINED | PROTOCOL MILESTONE | | 7/24/2024 |

| STUDYID | FORMNM | USUBJID | DSTERM | DSDECOD | DSCAT | DSSCAT | DSSTDTC |
|---|---|---|---|---|---|---|---|
| 1234-001 | PNA | 1234-001_040100003 | SCREEN FAILURE | SCREEN FAILURE | DISPOSITION EVENT | DISCONTINUED | 8/19/2024 |

| STUDYID | FORMNM | USUBJID | DSTERM | DSDECOD | DSCAT | DSSCAT | DSSTDTC |
|---|---|---|---|---|---|---|---|
| 1234-001 | DS01 | 1234-001_040100003 | SCREEN FAILURE | SCREEN FAILURE | DISPOSITION EVENT | DISCONTINUED | 8/19/2024 |

**Figure 12.**

**SDTM.DS**

| STUDYID | DOMAIN | USUBJID | DSGRPID | DSTERM | DSDECOD | DSCAT | DSSCAT | DSSTDTC | SUBJID |
|---|---|---|---|---|---|---|---|---|---|
| 1234-001 | DS | 1234-001_040100003 | DM | FULL STUDY INFORMED CONSENT OBTAINED | INFORMED CONSENT OBTAINED | PROTOCOL MILESTONE | | 2024-07-24 | 040100003A |
| 1234-001 | DS | 1234-001_040100003 | FBRC | GENETIC AND BIOMEDICAL CONSENT OBTAINED | INFORMED CONSENT OBTAINED | PROTOCOL MILESTONE | | 2024-07-24 | 040100003A |
| 1234-001 | DS | 1234-001_040100003 | PNA | SCREEN FAILURE | SCREEN FAILURE | DISPOSITION EVENT | STUDY PARTICIPATION | 2024-08-19 | 040100003A |
| 1234-001 | DS | 1234-001_040100003 | DS01 | SCREEN FAILURE | SCREEN FAILURE | DISPOSITION EVENT | STUDY PARTICIPATION | 2024-08-19 | 040100003A |

This evolution in subject identification practices illustrates the FDA's commitment to fostering robust data management protocols in clinical trials, thereby enabling researchers to maintain comprehensive and accurate records of each participant's journey throughout the duration of the study. As a result, the above implementation not only strengthens data quality and regulatory compliance but also improves the overall efficacy of clinical research by ensuring that participant data is handled with the highest standards of accuracy and consistency.

**P21 ISSUES – RATIONALE**

When discussing the issues encountered in the P21 tool related to a multiple screening approach, it's essential to provide both a detailed overview of the specific problems and the rationale behind the proposed solutions. The P21 tool is often utilized in clinical research to enhance data management and ensure compliance with regulatory standards; however, the use of a multiple screening approach can introduce various challenges that need to be addressed effectively.

The Clinical Study Data Request Guide (cSDRG) will include several critical issues related to the P21 validation tool, particularly concerning multiple screenings and enrollments. As specified below in Figure 13. Rule ID “SD9999” reveals a significant concern where the dataset's DC class is not recognized, resulting in a false positive within the P21 tool. This issue primarily stems from the tool's limited support for the Demographics as Collected (DC) domain, which it does not validate, creating challenges in accurately interpreting demographic data during the analysis phase. Furthermore, Rule ID “SD1367” addresses complications that arise from recording multiple disposition events for the same DSSCAT and EPOCH. When the DSDECOD indicates a 'SCREEN FAILURE' collected from two different case report forms (CRFs)—specifically, Disposition form (DS01) and PNA forms—this can lead to discrepancies in how participant dispositions are tracked. Additionally, Rule ID “SD0058” points out the presence of variables, such as SUBJID, in the dataset that are not included in the SDTM model. Despite their absence from the official standards, these variables are essential for analysis and are retained in the standard dataset to preserve data integrity. Collectively, these issues underscore the need for enhanced validation capabilities and clearer guidelines within the cSDRG framework to improve data consistency and quality across clinical trials.

Lastly, data analysis and interpretation can become cumbersome when dealing with results from multiple screenings. Variability in the data can complicate statistical analyses and affect the overall conclusions drawn from the study. To address this concern, utilizing advanced statistical techniques that account for the complexities introduced by multiple screenings can help ensure that the findings remain valid and reliable.

In summary, the issues encountered in the P21 tool related to a multiple screening approach can be effectively managed through the establishment of standardized screening criteria, the development of integrated data management systems, clear communication with participants, and the application of advanced analytical methods. By implementing these rationales, researchers can enhance the overall functionality of the P21 tool and improve the quality of data management in clinical studies.

**Figure 13.**

| Rule ID | Message | Rationale |
|---|---|---|
| SD9999 | Dataset DC class not recognized | False positive result within the P21 tool. P21 has limited support of Demographics as Collected (DC) domain and does not validate this domain. |
| SD1367 | Multiple disposition events for the same DSSCAT and EPOCH | DSDECOD= 'SCREEN FAILURE' is collected from two CRF pages (DS01 and PNA forms) |
| SD0058 | Variable appears in dataset, but is not in SDTM model | The variables listed below are required for analysis and hence are retained in the standard dataset.<br>- SUBJID |

## CONCLUSION

In conclusion, the implementation of a unique participation number for each screening period significantly enhances subject tracking, particularly for participants undergoing multiple screenings, by ensuring that all attempts are systematically recorded, thereby improving data accuracy and reliability. The introduction of a new Case Report Form (CRF) field promotes comprehensive data collection practices for previous screenings and enrollments, allowing for the capture of detailed information related to re-screening efforts and providing valuable insights for data analysis and reporting. Additionally, the new special purpose domain, DC, facilitates the mapping of re-screened data, enabling the separation of screening and re-screening records, which simplifies the process of mapping and appending DC.DCPATNUM as DM.SUBJID to other SDTM domains. This structured method of handling re-screening data aligns with the Technical Conformance Guide (TCG), reinforcing the importance of meticulous record-keeping in clinical trials and ensuring regulatory compliance. Furthermore, the separation of SUBJID from the randomization number aids in tracking participants throughout each visit in the trial, minimizing potential confusion associated with using multiple identification methods. Collectively, these enhancements contribute to a more efficient and transparent clinical trial process, ultimately improving data quality and the integrity of research outcomes.

By implementing these mapping strategies, we not only fulfill regulatory requirements but also improve the clarity and utility of the trial data. This approach reflects our commitment to maintaining high standards in data management throughout the clinical trial process.

## REFERENCES

1. StudyDataTechnicalConformanceGuide-v4-2-FINAL.pdf
2. CDISC SDTM Implementation Guide
3. FDA Guidance for Standardized Study Data
4. FDA Electronic Submissions Guidance

## ACKNOWLEDGMENTS

The authors would like to thank Janet Low, Shaveta Bansal, Kristi Garner and Amy Gillespie for their review and support.

## CONTACT INFORMATION

**Anahita Vyas,**
**Assoc Prin. Scientist, BARDS- Statistical programming**
**Merck & Co., Inc., Rahway, NJ, USA**
**anahita.vyas@merck.com**

**Rama Kudaravalli**
**Assoc Prin. Scientist, BARDS- Statistical programming**
**Merck & Co., Inc., Rahway, NJ, USA**
**rama.kudaravalli@merck.com**