



One eye open, one eye closed: How to approach masking in open-label oncology study

Abhay Kumar (GSK, PA, USA)
Tanya Dubinsky (GSK, London, UK)



Why and What to mask in open label trials?

- Primary goal is to maintain the integrity of results and minimize the possibility of sponsor's bias, mitigate both real and perceived bias.
- Enhances the credibility of study results.
- Improves regulatory acceptance.
- Masked to aggregated analysis and some aspects of subject-level data, during the trial conduct, especially concerning the generation of SDTMs, ADAMs, and statistical summaries.

Exploring Our Topics

- **Data Masking Techniques:** Outline methods for masking collected and central data.
- **Implementation Challenges:** Examine common challenges and present practical solutions.
- **Programming Approaches:** Review specific programming techniques for masking individual data elements.
- **CRF Scenarios & Risk Mitigation:** Discuss strategies for managing diverse Case Report Form scenarios and mitigating risks. Documentation through a data masking plan capturing masking rationale and specific methods for each data point, developed in collaboration with all stakeholders.

General Considerations

Data Masking Plan

- A good practice, developed in collaboration with the entire study team, including data management, clinical, safety, and biostatistics functions.
- Details Masking Rationale and specific methods for each data point, e.g., treatment scrambling, critical information can either be left blank, or a dummy value may be added.
- Finalize Masking Requirements, Data Transfer Mechanisms, and Validation processes.
- Meticulous Planning and capture outlines as agreed by stakeholders.

General Considerations

Handling Sensitive CRF and External Data

- To review all collected data elements to identify any information that could reveal the study treatment , CRFs may capture sensitive details about study treatment information.
- Definitive decision by all stakeholders on how to mask these data elements.
- To maintain uniformity, external data to deploy same methodology as per CRF data e.g., aligning treatment detail masking, using dummy values in place of sensitive information.
- Protocol and/or CRF amendment - The treatment sensitivity plan should be revised.

General Considerations

Quality Control and Validation

- Highly Recommended - Implementing a Verification Process.
- Validation programmer - Strict independence from the core study team.
- Determine validation strategy such as running frequency statistics and performing manual inspections.
- Ascertain all masking procedures performed correctly before releasing masking data.

Effective Data Masking Methods

- **Treatment Scrambling within Subjects**

- Effective method for data masking is to disrupt the expected treatment arm ratios, as defined by the study protocol e.g., study with a 2:1 protocol ratio could have new treatment assignments in a different ratio (e.g., 2:1:1 or 3:4), with treatments scrambled among subjects.
- May be applied either at subject-level analysis dataset or the demographics dataset.
- Case Scenario - Some subjects receive intravenous injections in treatment arm; others receive oral dosing in a control arm; scrambling subjects in a different ratio addresses these data elements without causing the TFLs to be completely misaligned or out of context.
- Minimize significant logical inconsistencies that can arise when CRFs are designed to accommodate various treatment schemes.

Effective Data Masking Methods

- The following snippet illustrates the dataset's appearance after scrambling, reflecting a change in the treatment ratio from 1:1 to 2:1:1. It may be carried out by programming team, using a Randomization Flag (e.g., 'Y' or 'N') macro variable set up -

Subject Identifier	Original Treatment Before Masking is applied	Masked Treatment Values after Treatment Scrambling in between Subjects
Study_A_001	Study Drug A	Control Drug B
Study_A_002	Control Drug B	Study Drug A
Study_A_003	Study Drug A	Control Drug B
Study_A_004	Study Drug A	Control Drug B
Study_A_005	Control Drug B	Study Drug A
Study_A_006	Control Drug B	Study Drug C
Study_A_007	Study Drug A	Control Drug B
Study_A_008	Control Drug B	Study Drug A
Study_A_009	Control Drug B	Study Drug C
Study_A_010	Study Drug A	Control Drug B

Study Drug Administration

Dose Levels and Frequencies

- Complexity of Study Drug Administration panels - High.
- Scheduled and actual current cycle dose levels (mapped to ECDOSE, EXDOSE).
- Dosing frequencies (EXDOSFRQ), Visits and Cycle patterns.
- Dose value captured consistently, but specific values differ per treatment arm (e.g., Treatment A: 1000 mg daily; Treatment B: 300 mg every alternate day), Differences can unintentionally reveal actual treatments.

Study Drug Administration

Drug Administration Method/Any Free text field

- Distinct administration methods can inadvertently reveal treatment assignments e.g.; experimental drug infused vs. oral comparator.
- Treatment-Specific Variables: Variables collected only for a specific treatment arm (e.g., "dose level changed from previous cycle" for an infused drug) may also compromise blinding.
- Reasons for dose changes, especially when captured as free text in the CRF, are highly sensitive.
- Leaving these details unmasked poses a high risk of unblinding the actual treatment assignment.

Adverse Events

- **Masking Key AE Variables**

- Variables like AERELn, (Adverse Event Relatedness) and AEACNn (Adverse Event Action Taken) and similar fields might need to be masked.
- Supplementary AE Data (SUPPAE): Even minor information in the SUPPAE dataset may be revealing and could be considered for masking.

- **Risk from AE Narratives & Comments**

- Free-text AE narratives and comments may inadvertently disclose treatment assignments.
- Example of Unblinding: An injection site AE report could unblind a subject if only the experimental drug is administered via infusion.

► Pharmacokinetic Analysis Datasets (ADPC & ADPP)

- **Sensitive ADPC Elements Requiring Masking:**
 - Concentration result values.
 - Below Limit of Quantitation (BLQ) / Lower Limit of Quantitation (LLOQ) indicators.
 - Sampling time references.
 - Free-text comment fields.
- **PK Sampling in Active Treatment Arm Only:**
 - In designs where PK sampling/analysis is exclusive to the active treatment arm. An independent programming team may be engaged to develop PK datasets and outputs.
 - This preserves study integrity and confidentiality, preventing unblinding.

Headline Results

- **Headline Results & Primary and Secondary Endpoints**
 - Essential to preserve key summary statistics related to primary or secondary objectives such as Kaplan Meier analyses for Progression-Free Survival (PFS) or Overall Survival (OS), Objective Response Rate (ORR), Duration of Response (DOR), and Disease Control Rate (DCR), Patient-Reported Outcomes (PROs) etc.
 - Special care should be taken to mask these data points such as treatment scrambling between subjects.
 - To ensure that even if aggregated statistical outputs are generated, the crucial primary and secondary efficacy endpoints are not revealed, thereby protecting blinding.

R code snippet to perform treatment scrambling (ADAM)

Emulating SAS 'DATA GEN1' step: Initial, unrandomized treatment assignment

```
GEN1 <- TIBBLE(UNIT = 1:NOBS1) %>% #  
MUTATE(  
  TREATMENT = IF_ELSE(UNIT <= NOBS1 * (1 / 2), 1, 2),  
  AB = LAG(TREATMENT)  
)
```

Used to link to the randomized assignments.

```
ADSL1_PROCESSED <- AD_RAND%>%  
ARRANGE(USUBJID) %>%  
MUTATE(  
  USUBJID = STR_TRIM(USUBJID),  
  UNIT = ROW_NUMBER())# Assigns a sequential 'UNIT' number (1 to nobs1) to each subject. This is good  
practice for reproducibility if the order matters for UNIT generation.
```

R code snippet to perform treatment scrambling (ADAM)

Emulating SAS PROC SQL to get NOBS2 and X (Unit_X)

NOBS2: Count of subjects initially assigned TREATMENT=2 in gen1

```
NOBS2 <- GEN1 %>%
```

```
FILTER(TREATMENT == 2) %>% # Filters for rows where TREATMENT is 2
```

```
nrow() # Counts the number of rows
```

x_val: The first UNIT number where TREATMENT changes from 1 to 2.

In a 1:1 split, this will be $(\text{nobs1}/2) + 1$.

```
X_VAL <- FLOOR (NOBS1 * (1 / 2)) + 1 # CALCULATES THE STARTING POINT OF TREATMENT=2
```

Unit_X: The UNIT number where the TREATMENT value first changes to This specifically looks for the first instance where previous treatment) is not equal to the current TREATMENT, and finds the transition point

```
UNIT_X <- GEN1 %>%
```

```
FILTER(AB != TREATMENT & TREATMENT == 2) %>%
```

```
PULL(UNIT) # Extracts the UNIT value as a vector
```

R code snippet to perform treatment scrambling (ADAM)

```
# Subdividing TREATMENT=2 into TREATMENT=2 and TREATMENT=3
#This logic splits the original '2's into '2' and '3' to create a 3-arm design (e.g., 2:1:1 or any ratio as decided by team)
GEN2 <- GEN1 %>%
MUTATE(
TREATMENT = CASE_WHEN(
# First half of original TREATMENT=2 block remains 2
TREATMENT == 2 & UNIT >= X_VAL & UNIT <= (X_VAL + FLOOR(NOBS2 / 2) - 1) ~ 2,
# Second half of original TREATMENT=2 block becomes 3
TREATMENT == 2 & UNIT >= (X_VAL + FLOOR(NOBS2 / 2))
& UNIT <= (X_VAL + NOBS2 - 1) ~ 3,
TRUE ~ TREATMENT # Keep other treatments (TREATMENT=1) as they are (default case)
) )
SET.SEED(12345) # Sets the random seed for reproducibility (matches SAS SEED) Emulating SAS PROC PLAN
SEED=12345; Generate the randomized 'UNIT' factor.
RANDOM_UNIT_ASSIGNMENT <- SAMPLE(1:NROW(GEN2)) # GENERATES A VECTOR OF SHUFFLED UNIT NUMBERS
```

Conclusion

- Robust data masking strategies are crucial for maintaining the integrity and validity of open-label oncology trials.
- Traditional blinding is often not feasible in these trials due to patient preferences for treatment knowledge, ethical considerations, and practical challenges.
- The inherent transparency of open-label designs necessitates meticulous planning to minimize bias.
- Establishing a comprehensive and detailed Data Masking Specification is essential.
 - This plan should be developed collaboratively with the entire study team.
 - It must outline the rationale, specific techniques, and scope of masking across all relevant data sources (e.g., source systems, SDTM, ADaM, TFLs).

Conclusion

- Successful implementation relies on clearly defined procedures for execution, transfer, and validation, agreed upon by all study participants in advance.
- A robust verification process, conducted by an independent validation programmer, is critical to confirm masking accuracy before data release.
- While data masking adds complexity to programming, its proper application:
 - Prevents premature revelation of treatment.
 - Maintains data traceability and reproducibility.
 - Ensures accurate and defensible masked summaries and regulatory outputs.
 - Safeguards study integrity.
 - Secures acceptance by regulatory authorities.

Thank You

