Paper PP27

Robust Programming Strategies for Exposure ADaM Datasets

Susann Grill, mainanalytics GmbH, Sulzbach, Germany

ABSTRACT

Building robust ADaM datasets for drug exposure is crucial for accurate safety and efficacy analyses in clinical trials. Depending on study design, dosing schedules, the potential for treatment changes and interruptions, as well as the need to combine multiple sources of data, deriving exposure variables can become increasingly complex and error-prone.

This paper explores anticipatory strategies to identify data inconsistencies or anomalies, such as missing, overlapping, or contradictory records. It focuses on minimizing errors through early assessment of input data quality, the application of logic checks to safeguard complex derivations, and cross-checks with other ADaM datasets.

These defensive approaches ensure data quality, enhance program stability, and promote efficient workflows in clinical trials by proactively addressing potential issues that may arise across a variety of study scenarios.

INTRODUCTION

The development of robust Analysis Data Model (ADaM) datasets for study drug exposure is essential to ensure accurate safety and efficacy analyses in clinical trials. A clear understanding of a participant's exposure to the protocol-specified treatment often depends on combining information from multiple sources. Creating reliable exposure indicators may involve complex derivations, and as study designs grow in complexity, so does the risk of error. The use of careful and consistent programming practices can help mitigate these risks and improve the overall quality of the datasets.

ROBUST PROGRAMMING

Robust or defensive programming strategies encompass a variety of techniques designed to detect and handle both anticipated and unanticipated data issues or programming errors. Such strategies are well documented in the general programming literature [1], in the context of programming with SAS® software [2, 3], and with respect to specific applications such as ADaM dataset programming [4] or defensive programming in ongoing clinical trials [5].

Among other practices, robust and defensive programming includes verifying the existence of datasets and performing dependent merges [5, 6], identifying duplicate records [2], detecting unexpected, implausible, or missing values [3, 4], and performing calculations only on non-missing values [3].

Handling identified issues – though not discussed further in this paper – depends largely on their nature and may involve generating informative warning and error messages in the log [3, 4], especially for unacceptable issues, capturing potentially incorrect records in spreadsheets for further review [4], providing information to data management or Study Data Tabulation Model (SDTM) programmers to query data points or update SDTM programming, modifying one's own algorithms, revising potentially inaccurate programming specifications or documents such as the statistical analysis plan (SAP), or addressing issues that cannot be resolved programmatically.

Early implementation of such practices in ongoing trials helps prevent extensive rework caused by database updates, safeguards data integrity, and ensures output quality while saving time and resources [4, 5].

OBJECTIVE

While the core principles of defensive programming remain unchanged and existing code can often be repurposed, exposure-specific considerations must be addressed to ensure the reliability of the applied algorithms. As these principles and robust coding practices have already been extensively documented, this paper focuses on data and plausibility checks specifically tailored to the development of exposure ADaM datasets.

The process of creating exposure ADaM datasets will be divided into a series of key tasks. These include obtaining input datasets, processing and transforming data, deriving complex variables, and verifying derived values against other ADaM datasets. For each task, potential issues and corresponding checks will be highlighted. The proposed checks are intended to be broadly applicable across different study designs; therefore, a solid understanding of both the study design and the underlying data is essential to determine which aspects are most appropriate to evaluate.

Before introducing these checks, the possible structures of exposure ADaM datasets and their input datasets are reviewed.

ADAM FOR EXPOSURE DATA

ADaM exposure datasets provide analysis-ready, traceable information on a participant's exposure to protocol-specified treatments, or it may serve as an intermediate dataset for the creation of other analysis-ready datasets or listings [7]. The dataset name is sponsor-defined but must adhere to ADaM naming conventions, commonly resulting in names such as ADEX, ADEXP, ADEXSUM, or ADTRT. Depending on analysis needs, ADEX (hereafter used as a general term for any exposure ADaM dataset in this paper) is typically structured according to either Basic Data Structure (BDS) or Structure for Occurrence Data (OCCDS), though considerable variability is possible. If neither the BDS nor OCCDS structures adequately meet the needs of the analysis or the requirements for an intermediate dataset, an ADaM OTHER structure may be used; for intermediate datasets, even a non-ADaM structure could be considered [8].

The first three examples illustrate single treatment administrations or treatment periods using OCCDS-structured datasets [9]. These types of ADEX datasets will hereafter be collectively referred to as period-based ADEX. However, only illustrative variables and records are shown; a complete OCCDS ADaM dataset would include additional required variables. The examples presented here can therefore be understood either as ADaM OTHER–structured datasets or as incomplete OCCDS datasets.

If single-dose traceability is required for analysis – for example, in a dose-escalation trial – ADEX can contain one record for each treatment administration. In this case, the ADaM dataset closely resembles the SDTM EX dataset but includes derived variables and, if needed, additional records depending on the analysis requirements. SDTM variables relevant for analysis or required for traceability are carried over directly from the SDTM EX domain.

An example of an unblinded ADEX dataset for a randomized, double-blind, placebo-controlled, parallel-design, dose-escalation study with three infusions is shown in Table 1. Participant 101 received three placebo infusions every 28 days as planned. Participant 102 received three increasing doses, with the last infusion both delayed and interrupted. Start and end datetimes define the infusion duration. For treatments administered over a brief period, such as tablets or injections, the start and end datetime would be identical. The residual effect period (REP) is not considered in this example.

USUBJID	EXTRT	TRTA	EXDOSE	EXDOSU	ASTDTM	AENDTM	ASTDY	AVISIT
101	Placebo	Placebo	0	mg	03MAR2025T09:05	03MAR2025T09:27	1	VISIT 2
101	Placebo	Placebo	0	mg	31MAR2025T10:31	31MAR2025T11:00	29	VISIT 4
101	Placebo	Placebo	0	mg	28APR2025T08:57	28APR2025T09:20	57	VISIT 6
102	IP	IP 20mg	20	mg	05MAR2025T12:22	05MAR2025T12:51	1	VISIT 2
102	IP	IP 50mg	50	mg	03APR2025T14:36	03APR2025T15:00	30	VISIT 4
102	IP	IP 80mg	55	mg	08MAY2025T10:15	08MAY2025T10:41	65	VISIT 6

Table 1: Example exposure ADaM dataset with one record per participant and treatment administration

If single treatment administrations and doses are not required for analysis or cannot be recorded at that level of granularity, ADEX can be collapsed to one record per participant and consistent dosing period. In this example of a randomized crossover study, participants take tablets daily for two weeks and then switch to the control after a minimum washout period of 28 days (see Table 2). After the second treatment period, another washout period of the same duration follows, corresponding to the REP, is added to capture the entire on-treatment time. For participant 1002, the REP record is shortened due to early study discontinuation. Actual dose and compliance are assessed based on the count of returned tablets.

USUBJID	APERIOD	ASPER	ASPERC	TRTP	TRTA	ECDOSP	EXDOSE	DOSEU	ASTDT	AENDT	COMPL	ADURN
1001	1	1	TRT	IP A 10mg	IP A 10mg	140	120	mg	08AUG2025	21AUG2025	85.7	14
1001	1	2	WAS	IP A 10mg	IP A 10mg	.	.		22AUG2025	18SEP2025	.	28
1001	2	1	TRT	IP B 20mg	IP B 20mg	280	200	mg	22SEP2025	05OCT2025	71.4	14
1001	2	2	WAS	IP B 20mg	IP B 20mg	.	.		06OCT2025	02NOV2025	.	28
1002	1	1	TRT	IP B 20mg	IP B 20mg	280	280	mg	11AUG2025	25AUG2025	100	15
1002	1	2	WAS	IP B 20mg	IP B 20mg	.	.		26AUG2025	22SEP2025	.	28
1002	2	1	TRT	IP A 10mg	IP A 10mg	140	130	mg	23SEP2025	06OCT2025	92.9	14
1002	2	2	WAS	IP A 10mg	IP A 10mg	.	.		07OCT2025	01NOV2025	.	26

Table 2: Example exposure ADaM dataset with one record per participant and treatment/washout period

The first example does not facilitate easy determination of on-treatment periods because REPs are not included. In contrast, the second example captures all on-treatment days. Another possible ADEX structure records every point in time from first contact to the end of study, allocating treatment information accordingly (see Table 3).

In this parallel-group study example, each participant has four records: one for screening, two for the treatment periods, and one for follow-up. Here, the REP is included within the treatment periods, though it could also be represented as separate records. Treatment periods can be further divided into subperiods if needed. If treatment interruptions are allowed, additional records may be included to reflect these interruptions, e.g., as subperiods. In this example, one phase or period ends one minute before the next begins; this interval could be further refined to seconds if such granularity is detectable and clinically meaningful.

USUBJID	APHASE	APERIOD	TRTP	TRTA	ASTDTM	AENDTM
201	SCREENING	.			13JAN2025T10:15	11FEB2025T08:34
201	TREATMENT	1	IP A 2mg	IP A 2mg	11FEB2025T08:35	08MAY2025T08:08
201	TREATMENT	2	IP A 5mg	IP A 5mg	08MAY2025T08:09	31JUL2025T10:09
201	FOLLOW-UP	.			31JUL2025T10:10	06NOV2025T14:25
202	SCREENING	.			15JAN2025T07:55	05FEB2025T11:40
202	TREATMENT	1	IP B 10mg	IP B 10mg	05FEB2025T11:41	02MAY2025T08:01
202	TREATMENT	2	IP B 20mg	IP B 20mg	02MAY2025T08:02	28JUL2025T15:37
202	FOLLOW-UP	.			28JUL2025T15:38	04NOV2025T12:04

Table 3: Example exposure ADaM dataset with one record per participant and study phase/treatment period

This example illustrates a particularly common use case for an intermediate dataset: determining, for instance, whether adverse events occur while a participant is on treatment. Although period start and end datetimes could also be represented in ADSL using subject-level timing variables, evaluating whether an event occurs during treatment is often more straightforward with this type of intermediate dataset – especially when multiple periods are involved. All three examples presented above can serve as intermediate datasets, for example, to support the creation of summary exposure datasets such as the one shown next.

The following example illustrates an ADaM BDS dataset [10], often referred to as ADEXSUM [7], which contains analysis-ready summary exposure metrics such as total dose or number of doses (see Table 4). Each record corresponds to a participant and an analysis parameter. While this example does not include time points, the BDS structure can also accommodate them if required. Within the BDS framework, the level of granularity and overall design may vary depending on the analysis objectives and study characteristics. When only a limited number of summary exposure statistics are required, these metrics may alternatively be incorporated directly into the ADSL dataset.

USUBJID	PARAM	PARAMCD	AVAL	AVALC
301	Total dose (mg)	DOSTOT	2000	
301	Number of administrations	DOSNUM	20	
301	Time at risk (days)	ATRISKD	127	
301	Time at risk (months)	ATRISKM	4.2	
301	Total duration of exposure (days)	EXPDURD	113	
301	Overall compliance	COMPLTOT	0.94	
301	Compliance treatment period 01	COMPL01	0.90	
301	Compliance treatment period 02	COMPL02	0.98	
301	Overall compliance >= 80%	COMPLFL		Y
302	Total dose (mg)	DOSTOT	1800	
302	...	...	...	...

Table 4: Example BDS exposure ADaM dataset with one record per participant and parameter

All examples presented reflect trial data from open-label studies or from blinded studies after unblinding. Prior to unblinding, treatment variables would contain dummy values, which may not be internally consistent within participants. In practice, both period-based and parameter-based exposure ADaM datasets are often produced to support different analysis needs – with OCCDS datasets frequently serving as the basis for creating BDS ADEX datasets.

INPUT DATASETS

In Clinical Data Interchange Standards Consortium (CDISC®)-compliant studies, ADaM datasets are developed from SDTM datasets. The complexity of the study design determines how many data sources are required to assemble a complete picture of a participant's treatment exposure. In an open-label, phase I, single-dose study, a single SDTM domain may be sufficient to capture all treatment information. In contrast, a complex, blinded, controlled, multi-arm phase III study may require several domains. According to SDTM Implementation Guide (SDTMIG) v3.4 [11], two exposure domains are available to address different aspects of treatment administration and data collection across study designs. Additional domains may also be needed to gather all relevant information. The following section briefly outlines these domains and datasets and their respective roles in supporting exposure data collection and derivation.

EX (EXPOSURE)

The SDTM EX domain is required for clinical trials that include protocol-specified study treatments. This interventions class domain captures information on a participant's exposure to the study treatment, with one observation recorded per participant, treatment, and sponsor-defined consistent dosing period. A dosing period may represent a single administration or multiple administrations at the same dose. The EX dose unit is aligned with the unit specified in the study protocol, making EX a derived dataset in most studies – particularly in blinded trials. Depending on sponsor definition, EX may be as granular as EC or presented in a more condensed form.

EC (EXPOSURE AS COLLECTED)

The SDTM EC domain is the second exposure domain, containing *as-collected* administration details for protocol-specified study treatments. This interventions class domain comprises one record per participant, treatment, collected dosing period and mode of administration. In contrast to EX, EC may also include planned treatment administrations. EC is used when the exposure information collected during the study differs from the protocol-specified representation of treatment – most commonly in the unit of measurement. Typically, EC captures product-level information (e.g., number of tablets or syringes), while the protocol specifies dose units such as mg. If the information in EC is identical to EX, the EC domain may be omitted.

DA (PRODUCT ACCOUNTABILITY)

The SDTM DA domain keeps track of amount and type of the protocol-specified products dispensed to and returned from the study participants. This findings domain comprises one record per participant and finding.

DM (DEMOGRAPHICS)

SDTM domain DM contains demographic and basic study information for each participant with one record per participant. DM connects exposure data to participant-level attributes, treatment assignment, and reference periods, and supports derivation of exposure analysis datasets. Actual and planned treatment information facilitate plausibility checks of administered treatments and reference dates allow for derivation of relative study days.

FA (FINDINGS ABOUT EVENTS OR INTERVENTIONS)

Additional findings about an intervention that cannot be recorded in EC or EX or a supplementary dataset can be included in SDTM FA domain, a findings domain with one row per participant, visit, object, time point, and finding. For investigational products given to treat acute events FA could contain degree of severity findings at several timepoints after administration.

RELREC (RELATED RECORDS)

If FA domain is utilized to capture findings for treatment administrations then RELREC describes the relation between FA and EC or EX records.

RANDOMIZATION AND MEDICATION LISTS

Randomization and medication lists are typically not needed in exposure ADaM derivation, since treatment allocation is handled during SDTM programming. However, after unblinding, these lists can be used to verify correct and plausible treatment arm assignments as well as coherent treatment administration. EX, EC and/or DA may include reference variables capturing identifiers for medication kits, bottles, vials, or boxes, enabling identification of the respective treatment during unblinding. During this process, the dummy lists are replaced with unblinded lists, allowing the actual treatment and dose to be assigned to the collected exposure records.

CHECKS FOR INPUT DATASETS

Early detection of missing, implausible, or inconsistent values prevents the propagation of errors into ADaM datasets, reduces rework, and helps ensure that analysis outputs accurately reflect the collected study data. Input dataset checks therefore play a key role in maintaining data integrity and enabling reliable derivation of analysis-ready variables. The scope and nature of these checks depend heavily on the study specifics. In blinded studies, checks must anticipate the unblinded data; however, when using blinded data, these checks may trigger numerous apparent errors due to discrepancies between blinded and unblinded information. In open-label studies, the readily interpretable treatment information simplifies data verification. The following section proposes input dataset checks, categorized by topic (see Table 5).

Table 5: Input dataset checks

Category	Condition	Check	Explanation/Examples
data availability	protocol-specified treatment	EX available	According to SDTMIG v3.4, the EX domain is required for trials with protocol-specified treatments.
	collected treatment information differs from protocol-specified treatment information	EC available	If exposure information for protocol-specified treatments cannot be collected exactly as defined in the protocol – as is commonly the case in double-blind trials, for example when the unit of collected exposure differs – the EC domain must be included.
	-	completeness of key variables	The variables EXTRT, ECTRT, DATESTCD, and DATEST must not be missing for any record. For each record, either EXDOSE or EXDOSTXT is expected to be populated. In addition, for active drug records (i.e., where ECOCCUR ≠ ”N”), either ECDOSE or ECDOSTXT should be populated.
	-	no partial or missing dates or datetimes	Date(time)s should not be missing, as they are essential for establishing treatment timelines, calculating durations, and ensuring consistency across domains. Partial or missing ECSTDTC and ECENDTC values should be queried. Partial or missing EXSTDTC and EXENDTC values may be retrieved from the EC domain. The time component of a datetime variable may be left missing or null if it is not relevant for the trial.
coherent treatment	-	plausible combinations of dose, unit, form, route, location, laterality, frequency and treatment values	Implausible combinations of variable values – such as treatment and route, particularly for combination treatments administered via different routes – may indicate underlying data issues or potential protocol deviations. This check applies to both EX and EC.
	-	(after unblinding) each participant’s EX, EC, and DA records correspond to a single treatment arm	After unblinding, or in open-label studies from the outset, treatment records for a given participant should be assignable to only one treatment arm. For example, in a placebo-controlled trial, a participant assigned to the placebo arm should generally have only placebo treatment records. In dose-escalation trials, however, certain doses may correspond to more than one treatment arm, as arms may differ only in their maximum dose levels. Deviations do not necessarily indicate data issues but may reflect treatment errors, such as a participant in the placebo group inadvertently receiving the investigational product.

Continued on next page

Table 5 continued from previous page

Category	Condition	Check	Explanation/Examples
plausible/ allowed values	-	ECTRT and EXTRT have sponsor-defined values	Only certain values are valid for ECTRT and EXTRT, as both variables contain the treatment name specified in the protocol. In double-blind trials, EXTRT contains dummy treatment names prior to unblinding, while ECTRT either contains the dummy treatment name available at the time of data collection or is completely masked.
	-	plausible EXDOSE and ECDOSE values	Since the treatment scheme is defined in the protocol, protocol-compliant administrations are restricted to certain dose values for each treatment. Deviations from these values may indicate dosing errors, data issues, or programming errors; however, if dose reductions are permitted, such deviations may be valid. Zero values should always be checked for plausibility: for placebo treatments, the dose in EX should typically equal 0, while for treatments not administered (ECOCCUR = ”N”), the dose in EC must be missing rather than 0. In addition, only certain units (EXDOSU and ECDOSU) are plausible. Units in EC usually correspond to the product level (e.g., number of tablets or syringes), whereas units in EX typically correspond to protocol-specified dose units such as mg.
duplicates	-	no exact duplicates	Exact duplicates in any input dataset may indicate underlying data issues.
	-	none within relevant variable combinations	Duplicates within specific variable combinations may indicate data issues. For example, a participant should not have more than one active treatment record at a given time point; therefore, only one record with the same values for USUBJID, ECTRT, ECOCCUR, and ECSTDTC or USUBJID, EXTRT, and EXSTDTC should exist.
number of records	-	record counts per USUBJID do not exceed the plausible maximum	Depending on study specifics, such as the treatment scheme, SDTM datasets containing exposure information generally have a maximum expected number of records. For example, if the maximum number of treatment administrations per participant is 10, no more than 10 records per USUBJID are expected in EX. Deviations from this may indicate data issues, treatment errors, or protocol deviations. In contrast, EC may also include planned administrations, so the total maximum number of records is less clearly defined, whereas the number of successfully performed administration records still has a defined maximum.

DATA PROCESSING AND TRANSFORMATION

Once all required SDTM datasets and variables needed for ADaM exposure dataset creation are confirmed to exist, and the data has been verified as clean and plausible – or at least appropriate checks are in place – the next milestone is to derive participant-level exposure data. This step first requires converting character datetime variables (e.g., EXSTDTC and EXENDTC) into numeric SAS dates or datetimes as needed, imputing incomplete dates according to pre-specified rules, and merging datasets to obtain additional treatment information, such as the treatment arm. The merged input datasets can then be used for additional plausibility checks (see Table 6).

Table 6: Checks after data transformation and input dataset merges

Category	Condition	Check	Explanation/Examples
coherent treatment	-	EX, EC and DA records can be linked plausibly	If the DA domain is available, each record in EC should correspond to at least one record in DA (e.g., dispensed). Returned medication recorded in DA may either not appear in EC or appear with ECOCCUR = ”N”. An active treatment must always have been dispensed; therefore, it must have a matching dispensed record in DA. Similarly, each EX record must correspond to an active exposure record in EC, although a many-to-one relationship may occur if EX is condensed.
	-	(after unblinding) participants have only EX, EC, and DA records consistent with DM.ACTARM	Treatment arm allocation defines the set of protocol-compliant treatments. Any deviation from these may indicate data issues, treatment errors, or protocol deviations.
	combination treatment	plausible mapping of combination treatments	A combination of different active substances can be administered either separately – for example, via two or more tablets or injections, potentially in different forms – or together in a single tablet or injection. In the EC domain, exposure information is recorded as collected: a combination treatment administered in one injection is recorded as a single observation, while a combination treatment administered separately is recorded in multiple observations. It may be necessary to split a single EC record into multiple EX records, with one record for each active substance and its respective dose, or to combine multiple EC records into a single EX record. Doses, units, forms, routes, and frequency must correspond to the associated treatment.
	blinded trial	dummy/unblinded medication code list merged correctly to EX, EC and/or DA	In double-blind trials, the actual treatment is not known to the investigator or the participant at the time of administration. Instead, the CRF captures labels or codes from the medication packaging, which can later be linked to the corresponding treatment and dose using an unblinded medication code list. These medication identifiers may be included in EX, EC, and/or DA.
dates and relative days	-	exposure start/end date(time)s plausible in relation to DM reference dates	All active exposure start and end date(time)s in EC and EX must be on or after informed consent (RFICDTC) and on or before end of participation (RFPENDTC) and death date (DTHDTC). The date(time)s of first and last study treatment in DM (RFXSTDTC and RFXENDTC) must correspond to the earliest EXSTDTC and the latest EXENDTC (or, if EXENDTC is not populated, the latest EXSTDTC), respectively. Reference start and end date(time)s in DM (RFSTDTC and RFENDTC) are study-specific and must be checked accordingly. Often, RFSTDTC equals the first exposure start date(time), while RFENDTC corresponds to the last exposure date(time) or the end of study.

Continued on next page

Table 6 continued from previous page

Category	Condition	Check	Explanation/Examples
dates and relative days	-	EXSTDTC ≤ EXENDTC	The treatment start datetime must always be equal to or earlier than the treatment end datetime. The same check applies to EC. Depending on the treatment form, the start and end datetime can be defined as identical — for example, this is typically the case for tablets or injections.
treatment periods and gaps	cross-over study	washout periods have minimum length	In cross-over trials, the duration of one or more washout periods is defined in the protocol. The actual duration is implicitely represented in EC and EX as the gap between two consecutive treatment period or administration records. Durations shorter than specified may indicate data issues or treatment errors; in the latter case, a protocol deviation record in DV should be present.
plausible/ allowed values	compliance data collected in CRF	compliance matches amount of dispensed, returned and/or administered medication	Compliance values recorded in the CRF by the investigator should be verifiable using dispensed, returned, and/or administered medication data from EX, EC, and DA. In cases where additional treatment is dispensed, returned medication may still yield a compliance value of 100%.
	collected unit of treatment differs from protocol-specified unit	doses in EX calculated correctly from EC	EC must be available when exposure is collected differently than specified in the protocol; that is, the dose and unit of collected exposure records must be transformed during EX derivation to match the protocol-specified dose unit.
number of records	-	all EX, EC and DA records have corresponding USUBJID in DM	If EX, EC, DA, or other relevant SDTM dataset records cannot be merged with DM, this indicates underlying data errors.

DERIVATION OF EXPOSURE VARIABLES AND ADAM DATASET CREATION

After merging SDTM datasets and performing necessary data transformations, development of the exposure ADaM dataset proceeds by creating ADaM-compliant variables according to the study specifications. This includes mapping key identifiers, such as STUDYID, USUBJID, and SUBJID, as well as deriving absolute and relative analysis dates and times (e.g., ADY). Additional steps involve summarizing dosing information and deriving analysis-specific variables required for efficacy or safety evaluations. All variables are assigned standardized names, labels, and formats, and controlled terminology is applied where appropriate. Once the exposure ADaM variables are created, further checks can be performed to ensure the integrity of these complex derivations (see Table 7).

Table 7: Plausibility checks for exposure ADaM variables

Category	Condition	Check	Explanation/Examples
treatment periods and gaps	multiple treatment administrations	plausible treatment gaps	Depending on trial specifications, treatment gaps may be obligatory, as in cross-over studies, tolerated, or prohibited, in which case they can lead to protocol deviations and/or early trial discontinuation. Treatment gaps – defined as the time between two administrations or treatment periods – can therefore be informative and may indicate missing data, particularly if they are not consistent with the protocol.

Continued on next page

Table 7 continued from previous page

Category	Condition	Check	Explanation/Examples
treatment periods and gaps	multiple treatment administrations	treatment interruptions may have a minimum lengths and do not include treatment start/stop date(time)s	If specified in the protocol, treatment interruptions may be allowed or even planned and can be represented in a period-based ADEX as separate records. Typically, a gap between two administrations or periods is only considered a treatment interruption once it exceeds a defined minimum duration. The start and end date(time)s of the preceding and subsequent treatment periods or administrations are not included in the interruption record.
		no overlapping treatment periods	For period-based ADEX datasets, treatment periods or administrations must not overlap; that is, the start date(time) of each subsequent period or administration must occur after the end date(time) of the preceding one.
	cross-over study	washout periods do not run into next treatment period	As with other treatment interruptions represented in a period-based ADEX, washout periods must not overlap with treatment periods or administrations.
duplicates	-	no exact duplicates	Exact duplicates may indicate data issues, programming errors, or flawed specifications.
	-	none within relevant variable combinations	Duplicates within specific variable combinations may indicate data issues, programming errors, or flawed specifications. For example, in a BDS-structured ADEX, a participant should not have more than one record per parameter – or per parameter and timepoint – so no duplicates should exist within USUBJID and PARAMCD (and, for example, ADY). In a period-based ADEX, a participant should only have a single record per treatment at each date or datetime.
dates and relative days	-	ASTDT ≤ AENDT	For period-based ADEX datasets, the administration or treatment period analysis start date must be on or before the corresponding analysis end date. The same check applies to datetime variables. Special caution should be exercised when working with imputed dates or datetimes.
	-	correct reference date used for relative days	Typically – but not always – the first date of exposure serves as the reference point for calculating relative analysis dates, such as ADY. If datetimes are required for analysis, this check can be refined to account for times as well.
	-	no missing or partial dates as per sponsor rules with flags for imputed variables	Depending on analysis needs, certain dates – such as the analysis start date (ASTDT) – are generally expected to be complete and, if necessary, are imputed according to the SAP. Imputed variables must be flagged using the corresponding imputation flag (e.g., ASTDTF for ASTDT). If times are relevant for analysis, the same principle applies to datetime variables.
	reference start date = first day of treatment	each participant has an active exposure record with ADY = 1	Typically – but not necessarily – the first exposure to the study treatment is set as the reference start date. Therefore, each treated participant must have at least one active treatment record in a period-based ADEX dataset with the analysis relative day (ADY) set to 1, and no records should have ADY values less than 1.
		no ADY < 1	

Continued on next page

Table 7 continued from previous page

Category	Condition	Check	Explanation/Examples
plausible/ allowed values	-	plausible variable values and combinations of variable values	A plausible and coherent ADEX dataset permits only certain values for individual variables as well as specific combinations of variable values, for various reasons. In ADaM-compliant datasets, certain variables – such as the sub-period variables ASPER and ASPERC – must maintain a one-to-one relationship. In a BDS-structured ADEX, only pre-specified PARAM and PARAMCD values are expected, and plausible values for AVAL and AVALC depend on the parameter; for example, the number of administrations would have a defined maximum plausible count, compliance would commonly have a value between 0 and 100%.
	multiple treatment phases	plausible, consecutive treatment phases, periods and subperiods	If required for analysis, treatment phases, periods, and subperiods can be specified. Periods should be plausibly nested within phases, and subperiods within periods. All phases and (sub)periods should be numbered consecutively, in chronological order according to the date variables, starting with 1.
	-	coherent summary metrics	A BDS ADEX can include numerous summary metrics, some of which may be very similar depending on analysis needs. For example, treatment duration can be derived in days, weeks, months, and/or years, while compliance might be represented both as a numerical value and as a yes/no variable indicating whether a minimum threshold was met. Such related metrics must be internally consistent.
coherent treatment	-	plausible combinations of dose, unit and treatment values	In an ADEX dataset based on treatment periods or administrations, the actual dose and unit must be consistent with the assigned treatment. Deviations are considered plausible only if dose reductions are permitted.
number of records	-	record counts per USUBJID do not exceed the plausible maximum	Both period- and parameter-based ADEX datasets have a maximum number of records per participant. In a BDS-structured ADEX, the number of parameters is prespecified, resulting in a clear maximum number of records per USUBJID. In an ADEX dataset with one record per administration, treatment period, or phase, the treatment scheme defines the maximum number of records.
traceability	-	each datapoint is traceable	The ADaM Implementation Guide [10] defines two approaches to traceability. Datapoint traceability ensures that the predecessor value in SDTM can be directly identified through variables such as sequence numbers carried over from the SDTM dataset. Metadata traceability supports traceability of derived variables, where the variable metadata describes how each value was derived, including its source and the algorithms applied.

CROSS-CHECKS AGAINST OTHER ADAM DATASETS

The final, logically checked exposure ADaM dataset – whether period- or parameter-based – should be compared with other ADaM datasets to ensure internal consistency across the study data. Discrepancies with datasets such as ADSL may indicate data issues or derivation errors. Cross-checking helps verify that exposure variables and summary metrics are consistent with subject-level information, including treatment assignments, thereby supporting the integrity and reliability of analysis outputs. Proposed checks for such comparisons are listed in Table 8.

Table 8: Cross-checks against other ADaM datasets

Category	Condition	Check	Explanation/Examples
coherent treatment	-	ADSL.ACTARM/ TRTxxA/ TRTSEQA match exposure records	Subject-level treatment variables in ADSL – such as the actual treatment arm (ACTARM), the actual treatment in each period (TRTxxA), or the actual treatment sequence (TRTSEQA) – must correspond to the exposure records in a period-based ADEX.
	-	ADSL.DOSExxA/ DOSExxU match exposure records	Subject-level dosing variables in ADSL – such as the actual dose per period (DOSExxA) together with the respective unit (DOSExxU) – must correspond to the exposure records in a period-based ADEX.
	-	consistency between exposure ADaM datasets	If both types of ADEX datasets are created, their content must be consistent. For example, the number of administrations in a summary ADEX should align with the number of records in an ADEX with one record per administration. These checks are particularly meaningful when both types of ADEX datasets are created independently, rather than when an period-based ADEX serves as an intermediate dataset for the summary ADEX.
dates and relative days	-	ADSL.TRTSDT equals earliest exposure start date	The first exposure date (ADSL.TRTSDT) must match the earliest exposure start date in ADEX. The same applies to the earliest start date for each treatment period (ADSL.TRxxSDT). If times are relevant for analysis, this check extends to the datetime variable TRTSDTM.
	randomized controlled trial	all exposure dates ≥ ADSL.ENRLDT/ RFICDT/RANDDT	The start and end dates of each treatment period or administration must occur on or after enrollment (ENRLDT), informed consent (RFICDT), and randomization (RANDDT).
	-	ADSL.TRTEDT equals last administration date	The date of last exposure (ADSL.TRTEDT) must match the latest exposure date in ADEX. This check also applies to the last exposure date for each treatment period (ADSL.TRxxEDT). If times are relevant for analysis, this check extends to the datetime variable TRTEDTM.
	-	all exposure dates ≤ ADSL.EOSDT/ DTHDT	The start and end dates of each treatment period or administration must occur on or before the end of study (EOSDT) and the date of death (DTHDT).
	residual effect period specified	if specified, ADSL.APxxEDT includes REP but is cut off at ADSL.DTHDT/ EOSDT	The last exposure date (ADSL.TRTEDT) does not include the REP. If treatment periods including REP are required for analysis, the variables APxxSDT and APxxEDT (period xx start and end date) can be defined to represent treatment periods including REP and added to ADSL. Both variables are also available as datetime variables (APxxSDTM and APxxEDTM).
treatment periods and gaps	multiple treatment administrations	unexpectedly short treatment durations/low number of treatment administrations match disposition data	A lower number of treatment administrations or a shorter treatment duration than specified in the protocol must be accompanied by a disposition record indicating early treatment discontinuation.

Continued on next page

Table 8 continued from previous page

Category	Condition	Check	Explanation/Examples
plausible/ allowed values	-	participant's population flags match treatment records	Subject-level population flags, such as the safety population flag (SAFFL) or the treated population flag (TRTFL), must accurately reflect a participant's actual exposure. That is, if a participant is flagged as part of the treated set, the period-based ADEX must contain at least one active record for that participant. Similarly, a parameter-based summary ADEX must include appropriate records, for example, a treatment duration greater than 0.
	-	ADxx.ONTRTFL = "Y" observation falls into active treatment period (+ REP)	Each observation with the on-treatment flag ONTRTFL = "Y" must fall within an active treatment period, taking REP into account. Depending on the ADEX structure, REP may either be included in each record or must be handled within the programming algorithm.
	-	treatment variables in other ADaM dataset match exposure records	Depending on the analysis needs, each BDS- or OCCDS-structured ADaM dataset contains at least one subject- or record-level treatment variable. The values of these variables must align with the treatment variables in ADSL and the exposure records in ADEX. Common examples include TRTP, TRTA, TRTxxA, and TRTxxP.
	compliance analysed	plausible ADSL.TRCMP values comply with exposure records	Overall compliance in ADSL.TRCMP is either identical to the compliance value in ADEX (for single treatment periods or BDS-structured ADEX) or is correctly calculated from multiple compliance values across treatment periods.

CONCLUSION

This paper proposes checks to safeguard the development of exposure ADaM datasets in CDISC-compliant clinical trials. These checks are intended to identify plausibility, data, and programming issues. Systematic implementation enhances transparency and confidence in the final dataset; however, performing an extensive set of checks can be time-consuming. A risk-based approach improves efficiency by focusing on critical and error-prone process steps – such as unblinding – guided by study specifics and a thorough understanding of the data.

REFERENCES

1. Nelson, G. S. & Zhou, J. Good Programming Practices in Healthcare Creating Robust Programs. PharmaSUG 2011 Conference Proceedings. https://www.lexjansen.com/pharmasug/2011/TT/PharmaSUG-2011-TT05.pdf (2011).
2. Franklin, D. Defensive Programming, Avoiding the Big Mistakes. PharmaSUG 2014 Conference Proceedings. https://pharmasug.org/proceedings/2014/BB/PharmaSUG-2014-BB19.pdf (2014).
3. Brucken, N. & Levy, D. E. Defensive Coding by Example: Kick the Tires, Pump the Brakes, Check Your Blind Spots, and Merge Ahead! MidWest SAS Users Group 2016 Conference Proceedings. https://www.mwsug.org/proceedings/2016/TT/MWSUG-2016-TT12.pdf (2016).
4. Sherman, T. & Ringelberg, A. Defensive Programming and Error-handling: The Path Less Travelled. PharmaSUG 2013 Conference Proceedings. https://pharmasug.org/proceedings/2013/TF/PharmaSUG-2013-TF24.pdf (2013).
5. Donthi, B. & Qi, L. Defensive programming techniques for ongoing clinical studies. Western Users of SAS Software 2018 Conference Proceedings. https://www.lexjansen.com/wuss/2018/120_Final_Paper_PDF.pdf (2018).
6. Li, S., Ma, S., Lan, B. & Li, R. From SDTM to ADaM. SAS Global Forum 2013 Conference Proceedings. https://support.sas.com/resources/papers/proceedings13/199-2013.pdf (2013).
7. Analysis Data Model (ADaM) Oncology Examples Version 1.0 (Provisional) Clinical Data Interchange Standards Consortium (CDISC) (2025). https://www.cdisc.org/system/files/members/standard/foundational/ADaM%20Oncology%20Examples%20v1.0_Provisional.pdf.
8. Troxell, J. What is the "ADAM OTHER" Class of Datasets, and When Should it be Used? PharmaSUG 2015 Conference Proceedings. https://pharmasug.org/proceedings/2015/DS/PharmaSUG-2015-DS16.pdf (2015).

9. ADaM Structure for Occurrence Data (OCCDS) Implementation Guide Version 1.1 Clinical Data Interchange Standards Consortium (CDISC) (2021). https://www.cdisc.org/system/files/members/standard/foundational/ADaM_OCCDS_Implementation_Guide%20v1.1.pdf.
10. Analysis Data Model Implementation Guide Version 1.3 Clinical Data Interchange Standards Consortium (CDISC) (2021). https://www.cdisc.org/system/files/members/standard/foundational/ADaMIG_v1.3.pdf.
11. Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.4 Clinical Data Interchange Standards Consortium (CDISC) (2022). https://www.cdisc.org/system/files/members/standard/foundational/SDTMIG%20v3.4-FINAL_2022-07-21.pdf.

ACKNOWLEDGMENTS

The author would like to thank Judith Kanz, Monika Kawohl, and Sven Greiner for their valuable input and constructive feedback during the preparation of this manuscript.

CONTACT INFORMATION

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

Author Name: Susann Grill
Company: mainanalytics GmbH
Address: Otto-Volger-Strasse 3c, 65843 Sulzbach/Taunus, Germany
Email: susann.grill@mainanalytics.de
Website: www.mainanalytics.de