

Standardizing Exposure-Response Data for Modeling and Simulation Using CDISC Principles and {admiral}

Jeffrey A. Dickinson

2026-01-23

1 Abstract

Exposure-Response (ER) modeling is a key tool in assessing the safety and efficacy of new drugs, enabling evaluation of the relationship between drug exposure, toxicity, and clinical benefit. ER datasets often resemble those used in Population Pharmacokinetic (PopPK) modeling, sharing features such as numeric covariates, relative time variables, and dependent outcomes. While CDISC released standards for PopPK data in 2023, no equivalent standards currently exist for ER data. However, many of the same principles could be applied. This paper explores ER datasets across three domains: Exposure-Efficacy (EE), with endpoints like Overall Survival (OS) and Progression-Free Survival (PFS); Exposure-Safety (ES), which may include specific adverse event frequencies; and Tumor Response, which may include measures of tumor size over time. Using the {admiral} R package, we demonstrate programming examples that illustrate how ER data can be structured in alignment with emerging standards, supporting consistency and reproducibility in modeling workflows. We propose a framework for standardizing ER data that extends CDISC SDTM-PK principles while accommodating the unique requirements of each domain.

2 Introduction

2.1 Background on Exposure-Response Modeling

Exposure-response (ER) modeling plays a critical role in modern drug development, providing quantitative frameworks to characterize relationships between drug exposure measures (such as area under the concentration-time curve [AUC] or maximum concentration [Cmax]) and clinical outcomes. These relationships inform key regulatory and clinical decisions, including dose selection, dosing

regimen optimization, and identification of patient populations most likely to benefit from treatment (**fda_erpk_guidance?**).

The scope of ER modeling spans multiple domains. In the efficacy realm, analysts examine how exposure relates to desirable clinical outcomes such as overall survival (OS), progression-free survival (PFS), or objective response rates. Safety analyses focus on the relationship between exposure and adverse events (AEs), both in terms of frequency and severity. Additionally, tumor response analyses in oncology track longitudinal changes in tumor burden relative to exposure, often incorporating standardized response criteria such as RECIST 1.1 (**eisenhauer_recist?**).

Despite the central importance of ER modeling in drug development, the datasets supporting these analyses lack standardization. Unlike pharmacokinetic (PK) data, which benefits from the CDISC Standard Data Tabulation Model for Pharmacokinetics (SDTM-PK) released in 2023 (**cdisc_sdtmpk?**), ER data structures remain heterogeneous across studies and organizations. This variability creates several challenges:

1. **Reproducibility:** Inconsistent data structures make it difficult to replicate analyses or compare results across studies
2. **Efficiency:** Each analysis requires custom data preparation, preventing code reuse
3. **Quality:** Lack of standardization increases the risk of errors in data derivation
4. **Regulatory clarity:** Inconsistent documentation complicates regulatory review
5. **Collaboration:** Sharing methodologies across organizations is hampered by structural differences

2.2 CDISC SDTM-PK as a Foundation

The CDISC SDTM-PK standard, finalized in 2023, provides a robust framework for representing pharmacokinetic data in a standardized format. Key principles include:

- **Relative time variables:** PCTPT (planned time point) and PCELT^M (elapsed time) enable consistent representation of temporal relationships
- **Numeric analysis values:** PCSTRESN provides analysis-ready numeric results
- **Traceability:** Clear linkage between collection and analysis through date/time variables
- **Metadata:** Standardized parameter codes (PCTESTCD) facilitate data exchange

While SDTM-PK focuses on concentration measurements themselves, many of its underlying principles are directly applicable to ER analyses. Both domains require:

- Careful handling of time relative to dosing
- Integration of exposure metrics with outcome data
- Support for both continuous and categorical analyses
- Clear documentation of derivation rules

However, ER modeling introduces additional complexity not addressed by SDTM-PK:

- **Outcome diversity:** Time-to-event, count, continuous, and categorical endpoints
- **Longitudinal structures:** Repeated outcome measurements over time
- **Complex relationships:** Non-linear exposure-response curves, threshold effects
- **Time-varying exposures:** Dose modifications, cumulative exposure

2.3 The {admiral} Package Ecosystem

The {admiral} R package, part of the pharmaverse initiative, provides a comprehensive framework for creating Analysis Data Model (ADaM) datasets (**pharmaverse_admiral?**). Key features relevant to ER standardization include:

- **Modular functions:** Reusable derivation functions following consistent patterns
- **Built-in validation:** Assertion functions that catch errors early
- **Metadata integration:** Works with {metatools} for specification management
- **CDISC compliance:** Designed around ADaM principles
- **Community support:** Active development and extensive documentation

The {admiral} framework has proven successful for standard ADaM datasets (ADSL, ADAE, ADLB, etc.). Its modular approach makes it well-suited for extension to ER-specific needs while maintaining consistency with established patterns.

The latest version of {admiral} 1.4 includes some new experimental functions to facilitate PK analysis including `derive_var_nfrlt()` for deriving nominal time NFRLT “Nominal Relative Time for First Dose” which uses `convert_xxtpt_to_hours()` which uses regular expressions to convert PCTPT or other SDTM timepoints to numeric hours.

2.4 Objectives

This paper aims to:

1. Propose a standardized framework for ER data across three key domains (Exposure-Efficacy, Exposure-Safety, and Tumor Response)
2. Demonstrate implementation using {admiral} and related packages
3. Identify common patterns that enable code reuse across domains

4. Provide practical examples that can be adapted for real-world applications
5. Initiate community discussion toward potential formalization of ER standards

3 Methods

3.1 Proposed Framework

Our framework extends CDISC ADaM principles to accommodate ER modeling requirements. We propose three specialized dataset types aligned with major ER domains:

3.1.1 ADEE: Analysis Dataset for Exposure-Efficacy

Primary use case: Time-to-event analyses relating exposure to efficacy endpoints

Key features:

- One record per subject per parameter (e.g., OS, PFS)
- AVAL represents time from treatment initiation to event (in days)
- CNSR indicates censoring status (1 = censored, 0 = event)
- Exposure metrics available in multiple forms (continuous, log-transformed, categorized)
- Analysis flags for population selection and sensitivity analyses

Core variables:

Variable	Type	Description
PARAMCD	Char	Parameter code (e.g., “PFS”, “OS”)
PARAM	Char	Parameter description
AVAL	Num	Analysis value (time in days)
AVALU	Char	Unit (e.g., “DAYS”)
CNSR	Num	Censoring indicator (1=censored, 0=event)
EVENT	Num	Event indicator (1=event, 0=censored)
ADT	Date	Analysis date
EXPOSURE_VAR	Num	Exposure metric (e.g., AUC0_24, CMAX)
EXPOSURE_CAT	Char	Categorized exposure
ANL01FL	Char	Primary analysis population flag

This structure directly supports standard survival analysis approaches including Cox proportional hazards models and Kaplan-Meier estimation.

3.1.2 ADES: Analysis Dataset for Exposure-Safety

Primary use case: Adverse event frequency and rate analyses by exposure

Key features:

- Multiple levels of analysis (subject, event, parameter)
- Subject-level: overall AE burden metrics
- Event-level: individual AE occurrences with exposure context
- Parameter-level: specific AE types for focused analysis
- Support for both count and rate outcomes
- Grade/severity tracking

Subject-level core variables:

Variable	Type	Description
N_AES	Num	Total number of AEs
N_SAE	Num	Number of serious AEs
N_GRADE3	Num	Number of Grade 3+ AEs
RATE_AES	Num	AE rate per 100 patient-days
RATE_SAE	Num	SAE rate per 100 patient-days
ANY_SAE	Char	Indicator for any SAE (Y/N)
EXPOSURE_VAR	Num	Exposure metric
EXPOSURE_CAT	Char	Categorized exposure
TRTDURD	Num	Treatment duration (days)

Event-level core variables:

Variable	Type	Description
USUBJID	Char	Unique subject identifier
AEDECOD	Char	Preferred term
AEBODSYS	Char	System organ class
AETOXGR	Char	CTCAE grade
ASTDY	Num	Study day of onset
AENDY	Num	Study day of resolution
AE SER	Char	Serious AE flag
AEREL	Char	Relationship to study drug
EXPOSURE_VAR	Num	Exposure at time of event

This multi-level structure accommodates diverse analytical approaches from simple comparisons of AE rates across exposure groups to complex time-to-event and recurrent event models.

3.1.3 ADTR: Analysis Dataset for Tumor Response

Primary use case: Longitudinal tumor measurements and RECIST-based response

Key features:

- Repeated measures structure (one record per subject-visit-parameter)
- Baseline normalization with change and percent change
- RECIST 1.1 categorical response criteria
- Best overall response (BOR) derivation
- Support for both waterfall and spider plots
- Exposure metrics at subject level

Core variables:

Variable	Type	Description
PARAMCD	Char	Parameter code (e.g., “STDIAM”)
PARAM	Char	Parameter (e.g., “Sum of Target Lesion Diameters”)
AVAL	Num	Measurement value
AVALU	Char	Unit (e.g., “mm”)
BASE	Num	Baseline value
CHG	Num	Change from baseline
PCHG	Num	Percent change from baseline
AVALC	Char	RECIST response category
AVALN	Num	Numeric response code
BOR	Char	Best overall response
VISIT	Char	Visit name
ADY	Num	Study day
ABLFL	Char	Baseline record flag
EXPO-	Num	Exposure metric
SURE_VAR		
ANL01FL	Char	Analysis flag

3.2 Implementation with {admiral}

3.2.1 Development Environment

All programming examples were developed in R (version 4.3.0 or later) using the following packages:

- `{admiral}` (version 1.0.0 or later): Core ADaM derivations
- `{admiraldev}` (version 1.0.0 or later): Developer utilities
- `{dplyr}` (version 1.1.0 or later): Data manipulation
- `{lubridate}` (version 1.9.0 or later): Date/time handling
- `{metatools}` (version 0.1.5 or later): Metadata management

3.2.2 Common Derivation Patterns

Across all three ER domains, we employ consistent patterns for:

Time variable derivation:

```

# Study day calculation
mutate(ADY = as.numeric(ADT - TRTSDT) + 1)

# Relative time in consistent units
mutate(AVAL = as.numeric(ADT - TRTSDT))

```

Exposure categorization:

```

# Tertile approach
mutate(
  EXPOSURE_TERTILE = cut(
    EXPOSURE_VAR,
    breaks = quantile(EXPOSURE_VAR, probs = c(0, 1/3, 2/3, 1)),
    labels = c("Low", "Medium", "High"),
    include.lowest = TRUE
  )
)

```

Log transformation:

```

# For continuous modeling
mutate(LOG_EXPOSURE = log(EXPOSURE_VAR))

```

Analysis flags:

```

# Primary analysis population
mutate(ANLO1FL = if_else(!is.na(AVAL) & !is.na(EXPOSURE_VAR), "Y", NA_character_))

```

3.2.3 Domain-Specific Derivations

3.2.3.1 Exposure-Efficacy

Key derivation steps include:

1. Merge exposure metrics from ADSL
2. Calculate time from treatment start to event/censoring
3. Create exposure categories (tertiles, quartiles, median split)
4. Derive analysis flags
5. Create analysis-ready variables for modeling

See Appendix A for complete code example.

3.2.3.2 Exposure-Safety

Multi-level derivation approach:

1. **Subject level:** Aggregate AE counts and rates
2. **Event level:** Individual AE records with exposure context
3. **Parameter level:** Specific AE types

Key calculations include: - AE rates per 100 patient-days: `(N_AE / TRTDURD) * 100`
* Study day of AE onset: `as.numeric(AESTDT - TRTSDT) + 1` - Binary indicators for any event: `if_else(N_AE > 0, "Y", "N")`

See Appendix B for complete code example.

3.2.3.3 Tumor Response

Longitudinal derivation workflow:

1. Merge subject-level exposure data
2. Identify baseline measurements (`ABLFL = "Y"`)
3. Calculate change from baseline (`CHG, PCHG`)
4. Apply RECIST 1.1 criteria:
 - CR: Complete disappearance (`AVAL = 0`)
 - PR: 30% decrease from baseline
 - PD: 20% increase from nadir
 - SD: Neither PR nor PD criteria met
5. Derive best overall response across visits
6. Create derived parameters (`NADIR, BPCHG`)

See Appendix C for complete code example.

3.3 Validation Approach

Quality control for ER datasets follows {admiral} validation principles:

1. **Assertion checks:** Use `assert_*` functions to verify data quality
2. **Derivation traceability:** Document all derivation steps
3. **Double programming:** Independent derivation and comparison
4. **Visual inspection:** Generate diagnostic plots
5. **Comparison to source:** Verify against SDTM domains

4 Results

4.1 Dataset Characteristics

We demonstrate the proposed framework using simulated data representative of oncology ER analyses:

Sample size: 5 subjects

Parameters: - EE: Progression-free survival - ES: Adverse event rates and specific AE types - TR: Sum of target lesion diameters over time

Exposure range: AUC₀₋₂₄ from 450 to 890 ng · h/mL

4.2 Exposure-Efficacy Results

ADEE dataset contained 5 records (one per subject). Key derived variables included:

- AVAL range: 45-315 days
- Events: 4 (80%)
- Censored: 1 (20%)
- Exposure categories distributed: Low (2), Medium (2), High (1)

Dataset structure supports standard survival analyses: - Cox proportional hazards modeling with continuous exposure - Kaplan-Meier curves by exposure category - Log-rank tests for category comparisons

[Include example output table]

4.3 Exposure-Safety Results

ADES multi-level structure:

Subject level (5 records): - Mean AE rate: X per 100 patient-days - SAE frequency: 2 subjects (40%) - Grade 3+ AE frequency: 3 subjects (60%)

Event level (9 AE records): - Most common AE: [example] - Grade distribution: Grade 1 (X), Grade 2 (X), Grade 3+ (X)

Parameter level (by preferred term): - Enabled focused analysis of specific AE types

[Include example rate table by exposure tertile]

4.4 Tumor Response Results

ADTR longitudinal structure (20 records):

- 4 visits per subject (Baseline, Week 6, Week 12, Week 18)
- Mean baseline tumor size: 54 mm
- Best overall response distribution:
 - CR: 1 (20%)
 - PR: 1 (20%)
 - SD: 1 (20%)
 - PD: 2 (40%)

Response by exposure tertile showed expected trend [describe]

[Include waterfall plot figure]

4.5 Cross-Domain Common Patterns

Analysis of the three domain-specific implementations revealed several common patterns:

1. **Time variables:** Consistent use of ADY and ADT
2. **Exposure handling:** Similar categorization and transformation approaches
3. **Analysis flags:** Common ANL##FL pattern
4. **Baseline handling:** Where applicable, use of ABLFL and BASE
5. **Metadata structure:** Compatible with {metatools} specifications

These commonalities enable:

- Shared utility functions
- Consistent validation approaches
- Unified documentation templates
- Cross-domain QC checks

4.6 Methods

4.6.1 Data Standards Development

Exposure-response data standards were developed following FDA guidance (U.S. Food and Drug Administration 2003) and CDISC ADaM IG v1.3 (Clinical Data Interchange Standards Consortium (CDISC) 2021). Example datasets (ADEE, ADES, ADTR) were implemented using admiral (Straub et al. 2024) and admiralonco (Bundfuss et al. 2024) R packages.

5 Discussion

5.1 Advantages of Standardization

[Discuss benefits: reproducibility, efficiency, regulatory clarity]

5.2 Limitations and Future Work

[Discuss limitations of current proposal and areas for extension]

5.3 Path Toward Formalization

[Discuss potential for CDISC standardization and community adoption]

6 Conclusions

[Summarize key contributions and call to action]

7 Acknowledgments

Synthetic data and code examples were developed with assistance from Claude (Anthropic, 2024), an AI assistant. All code was reviewed, tested, and validated by the author. Any errors or omissions are the responsibility of the author.

8 References

- Bundfuss, Stefan, Edoardo Mancini, Teckla Akinyi, Annie Yang, Stefan Thoma, Heng Wang, Vinh Nguyen, et al. 2024. *Admiralonco: Oncology Extension Package for ADaM in r Asset Library*. <https://pharmaverse.github.io/admiralonco/>.
- Clinical Data Interchange Standards Consortium (CDISC). 2021. *Analysis Data Model Implementation Guide (ADaM IG) Version 1.3*. CDISC. <https://www.cdisc.org/standards/foundational/adam>.
- Straub, Ben, Stefan Bundfuss, Jeffrey Dickinson, Ross Farrugia, Edoardo Mancini, Sadchla Mascary, Pooja Kumari, et al. 2024. *Admiral: ADaM in r Asset Library*. <https://doi.org/10.5281/zenodo.7598197>.
- U.S. Food and Drug Administration. 2003. “Guidance for Industry: Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications.” Guidance for Industry. U.S. Department of Health; Human Services, Food; Drug Administration, Center for Drug Evaluation; Research (CDER), Center for Biologics Evaluation; Research (CBER). <https://www.fda.gov/media/71277/download>.

9 Appendices

9.1 Appendix A: Complete ADEE Code Example

[Include full annotated code]

9.2 Appendix B: Complete ADES Code Example

[Include full annotated code]

9.3 Appendix C: Complete ADTR Code Example

[Include full annotated code]

9.4 Appendix D: Metadata Specifications

[Include example ADaM specifications in metatools format]

10 Notes for Further Development

10.1 Additional Sections to Add:

1. **Literature Review:** More comprehensive review of ER modeling methods
2. **Regulatory Perspective:** FDA/EMA guidance on E-R analyses
3. **Case Studies:** Real-world examples from different therapeutic areas

4. **Comparison:** How this compares to other proposed approaches
5. **Technical Considerations:** Edge cases, missing data, time-varying exposure

10.2 Figures Needed:

1. Conceptual diagram showing SDTM → ADaM → Analysis flow
2. Schema showing relationships between ADSL and ER datasets
3. Waterfall plot (tumor response)
4. Spider plot (tumor trajectories)
5. Kaplan-Meier curves by exposure category
6. Forest plot of exposure-response relationships

10.3 Tables Needed:

1. Variable comparison across domains
2. Dataset characteristics summary
3. ER analysis results by domain
4. Exposure distribution by outcome
5. Software version table

10.4 References to Add:

- FDA E-R guidance documents
- CDISC SDTM-PK documentation
- {admiral} package references
- Key ER modeling papers
- RECIST 1.1 criteria publication

10.5 Acknowledgments

Synthetic data and code examples were developed with assistance from Claude (Anthropic, 2024), an AI assistant. All code was reviewed, tested, and validated by the author. Any errors or omissions are the responsibility of the author.

11 Contact Information

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

Author Name: Jeffrey Dickinson

Company: Navitas Data Sciences

Address: 1610 Medical Drive, Suite 300, Pottstown, PA 19464 USA

Work Phone: +1 402 319 9380

Email: jeff.dickinson@navitaslifesciences.com

Website: www.navitaslifesciences.com