# Standardizing Exposure-Response Data for Modeling and Simulation

## Using CDISC Principles and {admiral}

Jeff Dickinson

Navitas Data Sciences

2026-01-27

# Overview

**The Challenge**

- ER modeling critical for drug development
- Datasets lack standardization
- CDISC SDTM-PK exists, but no ER equivalent
- Inconsistent structures hinder reproducibility

**Our Approach**

- Extend CDISC principles to ER data
- Three key domains:
  - Exposure-Efficacy (EE)
  - Exposure-Safety (ES)
  - Tumor Response
- Demonstrate with {admiral}

# The ER Modeling Landscape

**What is Exposure-Response Analysis?**

- Quantifies relationship between drug exposure (PK) and outcomes
- Key questions:
  - What exposure achieves target efficacy?
  - What exposure level increases toxicity risk?
  - How do we optimize dosing for subpopulations?

**Current State**

- Each analysis starts from scratch
- Variable naming inconsistent
- Difficult to share methodologies
- QC challenges

# Why Standards Matter

- **Consistency**: Same structure across studies and organizations

*"The beauty of standards is that we can build tooling that works across contexts"*

# Why Standards Matter

- **Consistency**: Same structure across studies and organizations
- **Traceability**: Clear lineage from raw data $\rightarrow$ analysis

"*The beauty of standards is that we can build tooling that works across contexts*"

# Why Standards Matter

- **Consistency**: Same structure across studies and organizations
- **Traceability**: Clear lineage from raw data $\rightarrow$ analysis
- **Reproducibility**: Others can verify and extend your work

"*The beauty of standards is that we can build tooling that works across contexts*"

# Why Standards Matter

- **Consistency**: Same structure across studies and organizations
- **Traceability**: Clear lineage from raw data $\rightarrow$ analysis
- **Reproducibility**: Others can verify and extend your work
- **Efficiency**: Reusable code and workflows

"*The beauty of standards is that we can build tooling that works across contexts*"

# Why Standards Matter

- **Consistency**: Same structure across studies and organizations
- **Traceability**: Clear lineage from raw data $\rightarrow$ analysis
- **Reproducibility**: Others can verify and extend your work
- **Efficiency**: Reusable code and workflows
- **Regulatory clarity**: Easier to document and defend analyses

  *"The beauty of standards is that we can build tooling that works across contexts"*

# SDTM-PK: Our Foundation

**Key Principles from SDTM-PK**

- Relative time variables (*TPT*, ELTM)
- Numeric analysis values (PCSTRESN)
- Analysis flags (*FL variables)
- Traceability (*DTC, –SEQ)
- Standardized parameter codes

**What We Can Leverage**

- Time calculation patterns
- Covariate structures
- Metadata approaches
- Validation frameworks

**The Gap**: SDTM-PK focuses on PK measurements, not PK-outcome relationships

# Three ER Domains: A Comparison

| Aspect | Exposure-Efficacy | Exposure-Safety | Tumor Response |
|---|---|---|---|
| **Primary Outcome** | Time-to-event (OS, PFS) | Event frequency/rates | Tumor size over time |
| **Data Structure** | One record per subject-event | Multiple levels (subject/event) | Longitudinal repeated measures |
| **Key Metrics** | Survival time, censoring | AE counts, rates, grades | Percent change, RECIST |
| **Analysis Type** | Cox regression, K-M curves | Poisson/negative binomial | Mixed models, waterfall plots |
| **Common Challenges** | Censoring handling | Time-varying exposure | Baseline normalization |

# Domain 1: Exposure-Efficacy

**Use Case**: Progression-Free Survival by AUC Quartiles

**Key Dataset Features**

- One record per subject-parameter
- AVAL = time from treatment start (days)
- CNSR = censoring indicator (1=censored, 0=event)
- Exposure available as continuous and categorical

```
adee %>%
  select(USUBJID, AVAL, CNSR, AUCO_24, AUC_TERTILE) %>%
  head(3)
```

# EE: Code Example

```r
# Derive time to event
adtte <- adtte_source %>%
  derive_vars_merged(
    dataset_add = adsl,
    by_vars = exprs(USUBJID)
  ) %>%
  mutate(
    # Analysis value = days from treatment start
    AVAL = as.numeric(ADT - TRTSDT),
    AVALU = "DAYS",
    # Event indicator (inverse of CNSR for modeling)
    EVENT = if_else(CNSR == 0, 1, 0)
  )

# Create exposure categories
exposure_cats <- adsl %>%
  mutate(
    AUC_TERTILE = cut(AUC0_24,
```

# EE: Analysis-Ready Output

```
# ADEE structure
USUBJID | PARAMCD | AVAL | AVALU | CNSR | EVENT | AUC0_24
--------|---------|------|-------|------|-------|---------
001-001 | PFS     | 104  | DAYS  | 1    | 0     | 450
001-002 | PFS     | 135  | DAYS  | 0    | 1     | 520
001-003 | PFS     | 213  | DAYS  | 0    | 1     | 680
```

**Ready for modeling:**

```
# Cox proportional hazards
coxph(Surv(AVAL, EVENT) ~ LOGAUC + AGE + SEX, data = adee)

# By category
survfit(Surv(AVAL, EVENT) ~ AUC_TERTILE, data = adee)
```

# Domain 2: Exposure-Safety

**Use Case**: Adverse Event Rates by Cmax Categories

**Unique Challenges**

- Multiple analysis levels (subject, event, parameter)
- Time-varying exposure considerations
- Grade/severity tracking
- Need both counts and rates

**Three Dataset Structures**

1. **Subject-level**: Overall AE burden per subject
2. **Event-level**: Individual AE occurrences
3. **Parameter-level**: Specific AE types

# ES: Subject-Level Structure

```r
# Subject-level AE summary
ades_subj <- exposure_cats %>%
  left_join(ae_summary, by = "USUBJID") %>%
  mutate(
    # Total events
    N_AES = n(),
    N_SAE = sum(AESER == "Y"),
    N_GRADE3 = sum(AETOXGR >= "3"),

    # Rates per 100 patient-days
    RATE_AES = (N_AES / TRTDURD) * 100,
    RATE_SAE = (N_SAE / TRTDURD) * 100,

    # Binary indicators
    ANY_SAE = if_else(N_SAE > 0, "Y", "N"),
    ANY_GRADE3 = if_else(N_GRADE3 > 0, "Y", "N")
  )
```

# ES: Event-Level for Detailed Analysis

```
# Event-level dataset
USUBJID | AEDECOD     | AETOXGR | ASTDY | CMAX | CMAX_TERTII
--------|-------------|---------|-------|------|------------
001-001 | Nausea      | 1       | 15    | 125  | Low
001-001 | Fatigue     | 2       | 32    | 125  | Low
001-002 | Neutropenia | 3       | 56    | 145  | Medium
```

**Enables analyses like:**

- Time to first Grade 3+ AE
- Recurring event models
- Exposure-toxicity relationships by AE type

# Domain 3: Tumor Response

**Use Case**: Longitudinal Tumor Measurements with RECIST 1.1

**Key Features**

- Repeated measures over time
- Baseline normalization critical
- Categorical response criteria (CR, PR, SD, PD)
- Best overall response (BOR) derivation

**Analysis Approaches**

- Waterfall plots (best percent change)
- Spider plots (individual trajectories)
- Response rate by exposure
- Time to response

# Tumor Response: Longitudinal Structure

```
adtr <- tr_raw %>%
  mutate(
    PARAM = "Sum of Target Lesion Diameters",
    PARAMCD = "STDIAM",
    AVAL = TRORRES, # Raw measurement
    ADY = as.numeric(ADT - TRTSDT) + 1
  ) %>%
  # Identify baseline
  group_by(USUBJID, PARAMCD) %>%
  mutate(
    ABLFL = if_else(VISITNUM == 0, "Y", NA_character_),
    BASE = AVAL[ABLFL == "Y"]
  ) %>%
  # Calculate changes
  mutate(
    CHG = if_else(ABLFL != "Y", AVAL - BASE, NA_real_),
    PCHG = if_else(ABLFL != "Y", (AVAL - BASE) / BASE * 100
  )
```

# RECIST Criteria Implementation

```r
adtr <- adtr %>%
  mutate(
    AVALC = case_when(
      ABLFL == "Y" ~ "BASELINE",
      AVAL == 0 ~ "CR", # Complete Response
      PCHG <= -30 ~ "PR", # Partial Response (>30% decrease
      PCHG >= 20 ~ "PD", # Progressive Disease (>20% increa
      TRUE ~ "SD" # Stable Disease
    )
  )

# Best Overall Response (BOR)
bor <- adtr %>%
  filter(ABLFL != "Y") %>%
  group_by(USUBJID) %>%
  summarise(
    BOR = case_when(
      any(AVALC == "CR") ~ "CR",
```

# Tumor Response: Output Structure

```
# Longitudinal measurements
USUBJID | VISIT    | ADY | BASE | AVAL | CHG  | PCHG    | AVA
--------|----------|-----|------|------|------|---------|----
001-002 | BASE     | 1   | 65   | 65   | NA   | NA      | BAS
001-002 | WEEK 6   | 43  | 65   | 55   | -10  | -15.4   | SD
001-002 | WEEK 12  | 85  | 65   | 35   | -30  | -46.2   | PR
001-002 | WEEK 18  | 127 | 65   | 32   | -33  | -50.8   | PR
```

**Each subject has:**

- Longitudinal trajectory (for spider plots)
- Best overall response (for response rate)
- Exposure metrics (for ER modeling)

# Common Patterns Across Domains

**Time Variables**

- ADT: Analysis date
- ADY: Study day (relative to TRTSDT)
- AVAL: Numeric outcome
- Consistent calculation methods

**Exposure Metrics**

- Raw values (AUC, Cmax, Cavg)
- Transformed (log, standardized)
- Categorized (tertiles, quartiles)
- Available at subject level

**Analysis Flags**

- ANL01FL: Primary analysis population
- ANL02FL: Sensitivity analyses
- Domain-specific (GRADE3FL, RESPFL, etc.)

**Traceability**

- Clear link to SDTM (via –SEQ, –DTC)
- Derivation rules documented
- Metadata specifications

# Why {admiral}?

### Advantages for ER Data

- **Modular functions**: `derive_vars_*`, `derive_param_*`

### Example Extension

```
# Hypothetical future function
derive_param_er_response <- function(
  dataset, exposure_var, outcome_var,
  method = c("tertile", "quartile", "continuous")
) { ... }
```

# Why {admiral}?

## Advantages for ER Data

- **Modular functions**: `derive_vars_*`, `derive_param_*`
- **Validation built-in**: `assert_*` functions catch errors early

## Example Extension

```
# Hypothetical future function
derive_param_er_response <- function(
  dataset, exposure_var, outcome_var,
  method = c("tertile", "quartile", "continuous")
) { ... }
```

# Why {admiral}?

### Advantages for ER Data

- **Modular functions**: `derive_vars_*`, `derive_param_*`
- **Validation built-in**: `assert_*` functions catch errors early
- **Metadata integration**: Works with {metatools}

### Example Extension

```
# Hypothetical future function
derive_param_er_response <- function(
  dataset, exposure_var, outcome_var,
  method = c("tertile", "quartile", "continuous")
) { ... }
```

# Why {admiral}?

## Advantages for ER Data

- **Modular functions**: `derive_vars_*`, `derive_param_*`
- **Validation built-in**: `assert_*` functions catch errors early
- **Metadata integration**: Works with {metatools}
- **Community-driven**: Pharmaverse ecosystem

## Example Extension

```
# Hypothetical future function
derive_param_er_response <- function(
  dataset, exposure_var, outcome_var,
  method = c("tertile", "quartile", "continuous")
) { ... }
```

# Why {admiral}?

## Advantages for ER Data

- **Modular functions**: `derive_vars_*`, `derive_param_*`
- **Validation built-in**: `assert_*` functions catch errors early
- **Metadata integration**: Works with {metatools}
- **Community-driven**: Pharmaverse ecosystem
- **Well-documented**: Functions follow consistent patterns

## Example Extension

```
# Hypothetical future function
derive_param_er_response <- function(
  dataset, exposure_var, outcome_var,
  method = c("tertile", "quartile", "continuous")
) { ... }
```

# Why {admiral}?

## Advantages for ER Data

- **Modular functions**: `derive_vars_*`, `derive_param_*`
- **Validation built-in**: `assert_*` functions catch errors early
- **Metadata integration**: Works with {metatools}
- **Community-driven**: Pharmaverse ecosystem
- **Well-documented**: Functions follow consistent patterns
- **CDISC-aligned**: Built with standards in mind

## Example Extension

```
# Hypothetical future function
derive_param_er_response <- function(
  dataset, exposure_var, outcome_var,
  method = c("tertile", "quartile", "continuous")
) { ... }
```

# Benefits: Reproducibility

**Before Standardization**

```
# Study A
data$time_days <- difftime(data$event_dt, data$trt_start, u
data$AUC_cat <- cut(data$AUC, breaks=c(0,500,1000,Inf))

# Study B
df$tte <- as.numeric(df$EventDate - df$TreatmentDate) + 1
df$exposure_grp <- ifelse(df$auc < median(df$auc), "Low", "
```

**After Standardization**

```
# All studies use consistent approach
adee <- derive_er_tte_data(
  adsl = adsl,
  adrs = adrs,
  exposure_var = "AUC0_24",
  param = "PFS"
)
```

# Benefits: Efficiency

**Time Savings**

- Dataset creation: 50% reduction with templates
- QC: Automated checks via {admiral} assertions
- Documentation: Metadata-driven specs
- Onboarding: New analysts can follow patterns

**Quality Improvements**

- Fewer manual errors
- Consistent validation approach
- Peer review easier
- Regulatory submissions smoother
  *"We reduced our ER dataset programming time from 3 weeks to 1 week per study" - [Example quote]*

# Path Forward: Community Adoption

## What We Need

1. **Feedback on proposed structures**
   - Variable naming conventions
   - Mandatory vs. optional elements
   - Domain-specific needs
2. **Real-world testing**
   - Pilot studies across organizations
   - Edge cases and exceptions
   - Tooling gaps
3. **Documentation**
   - Implementation guides
   - Worked examples
   - Best practices
4. **Potential formalization**
   - Could inform future CDISC guidance
   - Pharmaverse ER working group?

# Getting Started Today

## Resources Available

- GitHub repository: [https://github.com/jeffreyad/er-standards]
- Example datasets and scripts
- ADaM specifications templates
- Presentation materials

## Try It Yourself

```
# Install development version
remotes::install_github("your-username/er-standards-project

# Run examples
library(erstds)
run_ee_example()
run_es_example()
run_tumor_response_example()
```

## Connect

- Questions: [jeff.dickinson@navitaslifesciences.com]

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities
   - Exposure-Efficacy: time-to-event

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities
   - Exposure-Efficacy: time-to-event
   - Exposure-Safety: event rates

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities
   - Exposure-Efficacy: time-to-event
   - Exposure-Safety: event rates
   - Tumor Response: longitudinal measurements

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities
   - Exposure-Efficacy: time-to-event
   - Exposure-Safety: event rates
   - Tumor Response: longitudinal measurements
4. **{admiral} is well-suited** - modular, validated, community-supported

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities
   - Exposure-Efficacy: time-to-event
   - Exposure-Safety: event rates
   - Tumor Response: longitudinal measurements
4. **{admiral} is well-suited** - modular, validated, community-supported
5. **Benefits are real** - efficiency, reproducibility, quality

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent
2. **CDISC principles apply** - we can extend SDTM-PK patterns
3. **Three distinct domains** - but shared commonalities
   - Exposure-Efficacy: time-to-event
   - Exposure-Safety: event rates
   - Tumor Response: longitudinal measurements
4. **{admiral} is well-suited** - modular, validated, community-supported
5. **Benefits are real** - efficiency, reproducibility, quality
6. **Community input needed** - pilot, refine, formalize

# Questions?

**Contact Information**

- Email: jeff.dickinson@navitaslifesciences.com
- GitHub: [jeffreyad]
- LinkedIn: [https://www.linkedin.com/in/jeffreyad/]

**Resources**

- Repository: [https://github.com/jeffreyad/er-standards]
- Documentation: [https://github.com/jeffreyad/er-standards/specifications]
- Pharmaverse: https://pharmaverse.org
- Examples: https://pharmaverse.github.io/examples/

Thank you!

# Backup Slides

# Backup: SDTM-PK Variables Reference

**Key Variables from SDTM-PK**

| Variable | Description | ER Relevance |
|---|---|---|
| PCTPT | Time Point Name | Timing of outcome assessment |
| PCTPTNUM | Numeric Time Point | Analysis visit sequencing |
| PCELTM | Planned Elapsed Time | Relative timing calculations |
| PCSTRESN | Numeric Result | Maps to AVAL |
| PCSTRESU | Units | Maps to AVALU |

# Backup: Detailed Variable Specifications

**Proposed Core Variables for ADEE**

| Variable | Type | Description | Requ |
|---|---|---|---|
| PARAMCD | Char | Parameter Code (e.g., “PFS”, “OS”) | Yes |
| AVAL | Num | Analysis Value (days) | Yes |
| CNSR | Num | Censoring (1=censored, 0=event) | Yes |
| ADT | Date | Analysis Date | Yes |
| EXPOSURE_VAR | Num | Exposure metric | Yes |
| EXPOSURE_CAT | Char | Categorized exposure | No |
| ANL01FL | Char | Primary analysis flag | Yes |

# Backup: Handling Edge Cases

**Common Challenges**

1. **Missing Exposure Data**
   - Imputation strategies
   - Sensitivity analyses
   - Flagging for exclusion
2. **Competing Risks** (EE)
   - Death vs. progression
   - Multiple event types
3. **Time-Varying Exposure** (ES)
   - Dose changes
   - Cumulative exposure
4. **Unscheduled Assessments** (Tumor)
   - Off-schedule scans
   - Response confirmation

# 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.