# Standardizing Exposure-Response Data for Modeling and Simulation

## Using CDISC Principles and {admiral}

Jeff Dickinson

Navitas Data Sciences

2026-02-01

# 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
- Four key datasets:
  - ADER: Exposure-Response (foundational)
  - ADEE: Exposure-Efficacy
  - ADES: Exposure-Safety
  - ADTRR: Tumor Response for E-R
- 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 (AUC, AUC_SS, AUCSS, AUC0_24)
- Limited exposure metrics (typically 3-5 variables)
- 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"*

# 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
- **Comprehensive**: 20+ exposure metrics vs. 3-5 traditional

  *"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

# Four ER Datasets: A Comparison

| Aspect | ADER | ADEE | ADES | ADTRR |
|---|---|---|---|---|
| **Purpose** | Exposure foundation | Time-to-event | Safety events | Tumor response |
| **Structure** | Subject-level | BDS (one/subject/param) | BDS + OCCDS hybrid | BDS with visits |
| **Key Variables** | 20 exposure metrics | AVAL, CNSR, EVENT | TEAE, TEAESEV, event details | AVAL, BASE, CHG, PCHG |
| **Unique Features** | Comprehensive exposure | Censoring handling | Multiple analysis levels | RECIST 1.1, BOR |
| **Variables** | 65 | 74 | 78 | 80 |

# ADER: Exposure Foundation

**Core Exposure Metrics (20 variables)**

**Raw Metrics** - AUCSS: Steady-state AUC - CMAXSS: Steady-state Cmax - CAVGSS: Steady-state Cavg

**Transformed** - AUCSLOG: log(AUCSS) - CMXSLOG: log(CMAXSS) - CAVGLOG: log(CAVGSS)

**Standardized** - AUCSSSTD: Z-score - CMXSSSTD: Z-score

**Normalized** - AUCSSN: AUC/mean(AUC) - CMAXSSN: Cmax/mean(Cmax)

**Dose-Normalized** - AUCSSDOS: AUC/Dose - CMXSSDOS: Cmax/Dose

**Categorical** - AUCSSCAT: Tertiles/quartiles - AUCSCATN: Numeric category

# Variable Naming: 8-Character Compliance

**CDISC ADaM Requirement**: Variable names  8 characters

**Our Naming Strategy**

| Traditional | Updated ( 8 char) | Rationale |
|---|---|---|
| AUCSSLOG | **AUCSLOG** | Removed one "S" |
| CMAXSSLOG | **CMXSLOG** | Abbreviated CMAX |
| CAVGSSLOG | **CAVGLOG** | Removed SS |
| CMAXSSSTD | **CMXSSSTD** | Abbreviated CMAX |
| AUCSSDOSE | **AUCSSDOS** | Removed E |
| CMAXSSDOSE | **CMXSSDOS** | Abbreviated + removed E |
| AUCSSCATN | **AUCSCATN** | Removed one S |

# Domain 1: Exposure-Efficacy (ADEE)

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

**Key Dataset Features**

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

```
adee %>%
  select(USUBJID, PARAMCD, AVAL, CNSR, EVENT, AUCSS, AUCSSC
  head(3)
```

# ADEE: 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 (in ADER/exposure derivation)
exposure_cats <- adsl %>%
  mutate(
    AUCSSCAT = cut(AUCSS,
```

# ADEE: Analysis-Ready Output

```
# ADEE structure (74 variables)
USUBJID | PARAMCD | AVAL | AVALU | CNSR | EVENT | AUCSS | A
--------|---------|------|-------|------|-------|-------|--
001-001 | PFS     | 104  | DAYS  | 1    | 0     | 450.2 | 6
001-002 | PFS     | 135  | DAYS  | 0    | 1     | 520.5 | 6
001-003 | PFS     | 213  | DAYS  | 0    | 1     | 680.1 | 6
```

**Ready for modeling:**

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

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

# Domain 2: Exposure-Safety (ADES)

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

**Unique Challenges**

- Multiple analysis levels (subject, event, parameter)
- Time-varying exposure considerations
- Severity tracking (ASEV vs. AETOXGR)
- Need both counts and rates

**Dataset Structure (78 variables)**

1. **Subject-level parameters**: Overall AE burden
2. **Event-level records**: Individual AE occurrences with exposure context

# ADES: Subject-Level Parameters

```r
# Subject-level AE summary parameters
ades_subj <- adsl %>%
  left_join(ae_summary, by = "USUBJID") %>%
  mutate(
    PARAMCD = "TEAE",
    PARAM = "Treatment-Emergent Adverse Events",
    AVAL = n_aes, # Count of AEs
    AVALC = as.character(n_aes)
  ) %>%
  bind_rows(
    # Additional parameters: TEAESEV, TESAE, etc.
    create_param("TEAESEV", "Treatment-Emergent Severe AEs"
    create_param("TESAE", "Treatment-Emergent Serious AEs",
  )
```

# ADES: Severity Variables

**Important**: ADES uses `ASEV`/`ASEVN` (not `AETOXGR`/`AETOXGRN`)

| Variable | Values | Description |
|---|---|---|
| **ASEV** | MILD, MODERATE, SEVERE | Severity (character) |
| **ASEVN** | 1, 2, 3 | Severity (numeric) |
| **AESER** | Y, N | Serious flag |
| **AEREL** | NOT RELATED, UNLIKELY, POSSIBLE, PROBABLE, RELATED | Relationship |
| **AERELN** | 0, 1, 2, 3, 4 | Relationship (numeric) |

**Key Parameters**

# ADES: Event-Level Records

```
# Event-level dataset structure
USUBJID | AESEQ | AEDECOD     | ASEV      | ASEVN | AEREL
--------|-------|-------------|-----------|-------|----------
001-001 | 1     | NAUSEA      | MILD      | 1     | POSSIBLE
001-001 | 2     | FATIGUE     | MODERATE  | 2     | PROBABLE
001-002 | 1     | NEUTROPENIA | SEVERE    | 3     | RELATED
```

**Enables analyses like:**

- Time to first Severe AE (ASEVN $= 3$)
- Recurring event models
- Exposure-toxicity relationships by AE type
- Relationship analysis (AERELN as continuous)

# Domain 3: Tumor Response (ADTRR)

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

**Key Features (80 variables)**

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

**Analysis Approaches**

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

# ADTRR: Longitudinal Structure

```r
adtr <- tr_raw %>%
  mutate(
    PARAMCD = "TSIZE",
    PARAM = "Target Lesion Size",
    PARAMN = 1,
    AVAL = TRORRES, # Raw measurement (mm)
    AVALU = "mm",
    ADY = as.numeric(ADT - TRTSDT) + 1
  ) %>%
  # Identify baseline
  group_by(USUBJID, PARAMCD) %>%
  mutate(
    ABLFL = if_else(AVISITN == 1, "Y", ""),
    BASE = first(AVAL[ABLFL == "Y"])
  ) %>%
  # Calculate changes
  mutate(
    CHG = if_else(ABLFL != "Y", AVAL - BASE, NA_real_),
```

# ADTRR: Additional Parameters

## Beyond Target Lesion Size

```r
# Best Overall Response (BOR)
adtrr_bor <- adtrr %>%
  group_by(USUBJID) %>%
  summarise(
    PARAMCD = "BOR",
    PARAM = "Best Overall Response",
    PARAMN = 2,
    AVALC = derive_best_response(AVALC),
    BORN = case_when(
      AVALC == "CR" ~ 4,
      AVALC == "PR" ~ 3,
      AVALC == "SD" ~ 2,
      AVALC == "PD" ~ 1
    )
  )

# Nadir (minimum tumor size)
```

# RECIST Criteria Implementation

```r
adtrr <- adtrr %>%
  mutate(
    # RECIST 1.1 response at each visit
    AVALC = case_when(
      ABLFL == "Y" ~ "BASELINE",
      AVAL == 0 ~ "CR", # Complete Response
      PCHG <= -30 ~ "PR", # Partial Response (30% decrease
      PCHG >= 20 & CHG >= 5 ~ "PD", # Progressive Disease (
      TRUE ~ "SD" # Stable Disease
    )
  )

# Best Overall Response across all visits
bor <- adtrr %>%
  filter(ABLFL != "Y") %>%
  group_by(USUBJID) %>%
  summarise(
    BOR = case_when(
```

# ADTRR: Output Structure

```
# Longitudinal measurements (80 variables)
USUBJID | PARAMCD | AVISITN | AVISIT   | BASE | AVAL | CHG
--------|---------|---------|---------|------|------|-----
001-002 | TSIZE   | 1       | BASE    | 65.0 | 65.0 | NA
001-002 | TSIZE   | 2       | WEEK 6  | 65.0 | 55.0 | -10
001-002 | TSIZE   | 3       | WEEK 12 | 65.0 | 35.0 | -30
001-002 | BOR     | 99      | OVERALL | NA   | NA   | NA
001-002 | NADIR   | 99      | OVERALL | 65.0 | 35.0 | -30
```

**Each subject has:**

- Longitudinal trajectory (TSIZE records)
- Best overall response (BOR parameter)
- Nadir tumor size (NADIR parameter)
- Exposure metrics on all records

# Common Patterns Across Datasets

**Time Variables**

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

**Exposure Metrics (20 variables)**

- Raw: AUCSS, CMAXSS, CAVGSS
- Transformed: AUCSLOG, CMXSLOG
- Standardized: AUCSSSTD, CMXSSSTD
- Categorical: AUCSSCAT, AUCSCATN

**Traceability**

**Baseline Covariates (13 variables)**

- Demographics: AGE, SEX, RACE
- Vitals: WTBL, HTBL, BMIBL, BSA
- Renal: CREATBL, CRCLBL, EGFRBL
- Hepatic: ALTBL, ASTBL, TBILBL

**Analysis Flags**

- ANL01FL: Primary analysis
- ANL02FL: Sensitivity analyses
- Domain-specific flags

# 8-Character Compliance Summary

**All Variable Names 8 Characters**

| Category | Examples | Count |
|---|---|---|
| **Exposure (raw)** | AUCSS, CMAXSS, CAVGSS | 3 |
| **Exposure (log)** | AUCSLOG, CMXSLOG, CAVGLOG | 3 |
| **Exposure (std)** | AUCSSTD, CMXSSTD | 2 |
| **Exposure (norm)** | AUCSSN, CMAXSSN | 2 |
| **Exposure (dose)** | AUCSSDOS, CMXSSDOS | 2 |
| **Exposure (cat)** | AUCSSCAT, AUCSCATN | 2 |
| **Covariates** | CRCLBL, EGFRBL, BMIBL | 13 |
| **Response** | NADPCHG, NADVST | 2 |

**Total**: 20 exposure metrics + 13 covariates across all datasets

# Why {admiral}?

## Advantages for ER Data

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

## Example Extension

```
# Hypothetical future function
derive_param_exposure_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_exposure_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 {metacore}, {metatools}, {xportr}

### Example Extension

```
# Hypothetical future function
derive_param_exposure_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 {metacore}, {metatools}, {xportr}
- **Community-driven**: Pharmaverse ecosystem

### Example Extension

```
# Hypothetical future function
derive_param_exposure_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 {metacore}, {metatools}, {xportr}
- **Community-driven**: Pharmaverse ecosystem
- **Well-documented**: Functions follow consistent patterns

## Example Extension

```
# Hypothetical future function
derive_param_exposure_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 {metacore}, {metatools}, {xportr}
- **Community-driven**: Pharmaverse ecosystem
- **Well-documented**: Functions follow consistent patterns
- **CDISC-aligned**: Built with standards in mind

## Example Extension

```
# Hypothetical future function
derive_param_exposure_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 {metacore}, {metatools}, {xportr}
- **Community-driven**: Pharmaverse ecosystem
- **Well-documented**: Functions follow consistent patterns
- **CDISC-aligned**: Built with standards in mind
- **8-character compliance**: Follows naming conventions

## Example Extension

```
# Hypothetical future function
derive_param_exposure_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$AUC, breaks=c(0,500,1000,Inf))
data$log_AUC <- log(data$AUC)

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

**After Standardization**

```
# All studies use consistent approach and naming
adee <- derive_er_tte_data(
  adsl = adsl,
  adrs = adrs,
  exposure_vars = c("AUCSS", "AUCSLOG", "AUCSSSTD", "AUCSSC
  param = "PFS"
```

# Benefits: Comprehensive Exposure Metrics

**Traditional Approach**: 3-5 exposure variables

```
# Typical traditional dataset
c("AUC", "CMAX", "CAVG")
```

**New Framework**: 20 standardized exposure variables

```
# All transformations pre-computed
exposure_vars <- c(
  # Raw (3)
  "AUCSS", "CMAXSS", "CAVGSS",
  # Log-transformed (3)
  "AUCSLOG", "CMXSLOG", "CAVGLOG",
  # Standardized (3)
  "AUCSSSTD", "CMXSSSTD", "CAVGSTD",
  # Normalized (3)
  "AUCSSN", "CMAXSSN", "CAVGSSN",
  # Dose-normalized (3)
  "AUCSSDOS", "CMXSSDOS", "CAVGDOS",
  # Categorical (2)
```

# 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
- Analysis: No data prep needed

**Quality Improvements**

- Fewer manual errors (automated derivations)
- Consistent validation approach
- Peer review easier (standard structure)
- Regulatory submissions smoother
- 8-character compliance built-in

**Real Impact**

*"Standardized E-R datasets reduced programming time from 3 weeks to 1 week per study while improving quality"*

# Path Forward: Community Adoption

## What We Need

1. **Feedback on proposed structures**
   - Variable naming conventions (especially 8-char compliance)
   - 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 with code
   - 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 (simulated data)
- Complete derivation scripts (ADER, ADEE, ADES, ADTRR)
- ADaM specifications (P21 format + Excel)
- Metacore objects for metadata
- Presentation materials

## Try It Yourself

```
# Clone repository
git clone https://github.com/jeffreyad/er-standards.git
cd er-standards

# Generate example data
source("data-raw/S0_Generate_Example_Data.R")

# Run ADaM derivations
source("programs/run_all_adams.R")
```

# Key Takeaways

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

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)
   - ADES: Exposure-Safety events (78 variables)

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)
   - ADES: Exposure-Safety events (78 variables)
   - ADTRR: Tumor response longitudinal (80 variables)

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)
   - ADES: Exposure-Safety events (78 variables)
   - ADTRR: Tumor response longitudinal (80 variables)
4. **Comprehensive exposure coverage** - 20 metrics vs. traditional 3-5

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)
   - ADES: Exposure-Safety events (78 variables)
   - ADTRR: Tumor response longitudinal (80 variables)
4. **Comprehensive exposure coverage** - 20 metrics vs. traditional 3-5
5. **{admiral} is well-suited** - modular, validated, community-supported

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)
   - ADES: Exposure-Safety events (78 variables)
   - ADTRR: Tumor response longitudinal (80 variables)
4. **Comprehensive exposure coverage** - 20 metrics vs. traditional 3-5
5. **{admiral} is well-suited** - modular, validated, community-supported
6. **Benefits are real** - efficiency, reproducibility, quality, compliance

# Key Takeaways

1. **ER data needs standardization** - current state is inconsistent and limited
2. **CDISC principles apply** - we can extend SDTM-PK patterns with 8-character compliance
3. **Four specialized datasets** - each optimized for specific analyses
   - ADER: Exposure foundation (20 metrics, 65 variables)
   - ADEE: Exposure-Efficacy time-to-event (74 variables)
   - ADES: Exposure-Safety events (78 variables)
   - ADTRR: Tumor response longitudinal (80 variables)
4. **Comprehensive exposure coverage** - 20 metrics vs. traditional 3-5
5. **{admiral} is well-suited** - modular, validated, community-supported
6. **Benefits are real** - efficiency, reproducibility, quality, compliance

# 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: Complete Variable List (ADER)

**ADER: 65 Variables**

**Exposure Metrics (20)** - AUCSS, CMAXSS, CAVGSS - AUCSLOG, CMXSLOG, CAVGLOG - AUCSSSTD, CMXSSSTD, CAVGSTD - AUCSSN, CMAXSSN, CAVGSSN - AUCSSDOS, CMXSSDOS, CAVGDOS - AUCSSCAT, AUCSCATN - CMXSSCAT, CMXSCATN

**Baseline Covariates (13)** - WTBL, HTBL, BMIBL, BSA - CREATBL, CRCLBL, EGFRBL - ALTBL, ASTBL, TBILBL - ALBBL - ECOGBL, SMOKEBL

**Demographics (8)** - AGE, AGEGR1, AGEGR1N - SEX, SEXN - RACE, RACEN - ETHNIC, ETHNICN

**Treatment (8)** - ARM, ARMN, ARMCD - TRT01P, TRT01PN - TRT01A, TRT01AN - DOSE

**Other (16)** - Identifiers: STUDYID, USUBJID, SUBJID, etc. - Dates: TRTSDT, TRTEDT, TRTDURD - Flags: SAFFL, ITTFL, EFFFL

# 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 |

## Our Extensions

- Added comprehensive exposure transformations
- Categorical exposure variables
- Baseline covariate integration
- Domain-specific parameters

# Backup: Detailed Variable Specifications

**Proposed Core Variables for ADEE (74 total)**

| Variable | Type | Description | Required? | Notes |
|---|---|---|---|---|
| PARAMCD | Char | Parameter Code (PFS, OS, TTP, TTNT) | Yes | Standard codes |
| AVAL | Num | Analysis Value (days) | Yes | Continuous outcome |
| CNSR | Num | Censoring (1=censored, 0=event) | Yes | Standard TTE |
| EVENT | Num | Event (1=event, 0=censored) | Yes | For modeling |

# Backup: Handling Edge Cases

## Common Challenges

1. **Missing Exposure Data**
   - Imputation strategies
   - Sensitivity analyses
   - Flagging for exclusion (ANL01FL = "N")
2. **Competing Risks** (ADEE)
   - Death vs. progression
   - Multiple event types as separate parameters
3. **Time-Varying Exposure** (ADES)
   - Dose changes over time
   - Cumulative exposure metrics
   - Landmarking approaches
4. **Unscheduled Assessments** (ADTRR)
   - Off-schedule scans
   - Response confirmation windows
   - Handling missing visits
5. **8-Character Limit Conflicts**
   - Systematic abbreviation rules
   - Document in specifications

# Backup: Variable Name Abbreviation Rules

**Systematic Approach to 8-Character Compliance**

1. **Remove vowels from middle** (if >8 char)
   - BASELINE → BASELN (not used, but example)
2. **Abbreviate common terms**
   - MAXIMUM → MAX → MAX
   - AVERAGE → AVG → AVG
   - CATEGORY → CAT
3. **Remove repeated letters**
   - AUCSSLOG → AUCSLOG (one S removed)
   - AUCSSCATN → AUCSCATN (one S removed)
4. **Abbreviate exposure terms**
   - CMAX → CMX (when combined with other suffixes)
   - STEADY STATE → SS
5. **Remove last vowel from suffix**
   - DOSE → DOS (when combined: AUCSSDOS)

**Documentation**: All abbreviations documented in define.xml and specifications

# Backup: Comparison with Traditional

### Traditional Single-File Approach

```
# All data in one wide row per subject
C | PTNM | TRT | AUC | CMAX | OS | PFS | NADIR | BOR | TEAE
```

**Issues**: - Non-standard variable names - Limited exposure metrics (3-5 variables) - Hard to extend with new parameters - No transformation variants - Mixed parameter types

### New Framework (4 Datasets)

```
# ADER: 20 exposure metrics
# ADEE: Time-to-event + exposure (74 vars)
# ADES: Safety + exposure (78 vars)
# ADTRR: Tumor + exposure (80 vars)
```

**Benefits**: - CDISC-compliant (8-char limit) - Comprehensive exposure (20 metrics) - Flexible parameter addition - Pre-computed transformations - Domain-optimized structures

# 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.