PharmaSUG 2025 - Paper ST-189

Deciphering Exposure-Response Analysis Datasets: A Programmer's Perspective for Oncology Studies

Sabarinath Sundaram, Pfizer Inc. USA

ABSTRACT

Exposure-Response (E–R) evaluation is essential in drug development and regulatory reviews by informing decision-making towards optimized trial design, dose and regimen selection, and benefit-risk assessments in both early and late-stage trials. Analyzing the relationship between drug exposure and treatment outcomes using E-R data provides a level of granularity to support the primary evidence of a drug's safety (identifying negative effects) and/or efficacy (positive effects). The preparation of high-quality E-R datasets is a key step in this space, which can get challenging especially in oncology studies which are quite complex and involve multiple factors and mechanisms.

The creation of these E-R analysis datasets requires a comprehensive mix of data sources, including drug exposure, patient demographics, key covariates, PK/PD data, and key primary or secondary endpoints of the clinical trial. This intricate process demands strong programming expertise and a deep understanding of PK and PD, as it requires ongoing collaboration with PK modeling scientists to ensure accurate and meaningful analysis.

This paper will explore the role of E-R analysis datasets in regulatory submissions, address key challenges in their creation, and examine the FDA's guidance on E-R analysis. We will also discuss the development of ADaM standard E-R datasets, present masked dummy data and models to illustrate the practical application of E-R analyses. Ultimately, this paper emphasizes the importance of E-R evaluations in advancing drug development and optimizing therapeutic outcomes.

INTRODUCTION

Clinical trials are conducted to assess treatment efficacy and toxicity. Once the participant gets dosed, something happens, and we get results. The two main events (Figure 1) need to be dissected in detail to understand more about the drug and endpoint relationship.

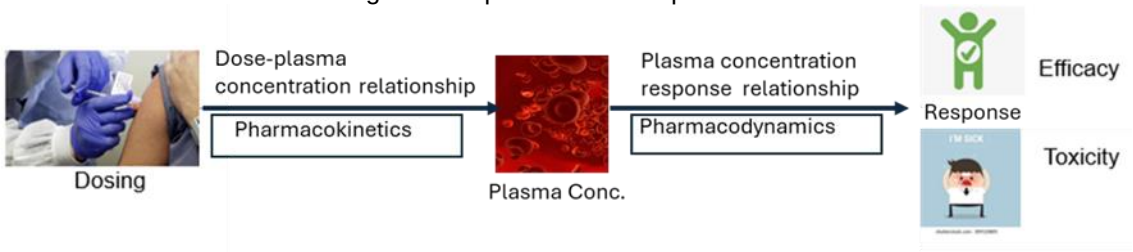

Figure 1: Overview of the PK and PD in a Clinical Trial.

Pharmacokinetics (PK) is defined as the study of the time course of drug absorption, distribution, metabolism, and excretion. It is the study of how the body interacts with administered drug for the entire duration of exposure and PK analysis is essential to fully investigate the pharmacokinetics of a new drug. Pharmacodynamics (PD) is the body's biological response to drugs and PD analysis allows us to quantify the relationship between the drug dose and the pharmacologic or toxicologic effect it has on patients. Exposure typically refers to the concentration of the drug in the bloodstream, while response could refer to efficacy or safety measures, such as symptom improvement or adverse events (AE). Exposure-response (E-R) analysis plays a critical role in understanding the relationship between a drug's pharmacokinetic exposure and the observed clinical responses in patients. Such analyses are pivotal in clinical research for determining appropriate dosing regimens, dose selection and optimization, and understanding efficacy or safety profiles and regulatory decisions during all stages of drug development (Dai *et al.,* 2020). In the context of SAS programming, creating and analyzing an E-R dataset involves

data preparation, integration of PK and PD data, and statistical modeling. In oncology, exposure-response analysis helps establish relationships between drug exposure (e.g., plasma concentration) and clinical response (e.g., tumor shrinkage, progression-free survival) (Wang *et al.,* 2020). Due to the complexity of cancer treatment, modeling approaches must consider interpatient variability, time-dependent effects, and toxicity risks. A well-structured programming approach ensures efficient data handling, preprocessing, and modeling. In clinical trials, a robust dataset is required to understand these relationships, identify optimal dosing regimens, and assess the therapeutic window of a drug.

EXPOSURE-RESPONSE DATA PROGRAMMING

E-R analysis data programming is the process of deriving a data-ready file based on clinical source data files, typically CDISC data sets such as SDTM and ADaM. The data definition table (DDT), the required variable list provided by the clinical pharmacologist, should be reviewed thoroughly by the programmer who also adds the necessary derivations for these specifications using the trial's ADaM DDT specification. We have standardized this process and have a basic specification (E-R DDT) for efficiency and accuracy of the E-R analysis data set. This E-R DDT contains detailed information about how the E-R data file shall be derived and consists of a complete list of required variables and their definitions. The DDT serves as the programming specification and to clarify the data file for the reviewers. This complete process requires high collaboration and information exchange between various functional teams such as Clinical Pharmacology, Biostatistics, sometimes the medical monitor, and our programming team. Deep understanding of the source data and PK analysis is necessary for this, along with effective communication.

EXPOSURE-RESPONSE INPUT DATA SET FLOW CHART

An E-R statistical analysis plan might be available only at a later stage after the clinical study report (CSR) programming and the statistical analysis plan are finalized. The clinical pharmacologists provide the lead statistical programmer with the draft scope for the E-R analysis. Using that, statistical programmer updates a study-specific copy of the standard specification based on the scope and creates the draft data for dry-run (or "test-the-waters") analysis. The E-R SAP may then be revisited based on the output and the potential models. The E-R-SAP may be evolving until the due date for the deliverables. The timeline for this analysis will usually be tight, so the statistical programmer working on creating E-R datasets needs to understand the challenges and importance of the analysis along with a clear understanding of PK/PD and the study protocol.

The study design and E-R-specific analysis requirements result in incorporation of various baseline variables/covariates, demographics, vital signs, exposure, biomarker data, response data, etc. Additional data records such as baseline labs, concomitant medications, AE, etc., can also be added if requested by the clinical pharmacologist for additional analysis. The clinical pharmacologist may request an E-R-specific analysis data set for an individual study or one that integrates several studies within a product. The construction of a final ADER dataset needs ADSL, ADPPK, safety datasets (e.g., ADAE, ADLB, ADVS), efficacy datasets (e.g., ADRS, ADTTE), and baseline datasets (e.g., ADBASE) as input (Figure 2). The exposure response (ADER) analysis dataset can be created for each study first and then combined for all studies from there.

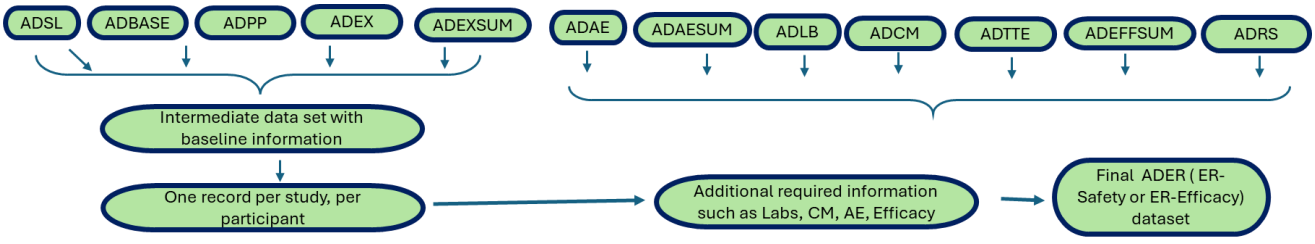

Figure 2: E-R (ADER) Input Data Set Flowchart.

Properly merging and organizing these data sources ensures that the dataset is suitable for statistical

analysis and modeling. A systematic approach to manage and analyze E-R data begins with documenting the data integration, data transfer, and validation processes.

(a)

Variable	Label	Data Type
SUBJID	Subject Identifier	text
SAFFL	Safety Population Flag	text
PKFL	PK Population Flag	text
RSBPOFL	Response Evaluable Pop. Flag per BICR	text
ERAPOP	ER Analysis Population Flag	text
RDI	Relative Dose Intensity (%)	float
EEFL	Efficacy Evaluble Population Flag	text
ERAPOP	ER Analysis Population Flag	text
IDI	Intended Dose Intensity (mg/kg/3wk)	float
RDI	Relative Dose Intensity (%)	float
DUR	Treatment Duration (Months)	float
HBA1C	Baseline HbA1c (%)	float
ECOGBL	Baseline ECOG Status	text
AGRHYGLY	All Grade Hyperglycemia During Treatment (0 No, 1 Yes)	integer
TAGHYGLY	Time to First All Grade Hyperglycemia (Months)	integer
TEAEDA	Dose Adjustments Due to All Grade TEAEs - 1 (Yes)/0 (No)	integer
TTDATE	Time to First Dose Adjustments Due to All Grade TEAEs - Months Relative to First Dose	integer
EVCAV1	ADC Cavg Cycle 1	text
MCAV1	MMAE Cavg Cycle 1	text
EVCAV3	ADC Cavg Cycle 3	text

(b)

Variable	Label	Data Type
SUBJID	Subject Identifier	text
SAFFL	Safety Population Flag	text
PKFL	PK Population Flag	text
EEFL	Efficacy Evaluable Flag	text
ERAPOP	ER Analysis Population Flag	text
ADI	Absolute Dose Intensity (mg/kg/3wk)	float
IDI	Intended Dose Intensity (mg/kg/3wk)	float
RDI	Relative Dose Intensity (%)	float
HBA1CBL	Baseline HbA1c (%)	float
ECOGBL	Baseline ECOG Status	text
TTORR	Time to ORR	float
DOR	Duration of Response	float
CNSROS	Censor for OS	integer
EDURPFS	Duration of Exposure up to Either FIRST Disease Progression or DEATH in Months	float
EVCAV1	ADC Cavg Cycle 1	text
MCAV1	MMAE Cavg Cycle 1	text
EVCAV3	ADC Cavg Cycle 3	text

Table 1: a) Sample E-R safety analysis dataset specification containing baseline and exposure variables, and time-to-event endpoints for AEs of interest plus relevant pharmacokinetic variables. **b)** Sample E-R efficacy analysis dataset specification containing baseline and exposure variables, and efficacy endpoints plus relevant pharmacokinetic variables.

DATA SET STRUCTURE

The E-R input data set is typically a subject-level dataset with one row per study, per subject. Three basic components are dosing records, PK/PD results, and covariates (PD baseline and/or demographics covariates). This integrates data from various sources, including pharmacokinetics, pharmacodynamics, efficacy endpoints, safety data, and patient demographics. The main components typically include:

a) Exposure variables, which are derived from pharmacokinetic data, such as area under the concentration-time curve (AUC), maximum concentration (Cmax), and C trough concentration.
b) Patient information, which includes unique patient identifiers, demographics (age, weight, sex, etc.), and baseline characteristics.
c) Covariates, which are the factors that may influence exposure or response, such as renal function, hepatic function, or concomitant medications.
d) Time variables, which refer to time since dose, sampling time, or other temporal factors relevant to exposure and response.
e) Response variables, which are safety endpoints such as AE incidence and lab abnormalities.

For E-R analysis, the response variable could be continuous, categorical, or time-to-event data. Efficacy outcomes may include change in disease-specific biomarkers and clinical scores. For event analyses, whether logistic regression with binary endpoints (e.g., objective response rate for efficacy or treatment-emergent adverse events (TEAEs) for safety) or survival analysis with time-to-event (TTE) endpoints (e.g., progression-free survival), the choice and derivation of exposure metric may influence key decisions during E-R model development. Typically, several exposure metrics are selected when investigating E-R relationships. The most common exposure metrics that are suggested in regulatory guidance documents (Overgaard *et al.,* 2015; Food and Drug Administration, 2003) include the maximum concentration, minimum concentration, and area under the concentration-time curve (AUC) after the first dose or cycle 1 (e.g., for oncology drugs).

Case I)

		a							b					c				d		
	SUBJID	SAFFL	ITTFL	PKFL	RSBPOFL	ERAPOP	BLHT	SEX	ADI	IDI	RDI	HBA1CBL	ECOGBL	TTORR	DOR	DCTRL	CNSROS	EVCAV1	MCAV1	EVCAV3
1	10001-	Y	Y	Y	Y	1	178	M	1.57669	2.5	63.06773	5.5	1	-99	-99	N	0	2.7	4.2	1.37
2	10001-	Y	Y	Y	Y	1	150.5	F	2.19911	2.5	87.96475	-99	2	2.07	4.17	Y	0	3.19	1.32	3.33
3	10005-	Y	Y	Y	Y	1	169.6	M	1.25186	2.5	50.07451	5.9	2	-99	-99	N	0	1.45	1.55	1.45
4	10005-	Y	Y	Y	Y	1	193.6	M	2.31617	2.5	92.64705	5.8	0	2.27	13.9	Y	1	3.63	1.21	3.57
5	10005-	Y	Y	Y	Y	1	185.6	M	1.54655	2.5	61.86216	-99	0	1.91	6.51	Y	1	3.16	1.44	3.19

Case 2)

		a						b					e				d		
	SUBJID	SAFFL	PKFL	EEFL	ERAPOP	BLBSA	BMI	AGE	IDI	RDI	DUR	HBA1C	ECOGBL	AGRHYGLY	TAGHYGLY	TEAEDA	TTDATE	EVCAV1	MCAV1	EVCAV3
1	10001-	Y	Y		0	2.06	36.42	52	2.5	100.01	8.2793	6.1	1	0	-99	0	-99	NA	NA	NA
2	10001-	Y	Y		1	1.98	25.09	65	2.5	63.067	4.1068	5.5	1	0	-99	1	0.7885	2.7	4.2	1.37
3	10001-	Y	Y		1	1.24	16.11	66	2.5	87.964	8.5749	-99	2	0	-99	0	-99	3.19	1.32	3.33
4	10001-	Y	Y	Y	1	1.94	23.15	65	2.5	94.073	6.2094	5.9	1	0	-99	1	3.0554	2.82	6.4	2.86
5	10001-	Y	Y	Y	1	2.01	37.53	82	2.5	54.039	3.6797	5.9	0	1	3.5811	1	0.2628	1.96	3.51	2.63

Case I is an example of a typical E-R-efficacy analysis dataset and **Case 2** is an example of a typical E-R-safety analysis dataset. a. population flag, b. exposure variable, c. efficacy endpoints, d. pharmacokinetic variables, and e. AE of special interest (AESI).

POST-PROCESSING STEPS AFTER CREATING E-R DATA SETS

Understanding the post-processing steps helps programmers better understand the E-R data set. Model-based exposure endpoints are often utilized for subject-level exposure derivation in E-R analyses. Note that exposure data may not be available from all patients depending on the trial design, and E-R population may be a subset of the full analysis set. Multiple exposure metrics may be evaluated to obtain the best PK driver for response.

If the exposure variable is model-derived, a round of iteration may occur between the pharmacometrician and the programmer. Individual PK exposure metrics of interest (e.g., Cmax, AUC, Ctrough, Cavg) will be integrated into the E-R dataset typically in a wide data format. In general, the response variables in E-R analysis are consistent with the pre-specified study endpoints and can be readily taken from ADaM datasets developed for the CSR. However, there are cases where further derivations are needed. Examples include a categorical variable that may be mapped into a binary variable to facilitate logistic regression analysis. In cases where biomarkers are considered as an acceptable surrogate endpoint for efficacy/safety, inclusion of such data may add to the weight of evidence, and communication between the statistical programming lead and biomarker group is key.

Once the E-R dataset is complete, E-R analyses are performed. A common type of E-R analysis is based on primary efficacy endpoints and safety endpoints of interest to support the proposed dose regimen (Figure 3).

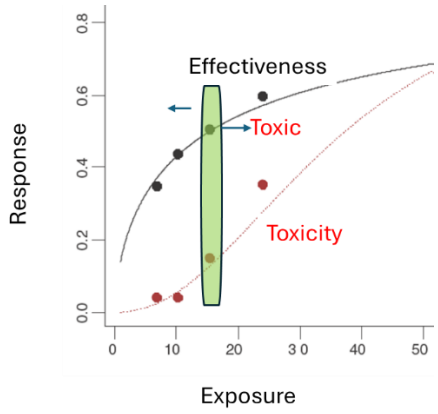

Figure 3: Dose-Response Curve. A graph demonstrating how changes in drug dose (x-axis) influence the therapeutic response (y-axis). This helps in determining the optimal dosing regimen.

While demonstrating significant benefit in progression-free survival, objective response rate, and overall survival, sometimes AESI are identified as the most common reasons for dose reduction and discontinuation. E-R analysis can be used as a tool to balance out safety and efficacy and identify an optimal therapeutic window. Efficacy (OR) and safety (AESI) response curves can be overlaid to visualize this therapeutic window [purple in figure 4].

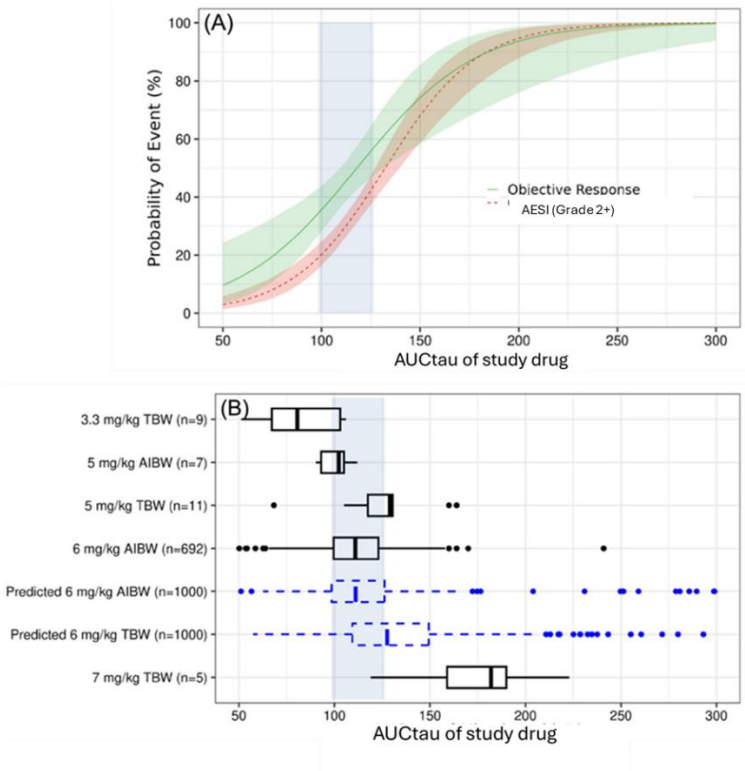

Figure 4: Dose-Response Relationship. (A) Model-predicted relationships between AUCtau of study drug and responses in efficacy (objective response rate) and safety (grade ≥ 2 AESI). (B) Box plots of AUCtau for various dose regimens. The vertical gray band represents the interquartile range of AUCtau predicted for 6 mg/kg AIBW (n = 1000). AE, adverse event; AIBW, adjusted ideal body weight; AUCtau, area under the concentration–time curve over the dosing interval (21 days); TBW, total body weight.

REGULATORY GUIDANCE

The U.S. Food and Drug Administration (FDA) provides guidance on the role of E-R analysis in assessing drug efficacy and safety, optimizing dosing strategies, and informing labeling decisions. The FDA has issued several guidance documents emphasizing the importance of E-R analysis. The primary reference is the guidance titled Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications, which outlines best practices for conducting and interpreting E-R analyses (FDA, 2003). Additional guidance documents, such as those on population pharmacokinetics and model-informed drug development (MIDD) (FDA, 2018), further support the integration of E-R findings in regulatory submissions. FDA recommends using nonlinear mixed-effects modeling (e.g., NONMEM, Monolix) and employing Bayesian or frequentist approaches to quantify variability and uncertainty in E-R relationships. The FDA's guidance provides a structured approach to designing, analyzing, and applying E-R models to optimize drug therapy, improve patient outcomes, and streamline regulatory approval processes.

E-R analysis helps determine the optimal dose that maximizes efficacy while minimizing toxicity. It supports dose adjustments for special populations, such as pediatric or renal impaired patients. By

characterizing the relationship between drug exposure and adverse events, E-R models contribute to a comprehensive benefit-risk evaluation. E-R analysis aids in extrapolating efficacy and safety data from one population to another, reducing the need for extensive clinical trials. The FDA relies on E-R data to inform dosing recommendations and justify label claims.

CONCLUSION

Statistical programmers create datasets that support analyses which help interpret the outcome of clinical trials on investigational products. The quality and accuracy of the E-R analysis data is critical, as the exposure-response analysis provides knowledge of the relationship between exposure, efficacy, and toxicity. Basic knowledge about PK/PD plus understanding the clinical trial design and background knowledge in safety and efficacy data are always an added advantage for statistical programmers working on creating ADER data. In addition, extensive cross-functional collaboration between Statistical Programmers, Biostatisticians, Clinical Pharmacologists, and Medical Monitors is needed to understand the modeling purpose, which helps provide the required data in E-R data set format. This paper covered the importance of E-R analysis and the efficient way of creating its dataset.

REFERENCES

1. U.S. Food and Drug Administration. (2003). *Guidance for Industry: Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications*. https://www.fda.gov/media/71277/download
2. U.S. Food and Drug Administration. (2018). *Model-Informed Drug Development Pilot Program*. https://www.fda.gov/drugs/development-resources/model-informed-drug-development-pilot-program
3. Dai *et al.,* (2020). Characterizing Exposure–Response Relationship for Therapeutic Monoclonal Antibodies in Immuno-Oncology and Beyond: Challenges, Perspectives, and Prospects. Clin Pharmacol Ther. 2020 Aug 2;108(6):1156–1170.
4. Wang, B *et al.,* (2022). Population pharmacokinetic and exploratory exposure-response analysis of the fixed-dose combination of pertuzumab and trastuzumab for subcutaneous injection in patients with HER2-positive early breast cancer in the FeDeriCa study. *Cancer Chemother. Pharmacol.* 89 (4), 499–512.
5. Overgaard, R. V., *et al.,* (2015). Establishing good practices for exposure-response analysis of clinical endpoints in drug development. *CPT Pharmacometrics Syst. Pharmacol.* 4 (10), 565–575.

ACKNOWLEDGMENTS

We would like to thank Faye Zhang, Shefalica Chand, and Michiel Hagendoorn for their valuable feedback, constant support, and guidance.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Sabarinath Sundaram
Pfizer Inc.
Sabarinath.sundaram@pfizer.com