Received: 28 September 2022 | Revised: 15 November 2022 | Accepted: 9 December 2022

DOI: 10.1002/psp4.12905

PERSPECTIVE

Variable or variate? A conundrum in pharmacometrics exposure–response models

Ana-Marija Grisic[1] | **Karthik Venkatakrishnan**[2] | **Jonathan French**[3] | **Akash Khandelwal**[1]

[1]The Healthcare Business of Merck KGaA, Darmstadt, Germany

[2]EMD Serono, Billerica, Massachusetts, USA

[3]Metrum Research Group, Tariffville, Connecticut, USA

Correspondence
Akash Khandelwal, The Healthcare Business of Merck KGaA, Darmstadt, Germany.
Email: akash.khandelwal@emdgroup.com

Abstract

Key elements of scientific writing—consistency and clarity—can be compromised in case of inaccurate use of methodological terms, especially in complex and multidisciplinary scientific fields. Such is the case in reports of pharmacometrics exposure–response analyses with the use of the terms *univariate*/*multivariate* and *univariable*/*multivariable*. This perspective outlines the issues in the use of these terms, clarifies their definitions, provides examples, and makes recommendations for authors, reviewers, and journals in the fields of clinical pharmacology and pharmacometrics.

Consistency and clarity are vital in scientific writing and communication, and a lack thereof can reduce the impact and reproducibility of the reported research. However, instances of incorrect or inconsistent use of methodological terms do occur, and when they are repeatedly published—and subsequently referenced—they generate confusion and contribute to miseducation. One such example calling for broader awareness and consensus is found in the reporting of the pharmacometrics exposure–response analyses with the use of the terms *univariate*/*multivariate* and *univariable*/*multivariable*.

The number of outcomes characterizes the model as univariate or multivariate; however, in the context of pharmacometrics models, usually the intent is to differentiate models based on the number of independent variables as univariable or multivariable. This perspective aims to address the issue of incorrectness and inconsistency in the use of these terms, provide clarity on their definitions and correct use, and call for incorporation of clear guidelines for authors and peer reviewers of scientific publications to mitigate further inconsistent and/or incorrect use with a focus on exposure–response models.

In exposure–response models, the left-hand side (LHS) terms, that is, "Y," are most commonly referred to as end points, outcomes, response, or dependent variables, whereas the right-hand side (RHS) terms, that is, "X," are referred to as predictors, covariates, or independent variables. Throughout this perspective, the following terms will be used: (1) *outcome* to denote dependent variable/end point/response variable, that is, LHS terms; and (2) *independent variable* to denote independent variable/covariate/predictor, that is, RHS terms.

Exposure–response analyses are routinely performed in drug development to investigate the link between independent variables (exposure metrics and covariates) and efficacy (exposure–efficacy) or safety (exposure–safety) outcomes. Methodologically, these often comprise logistic regression or time-to-event (aka, survival) Cox or parametric time-to-event models. An exposure–response

analysis typically includes a single outcome (e.g., therapeutic response to treatment, overall survival, occurrence of an adverse event)—it is univariate—and one or multiple independent variables, making it univariable or multivariable, respectively. Despite the seemingly straightforward definitions, these terms are often incorrectly used, whereby they are either used interchangeably or one of the words is incorrectly used instead of the other. In addition, "hybrid" terms (e.g., "bivariate" for a model with two variables in total; i.e., one independent and one outcome measure) also appear in the literature, potentially contributing to the unclarity.

This observation of the incorrect use of these terms has been reported previously. For example, Hidalgo and Goodman[1] underlined that "the terms multivariate and multivariable are often used interchangeably in the public health literature although these terms actually represent 2 very distinct types of analyses." The most common mistake is use of the term *uni-* or *multivariate* when the correct term would be *uni-* or *multivariable*. Reboldi et al.[2] noted that "the term 'multivariate analysis' is often used when one is referring to a multivariable analysis." In pharmacometrics models, this can, for example, be the use of the term *univariate* logistic regression for an exposure–response model with exposure as the sole independent variable and the description of a "multivariate" model when the logistic regression exposure–response model includes covariate effects to account for intrinsic or extrinsic factors. In these cases, it is clear from the text that the authors wanted to differentiate models by the number of independent variables, for which only the term *univariable* (or *multivariable* in the case of multiple independent variables) appropriately conveys the message.[3] In other instances, the terms *multivariable* and *multivariate* are used interchangeably.[4] There are also examples in which the terms *univariable* and *multivariable* are used appropriately as the authors wanted to differentiate models based on the numbers of independent variables and the model was "univariate" by definition, for example, a report of Cox regression progression-free survival analyses in non-small cell lung cancer patients treated with anaplastic lymphoma kinase inhibitors.[5] Similarly, Ogasawara et al.[6] appropriately reported a logistic regression analysis in lymphoma patients receiving a chimeric antigen receptor T cell therapy, wherein first the univariable analysis investigated the relationship between in vivo cellular expansion parameters and the probability of clinical outcomes (overall response, complete response, cytokine release syndrome, any-grade and grade≥3 neurological events), followed by multivariable analysis to control for potential confounders. In the following part, we clarify the correct use of these terms with regard to the outcome(s) and independent variable(s) and provide examples of the correctly classified analyses (Table 1). Of note, in all examples, the independent variables for longitudinal models are assumed to be observed at baseline or are model predicted in case of time varying to mitigate the impact of immortal time bias.[7]

A general guidance to keep in mind is that the number of outcomes in the model determines whether an analysis is univariate or multivariate, whereas the number of independent variables determines whether an analysis is univariable or multivariable. In the context of exposure–response analyses, examples of outcomes are overall response rate (ORR), overall survival, and progression-free survival in oncology, renal outcome in diabetes, annualized relapse rate in multiple sclerosis, need for intubation in coronavirus disease (COVID), and occurrence of an adverse event, whereas examples of independent variables are exposure metrics (e.g., minimum drug concentration [C_{min}], area under the curve) and covariates such as body weight, sex, baseline disease severity, and concomitant therapies.

The simplest example is a model that contains a single independent variable and a single outcome, for example, a parametric time-to event model linking C_{min} as an exposure metric to overall survival, or logistic regression model for predicting the need for intubation in COVID patients with vaccination status as an independent variable. These are univariate univariable analyses. If multiple independent variables are related simultaneously to one outcome, for example, a logistic regression model for ORR with exposure, performance status, and number of prior lines of treatment as independent variables, the analysis is univariate multivariable. As most of the confusion seems to be related to use of the terms *multivariate* and *multivariable*, we wish to reinforce that both analyses are univariate as they have only one outcome.

Multivariate models are those with multiple outcomes. The multiple outcomes may comprise multiple outcome measures or they can arise as repeated measurements of one outcome construct (e.g., an outcome measured at multiple timepoints), such as repeated time-to-event models, for example, for predicting time to relapse in patients with remitting–relapsing multiple sclerosis.[8] As noted previously, an analysis is fully described with both univariate/multivariate and univariable/multivariable designations, but as common "static" logistic regression and time-to-first event analyses are usually univariate, omitting the first term may be acceptably clear. In other words, when reporting "static" logistic regression and time-to-(first) event analyses, it is sufficient to explicitly state only whether the analysis is univariable or multivariable unless the analysis indeed included multiple outcomes, in which case it should be explicitly stressed that it was also multivariate. One such example is the exposure–response logistic regression model that is "dynamic," that is, considers the outcome longitudinally over time in relation to a longitudinal exposure predictor to account for the time course of drug exposure. Such models are particularly useful when an assumption of reasonably constant exposure over time in a patient cannot be supported

TABLE 1 Examples of variables and variate exposure–response analyses

Example	Regression model	Analysis (dependent variable ~ independent variable[s])	Classification per number of independent variable(s)	Classification per number of outcomes/ dependent variables
1a	Logistic regression	AE occurrence ~ C_{max}	Univariable	Univariate
1b		AE occurrence ~ C_{max}, age, race, cotherapy	Multivariable	Univariate
1c	Longitudinal logistic regression	AE occurrence over time ~ C_{max} preceding AE event (longitudinal)	Univariable	Multivariate
1d		AE occurrence over time ~ C_{max} preceding AE event (longitudinal), age, race, cotherapy	Multivariable	Multivariate
2a	Logistic regression	BOR ~ C_{min}	Univariable	Univariate
2b		BOR ~ C_{min}, tumor size, performance status	Multivariable	Univariate
2c	Longitudinal logistic regression	Response status over time ~ C_{min}	Univariable	Multivariate
2d		Response status over time ~ C_{min}, tumor size, performance status	Multivariable	Multivariate
3a	Parametric TTE	PFS ~ C_{min}	Univariable	Univariate
3b		PFS ~ C_{min}, number of nontarget lesions, smoking status	Multivariable	Univariate
3c	Parametric repeated TTE	Occurrence of relapses in multiple sclerosis ~ model-predicted absolute lymphocyte count	Univariable	Multivariate
3d		Occurrence of relapses in multiple sclerosis ~ model-predicted absolute lymphocyte count, age	Multivariable	Multivariate
4a	Parametric TTE	OS ~ C_{min}	Univariable	Univariate
4b		OS ~ C_{min}, tumor size, neutrophil-to-lymphocyte ratio	Multivariable	Univariate
5a	NLME	Plasma glucose and HbA1c ~ drug concentration over time	Univariable	Multivariate
5b		Plasma glucose and HbA1c ~ drug concentration over time, anemia	Multivariable	Multivariate
6a	Linear or nonlinear regression	Change in tumor size from baseline at Week X ~ C_{min}	Univariable	Univariate
6b		Change in tumor size from baseline at Week X ~ C_{min}, body weight, performance status	Multivariable	Univariate

Abbreviations: AE, adverse event; BOR, best overall response; C_{max}, maximum drug concentration; C_{min}, minimum drug concentration; HbA1c, hemoglobin A1c; NLME, nonlinear mixed effects; OS, overall survival; PFS, progression-free survival; TTE, time to event.

(e.g., in the setting of titration dosing regimens or when there is a meaningful extent of dose reductions in response to treatment-emergent toxicities). Of note, classical population pharmacokinetics (PK) models are not the focus of this perspective, but the authors briefly underline that as per these definitions, population PK models can be considered multivariate (repeated measures) and multivariable (even the simplest model will have two independent variables: time and dose). However, the differentiation of population PK models in terms of univariable or multivariable can be done if dose and time are considered as an inherent part of the model and any further addition of independent variables would determine whether the model is univariable (one additional covariate) or multivariable (multiple covariates). The misuse of these terms is clearly inadvertent and not entirely surprising. Many scientists are not taught the difference during their training. For those who were, remembering the difference can be difficult because of the other terms we typically use. For example, it may seem logical that a model with multiple covariates should be a "multivariate" model, even though, as

we have shown, the correct terminology for such a model is a "multivariable" model. Further reasons for the incorrect use of scientific terms may be related to the complexity and multidisciplinarity of the field, or perhaps be fueled by linguistic similarity of terms. Regardless of the underlying reasons, it is of utmost importance to openly address them and take steps toward future correct and consistent use.

Based on aforementioned considerations, we would first like to make an appeal to the reader to be mindful of the correct definitions and use of the terms *univariate/multivariate* and *univariable/multivariable*. Second, we would like this perspective to serve as a call for journals and the overall community of clinical pharmacology and pharmacometrics to address this nomenclature issue. Notably, some journals, including *JAMA Pediatrics*[9] and *Pediatric and Perinatal Epidemiology* journals,[10] have made the first steps by incorporating hints on the correct use of the terms *univariate/multivariate* and *univariable/multivariable* to their guidelines to authors: Identify regression models with more than 1 independent variable as multivariable and regression models with more than 1 dependent variable as multivariate.[9]

> Regression models of all kinds (standard, logistic, etc) that involve a single outcome are "univariate" regardless of how many explanatory variables are included in the model. The term "multivariate" regression should be restricted to those cases where there is more than one outcome (strictly speaking, a more general specification is where the model requires the assumption of a joint distribution of some kind, including certain applications of repeated measures regression).[10]

Our recommendation for scientific publishing in the disciplines of clinical pharmacology and pharmacometrics is to follow suit of the aforementioned journals by implementing clear guidance in the author guidelines on the use of univariate/multivariate and univariable/multivariable terminology. This could include a statement directing authors to (1) classify an analysis as univariable/multivariable based on the number of independent variables and (2) in the case of standard exposure–response analyses (e.g., logistic regression) omit the term *univariate* as it is implied unless the model requires otherwise (e.g., a multivariate regression model). Such a concise and clear guidance would undoubtedly contribute to a more widespread understanding and consensus on the use of these terms, therefore supporting good scientific writing practices.

FUNDING INFORMATION

No funding was received for this work.

CONFLICT OF INTEREST

As an Associate Editor of CPT: Pharmacometrics & Systems Pharmacology, Jonathan French was not involved in the review or decision process for this paper.

ORCID

Ana-Marija Grisic https://orcid.org/0000-0002-1093-6879
Karthik Venkatakrishnan https://orcid.org/0000-0003-4039-9813
Jonathan French https://orcid.org/0000-0001-9649-169X
Akash Khandelwal https://orcid.org/0000-0002-0472-7173

REFERENCES

1. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health*. 2013;103:39-40.
2. Reboldi G, Angeli F, Verdecchia P. Multivariable analysis in cerebrovascular research: practical notes for the clinician. *Cerebrovasc Dis*. 2013;35:187-193.
3. Hess B, Townsend W, Ai W, et al. Efficacy and safety exposure-response analysis of loncastuximab tesirine in patients with B cell non-Hodgkin lymphoma. *AAPS J*. 2022;24:11.
4. Bellesoeur A, Ollier E, Allard M, et al. Is there an exposure-response relationship for nivolumab in real-world NSCLC patients? *Cancer*. 2019;11:1784.
5. Groenland SL, Geel DR, Janssen JM, et al. Exposure-response analyses of anaplastic lymphoma kinase inhibitors crizotinib and alectinib in non-small cell lung cancer patients. *Clin Pharmacol Ther*. 2021;109:394-402.
6. Ogasawara K, Lymp J, Mack T, et al. In vivo cellular expansion of lisocabtagene maraleucel and association with efficacy and safety in relapsed/refractory large B-cell lyphoma. *Clin Pharmacol Ther*. 2022;112:81-89.
7. Khandelwal A, Grisic AM, French J, Venkatakrishnan K. Pharmacometrics golems: exposure-response models in oncology. *Clin Pharmacol Ther*. 2022;112:941-945. doi:10.1002/cpt.2564
8. Novakovic AM, Thorsted A, Schindler E, Jönsson S, Munafo A, Karlsson MO. Pharmacometric analyses of the relationship between absolute lymphocyte count and expanded disability status scale and relapse rate, efficacy end points, in multiple sclerosis trials. *J Clin Pharmacol*. 2018;58:1284-1294.
9. Instructions for authors. *JAMA Pediatr*. https://jamanetwork.com/journals/jamapediatrics/pages/instructions-for-authors. Accessed August 1, 2022.
10. Peters TJ. Multifarious terminology: multivariable or multivariate? Univariable or univariate? *Paediatr Perinat Epidemiol*. 2008;22:506.

How to cite this article: Grisic A-M, Venkatakrishnan K, French J, Khandelwal A. Variable or variate? A conundrum in pharmacometrics exposure–response models. *CPT Pharmacometrics Syst Pharmacol*. 2023;12:144-147. doi:10.1002/psp4.12905