



Standard Guide for Multivariate Data Analysis in Pharmaceutical Development and Manufacturing Applications¹

This standard is issued under the fixed designation E2891; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide covers the applications of multivariate data analysis (MVDA) to support pharmaceutical development and manufacturing activities. MVDA is one of the key enablers for process understanding and decision making in pharmaceutical development, and for the release of intermediate and final products.

1.2 The scope of this guide is to provide general guidelines on the application of MVDA in the pharmaceutical industry. While MVDA refers to typical empirical data analysis, the scope is limited to providing a high level guidance and not intended to provide application-specific data analysis procedures. This guide provides considerations on the following aspects:

1.2.1 Use of a risk-based approach (understanding the objective requirements and assessing the fit-for-use status),

1.2.2 Considerations on the data collection and diagnostics used for MVDA (including data preprocessing and outliers),

1.2.3 Considerations on the different types of data analysis and model validation,

1.2.4 Qualified and competent personnel, and

1.2.5 Life-cycle management of MVDA.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*²

C1174 Practice for Prediction of the Long-Term Behavior of Materials, Including Waste Forms, Used in Engineered

¹ This guide is under the jurisdiction of ASTM Committee E55 on Manufacture of Pharmaceutical and Biopharmaceutical Products and is the direct responsibility of Subcommittee E55.01 on Process Understanding and PAT System Management, Implementation and Practice.

Current edition approved Nov. 1, 2013. Published November 2013. DOI: 10.1520/E2891-13.

² For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

Barrier Systems (EBS) for Geological Disposal of High-Level Radioactive Waste

E178 Practice for Dealing With Outlying Observations

E1355 Guide for Evaluating the Predictive Capability of Deterministic Fire Models

E1655 Practices for Infrared Multivariate Quantitative Analysis

E1790 Practice for Near Infrared Qualitative Analysis

E2363 Terminology Relating to Process Analytical Technology in the Pharmaceutical Industry

E2474 Practice for Pharmaceutical Process Design Utilizing Process Analytical Technology

E2476 Guide for Risk Assessment and Risk Control as it Impacts the Design, Development, and Operation of PAT Processes for Pharmaceutical Manufacture

E2617 Practice for Validation of Empirically Derived Multivariate Calibrations

2.2 *ICH Standards:*³

ICH-Endorsed Guide for ICH Q8/Q9/Q10 Implementation ICH Quality Implementation Working Group Points to Consider (R2)

ICH Q2(R1) Validation of Analytical Procedures: Text and Methodology

3. Terminology

3.1 *Definitions*—Common term definitions can be found in Terminology E2363 for pharmaceutical applications and some terms can be found in other standards and are cited when they are mentioned.

4. Significance and Use

4.1 A significant amount of data is being generated during pharmaceutical development and manufacturing activities. The interpretation of such data is becoming increasingly difficult. Individual examination of the univariate process variables is relevant but can be significantly complemented by multivariate data analysis (MVDA). Such methodology has been shown to be particularly efficient at handling large amounts of data from

³ Available from International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), ICH Secretariat, c/o IFPMA, 15 ch. Louis-Dunant, P.O. Box 195, 1211 Geneva 20, Switzerland, <http://www.ich.org>.

multiple sources, summarizing complex information into meaningful low dimensional graphical representations, identifying intricate correlations between multivariate datasets taking into account variable interactions. The output from MVDA will generate useful information that can be used to enhance process understanding, decision making in process development, process monitoring and control (including product release), product life-cycle management and continual improvement.

4.2 MVDA is a widely used tool in various industries including the pharmaceutical industry. To generate a valid outcome, MVDA should contain the following components:

- 4.2.1 A predefined objective based on a risk and scientific hypothesis specific to the application,
- 4.2.2 Relevant data,
- 4.2.3 Appropriate data analysis techniques, including considerations on validation,
- 4.2.4 Appropriately trained staff, and
- 4.2.5 Life-cycle management.

4.3 This guide can be used to support data analysis activities associated with pharmaceutical development and manufacturing, process performance and product quality monitoring in manufacturing, as well as for troubleshooting and investigation events. Technical details in data analysis can be found in scientific literature and standard practices in data analysis are already available (such as Practices E1655 and E1790 for spectroscopic applications, Practice E2617 for model validation and Practice E2474 for utilizing process analytical technology).

5. Concepts of MVDA Model and MVDA Method

5.1 When implementing MVDA it is important to understand the differentiation between a multivariate model and a multivariate method. This is especially true as an MVDA application reaches the validation stage.

5.2 MVDA Model:

5.2.1 As defined in Practice C1174, a model is a simplified representation of a system or phenomenon with multiple variables based on a set of hypotheses (assumptions, data, simplifications, or idealizations, or a combinations thereof) that describe the system or explain the phenomenon, often expressed mathematically. In the context of this guidance the term MVDA model is to be taken in a broad sense covering, for example multivariate regression as well as latent variable-based techniques—such as, but not limited to, Principal Component Analysis (PCA) and Partial Least Squares (PLS) Regression. These models often relate observational data to a known property or set of properties from a process. The mathematical relationship is established for a sufficient number of cases—preferably derived from experimental designs. The model can then be applied to a similar set of observational data in order to predict the targeted property/properties.

5.2.2 MVDA is not limited to such multivariate calibrations and predictions, and similar considerations as the ones described in this guidance are applicable to direct and indirect calibration, as well as PCA-based approaches used for example for exploratory data analysis.

5.3 MVDA Method:

5.3.1 The MVDA method uses the output from the MVDA model to define the targeted and predefined process characteristic of interest. The MVDA model is one component of the broader concept that is an MVDA method. Such method should typically be characterized by the collection of data, the input data to the calculation, the data analysis, and some potential transformation from the MVDA model output to generate the pre-defined MVDA method characteristic of interest. (See Fig. 1.)

5.3.2 Note that an MVDA method can incorporate multiple MVDA models (for example, across multiple unit operations, from multiple pieces of equipment, etc.) that can be running in parallel or feeding sequentially into one another to provide the

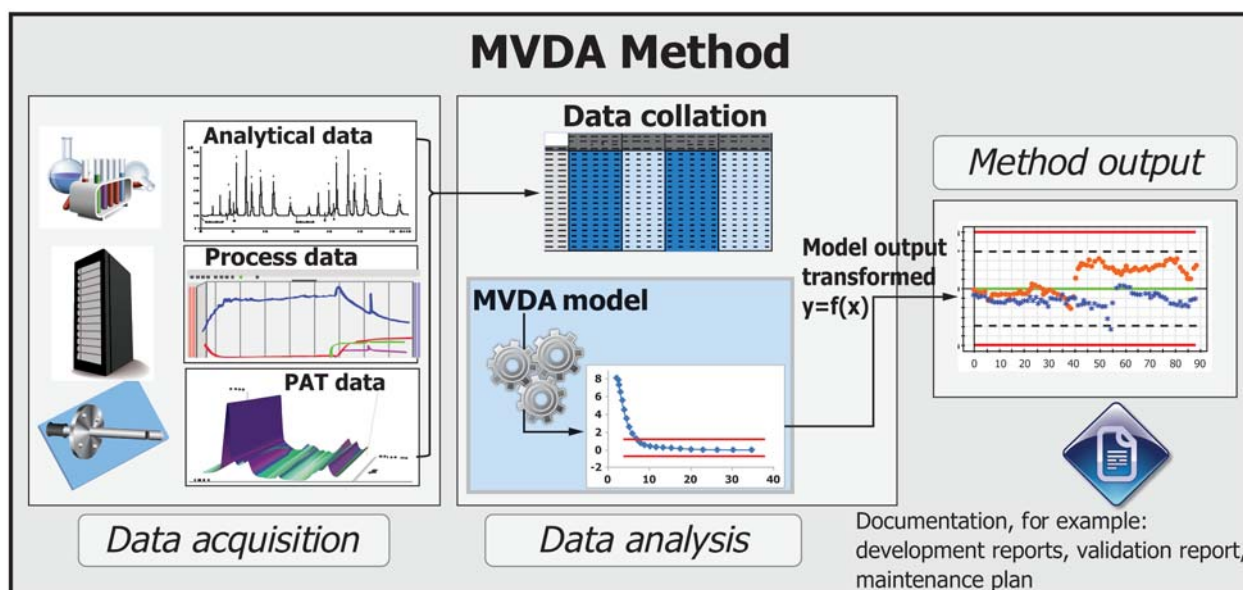


FIG. 1 Relationship Between an MVDA Method and an MVDA Model