MODEL: MECHANISTIC vs EMPIRICAL

Ajit K. Thakur

Hazleton Washington, Inc.
Vienna, VA 22182

INTRODUCTION

According to the Oxford English Dictionary, the word Model (French *modèle*, Italian *modello*, Latin *modulus*) has many meanings. In our context, the word could mean 'something that accurately resembles something else', 'an object of imitation', or 'a perfect exemplar of some excellence'. The Sanskrit equivalent of Model is 'Pratirupa', i.e. a perfect copy or imitation. From sociology to science, models have been used for centuries. National heroes like George Washington, Mahatma Gandhi and many others are models that many parents wish their children to follow (obviously there are hundreds of counter examples as well!). In science, the ancient Hindu mathematicians used picture models to calculate astronomical parameters. And of course, we have the model of gravity in Isaac Newton's Apple! There is no branch of science today that does not employ models to understand the system under study. Unfortunately, even with such a long tradition, the word model still brings out a lot of confusion and disagreement. This is particularly true when the statisticians are asked to make certain statements regarding observations made in various fields. The purpose of this discussion is not to settle any controversies, but to open the door for understanding such a broad concept.

From a systems analytic standpoint, we need some terms to be defined before we can discuss the concept of a model. These definitions are elegantly expressed in Rescigno and Beck [1987]. Accordingly:

(1) A system under study is the primary system.
(2) Any aspect of this primary system that an investigator uses to study the system is a secondary system. The data or a graph representing the primary system will then be secondary system.
(3) A model is a secondary system used to verify any hypotheses on the primary system.

Obviously, if one knew everything about the primary system, one would not need a model. Unfortunately, many of the real systems one wishes to study may not be wholly accessible. As a result, one has to use models to explain or verify certain aspects of such systems. Legal, social, religious, and ethical practices of the civilized world do not always allow us to pry into living subjects (some times not even dead subjects) to understand all intricate biological functions. So there must be needs for

New Trends in Pharmacokinetics, Edited by A. Rescigno and A.K. Thakur
Plenum Press, New York, 1991

models. Besides, modeling is like magic. It is so challenging to predict what is happening inside a black box, and then find out that many or all of the predictions were true!

MODELS: MECHANISTIC VS EMPIRICAL

For the purpose of our discussion, we will design two types of models: systems analytic or mechanistic and empirical or often designated as statistical (although the present author has some problem in using 'statistical' as a synonym for 'empirical'). A mechanistic model, as the name implies, should have as many features of the primary system built into it as observations or data will allow. Such a model should be consistent with the observed behavior of the system — retrodiction — [Rescigno and Beck, 1987]; it should further be predictive of the system's future behavior or behavior under perturbation — prediction — [Rescigno and Beck, 1987]. One must have some knowledge of the primary system in terms of structural connectivity and functional mechanisms. Some prefer to call this type of models realistic, intrinsic, and various other names. Many great discoveries in biology, medicine, and other branches of science have been made using such models. In this context one must remember that such models do not necessarily have to have an explicit mathematical expressions; they could be just conceptualizations.

On the other hand, when the system under study is complex and hardly anything is known about its structural connectivity and functional mechanisms, yet one has to produce hypotheses about it based on some external characteristics such as a dose-response (secondary system), one often relies on mathematical functional forms for such a system. These mathematical functions are empirical models. They may incorporate some mechanistic assumptions so that they may look realistic. Numerically, these models are generally easier to handle as opposed to many mechanistic models. Most normal theory based statistical hypothesis testing and confidence interval procedures are based on such models. One should not get the wrong impression that mechanistic models are not useful for such statistical techniques; they may be more difficult to handle numerically from estimation standpoints. Some people would call empirical models extrinsic because they are based purely on the external behavior of the system. Some call them statistical models. As mentioned earlier, it is unfair to assume that statisticians always like to use empirical models for their purposes. The reasons why there are abundance of this type of models in literature are obvious. Our knowledge about the primary system may be inadequate-to-none to allow us the formulation of a mechanistic model or one may not be interested in understanding the inherent structure of the system. In the present author's mind, the phrase statistical model includes both types of models. One must remember that an empirical model may be 'retroactive' (explaining what happened from a secondary system) and even locally 'predictive' (i.e. interpolation may be performed within the range of observations), but it is, in general, not globally 'predictive' (indicating outcome of future experiments). In fact, empirical models should never be used with any authority for extrapolative purposes.

According to Fisher [1925], K.F. Gauss in the early 1800's may have been instrumental in developing empirical modeling concept with his work on maximum likelihood and least squares theories. 'Gauss, further, perfected the systematic fitting of regression formulae, simple and multiple, by the method of least squares, which, in the cases to which it is appropriate, is a particular example of the method of maximum likelihood' [Fisher, 1925]. A slight variation of empirical modeling is defined by Ashby [1958]. In this form, one takes the system and examines its individual components. One makes hypotheses on these individual components with models and finally one tries to draw a global conclusion about the system. According