Geometric Statistics in PK analysis - Programmer's perspective
Niraj J. Pandya, Element Technologies Inc., NJ
Smitha Mullapudi, Independent Consultant, CT

ABSTRACT
For any SAS® statistical programmer working on PK analysis, it is an obvious question: “Why geometric statistics are required and why arithmetic statistics alone are not enough?” The answer to this question lies in the fact that most drug concentration and pharmacokinetic parameters data follow a log-normal distribution. In such case, geometric statistics are more appropriate summary descriptors. At present, none of the SAS/STAT® procedures directly compute geometric statistics and so it is of greater interest to understand the concept and implement it in actual programming. This article would explain more on log-normal distribution and programming approach to calculate geo-stats using the example of PK parameters data.

INTRODUCTION
The drugs taken by patients can range from a single dose of a drug for an occasional headache to drugs every day for a life time those have chronic ailments. To achieve optimal therapy, the appropriate ‘drug of choice’ must be selected. This decision implies not only an accurate diagnosis of the disease, knowledge of the clinical state of the patient, but also on the sound understanding of the pharmacotherapeutic management of the disease. To understand this, the drug administration can be divided into two phases, a pharmacokinetic phase, in which the adjustable elements of dose, dosage form, frequency, and route of administration are related to drug leve-time relationships in the body and a pharmacodynamic phase, in which the concentration of drug at the site of action is related to the magnitude of the effects produced. The advantage of this approach is it would give the ‘therapeutic window’.

Above figure explains the flow of information in PK/PD analysis very broadly.

The magnitude and relative contribution of pharmacokinetic and pharmacodynamics to variability in response to a given dose within a patient population vary with the drug and to some extent, the condition being treated. To administer drugs optimally, therefore, knowledge is needed not only of the mechanics of drug administered but also of the pharmacokinetics.
PK/PD analysis has many useful applications throughout the life cycle of drug development. To understand the pharmacokinetics of drug and to perform statistical analysis of such data, understanding of concepts of drug absorption and disposition in a human body is important. We will briefly go through some important concepts first in next few sections before we start discussing about the statistical analysis part in detail.

**ABSORPTION**

Absorption is defined as the process by which unchanged drug proceeds from site of administration to site of measurement within the body. Factors such as metabolized by enzymes by gastrointestinal or metabolized completely by liver, makes the absorption dependent on elimination. Absorption critically determines the compound’s bioavailability. Drugs that absorb poorly when taken orally, intramuscular, subcutaneous and other extra vascular routes of administration can be considered. This would help in drug monitoring in different body fluids and useful means of assessment.

**DISTRIBUTION**

Distribution is the process of reversible transfer of a drug to and from the site of measurement by body fluid. From there, the compound may distribute into tissues and organs, usually to differing extents. The pathway for return of drug need not be the same as that leaving the circulation.

**METABOLISM**

Metabolism is the conversion of one chemical species to another. Compounds begin to break down as soon as they enter the body. When metabolites are pharmacologically inert, metabolism deactivates the administered dose of parent drug and this usually reduces the effects on the body. Occasionally, metabolites are converted back to the drug. This is a route of elimination only to the extent that the metabolite is excreted or otherwise irreversible lost.

**EXCRETION**

Elimination is the irreversible loss of drug from the site of measurement. Compounds and their metabolites removed from the body either by excretion, usually through the kidneys (urine) or in the feces. Biliary excretion or fecal excretion is the process that initiates in the liver and passes through to the gut until the products are finally excreted along with waste products or feces. The last method of excretion is through the lungs e.g. anesthetic gases.

**ARITHMETIC MEAN**

If we go by the definition of the arithmetic mean, it is applicable when several observations are added together to produce a total. In this case, we seek to determine if all the observations had the same value, what that value would have to be to get the same total. According to Gaussian statistical model, arithmetic mean simply adds a series of observations and divides the sum by total number of observations. This mean is the most widely known and frequently referenced.

The arithmetic mean can be illustrated by the following formula:

\[ \bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i = \frac{1}{n} (r_1 + r_2 + \ldots + r_n) \]

**GEOMETRIC MEAN**

Geometric average is established when various observations are multiplied together to produce a product. In this case, we seek to determine of all the observations had the same value, what that value would have to
be to get the same product. This can be best described by log-normal distribution rather than a Gaussian statistical model. When the purpose is to account for compounding effect, geometric mean is the best descriptor.

The geometric mean can be illustrated by the following formula:

\[ G_m = \left( \prod_{i=1}^{n} a_i \right)^{1/n} - 1 = \sqrt[n]{a_1 \times a_2 \times \cdots \times a_n} - 1 \]

WHY GEOMETRIC STATISTICS IN PK ANALYSIS?

Both drug concentration data and pharmacokinetic parameters usually follow a skewed distribution among patient population. Reaction rate or a transfer rate of drug in a patient’s body are proportional to the amount of substrate available on a given time point. In most cases, an enzyme inhibitor will increase the concentration of drug by a given percentage rather than by a fixed amount. Because of this, different study factors are often assumed to impact multiplicatively rather than additively on the variable of interest. On any specific given time point, the drug concentration amount is not independent to the drug concentration measured at previous time point. This fact gives importance to relative change in the concentration and drug transfer rate rather than absolute change. Most pharmacokinetic processes are governed by first-order rates which also give importance to relative changes and it can be best described by log-normal distribution.

In this situation, it is not appropriate to summarize the concentration results using arithmetic mean and standard deviation. When the concentration data distribution is plotted, the thin tail of the distribution drags the mean above the density of typical values and increases the magnitude of standard deviation which exceeds the size of mean. In this situation, more suitable summary descriptor is geometric mean.

Geometric mean can be defined as the \( n^{th} \) root of the product of \( n \) individual values. Practically, in programming, geometric mean can be obtained by transforming individual concentration values into log values. Then take the arithmetic mean of log-transformed values and calculating back the antilog of this mean. Keeping this in mind, it will be noteworthy that any observation with concentration value of 0 should be ignored while deriving the geometric mean.

The relation of arithmetic mean and geometric mean is that arithmetic mean is always greater than or equal to geometric mean. This fact indicates that arithmetic mean will always overestimate the average concentration among the patient population. In addition, if the concentration values are observed for longer time period, larger the difference between geometric and arithmetic means will be.

GEOMETRIC STANDARD DEVIATION (SD) AND COEFFICIENT OF VARIATION (CV)

In Gaussian distribution model, arithmetic standard deviation around the arithmetic mean is the difference either added or subtracted from the mean, which encompasses about two thirds of the complete set of data.

The same way, in log-normal distribution, geometric standard deviation is a factor by which if geometric mean is multiplied or divided will cover the two third of the complete set of data.

Converting the normal concentration values into log-normal values, calculating the typical standard deviation of log transformed concentrations and then taking its antilog provides the geometric standard deviation. This standard deviation provides an estimate of the relative spread of the log-normally distributed values. This relative spread simply gives the measure of variability existing in the data values.

When above mentioned relative spread is presented in terms of percentages, it is referred as geometric coefficient of variation.
While arithmetic coefficient of variation is defined by arithmetic standard deviation divided by arithmetic mean, geometric coefficient of variation can be easily obtained by simply subtracting 1 from the geometric standard deviation and multiplying it by 100. It simply gives the measure of percentage of variability existing in the data values.

**GRAPHICAL REPRESENTATION**

When drug concentration data is required to be presented graphically, it is always recommended to present geometric means along with error bars. Error bars show multiplication and division factor corresponding to geometric standard deviation and hence error bars appear symmetric along Y axis on the plot.

Likewise, when PK parameter data statistics are required to be used to compare multiple treatment groups, data is first translated to log-normal values.

**CONCLUSION**

One can always argue that additional work involved in obtaining geometric mean is not justifiable knowing the fact that difference between arithmetic and geometric mean is usually too small. But it is also a fact that the difference grows significantly when drug concentration and transfer rates are widely dispersed and observed over a longer time period. Nevertheless, it is important to be mathematically accurate and use geometric means and related statistical descriptors when variable of interest changes relatively and not absolutely. For a SAS programmer working on PK data reporting, SAS does not offer much help in calculating geometric statistics as there is no inbuilt procedure for them yet. But, with the knowledge of basic principles of geometric statistics and practical way of defining them which are explained in this article, one can calculate geometric statistics easily and efficiently.

**REFERENCES**

- Clinical Trials Handbook by Shayne Cox Gad; John Wiley and Sons, 2009

**CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the authors at:

Name: NIRAJ J PANDYA
Phone: 201-936-5826
E-mail: npandya@elementtechnologies.com

Name: SMITHA MULLAPUDI
Phone: 718-839-4884
E-mail: smitha.mullapudi@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.