It is also the case that some clinical trials are concluded before the event of interest has occurred in every single one of the subjects. This situation can lead to bias, where the physiological properties of patients enrolled early in the trial differ from those enrolled late in the trial. As articulated by Bland and Altman (35), "we assume that the survival probabilities are the same for subjects recruited early and late in the study. In a long term observational study of patients with cancer, for example, the case mix may change over the period of recruitment."

V. SAMPLE VERSUS POPULATION

The terms *sample* and *population* are standard terms in statistics. The term *sample* refers to data acquired by actual measurements. The investigator has the option of testing one sample, taken from a population, or of testing more than one sample, taken from the population. In discussions of statistics, it is the case that the term *sample* refers to a group of objects, for example, 50 drug tablets, while the term *population* refers to the entire batch of 10,000 drug tablets that was manufactured. In statistics, it is the case that the term *sample* refers to 100 subjects enrolled in a clinical trial, while the term *population* refers to the entire world's population of people with the disease of interest. The sample needs to be representative of the population.

The term *population* can refer to a hypothesized, underlying value or to an imaginary, idealized value. In some situations, it is possible for the researcher to measure a parameter of interest from all members of the population. But often, it is impractical or impossible to measure the parameter in all members of the population. Data acquired by analyzing a sample are subject to variations in the properties of the sample and to variations in the techniques used by the investigator. For example, in a study of 50 human subjects, 15 of the subjects may have a mutation in a growth factor receptor gene, while the other 35 subjects have the wild-type gene. Or, in a study of 50 human subjects, five of the subjects may have forgotten to take two of their drug doses, while 45 of the subjects had remembered to take all of the drug doses. But data acquired by analyzing a population take into account these and all other variations. But data acquired by analyzing only a 15-person sample taken from this population of 50 human subjects may be drastically biased, for example, where all of the subjects in the 15-person sample have the genetic mutation.

Researchers may be interested in measuring a parameter from a sample, where the goal is to predict the same parameter in the entire population, that is, where it is not practical or not possible to determine that parameter in the population. An example can be found in the manufacture of tablets or pills. Where the goal is to determine the weight of the tablets, and to determine whether the range of weights is within manufacturing specifications, the analyst can measure the weights of 100 tablets, taken from a population of 1 million tablets that was manufactured in a specific batch. In this situation, the 100 tablets constitute a "sample," while the 1 million tablets manufactured in a specific batch constitute the "population." Batchwise manufacture is distinguished in that each component has a specific lot number, and by the fact that the machinery was cleaned and calibrated specifically for the manufacture of that batch.

In this scenario, the various statistical parameters of the sample (mean; standard deviation) are known, and the statistical parameters of the population (mean; standard deviation) are also known. Instead of measuring a parameter of 1 million tablets, the researcher can refer to standards set forth by the pharmaceutical industry. These standards may relate to mean weight and standard deviation.

[35]Bland JM, Altman DG. Survival probabilities (the Kaplan–Meier method). Brit. Med. J. 1998;317:1572.