Mohini P. Barde, Prajakt J. Barde[1]

Shrimohini Centre for Medical Writing and Biostatistics Pune, Maharashtra, [1]Glenmark Pharmaceutical Ltd., Mumbai, Maharashtra, India

Address for correspondence:
Dr. Prajakt J. Barde,
Glenmark Pharmaceutical Ltd.,
Mumbai, Maharashtra, India
E-mail: drprajaktb@gmail.com

What to use to express the variability of data: Standard deviation or standard error of mean?

Abstract

Statistics plays a vital role in biomedical research. It helps present data precisely and draws the meaningful conclusions. While presenting data, one should be aware of using adequate statistical measures. In biomedical journals, Standard Error of Mean (SEM) and Standard Deviation (SD) are used interchangeably to express the variability; though they measure different parameters. SEM quantifies uncertainty in estimate of the mean whereas SD indicates dispersion of the data from mean. As readers are generally interested in knowing the variability within sample, descriptive data should be precisely summarized with SD. Use of SEM should be limited to compute CI which measures the precision of population estimate. Journals can avoid such errors by requiring authors to adhere to their guidelines.

Key words: Standard deviation, standard error of mean, confidence interval

INTRODUCTION

Statistics plays a vital role in biomedical research. It helps present data precisely and draws meaningful conclusions. A large number of biomedical articles have statistical errors either in presentation[1-3] or analysis of data. The scathing remark by Yates *"It is depressing to find how much good biological work is in danger of being wasted through incompetent and misleading analysis."* highlights need of proper understanding of statistics and its appropriate use in medical literature.

In late nineties, biomedical journals have made a concerted effort to improve quality of statistics.[4-6] Despite this, errors are still present in published articles. One such common error is use of SEM instead of SD to express variability of data.[7-10] Negele *et al*, also showed clearly that a significant number of published articles in leading journals had misused SEM in descriptive statistics.[11] In this article, we discussed the concept and use of SD and SEM.

CONCEPT OF SD AND SEM

To study the entire population is time and resource intensive and not always feasible; therefore studies are often done on the sample; and data is summarized using descriptive statistics. These findings are further generalized to the larger, unobserved population using inferential statistics.

For example, in order to understand cholesterol levels of the population, cholesterol levels of study sample, drawn from same population are measured. The findings of this sample are best described by two parameters; mean and SD. Sample mean is average of these observations and denoted by $\overline{X}$. It is the center of distribution of observations (central tendency). Other parameter, SD tells us dispersion of individual observations about the mean. In other words, it characterizes typical distance of an observation from distribution center or middle value. If observations are more disperse, then there will be more variability. Thus, a low SD signifies less variability while high SD indicates more spread out of data. Mathematically, the SD is[12]

$$s = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}}$$

s = sample SD; X - individual value; $\overline{X}$ - sample mean; n = sample size.

Figure 1a shows cholesterol levels of population of 200 healthy individuals. Cholesterol of the most of individuals is between 190-210mg/dl, with a mean (μ) 200mg/dl and SD (s) 10mg/dl. A study in 10 individuals drawn from same population with cholesterol levels of 180, 200, 190, 180, 220, 190, 230, 190, 190, 180mg/dl gives $\overline{X}$ = 195 mg/dl and SD (s) = 17.1 mg/dl.

These sample results are used to make inferences based on the premise that what is true for a randomly selected sample will be true, more or less, for the population from which the sample is chosen. This means, sample mean ($\overline{X}$) estimates the true but unknown population mean (μ) and sample SD (s) estimates population SD (s). However, the precision with which sample results determine population parameters needs to be addressed. Thus, in above case $\overline{X}$ = 195 mg/dl estimates the population mean μ = 200 mg/dl. If other samples of 10 individuals are selected, because of intrinsic variability, it is unlikely that exactly same mean and SD [Figures 1b, c and d] would be observed; and therefore we may expect different estimate of population mean every time.

Figure 2 shows mean of 25 groups of 10 individuals each drawn from the population shown in Figure 1. If these 25 group means are treated as 25 observations, then as per the statistical "Central Limit Theorem" these observations will be normally distributed regardless of nature of original population. Mean of all these sample means will equal the mean of original population and standard deviation of all these sample means will be called as SEM as explained below.

SEM is the standard deviation of mean of random samples drawn from the original population. Just as the sample SD (s) is an estimate of variability of observations, SEM is an estimate of variability of possible values of means of samples. As mean values are considered for calculation of SEM, it is expected that there will be less variability in the values of sample mean than in the original population. This shows that SEM is a measure of the precision with which sample mean $\overline{X}$ estimate the population mean μ. The precision increases as the sample size increases [Figure 3].

Thus, SEM quantifies uncertainty in the estimate of the mean.[13,14] Mathematically, the best estimate of SEM from single sample is[15]
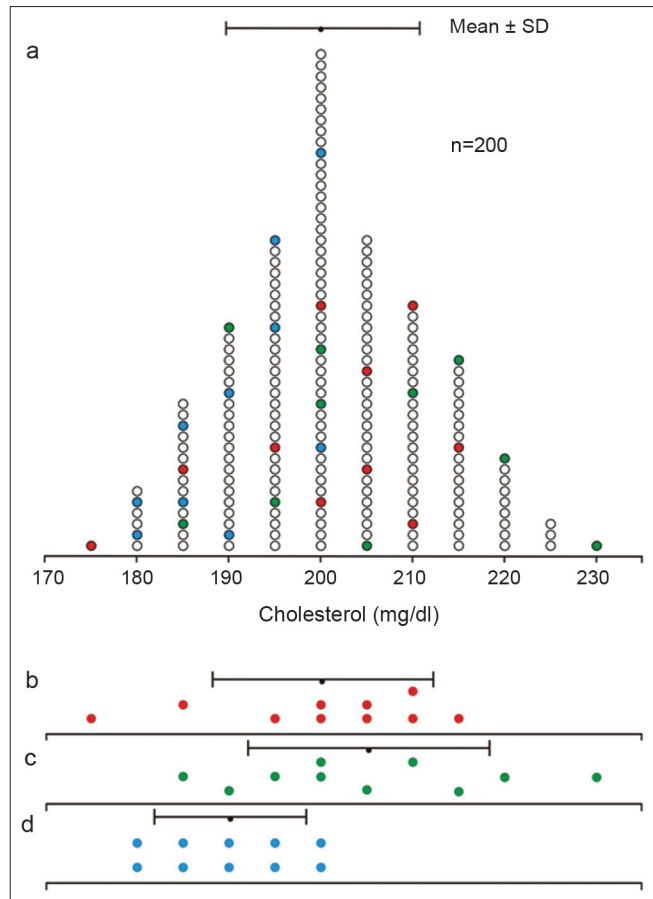

Figure 1: If one draws three different groups of 10 individuals each, one will obtain three different mean and SD. (Adapted from Glantz, 2002)
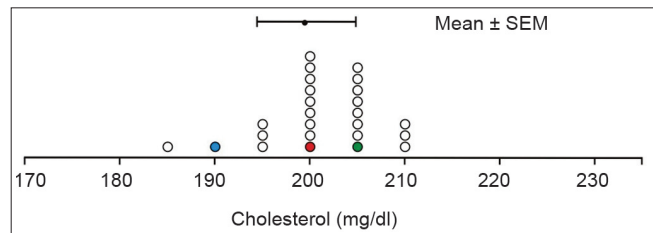

Figure 2: This figure illustrates the mean of 25 groups of 10 individuals each drawn from the population of 200 individuals shown in the Figure 1. The means of three groups shown in Figure 1 are shown using circles filled with corresponding patterns

$$\sigma_M = \frac{s}{\sqrt{n}}$$

σ_M = SEM; s = SD of sample; n = sample size.

However, SEM by itself doesn't convey much useful information. Its main function is to help construct confidence intervals (CI).[16] CI is the range of values that is believed to encompass the actual ("true") population value. This true population value usually is not known, but can be estimated from an appropriately selected sample. If samples are drawn repeatedly from population and CI is
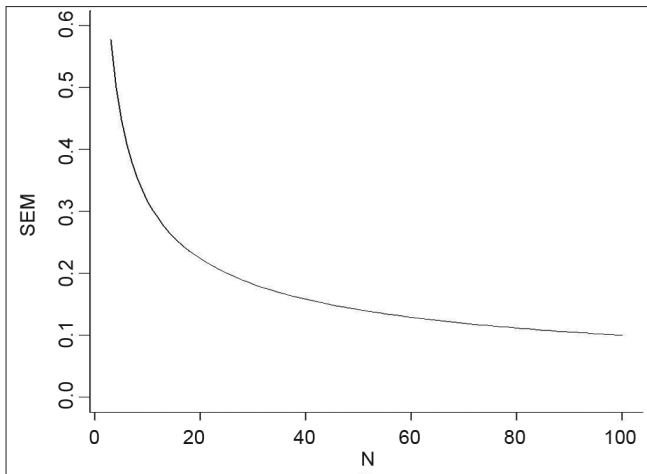
Figure 3: The figure shows that the SEM is a function of the sample size

constructed for every sample, then certain percentage of CIs can include the value of true population while certain percentage will not include that value. Wider CIs indicate lesser precision, while narrower ones indicate greater precision.[17]

CI is calculated for any desired degree of confidence by using sample size and variability (SD) of the sample, although 95% CIs are by far the most commonly used; indicating that the level of certainty to include true parameter value is 95%. CI for the true population mean μ is given by[12]

$$95\% \, CI = \overline{X} \pm Z \frac{s}{\sqrt{n}}$$

s = SD of sample; n = sample size; z (standardized score) is the value of the standard normal distribution with the specific level of confidence. For a 95% CI, Z = 1.96.

A 95% CI for population as per the first sample with mean and SD as 195 mg/dl and 17.1 mg/dl respectively will be 184.4 - 205.5 mg/dl; indicating that the interval includes true population mean m = 200 mg/dl with 95% confidence. In essence, a confidence interval is a range that we expect, with some level of confidence, to include the actual value of population mean.[17]

APPLICATION

As explained above, SD and SEM estimate quite different things. But in many articles, SEM and SD are used interchangeably and authors summarize their data with SEM as it makes data seem less variable and more representative. However, unlike SD which quantifies the variability, SEM quantifies uncertainty in estimate of the mean.[13] As readers are generally interested in knowing the

variability within sample and not proximity of mean to the population mean, data should be precisely summarized with SD and not with SEM.[18,19]

The importance of SD in clinical settings is discussed below. In a atherosclerotic disease study, an investigator reports mean peak systolic velocity (PSV) in the carotid artery, a measure of stenosis, as 220cm/sec with SD of 10cm/sec.[20] In this case it would be unusual to observe PSV less than 200 cm/sec or greater than 240cm/sec as 95% of population fall within 2SD of the mean, assuming that the population follows a normal distribution. Thus, there is a quick summary of the population and the range against which to compare the specific findings. Unfortunately, investigators are quite likely to report the PSV as 220cm/sec ± 1.6 (SEM). If one confused the SEM with the SD, one would believe that the range of the population is narrow (216.8 to 223.2cm/sec), which is not the case.

Additionally, when two groups are compared (e.g. treatment and control groups), SD helps in visualizing the effect size, which is an index of how much difference is there between two groups.[12] Effect size gives an idea of magnitude of difference to help differentiate between statistical significance and practical importance. Effect size is determined by calculating the difference between the means divided by the pooled or average standard deviation from two groups. Generally, effect size of 0.8 or more is considered as a large effect and indicates that the means of two groups are separated by 0.8SD; effect size of 0.5 and 0.2, are considered as moderate or small respectively and indicate that the means of the two groups are separated by 0.5 and 0.2SD.[12] However, same can't be interpreted with SEM. More importantly, SEMs do not provide direct visual impression of the effect size, if number of subjects differs between groups.

Exceptionally the SD as an index of variability may be a deceptive one in many experimental situations where biological variable differs grossly from a normal distribution (e.g. distribution of plasma creatinine, growth rate of tumor and plasma concentration of immune or inflammatory mediators). In these cases, because of the skewed distribution, SD will be an inflated measure of variability. In such cases, data can be presented using other measures of variability (e.g. mean absolute deviation and the interquartile range), or can be transformed (common transformations include the logarithmic, inverse, square root, and arc sine transformations).[17]

Some journal editors require their authors to use the SD and not the SEM. There are two reasons for this trend. First, the SEM is a function of the sample size, so it can

be made smaller simply by increasing the sample size (n) [Figure 3]. Second, the interval (mean ± 2 SEM) will contain approximately 95% of the means of samples, but will never contain 95% of the observations on individuals; in the latter situation, mean ± 2 SD is needed.[21]

In general, the use of the SEM should be limited to inferential statistics where the author explicitly wants to inform the reader about the precision of the study, and how well the sample truly represents the entire population.[22] In graphs and figures too, use of SD is preferable to the SEM. Further, in every case, standard deviations should preferably be reported in parentheses [i.e., mean (SD)] than using mean ± SD expressions, as the latter specification can be confused with a 95% CI.[17]

CONCLUSION

Proper understanding and use of fundamental statistics, such as SD and SEM and their application will allow more reliable analysis, interpretation, and communication of data to readers. Though, SEM and SD are used interchangeably to express the variability; they measure different parameters. SEM, an inferential parameter, quantifies uncertainty in the estimate of the mean; whereas SD is a descriptive parameter and quantifies the variability. As readers are generally interested in knowing variability within the sample, descriptive data should be precisely summarized with SD. Use of SEM should be limited to compute CI which measures the precision of population estimate.

REFERENCES

1. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials – a survey of three medical journals. N Engl J Med 1987;317:426-32.
2. García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 2004;4:13-7.
3. Cooper RJ, Schriger DL, Close RJ. Graphical literacy: The quality of graphs in a large-circulation journal. Ann Emerg Med 2002;40:317-22.
4. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals. J Gen Intern Med 1998;13:753-6.
5. Gore SM, Jones G, Thompson SG. The Lancet's statistical review process: Areas for improvement by authors. Lancet 1992;340:100-2.
6. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. BMJ 1983;286:1489-93.
7. Viswanatha Swamy AH, Wangikar U, Koti BC, Thippeswamy AH, Ronad PM, Manjula DV. Cardioprotective effect of ascorbic acid on doxorubicin-induced myocardial toxicity in rats. Indian J Pharmacol 2011;43:507-11.
8. Bihaqi SW, Singh AP, Tiwari M. *In vivo* investigation of the neuroprotective property of Convolvulus pluricaulis in scopolamine-induced cognitive impairments in Wistar rats. Indian J Pharmacol 2011;43:520-5.
9. Adenubi OT, Raji Y, Awe EO, Makinde JM. The effect of the aqueous extract of the leaves of boerhavia diffusa linn. on semen and testicular morphology of male Wistar rats. Sci World J 2010;5:1-6.
10. Banji D, Pinnapureddy J, Banji OJ, Kumar AR, Reddy KN. Evaluation of the concomitant use of methotrexate and curcumin on Freund's complete adjuvant-induced arthritis and hematological indices in rats. Indian J Pharmacol 2011;43:546-50.
11. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. Br J Anaesth 2003;90:514-6.
12. Dawson-Sanders B, Trapp RG. Basic and clinical biostatistics. Norwalk, Connecticut: Appleton & Lange; 1990.
13. Glantz SA. How to summarize data. "Primer of Biostatistics." 5th ed. Philadelphia: McGraw-Hill; 2002. p. 10-30.
14. Lang TA. How to report statistics in medicine: Annotated guidelines for authors, editors, and reviewers. Philadelphia: American College of Physicians; 1997.
15. Lee HB, Comrey AL. Elementary statistics: A Problem Solving Approach. 4th ed. UK: William Brown; 2007.
16. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Cambridge, MA: Blackwell Scientific; 1994.
17. Curran-Everett D, Taylor S, Kafadar K. Fundamental concepts in statistics: Elucidation and illustration. J Appl Physiol 1998;85:775-86.
18. Jaykaran, Yadav P, Chavda N, Kantharia ND. Some issue related to the reporting of statistics in clinical trials published in Indian medical journals: A Survey. Int J Pharmacol 2010;6:354-9.
19. Tom L. Twenty statistical error even YOU can find in biomedical research articles. Croat Med J 2004;45:361-70.
20. Medina LS, Zurakowski D. Measurement variability and confidence intervals in medicine: Why should radiologists care? Radiology 2003;226:297-301.
21. Bartko JJ. Rationale for reporting standard deviations rather than standard errors of mean. Am J Psychiatry 1985;142:1060.
22. Strasak AM, Zaman Q, Pfeiffer KP, Gobel G, Ulmer H, Statistical errors in medical research--a review of common pitfalls. Swiss Med Wkly 2007;137:44-9.