doi:10.1111/anae.13565

Statistically Speaking

It's just a standard deviation!

To study an entire population is time-consuming, and usually not feasible. Therefore, studies are usually conducted on a sample of the population, and inferences about the population are made based on the data obtained from the sample. When reporting experimental data in medical manuscripts, authors often use descriptive statistics to describe the data, for example, ages of patients in the propofol group were mean (standard deviation) 38.5 (12.3), whereas those in the sevoflurane group were 31.9 (9.4) yrs [1]. Descriptive statistics used for normally distributed data are mean and standard deviation, whereas for data which is not normally distributed, such as pain scores and sedation scores, the median, IQR and range would be given.

The standard deviation (SD, sometimes given the Greek letter σ) of a sample is an estimate of the variability of the population from which the sample was drawn, and is given in the same units as the variable. It is calculated by taking the square root of the average of the squared deviations of the values from the mean for that series (for a simple worked example, please see Box 1). A set of values that are closely clustered near the mean will have a low SD, whereas a set of numbers that are widely apart will have a higher SD. For normally distributed data, around 95% of individuals will have values within two SDs of the mean, with the remaining 5% being equally distributed above and below these limits (Fig. 1) [2]. The standard deviation is a valid measure of variability, regardless of the distribution, with around 95% of individuals falling within two SDs of the mean, though the remaining 5% might not be so equally distributed above and below these limits (Fig. 2) [3].

The standard error of the mean (SEM) is an example of inferential statistics. It is given by SD/$\sqrt{}$sample size and is an estimate of how close your sample mean is to your population mean. Like the standard deviation, the SEM is also given in the same units as the variable it describes. The SEM will become smaller as the sample size increases, as the extent of chance variation decreases [3], whereas the SD will not change predictably with sample size.

Standard deviation versus standard error of the mean

So which statistic should be used when presenting our results in manuscripts? In nearly all instances, and especially with experimental data, you are interested in how

Box 1: How to calculate the standard deviation

You have six values, 4, 7, 8, 12, 13, 16

1 The mean of the six values is 10

2 Minus the mean from each value, and square the result.

$4–10 = (-6)^2 = 36$

$7–10 = (-3)^2 = 9$

$8–10 = (-2)^2 = 4$

$12–10 = (2)^2 = 4$

$13–10 = (3)^2 = 9$

$16–10 = (6)^2 = 36$

3 Sum up the squared differences = 98

4 Divide the sum of squared differences by n – 1
= 98/5 = 19.6 = variance

5 The SD is square root of the variance = **4.43**

widely scattered your measurements are, and so the SD should be used. Unfortunately though, it is often the case that authors will use the SEM when they want to emphasise the lack of variation of a particular marker when biological or technical repeat experiments are performed [4]. Figure 3 is an example of when the SEM is used incorrectly. The investigators measured the ratio of IL-1β to β-actin in six samples from the same mouse, and displayed their data using a bar chart showing mean and SEM. In this instance, we are interested in knowing how varied that particular marker (IL-1β) is in the same tissue type of the same animal (any variability seen would be attributed to technical issues), and so, SD would be the correct descriptive statistic to use. Because SEM is given by the SD divided by the square root of the sample size, by definition, it must be smaller than the SD. When looking at graphical depictions of experimental data, the reader should ascertain what the error bars actually show, do not mistake a small error bar showing SEM for little variability among the data.

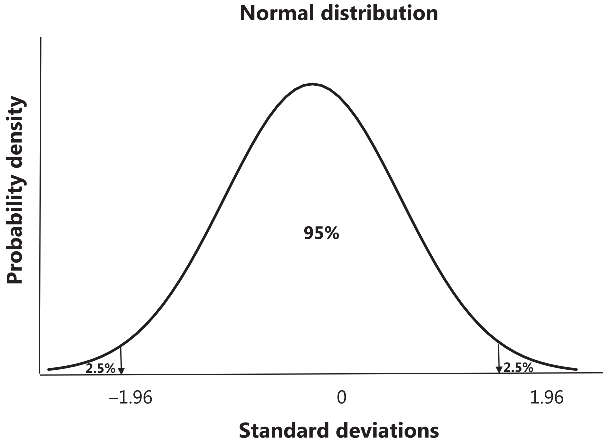

Figure 1 In a normal distribution, 95% of all values are within ±1.96 standard deviations.

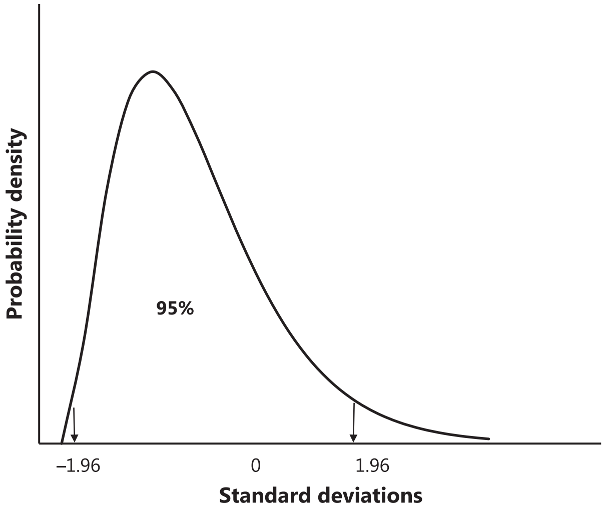

Figure 2 Even in skewed distributions, 95% of all values are within ±1.96 standard deviations, though the remaining 5% are not evenly distributed on either side of the curve.

Statistical significance and error bars

When scanning the results section of a manuscript, many of us are tempted to look for ‘significance’ by eyeballing the error bars on bar charts. We vaguely remember a rule which says that if error bars do not overlap (e.g. the error bars of CFA and inhibitor in Fig. 3), then there is a statistically significant difference between the two groups. It is true that if two errors bars showing SEM do overlap, there is no significant difference between the two groups, but the converse is not true. When SEM error bars do not overlap, we cannot automatically assume significant differences. Even less information can be obtained by eyeballing SD error bars. When the two means are significantly different, the SD error bars may, or may not overlap; this is also true for when there is no significant difference, the SD error bars may, or may not overlap [5].

Confidence intervals

The confidence interval is another example of inferential statistics which we often confuse. A confidence interval gives an estimated range of values which is likely to include an unknown population parameter (the parameter most often used in this context is the population mean), the estimated

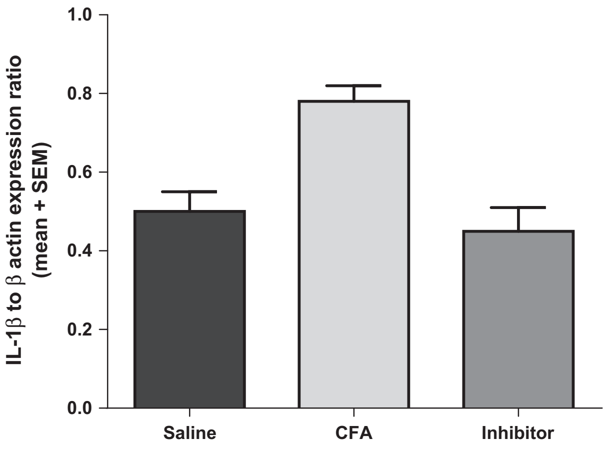

Figure 3 An example of how standard error of the means should not be used.

Table 1 General rules for quick decisions on significance.

Error bar shown	Overlap	No overlap
Standard deviation	No conclusion	No conclusion
Standard error of the mean	p > 0.05	No conclusion
95% confidence interval	No conclusion	p < 0.05

range being calculated from a given set of sample data. The probability that the confidence interval (CI) encompasses the true value is called the confidence level of the CI. The 95% confidence level is the most often used, although confidence intervals can be calculated for 90%, 95% or 99%. Confidence intervals would be cited thus, in the literature "The number needed to treat (NNT) to prevent laryngospasm in children is 7 (95% CI 5–12)" [6]. This is interpreted as having 95% confidence that the true number needed to treat (in the population) is between 5 and 12. To calculate the 95% confidence interval of the mean, (e.g. if your mean age was 54, SEM was 6) then you multiply the SEM by 1.96 (1.96 is the magic number where in a normal distribution, 95% of the values would lie within 1.96 standard deviations of the mean [2]), $6 \times 1.96 = 11.76$ and you would express your data as mean 54, 95% CI 42.2–65.6, meaning that there is a 95% probability that the true population mean lies between 42.2 and 65.6.

Although seldom used in the medical literature [7], confidence intervals (for the mean) are more informative than standard error of the mean, and should always be given. When bar charts are drawn using confidence interval error bars, you can roughly eyeball the data for significance. If the CI error bars do not overlap (and presuming that multiple comparisons were not conducted), you can conclude that there is a statistically significant difference between the means of two groups at $p < 0.05$ (Table 1).

Acknowledgements

No external funding or conflict of interest declared.

S. W. Choi
Postdoctoral Fellow
G. T. C. Wong
Clinical Associate Professor
Department of Anaesthesiology,
The University of Hong Kong,
Hong Kong Special Administrative Region
Email: htswchoi@hku.hk

References

1. Kenwright DA, Bernjak A, Draegni T, et al. The discriminatory value of cardiorespiratory interactions in distinguishing awake from anaesthetised states: a randomised observational study. *Anaesthesia* 2015; **70**: 1356–68.
2. Choi SW, Lam DMH. An alarm for a false alarm. *Anaesthesia* 2016; **71**: 106–8.
3. Altman DG, Bland JM. Standard deviations and standard errors. *British Medical Journal* 2005; **331**: 903.
4. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia* 2003; **90**: 514–6.
5. Cumming G, Fidler F, Vaux DL. Error bars in experimental biology. *Journal of Cell Biology* 2007; **177**: 7–11.
6. Mihara T, Uchimoto K, Morita S, Goto T. The efficacy of lidocaine to prevent laryngospasm in children: a systematic review and meta-analysis. *Anaesthesia* 2014; **69**: 1388–96.
7. McCormack J, Vandermeer B, Allan GM. How confidence intervals become confusion intervals. *BMC Medical Research Methodology* 2013; **13**: 134.