INTRODUCTION
Knowledge about the concepts of research methodology and statistics is essential for head and neck oncologists. Understanding the literature help them to practice evidence-based medicine as applied to head and neck surgery and oncology. This article plans to give an overall summary of the basic concepts in the study design, measures of treatment effects, hypothesis testing, diagnostic tests, assessment of outcomes, and statistical tests.

LEVELS OF EVIDENCE
A systematic review summarizes the medical literature on a topic, with explicit methodology and gives a critical appraisal of studies. Meta-analysis is a type of systematic review where quantitative methods are used to combine the results of independent studies to get summary data and statistics. A meta-analysis of randomized clinical trials carries the highest weightage in terms of evidence. Table 1 shows a summary of different types of studies and the levels of evidence they offer.

STUDY DESIGNS
There are two basic designs in studies, namely observational studies and experimental studies. In observational studies, participants are observed without any interventions, whereas in experimental studies, an intervention is involved. Those experimental studies involving human subjects are called trials.

Research studies may be prospective when the direction of study is forward from the inception of the cohort and the events of interest happen after the onset of the study. In retrospective studies, the direction of the study is backward from the cases, and the events of interest happen before the onset of the study. Cross-sectional studies are surveys conducted at a single point in time. Longitudinal studies follow the same individuals over multiple points in time.

Research studies are susceptible to bias, confounding and chance. Bias includes a nonrandom systematic error in the design or the conduct of the study. It’s not usually intentional. Bias can affect a study at any phase, including patient’s selection (selection bias), follow up (nonresponders’ bias), the outcomes (detection, recall, interviewer bias). Selection bias is common in head and neck literature when dissimilar groups are compared. A confounder is a variable that has got independent associations with both the predictor (independent) and the outcome (dependent) variables. This may potentially distort their relationship.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms. For reprints contact: reprints@medknow.com

How to cite this article: Thankappan K. Basic concepts in research methodology and biostatistics for head and neck oncologists. J Head Neck Physicians Surg 2019;7:37-44.

Submitted: 28-Dec-19 Accepted: 19-Jan-20 Published: 21-Feb-20
The common confounders in clinical research include age, gender, socioeconomic status, and comorbidities. Chance can lead to conclusions that are not valid based on the probability of errors.

The problems of bias, confounding and chance can be reduced by proper study design and statistical analysis. Randomization minimizes the selection bias and distributes the confound equally. Blinding and matching can also minimize the confounding. Post hoc analysis like stratified or multivariate methods can also rectify the confounders. Chance can be minimized by adequate sample size and our calculations. Proving a cause–effect relationship is more difficult than to suggest an association. Inference of causation requires data from nonobservational studies like randomized controlled studies, providing a biologically plausible explanation, showing a large effect size, reproducibility of findings, a temporal relationship between the cause and effect and a dose–response relationship.

**OBSERVATIONAL STUDIES**

This study design includes case series, case–control studies, cross-sectional surveys, and cohort studies.[3]

A case series is a retrospective descriptive study of a group of patients or individuals with some notable characteristics or a series of patients who have undergone a particular treatment. A case series with single patient details and outcomes are often referred to as case reports. His series are easy to conduct. But they are often anecdotal and subject to multiple biases and lack a hypothesis. They are often considered hypothesis-generating studies and are not conclusive.

A case–control study is when the investigator compares patients without an outcome of interest (Cases) and those

without the outcome (control) in terms of possible risk factors. The effects of such a study are often reported as odds ratio. They are efficient for the evaluation of unusual conditions and outcomes. They are relatively easy to perform. However, the challenge is to identify an appropriate control group. Existence of high-quality medical records is also essential for the conduct of the case–control study. They are susceptible to selection and detection bias.

Cross-sectional surveys are used to determine the prevalence of the disease or to bring out the existing associations in patients with particular condition at a certain point of time. The prevalence is the number of individuals with the condition divided by the total number of individuals at one point in time. Incidence refers to the number of individuals divided by the total number of individuals with the condition over a defined period. The prevalence data are obtained from a cross-sectional study as a proportion, but incidence data are obtained from a prospective cohort study and time value is necessary for the denominator. Surveys are also performed to find the preferences and treatment patterns. They may have unique challenges in terms of adequate response rate, representative samples and acceptability bias.

A cohort study is one in which the population of interest is followed prospectively to determine the outcomes and the associations with risk factors. Retrospective Cohort studies are when the members are identified based on medical records. The follow-up period for a retrospective cohort study is in the past. Cohort studies are ideal for analyzing the incidence and course of a disease because they are longitudinal. The effects of a cohort study are reported as Relative Risk (RR). Since the common cohort studies are prospective, the follow-up and the data quality can be optimized. Selection bias can be minimized. These studies are logistically difficult, takes long duration, expensive and inefficient for assessment of unusual outcomes or diseases.

**EXPERIMENTAL STUDIES**

There is a control arm in experimental studies. Controls can be concurrent, sequential (cross-over trials) or historical.

A Randomized controlled trial (RCT) is the gold standard in the field of clinical evidence, as it provides the most valid conclusions by minimizing the bias and confounding. A RCT necessitates the need for a protocol document that provides the eligibility criteria, sample size, informed consent, randomization, conditions for stopping the trial, blinding, measurement, monitoring of the compliance, assessment of safety, and the data analysis. The allocation is at random, and hence, it avoids selection bias. The confounders are theoretically equally distributed among the

| Grade of recommendation | Level of evidence | Type of study                                      |
|-------------------------|-------------------|---------------------------------------------------|
| A                       | 1a                | Individual randomized controlled trials            |
| B                       | 2a                | Systematic review of (homogeneous) cohort studies  |
| B                       | 2b                | Individual cohort study/low-quality randomized studies |
| C                       | 4                 | Case series, low-quality cohort or case-control studies |
| D                       | 5                 | Expert opinions based on nonsystematic reviews of results or mechanistic studies |
groups. The procedure of blinding reduces the performance, detection, interviewer, and acceptability bias. Blinding may be applied at four levels, the participants, the investigators, outcome assessors, and the analysts. Intention-to-treat analysis principle minimizes the nonresponder and transfer bias. This principle states that all patients should be analyzed within the treatment group to which they were randomized, to preserve the goals of randomization. Sample size determination ensures adequate power. Though RCT is considered the best model, the disadvantages include the expense, the logistics, and the time required for completion. RCT requires the clinical equipoise or the equality of treatment options in the clinician’s judgment and the interim stopping rules should be stated to avoid harm. The adverse events need to be evaluated and taken care; informed consent should be applied. RCTs, though considered to have an excellent internal validity, the external validity or the generalizability in the population may be less.

## RESEARCH ETHICS

The design and conduct of the research studies should be done under strict ethical premises. Informed consent is an integral part of all studies. The Nuremberg code and the Declaration of Helsinki should be familiar with the investigators as applicable to ethical issues of risks and benefits, privacy protection, and respect for autonomy.

## HYPOTHESIS TESTING

It allows generalizations from a sample to the population from which it was derived. It confirms or refutes the statement that the observed findings are not by chance alone, but there exists a true association between the variables under consideration. The Null hypothesis says that there is no significant association between the variables and the alternative hypothesis states that there exists a significant association. We cannot reject the hypothesis if the findings of the study are not significant. If the findings are significant, we reject the null hypothesis and accept the alternative hypothesis.

A 2 × 2 table can be constructed about the possible outcomes of the study. The study inference is correct if a significant association is found when there is a true association or if no association is found when there is no true association. There can be two types of errors. A Type-I error or alpha (α) error, when a significant association is found when there is no true association (a false positive study that rejects a true null hypothesis).

A Type-II error or beta (β) error wrongly conclude that there is no significant association (a false-negative study, that rejects an alternative hypothesis that is true) [Table 2].

### Table 2: Hypothesis testing

| Study                  | Truth                      |
|------------------------|----------------------------|
|                        | No association | Association present |
| No association         | Correct         | Type-II (β) error |
| Association present    | Type-I (α) error | Correct           |

The α level refers to the probability of a Type-I error. The alpha level of significance is usually set at 0.05. This means that we accept the finding of a significant association if there is <5% possibility that the observed association was due to chance alone. The P value that is derived by a statistical test is a measure of the strength of the evidence provided by the data favoring null hypothesis. When the P value is less than the alpha level, the null hypothesis may be rejected, and the study result is significant.

## THE P VALUES AND 95% CONFIDENCE INTERVAL

Alternatives to the traditional “P value” is the use of 95% confidence interval (CI). These intervals convey the information about the significance of the study data. There will be no overlap in 95% CIs if they are significantly different. The magnitude of the differences and the precision of measurement is indicated by the range of the 95% CI. The P value is often interpreted as either significant or not; but, the 95% CI provides a range of values that allows interpreting the implications of the result. The P values do not have any units, but the CI takes the unit of the variable of interest. The P values convey the statistical significance only, whereas (the CI conveys the statistical significance (the CIs do not overlap), clinical significance (the magnitude of the values) and the precision (the range of the CIs).

## POWER

It is the probability of getting a significant association if it genuinely exists. It is defined as 1-the probability of a Type II error (β). Usually, the power is set at ≥80%. This means that there is ≤20% chance that the study will show no significant association when there is a true association.

When a study shows a significant association, the error of concern is the α error, and when a study shows no significant association, the error of concern is β error expressed by the power. In a study that shows no significant effect, there may genuinely be (1) no significant effect, (2) there may be a significant effect, but the study was underpowered because the sample size was too small, or (3) the measurements were too imprecise. Thus, it is essential to report the power when the study demonstrates no significant effect.

The four critical factors relevant for power analysis are α, β, effect size, and the sample size. Effect size is the difference...
to be detected with the given $\alpha$ and $\beta$. It is based on a clinically meaningful difference. For comparing two groups, effect sizes are usually defined in dimensionless terms, based on the difference in means divided by the pooled standard deviation. The power of the study is diminished by small sample sizes, small effect sizes, and large variances.

SAMPLE SIZE AND POWER

Sample size calculations are essential when a study is being planned. Conventionally, the power is taken as 80%; alpha is at 0.05, the effect size and variance are derived from pilot data or the literature.

DIAGNOSTIC PERFORMANCE OF A TEST

There are four possible situations in diagnostic testing [Table 3]:

1. True positive: When the test is positive, and the disease is present
2. False positive: When the test is positive and the disease is absent
3. True negative: When the test is negative and the disease is absent
4. False negative: When the test is negative, and the disease is present.

The sensitivity of a test is defined as the percentage of patients with the disease, who are tested positive (the true-positive rate). A test with 98% sensitivity means that, of 100 patients with the disease, 98 will have a positive test. Sensitive tests will have a low false negative rate. A negative test result of a highly sensitive test rules out the disease.

The specificity of a test is the percentage of patients without the disease, who are tested negative (the true-negative rate). A test with 90% specificity means that of 100 patients without the disease, ninety will have the test negative. Specific tests have a low false positive rate. A positive test result of a highly specific test rules in the disease. Sensitivity and specificity can be calculated when the results of a diagnostic test are compared to that of the “gold standard” test for the diagnosis in the same set of patients.

### Table 3: Diagnostic test performance

| Disease positive | Disease negative |
|------------------|------------------|
| Positive test    | a (true positive) | b (false positive) |
| Negative test    | c (false negative) | d (true negative) |

Sensitivity: $a/(a + c)$, Specificity $= d/(b + d)$, Accuracy $= (a + c)/(a + b + c + d)$, False-negative rate $= 1 −$ sensitivity, False-positive rate $= 1 −$ specificity,

Likelihood ratio $(+)= \text{Sensitivity}/\text{false-positive rate}$, Likelihood ratio $(-)= \text{False-negative rate}/\text{specificity}$,

Positive predictive value $= ([\text{Prevalence}](\text{sensitivity})/([\text{prevalence}](\text{specificity}) + (1−\text{prevalence})(1−\text{specificity}))$, Negative predictive value $= ([1−\text{prevalence}](\text{specificity})/([1−\text{prevalence}](\text{specificity}) + (\text{prevalence})(1−\text{specificity}))$.

Figure 1: The relationship between sensitivity and specificity of a test can be shown on a receiver operating characteristic (ROC) curve

Likelihood ratio

Sensitivity and specificity are combined into a single parameter. It is the probability of a true positive divided by the probability of a false positive.

Positive predictive value

It is the probability of the patient having the disease when the test is positive.

Negative predictive value

It is the probability of the patient not having the disease when the test is negative. Positive and negative predictive values require an estimate of prevalence in the population.

There is a trade-off between sensitivity and specificity. A positivity criterion can be selected with a low false negativity rate, to optimize sensitivity or a low false positivity rate, to optimize specificity. Practically, positivity criteria are selected based on the consequences of a false positive or a false negative diagnosis. When the consequences of a false-negative diagnosis outweigh that of the false-positive diagnosis of a condition, a more sensitive criterion is chosen.

The relationship between sensitivity and specificity of a test can be shown on a receiver operating characteristic (ROC) curve [Figure 1]. It has the true positive rate (sensitivity) on the y-axis, and the false-positive rate $(1 −$ specificity) on the x-axis plotted at each possible cutoff. The area under the ROC curve gives the overall diagnostic performance of the test.

MEASURES OF LIKELIHOOD

Probability indicates how likely an event is to occur. It is a number between 0 and 1. It is the number of events per number of trials. The probability of tails on tossing a coin is 0.5.
Odds
It is the ratio of the probability of an event occurring to the probability of the event not occurring. The odds of getting tails on a coin toss is 1 (0.5/0.5).
Odds and probability are related, Odds = probability/1− probability.

Relative risk
It is determined from a prospective cohort study. It is calculated as the incidence of the disease in the exposed/incidence of the disease in the nonexposed cohort [Table 4].

Odds ratio
It is determined from a retrospective case–control study. The incidence cannot be calculated in such a study. It is the ratio of the odds of having the disease in the study group to the odds of having it in the control group.

Risk factors are the factors that are likely to increase the incidence, prevalence, morbidity, or mortality of the disease. The RR reduction, the absolute risk reduction, and the number needed to treat help to quantify the effect of a factor to reduce the probability of an adverse event. The RR increase, the absolute risk increase, and the number needed to harm quantify the effect of a factor that increases the probability of an adverse outcome.

QUALITY OF LIFE STUDIES AND ASSESSMENT OF OUTCOMES
The measures could be generic or condition specific.[9] An example of a generic measures instrument is Short Form-36 used to evaluate health status or health-related quality of life. Examples of head and neck cancer-specific measures include EORTC QLQ-H and N35 and FACT-HN.[10]

Outcomes’ assessment measure/tool development
The steps include identifying the construct, devising items, scaling responses, selecting items, forming factors, and creating scales. Validation of the instrument has to be carried out to test the reliability and validity.

Reliability denotes the repeatability of an instrument. Interobserver reliability and intraobserver reliability have to be tested. Interobserver reliability refers to the repeatability of the instrument when used by different observers, and intraobserver reliability refers to the repeatability by the same observer at different time-points. Test–retest reliability can be evaluated by using the instrument to assess the same patient on two different occasions. These results can be statistically tested using kappa statistic or intraclass correlation coefficient.

Validity sees whether the instrument measures what it is intended to measure. Content validity evaluates the instrument to see if it is representative of the characteristic being measured. An expert consensus opinion is often resorted to (face validity). Construct validity evaluates whether an instrument follows accepted hypotheses (constructs) and produces results consistent with theoretical expectations. Criterion validity compares it with an accepted, “gold-standard” instrument.

BASIC CONCEPTS IN BIOSTATISTICS
Types of data
Data can be classified as given in Figure 2.

Categorical
Categorical data represents characteristics. They indicate types or categories. Examples for categorical data include gender and a dichotomous (yes/no, success/failure) outcome. Categorical data are usually described in terms of proportions or percentages. They are reported in tables or bar charts. Categorical data can be nominal or Ordinal.

Nominal data represent discrete units and are used to label variables that have no quantitative value. There is no order. Therefore if you changed the order of its values, the meaning would not change. Examples: Gender (Male/female), languages spoken (English/Hindi/French).

Interval
Ratio
Continuous
discrete
numerical
nominal
ordinal

table 4: Measures of Likelihood
|          | Adverse events | No adverse events |
|----------|----------------|-------------------|
| Experimental group | a              | b                 |
| Control group | c              | d                 |

CER= c/(c + d), EER= a/(a + b), CEO= c/d, Experimental event odds (EEO)= a/b, RR= EER/CER, OR= CEO/CER, RRR = (EEO− CER)/CER, ARR = EER - CER, NNT= 1/ ARR, CER: Control event rate, EER: Experimental event rate, CEO: Control event odds, RR: Relative risk, OR: Odds ratio, RRR: Relative risk reduction, ARR: Absolute risk reduction, NNT: Number needed to treat
Ordinal data takes an order. Example is cancer staging. Ordinal data are also reported as proportions or percentages and are reported in tables or bar charts.

Numerical data can be Discrete and Continuous. The scale of measurement is called discrete when a numerical observation can have only integer values.

Discrete data values are distinct and separate. This type of data cannot be measured, but it can be counted. For example, the number of tails in 100 coin flips.

Continuous Data values cannot be counted, but they can be measured. For example, the height of a person. Continuous data are usually described in terms of mean and standard deviation and can be reported in tables or graphs. Continuous data may be interval data or ratio.

Interval values represent ordered units that have the same difference. Therefore we speak of interval data when we have a variable that contains numeric values that are ordered and where we know the exact differences between the values – for example, temperature. The problem with interval data is that they do not have a “true zero.” There is no such thing as no temperature. We can add and subtract with interval data, but we cannot multiply, divide or calculate ratios. Because there is no true zero, a lot of descriptive and inferential statistics can’t be applied.

Ratio values are also ordered units that have the same difference. Ratio values are the same as interval values, with the difference that they do have an absolute zero. Good examples are height, weight, and length.

**MEASURES OF CENTRAL TENDENCY AND DISPERSION**

Data can be summarized in terms of measures of central tendency. Mean, median, and mode are commonly used. Measures of dispersion include range, standard deviation, and percentiles. Data can be in the form different distributions, such as the normal (Gaussian) distribution, skewed distributions, and bimodal distributions.

**UNIVARIATE ANALYSIS**

Univariate, or bivariate, analysis studies the relationship of a single independent and a single dependent variable.

**Comparison of means**

Statistical tests for comparing means of normally distributed continuous variables include the Student’s t-test for two independent groups. For paired samples, it is the paired t-test. In case of continuous or categorical variables, when they are not normally distributed, nonparametric tests are used. Nonparametric tests for comparing medians in two independent groups include the Mann-Whitney U test (also called the Wilcoxon sum test). The Wilcoxon signed-rank test is used to compare paired samples [Table 5].

Analysis of variance (ANOVA) is used for comparison of the means of three or more independent groups when the data are normally distributed. F-test is the result, P value indicates whether there is an overall significant difference. To analyze whether there are significant differences between individual groups, Post hoc tests are used to perform multiple pairwise comparisons between the groups (tests like Bonferroni, Tukey). The Kruskal–Wallis test is used to compare medians of three or more independent groups in situations where the data do not follow a normal distribution. It is a nonparametric alternative to ANOVA. Repeated-measures ANOVA is used for normally distributed variables when the study has matched individuals included. The nonparametric test to compare medians among three or more matched groups is called the Friedman test.

**Comparison of proportions**

Pearson Chi-square test can be used to compare proportions for categorical or ordinal variables, for two or more independent groups and the Fisher exact test when expected

---

**Table 5: Summary of Statistical tests**

| Type of data | Number of groups | Independent groups | Paired samples |
|--------------|------------------|--------------------|---------------|
| Continuous   |                  |                    |               |
| Normal       | 2                | Student t-test     | Paired t-test |
| Nonnormal    | 2                | Mann-Whitney U-test| Wilcoxon signed-rank test |
| Normal       | 3 or more        | ANOVA              | Repeated-measures ANOVA |
| Nonnormal    | 3 or more        | Kruskal-Wallis test| Friedman test |
| Ordinal      | 2                | Mann-Whitney U-test| Wilcoxon signed-rank test |
|              | 3 or more        | Kruskal-Wallis test| Friedman test |
| Nominal      | 2                | Fisher’s Exact test| McNemar test |
|              | 3 or more        | Pearson Chi-square test| Cochran Q test |
| Survival     | 2                | Log-rank test      | Conditional logistic regression |

ANOVA: Analysis of variance
cell frequencies are small (five or less). For matched samples, the McNemar test can be used for two variables, and the Cochran Q test is used for three or more variables. Table 5 gives a summary of statistical tests.[11]

**Determination of associations**
Pearson product-moment correlation can be used for normally distributed continuous variables, Spearman rank-order correlation (rho) for nonparametric variables, and Kendall rank-correlation for ordinal variables.

**SURVIVAL ANALYSIS**
It is used to analyze data when the outcome of interest is “time to event.”

**Event**
A group of patients is followed up to see if they come across the event. In cancer studies, the events may be death, recurrence, progression. It is essential to state the start time, which is usually the date of diagnosis, date of surgery, or date of randomization.

**Censoring**
An individual is censored when the event of interest does not occur during the study period. Survival is the time from an individual’s entry into the study until the event or until the time of censoring.

**Kaplan–Meir survival analysis**
Survival data are commonly analyzed with the use of the Kaplan–Meier method. The survivorship is calculated every time that an event occurs but not at time of censoring. It is used when the actual date of the endpoint is known. Endpoints that have not been reached are treated as censored at the date of the last follow-up for the analysis. Survival analysis produces a life table showing the number of events occurring within the time intervals and the number of individuals withdrawn during the interval. A survival curve can be plotted to show the percentage of individuals that remain event free on the vertical axis and the follow-up time on the horizontal axis. Figure 3 shows a Kaplan–Meier survival curve showing overall survival for 100 patients of Basaloid squamous cell carcinoma of the larynx. X-axis shows the survival duration in months and the Y-axis shows the overall survival probability.[12]

The graph plotted between estimated survival probabilities/estimated survival percentages (on Y axis) and time past after entry into the study (on X axis) consists of horizontal and vertical lines. The survival curve is drawn as a step function: the proportion surviving remains unchanged between the events, even if there are some intermediate censored observations. It is incorrect to join the calculated points by sloping lines. Survival for different groups can be compared (univariate analysis) by the log-rank test.

In cancer studies, the common survival outcomes are overall survival, disease-free survival, recurrence-free survival and progression-free survival depending on the endpoints chosen.

**MULTIVARIATE ANALYSIS**
Multivariate analysis studies. Studies the relationship between multiple variables. Regression is the standard method used. There would be one outcome or dependent variable (Y), an explanatory (X) and a set of independent variables (X’s). When the outcome variable is continuous, linear regression is used. Multiple regression fits data to a model that defines Y as a function of two or more explanatory variables or predictors. Logistic regression is used when the outcome variable is binary or dichotomous and is commonly used for multivariate analysis for nontime-related outcomes. Cox proportional-hazards regression is commonly used for time-to-event data or survival analysis. Such modeling is commonly used to predict outcomes or to establish independent associations (controlling for confounding and colinearity) among predictor or explanatory variables. The goal of multivariate analysis is to identify, from among the many patient and surgical variables observed and recorded, those most related independently to the outcome.

**SUMMARY**
Knowledge about the basic concepts and its application in oncology is essential for the practice of an academic head and neck surgeon. This article is a primer for understanding the research methodology and biostatistics.
Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

Disclosure
This material has never been published and is not currently under evaluation in any other peer reviewed publication.

Ethical approval
The permission was taken from Institutional Ethics Committee prior to starting the project. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent
Informed consent was obtained from all individual participants included in the study.

REFERENCES
1. Delgado-Rodríguez M, Sillero-Arenas M. Systematic review and meta-analysis. Med Intensiva 2018;42:444-53.
2. Panesar SS, Bhandari M, Darzi A, Athanasiou T. Meta-analysis: A practical decision making tool for surgeons. Int J Surg 2009;7:291-6.
3. Martins RP, Buschang PH. What is the level of evidence of what you are reading? Dental Press J Orthod 2015;20:22-5.
4. Concato J. Study design and “evidence” in patient-oriented research. Am J Respir Crit Care Med 2013;187:1167-72.
5. Dowrick AS, Tornetta P 3rd, Obremskey WT, Dirschl DR, Bhandari M. Practical research methods for orthopaedic surgeons. Instr Course Lect 2012;61:581-6.
6. Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Practical tips for surgical research: Research questions, hypotheses and objectives. Can J Surg 2010;53:278-81.
7. Boushey CJ, Harris J, Bruemmer B, Archer SL. Publishing nutrition research: A review of sampling, sample size, statistical analysis, and other key elements of manuscript preparation, Part 2. J Am Diet Assoc 2008;108:679-88.
8. Bruemmer B, Harris J, Gleason P, Boushey CJ, Sheean PM, Archer S, et al. Publishing nutrition research: A review of epidemiologic methods. J Am Diet Assoc 2009;109:1728-37.
9. Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, de Vet HC. Outcome instruments: Rationale for their use. J Bone Joint Surg Am 2009;91 Suppl 3:41-9.
10. Ringash J. Survivorship and quality of life in head and neck cancer. J Clin Oncol 2015;33:3322-7.
11. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: A primer for orthopaedic surgeons. J Bone Joint Surg Am 2004;86:607-20.
12. Thankappan K. Basaloid squamous cell carcinoma of the larynx – A systematic review. Auris Nasus Larynx 2012;39:397-401.