Confounding and Effect Modification in Research

Dharmashree Satyarup[1], Manish Kumar[2], Radha Prasanna Dalai[3], Karishma R. Rathor[4]

[1]Professor, Department of Public Health Dentistry, Institute of Dental Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, [2]Professor & Head, Department of Public Health Dentistry, GITAM Dental College and Hospital, Visakhapatnam, Andhra Pradesh, India, [3]Senior lecturer, Department of Public Health Dentistry, Institute of Dental Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, [4]Tutor, Department of Public Health Dentistry, Institute of Dental Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India

Abstract

An association that is not causal (Alternate explanations) is seen between outcome and exposure whenever the study has confounding factors, bias and random error/chance error. Bias in any study leads to systematic variation of inferences or results or interpretations from the true picture. One of the most important bias encountered during research is confounding. Confounding is seen when a variable is observed to be associated with both the exposure and the outcome but is not a part of the causal pathway. Confounding factors, when present, in any study, are the "Nuisance" and can be the cause in part or in full of the observed association between the disease and exposure. Effect modification, on the other hand, is seen when various effects are brought about among different subgroups by an exposure and this can be handled by doing stratification. It is the outcome that is linked to the effect modification and not the exposure. This article will define and discuss in detail confounding, and the concept of effect modification so that drawing conclusions of a study can be done after considering these errors.

Keywords: *Effect modification, Matching, Restriction, Confounding factors.*

Introduction

Bias is defined as any deviation of results or inferences from the truth, or processes leading to such deviation. Any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth.[1] There are a plethora of biases mentioned in the literature which an investigator might encounter during research.[2] All types of study design, be it case-control, cohort,or experimental are prone to one or more types of bias. A false picture of association is observed due to bias or they can even hide a true association when it is present.

Internal validity of a study is threatened by the presence of confounding, random error/chance error and bias. A researcher's unintentional mistake usually leads to bias, whereas, confounding occurs due to the presence of an extra variable that affects the outcome positively or negatively apart from the exposure. Random variation/ chance can be reduced to a greater extent by increasing the sample size; but, on the contrary, bias and confounding cannot be reduced by merely increasing the sample size.[3] This paper emphasizes only on confounding and the bias associated with it. The paper also attempts to throw some light on effect modification.

Confounding: The word is derived from Latin "confundere" meaning to mix or muddle. It means a distortion in an association that is seen when the exposal factor of interest is muddled with other factors that are

Corresponding Author:
Dr. Dharmashree Satyarup
Professor, Department of Public Health Dentistry, Institute of Dental Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India
e-mail: dharmashree_s@yahoo.com

related to the outcome.[1] As a researcher, we ought to be careful to avoid confounders in a study as much as possible otherwise it may lead to "False Positive" (Type–1 Error) ie, conclusion of a spurious relationship. In simple terms, it undermines the internal validity of the study, as the confounder rather than the exposal agent seems to be associated with all the effects observed.

Example: A researcher wanted to find out the association (primarily) between physical activity (Exposure) and lung cancer (Outcome). Confounding factor in the given example is age as it is associated with exposure (Old people tend to be more physically inactive), and also associated with outcome (Lung cancer is at higher risk to develop among old age).However, race cannot be considered a risk factor for lung cancer. Adjustments to accommodate variables that might not be cofounders can bring in bias and compromise the precision of a study.

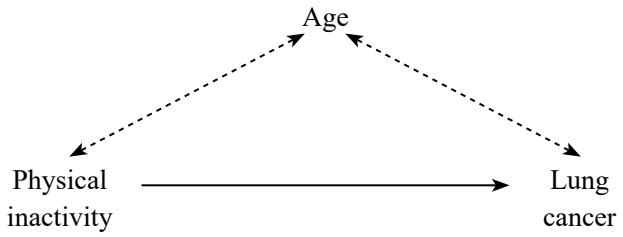


Figure 1: Understanding Confounding:

To understand confounding in a simplified manner, Rothman and others have investigated and illustrated an association between Down's Syndrome and birth order.[4]

Figure 1 shows that with an increase in birth order, the chances to develop Down's syndrome among the children also increases roughly around 4 – fold. This type of analysis did not take other potential risk factors/ confounders into consideration, making a researcher believe that birth order is associated with Down's syndrome.

But, when maternal age was used for children with Down's syndrome, it demonstrated that the age of the mother was more important in the causation of the disease i.e. as the age of the women increased to give birth, the prevalence of Down's syndrome also increased proportionally. (Figure 2)

Women who are mothers for the first time are much younger than women having their fifth child. In other words, maternal age is mixed up with the birth order of the child. There is a stronger correlation between maternal age and Down's syndrome prevalence than the birth order of the children and Down's syndrome. In lieu of this, it can be stated that the occurrence of Down's syndrome in relation to the order of birth is confounded by the age of the mother.

Investigators also tried to find if maternal age was confounded by birth order i.e. if birth order had an independent effect on the causation of Down's syndrome. Rothmansorted this by putting both effects i.e. maternal age and birth order v/s Down's syndrome prevalence. (Figure 3) (4).

The figure amply demonstrates that Down's syndrome prevalence is not associated (has no impact) with increasing birth order when observed from the side. But, when the figure is observed from forwards to backward, for every order of birth, an increase in the age of the mothers at all five levels of birth order can be observed, i.e. even after controlling for birth order, there is a strong association between increased maternal age and Down's syndrome among the children. Hence, the apparent association between birth order and Down's syndrome observed in Figure-1 was due to the confounding effect of maternal age completely. On the contrary, the association between Down's syndrome prevalence and increasing maternal age is not confounded by the birth order of the children.

Confounding Factor Criteria:

For a factor to be termed confounder, the following three criteria have to be met:[5]

1. Both the risk factor under study and outcome should be related to the confounder.
2. The groups under comparison should have the confounding factor among them, albeit it is distributed unequally.
3. A confounder cannot be an intermediary step in the causal pathway from the exposure of interest to the outcome of interest. In the pathway of causation from exposure to the outcome, the confounding factor cannot be a transitionary step.

The first two of these conditions can be tested with data. The third is more biological and conceptual.

To ascertain if the association shown between the exposure under study and the variable is independent, there should be a plausible biological or social explanation. These decisions should be best made

based on the basic science, psychological, sociological, or clinical data (i.e. non – epidemiological). In a nutshell, the confounding factor should precede the outcome of interest and not be an intermediary step. For example: To prove a relationship betweenlack of exercise and cardiovascular diseases, overweight might be an intermediary step between lack of exercise and cardiovascular diseases. Hence, bodyweight should not be controlled.[6]

However, bodyweight can be considered as a confounder if it is not a part of the causal pathway. Hence, controlling these intervening variables during analysis may mask a true relationship between exposure and disease.

Most of the diseases have multiple risk factors i.e. confounders. Confounder in a research study can be of any of the types mentioned below.[7,8]

- **Confounder as risk factor:** For example – In an epidemiological study to associate diabetes and exercise, age as it is a risk factor for diabetes, can be a confounding factor.
- **Confounder as a preventive factor:** For example – People who exercise regularlyon the prescription of aspirin, and being on aspirin decreases the risk of heart disease. Thus, the benefit of exercise would be exaggerated for reducing heart disease by regular use of aspirin (Confounder).
- Confounder as ***a marker or a surrogate***: For example – For most of the systemic diseases, socio-economic status would act as a confounder. Lower socioeconomic status (Marker) forms a collection of factors that are not understood completely and hence pose a high risk for many systemic diseases.

Effects caused due to confounding:

- An apparent association may be in part or wholly caused by a confounder.
- Can cause over-estimation of true association (Positive Confounding) or under-estimation of true association (Negative Confounding)

Identifying Confounders:[9]

- The best way to is to observe for the difference in the association before and after modifying for the confounding agent in a study. If the association between the two measures differs by ten percent or more, it confirms the presence of a confounder. On the contrary, if the difference is lesser than ten percent, then there is little or no role of the confounding agent.
- With the prior knowledge or experience of the investigator, if the variable is common to the exposure and outcome of interest, then the variable is believed to be a confounder.
- Test of hypothesis can be done to establish if the confounder is present in exposure and outcome.

Controlling for confounding during research:

The strategies employed to control confounding are:

1. Restriction
2. Matching
3. Randomization

} During the designing of the study

4. Stratified analysis
5. Multi-variable regression analysis

} During the analysis of the study

1. Restriction: A factor must be unequally present among groups being compared to be considered a confounding factor. Hence, a common method used to avoid confounding bias in the research is to restrict the selection and entry of participants by adhering to strict inclusion and exclusion criteria having the same levels of confounding factors.[10]

To understand simply, let us take an example of smoking and lung cancer, with age and gender being confounders of concern to the investigation. Restriction in this example, would be to select only male participants in the age group of 50 – 60 years. This will minimize the confounding to a greater extent as both the groups are comparable with similar characteristics.

Limitations to restriction technique:

- The problem of sample size might arise as the eligibility of the number of subjects will be reduced.
- If restriction is not done narrowly, then the problem of residual confounding might arise. The distortion that is still observed in study design or analysis even after controlling of the confounder is called residual confounding. For example, in an investigation of the association between smoking and lung cancer, if the investigators choose male participants between the age of 50 – 75 years, the occurrence of lung cancer varies widely in this age range as cancer might be more among the older adults.

- The effect of factors that have been adjusted for cannot be evaluated. For instance, the effect of men aged 50 – 60 years on lung cancer cannot be evaluated.
- Generalizability is limited by restriction considerably. For example, if a study only men are included, then the results cannot be generalized to women.[10]

2. **Matching[11,12]:** Matching is another straight-forward approach and commonly used method for reducing confounders in the research design especially in case-control studies. For instance, age groups among the cases and controls have to be matched in research to study the association between physical inactivity and cardiovascular disease. If not done, younger aged people might be mainly controlled strata and have fewer cases (suffering from heart disease) among them, on the contrary, older aged individuals might be forming up cases strata with fewer controls. Thus, by matching the age stratum, i.e. for every case aged 40–50 years, a matched control aged 40 years is chosen, thereby reducing the confounding to a great extent. As the analysis of unmatched study groups is different from than when matching is done, researchers should include matching during the study design to prevent confounding.

Applications of matching:

- Multi-faceted and complex variables need to be controlled (Example: Environmental factors, genetic factors, etc)
- In a case-control study, where the sample size for cases is less with many possible matched controls.

Limitations of Matching:

- The sample size is sometimes inadequate.
- Might be expensive and takes a long time to find exact matches
- Effect of the factors that are matched cannot be evaluated

3. **Randomization[13]:** To control both known and unknown confounders, the most powerful and ideal method is randomization. Selection bias by the investigator is prevented/reduced considerably by randomization, as it allocates subjects to control or experimental groups. Before randomization is conducted, the investigator should ensure all the participants are equal at baseline, to create comparability across groups. Similar distribution of known and unknown confounders is ensured by randomization to the control and experimental groups. Randomization also guarantees that any differences found between the two groups is due to chance and not due to the investigator.

4. **Stratified Analysis[13,14]:** Stratification permits an investigator to examine the difference of the confounder among the strata between outcome and exposure. For example: alcohol and smoking consumption, socioeconomic status, age, etc. In a hypothetical study, to find the association between air pollution and Chronic Obstructive Pulmonary Disease (COPD), smoking was the confounder to be controlled. The population under study was stratified based on their smoking status. Assessment of the association between COPD and air pollution can then be made within each stratum. Stratification allows us to examine the effect of smoking on the association between air pollution and COPD.

   The issue of simply creating strata, in general, is that, there will be more individuals in some strata giving an accurate estimate of the association. But strata with fewer individuals, the association estimate might not be accurate. Hence, the weighted average is calculated so that strata with more data gets greater weight. Mantel-Haenszel Method is the most common weight scheme used during the analysis. A single, summary measure of association providing a weighted average of the relative risk or odds ratio is produced by the Mantel-Haenszel method across the different strata of the confounding factor.All the usual statistical packages allow calculation of the estimates by the Mantel-Haenszel method. The major limitation of the stratified analysis is, if there are multiple confounding factors then it is very difficult to control all of them simultaneously.

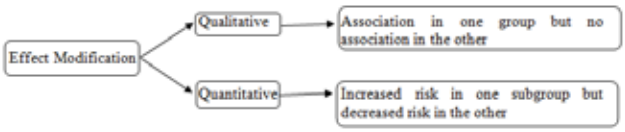
5. **Multi-Variable Regression Analysis[10]:** If only one or two confounders have to be controlled, the best way is to conduct stratified analysis. Multivariate analysis is the only possible way to control simultaneously multiple confounders or if their grouping is large. For example: In a hypothetical study to find an association between smoking and oral cancer, we can control simultaneously other confounders (covariates) in the same model like age, gender, alcohol consumption, gutkha consumption, exposure to sunlight, sharp teeth, sepsis, etc. In

regression models (Linear and logistic), multiple confounders can be investigated at the same time by adjusting the odds ratio. Adjustment is the process of accounting for confounders followed by comparing simple and multiple regression results to get a clear picture of the quantum of distortion the confounder has made in the relationship between the exposure and outcome.

Effect Modification (Interaction): Effect modification, is seen when various effects are brought about among different subgroups by an exposure and this can be handled by doing stratification. Effect modification is associated only with the outcome of the study, but not the exposure. It is not considered a nuisance as it often brings to light significant information. The presence of a third factor determines the strength of effect an exposure on an outcome. For example, when a new drug is introduced called drug A. If this drug A works only in males but does not work in females, then gender is the effect modifier and its an example of effect modification.

Confounding is considered as muddling of effects, on the other hand, effect modification is a property of the effect.[15] Effect modifier is defined as a factor that modifies the effect of a putative causal factor under study.Effect modification is detected by varying the selected effect measure for the factor under study across levels of another factor. Hence, effect modification is a change in the effects caused by one or more factors due to the presence of the remaining factors.[1]

Types of effect modification[16]: Effect modification is generally classified into two types: Qualitative and Quantitative.



When the direction of effect (nature or quality) depends on the effect modifier, it is qualitative effect modification. On the other hand, if the strength (quantity) of association differs across strata of the effect modifier without a change in the direction of the effect modifier, it is quantitative effect modification.

Reasons for finding effect modification during research:

- For preventive actions, high-risk sub-groups can be defined appropriately.
- The precision of effect estimation is increased among the groups that would have been differently affected.
- The ability to compare across studies having different proportions of effect modifying groups is increased.
- Causal – The hypothesis of a disease can be developed.

Effect Modification Identification: The potential important effect modifier should not be matched during the study, as its effect on the outcome then cannot be determined. During the design and conducting a research study, the method employed to identify effect modifiers are: prior knowledge of potential effect modifiers and power of the study. The method usually employed to identify effect modifiers in research is generally done during the statistical analysis of the study i.e. stratification and regression (Already mentioned).

Conclusion

When a variable is observed to be associated with both exposure and the disease, confounding occurs causing confusion regarding the outcome. Hence it is imperative on the investigator's part to identify these confounders, otherwise, the study will be replete with bias. Randomization atthe design stage and regression models during the analysis stage are the best ways to prevent/reduce confounding in the research. On the other hand, effect modification does not create 'nuisance' in the research. In fact, effect modification helps the researcher to find the effect of one more variable on the outcome and hence are beneficial to the research.

Conflict of Interests: None

Ethical Permission: Approved

Funding: Nil

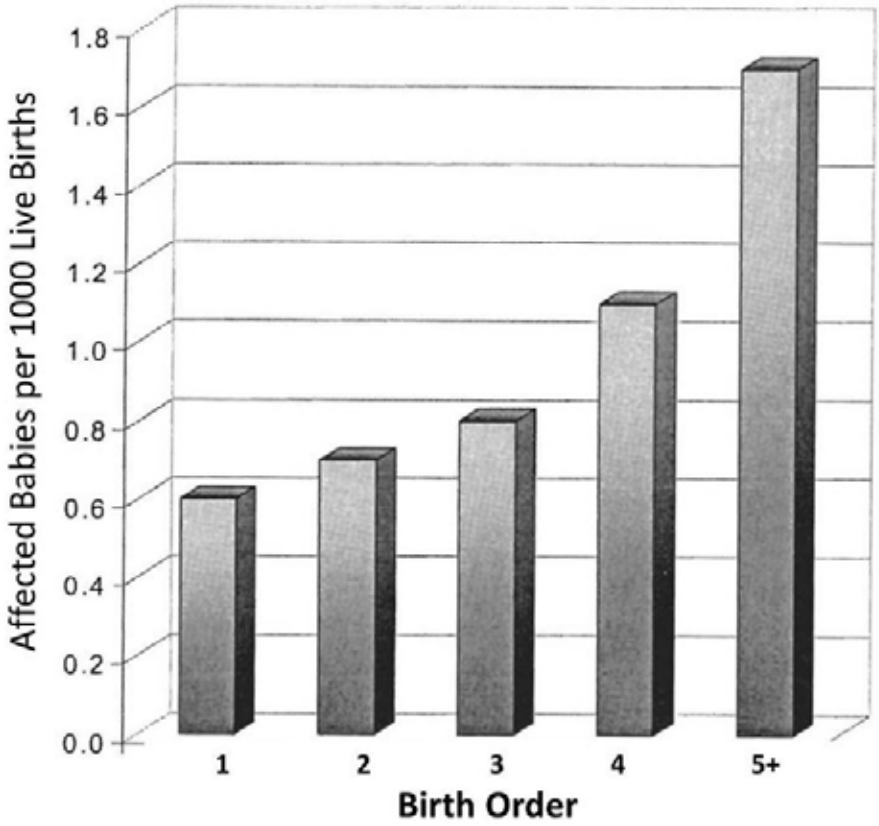


Figure 1: Relationship between the occurrence of Down's syndrome and order of birth.

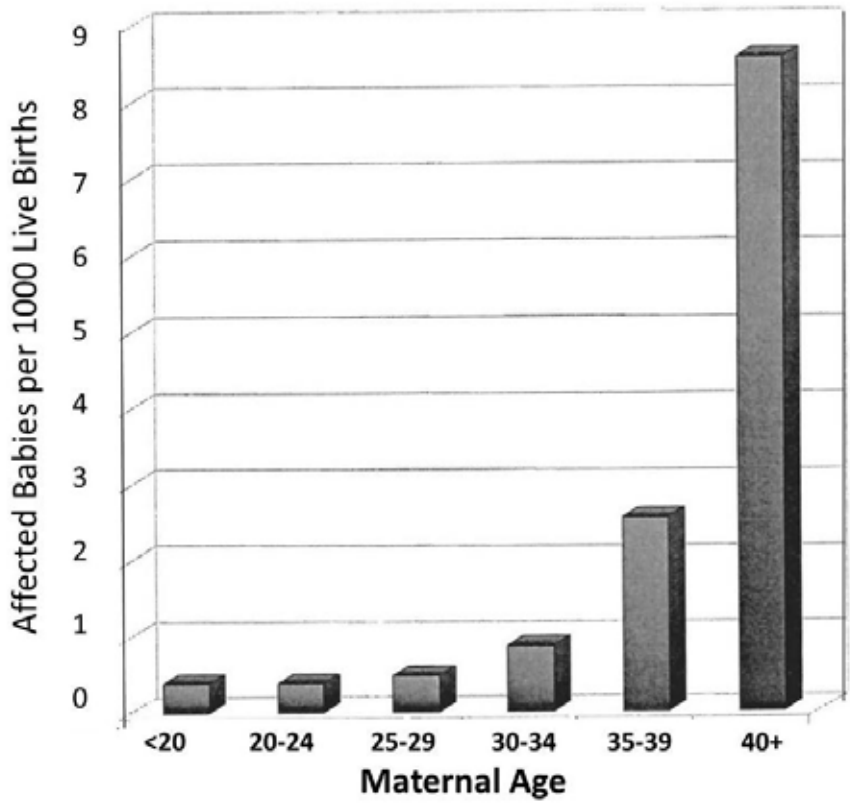


Figure 2: Relationship between maternal age and the occurrence of Down's syndrome

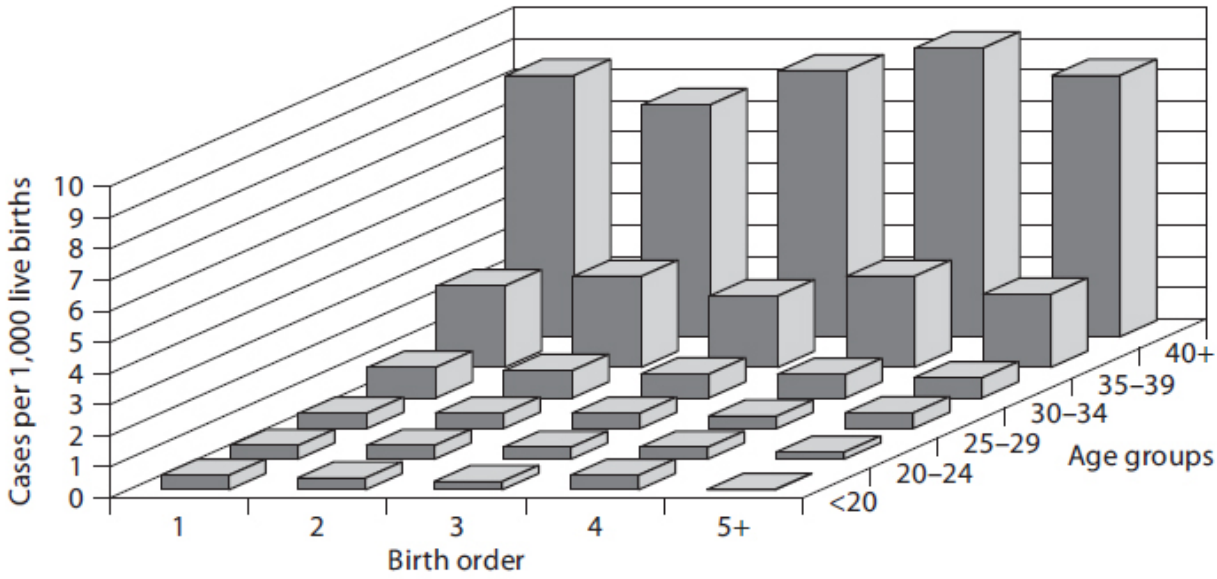


Figure 3: Association between maternal age and birth order v/s Down's syndrome

References

1. John M Last. A Dictionary of Epidemiology: Oxford University Press:4th Edition: P. No. 14, 37, 57.
2. Sackett DL. Bias in analytic research. J Chronic Dis 1979:32:51-63.
3. Stephen E Glasser. Essentials of Clinical Research: Springer Publication: 1stEdn: 295 – 302
4. K.J. Rothman, S. Greenland, Timothy L Lash. Modern Epidemiology: Lippincott, Williams & Wilkins: 2008: 128-135.
5. Pearl, Judea. Causality – Models, Reasoning, and Inference. Cambridge University Press, Cambridge: 2000: ISBN 0-521-77362-8.
6. Lorraine K. Alexander, Brettania Lopes, Kristen Ricchetti-Masterson, Karin B. Yeatts. UNC Gilling's School of Global Public Health: Confounding Bias, part I: https://sph.unc.edu. Accessed on 25-04-2020 at 17:15.
7. Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. BMJ 2002: 325: 1437–1438.
8. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology: Epidemiology 2000: 11(5): 550–60.
9. Robins, J. M. Data, design, and background knowledge in etiologic inference. Epidemiology 2001; 12: 313–320.
10. Mohamad Amin Pourhoseingholi, Ahmad Reza Baghestani, Mohsen Vahedi. How to control confounding effects by statistical analysis: GastroenterolHepatol Bed Bench 2012; 5(2):79-83
11. Pearce N.Analysis of matched case-control studies: BMJ 2016;352:i969 | doi: 10.1136/bmj.i969
12. Michiel A. de Graaf, Kitty J. Jager, Carmine Zoccali, Friedo W. Dekker. Matching, an Appealing Method to Avoid Confounding: Nephron ClinPract 2011;118:c315–c318.
13. KJ Jager, C Zoccali, A MacLeod, FW Dekker. Confounding: What it is and how to deal with it: Kidney International (2008); 73: 256–260.
14. Giovanni Tripepi, Kitty J. Jager, Friedo W. Dekker, Carmine Zoccali. Stratification for Confounding – Part 1: The Mantel-Haenszel Formula: Nephron Clin Pract 2010; 116:c317–c321
15. Ylostalo PV, Knuuttila ML. Confounding and effect modification: possible explanation for variation in the results on the association between oral and systemic diseases. J Clin Periodontol 2006; 33: 104–108.
16. Kamangar F. Effect modification in epidemiology and medicine. Arch Iran Med. 2012; 15(9): 575 – 582.