

Predictors of missing data when asking parents about their children's diet based on "Oral Health Situation of Iranian Children" Survey

Arash Shahravan, Hossein Hessari¹, Mohammad Reza Baneshi², Maryam Rad³,
Ali Akbar Haghdoost^{2,4}

Department of Endodontics, ³Researcher, Oral and Dental Diseases Research Center, ²Department of Epidemiology and Biostatistics, ⁴Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, ¹Department of Community Oral Health, School of Dentistry, Tehran University of Medical Sciences, Tehran, Iran

ABSTRACT

Background: The predictors of missing data when parents fill out questionnaire about their children's diet are not defined. The aim of this study was to evaluate predictors which affect unresponsiveness to diet questions based on "Oral Health Situation of Iranian Children" Survey carried out in 1998. **Materials and Methods:** A dummy variable was created as dependent variable according to responding eight questions relating to diet. Then predictors of missing data were defined using multivariable logistic regression and classification tree method. To evaluate goodness of fit of logistic regression model, sensitivity and specificity were assessed. Classification tree analysis was done by Quest growing method. Significance level was set at 0.05 in logistic regression analysis. **Observations and Results:** "Missing data" variable was marked as missing in 616 (7.2%) of questionnaires. In logistic regression model revealed that, gender, mother's education level and father's education level didn't affect "missing data" variable ($P > 0.05$). But, dmf index (OR = 0.94), Area of living (OR = 0.49), number of children in the family (OR = 1.19), sibling order (OR = 0.85), brushing (OR = 0.84) and visiting dentist (OR = 0.59) had statistically significant association with dependent variable ($P < 0.05$). Classification tree analysis showed that questionnaires related to children in urban area whose dmf index is under 5 and haven't visited dentist are predicted to have missing data more than other groups. **Conclusions:** Area of living, dmf index, number of children in the family, sibling order, brushing and visiting dentist are significant variables for predicting the risk of missing data when asking parents about their children's diet.

KEYWORDS: Missing data, oral health, predictor

Address for correspondence:

Prof. Ali Akbar Haghdoost,
Avicenna Ave. Jahad Blvd-7619813159, Kerman, Iran.
E-mail: ahaghdoost@kmu.ac.ir

Access this article online

Quick response code



Website:

www.jisppd.com

DOI:

10.4103/0970-4388.155117

Introduction

Dental caries is one of the most common infectious diseases that the majority of people may experience.^[1] Diet plays a central role in the development of dental caries unquestionably.^[2] Thus, decreasing consumption of cariogenic foods to prevent dental caries is backbone of most dental caries preventive programs. Usually information about children's diet is obtained from their parents. It's so important to design these studies to have most accurate information about children's diet.

Data values that have not been measured for some participants are called missing values or missing data. Missing data are common.^[3] However, they are usually inadequately handled in both observational and experimental studies. Chan *et al.*, estimated that 65% of studies in PubMed journals do not report the handling technique of missing data.^[4] Missing values create serious problems in data analyses. The seriousness of the problem depends largely on the pattern of missing data, how much are missing, and why they are missing.^[5] There are some imputation methods to decrease errors created by missing values. One simple method is

to substitute missing data by fixed values such as the mean (for data with a normal distribution) or median of observed values (for data with a skewed distribution). Recent Improvements in the field of analysis of missing data, including Expectation Maximum (EM) algorithm and Multiple Imputation via Chained Equations (MICE), provided techniques to deal adequately with missing data.^[6]

In an oral health survey, usually some questions about cariogenic diets are asked. Obviously, it is possible to have missing values whenever parents respond these questions about their children's diet. To decrease the percentage of missing values, knowing the predictors of missing values are helpful during data collection phase. The predictors of missing data in dental studies are not addressed before.

The aim of present study was to determine the predictors for missing data when asking parents about their children's diet based on Iranian Oral Health National Survey carried out in 1998.

Materials and Methods

The first national oral health survey "Oral Health Situation of Iranian Children" was carried out in 1998 including 8569 6-year old participants.^[7] The survey was done according to WHO criteria.^[8] Accordingly, to determine dmft index of the children dental examinations were done, and some information related to demographic and socio-economic situation and eating cariogenic diets of the participants were gathered.

A binary variable was made according to the responses to 8 questions about cariogenic diet consumption. Whenever, a respondent didn't respond to 4 questions or more, his/her questionnaire was marked as having missing data. The binary variable named "missing data" was considered as dependent variable. In univariate logistic regression and a multivariate logistic regression, adjusted effects of gender, place of residence, mother's education level, father's education level, number of children in the family, sibling order, brushing, visiting dentist and dmft index on "missing data" variable were evaluated. To evaluate goodness-of-fit, the sensitivity and specificity of the model were assessed. The cut-of-point value was set at 0.07. Classification tree analysis were done to predict values of "missing data" variable based on values of independent (predictor) variables and to construct rules for making predictions about individual cases. Classification tree analysis was done by Quest growing method. Minimum number of cases for parent nodes was set at 100 and minimum number of cases for child nodes was set at 50. SPSS 16 (SPSS, Chicago, IL, USA) was used for data analysis.

Results

"Missing data" variable (having 4 or more unanswered questions) was marked as missing in 616 (7.2%) of questionnaires. Diet questions and their missing percentage values are shown in Table 1.

The results of univariate and multivariable binary logistic regression is shown in Table 2. In univariate analysis, there was statistically significant association between dmft index ($P = 0.00$), area ($P = 0.00$) and visiting dentist ($P = 0.00$) with missing variable. In multivariate model, there was no association between gender ($P = 0.40$), mother's education level ($P = 0.17$) and father's education level ($P = 0.42$) with "missing data" variable. Place of residence ($P = 0.00$), number of children in the family ($P = 0.00$), sibling order ($P = 0.00$), brushing ($P = 0.05$) and visiting dentist ($P = 0.00$) had statistically significant association with dependent variable, missing data.

Hosmer and Lemeshow test showed the logistic model has appropriate goodness of fit ($P = 0.40$). The sensitivity of the model to predict dependent variable was 63.8% and the specificity was 56.8%.

Classification tree analysis showed that questionnaires related to children in urban area whose dmft was under 5 and didn't have visited dentist were predictors for missing data (Risk estimate = 0.38, Sensitivity = 80%, Specificity = 35.7%) [Figure 1]. According to low frequency of missed cases, classification tree can be used in non-missing cases better. Therefore, most of the terminal nodes consisted of non-missing cases.

Table 1: Frequencies of valid and missing data for questions about cariogenic diet consumption in the last 24 hours in Oral Health Situation of Iranian Children Survey (1998)

Questions	Valid (N)	Missing	
		(N)	%
How many times did your child drink soft drinks in the last-24-hours?	8031	566	7
How many times did your child drink sweet beverage in the last-24-hours?	8027	570	7
How many times did your child eat ice cream in the last-24-hours?	7969	628	4
How many times did your child eat biscuits in the last-24-hours?	8245	352	7
How many times did your child drink fruit juice in the last 24 hours?	8005	592	6
How many times your child you eat cookie in the last-24-hours?	8092	505	6
How many times did your child eat candies in the last-24-hours?	8068	529	6
How many times did your child drink cocoa milk in the last-24-hours?	7957	640	7

Table 2: Univariate and multivariate regression models to predict variables relating to missing data, P-values lower than 0.05 are statistically significant

Univariate model					Multivariate model				
Independent variables	P-value	OR	95% CI for OR		Independent variables	P-value	OR	95% CI for OR	
			Lower	Upper				Lower	Upper
dmf	0.00	0.94	0.92	0.96	dmf	0.00	0.94	0.92	0.96
Area	0.00	0.60	0.49	0.71	Area	0.00	0.50	0.41	0.60
Brushing	0.06	0.86	0.73	1.01	Brushing	0.05	0.84	0.71	1.00
Number of children	0.36	1.02	0.98	1.06	Number of children	0.00	1.20	1.10	1.31
Sibling order	0.56	0.99	0.95	1.03	Sibling order	0.00	0.85	0.78	0.93
Visiting dentist	0.00	0.57	0.47	0.68	Visiting dentist	0.00	0.58	0.48	0.71
Education of mother	0.12	1.08	0.98	1.2	Education of mother	0.17	1.11	0.96	1.30
Education of father	0.79	1.01	0.93	1.11	Education of father	0.42	0.95	0.83	1.08
Sex	0.44	0.94	0.79	1.1	sex	0.40	0.93	0.79	1.10

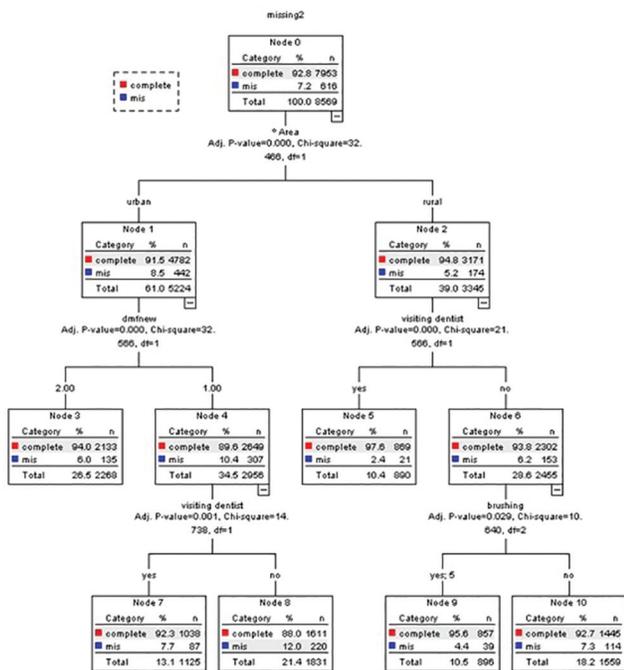


Figure 1: Classification tree to predict missing variable by Quest method. Node 8 shows the situation in which the possibility of data missing is higher than other nodes

Discussion

In this study it was shown that in univariate analysis, dmf index, visiting dentist and area, are correlated with missing of responding the questions. In multivariate regression analysis dmf index, area of living (urban, rural), number of children in a family, sibling order, brushing teeth by children and history of visiting dentist are some predictors for missing value in responding questions about cariogenic diet by parents of children. But, gender, mother’s education level and father’s education level are not significant predictors for missing data.

As shown in multivariate models in comparison to univariate models, some variables (number of children

in a family, sibling order and brushing teeth) became statistically significant correlated.

As this is the first article which evaluating predictors of missing data in cariogenic diet questions, we couldn’t compare the results with other findings but, our findings seem normal.

By increasing one point to dmf value of the children, risk of missing data lessens 6%. In questionnaires which are filled out by parents living in urban area, there were nearly two-fold chance of missing data comparing to rural area. By increasing one child to family size, the probability of missing data increase 19%. Thus in crowded families, the probability of missing data for questionnaires asking about diet will be increased. Simultaneously, the probability of missed data will be decreased 15% in questionnaire which is related to younger child. In questionnaires which are related to children who mentioned they brush their teeth and have visited dentist, the chance of missing data were 16% and 41% less than questionnaires related to children which didn’t brush their teeth.

According to remarkable impact of missing data on accuracy of data analysis, knowing predictors of missing data could help to do more attention in data collection phase of the research to prevent it as possible. Knowing predictors could improve efforts to intensify data completion and adjust analyses for bias caused by missing data.^[9]

It should be noted that despite low frequency of missing data was found in this study, bias resulting from non-random missing data is still a threat for validity of the research. Thus missing data must be concerned in designing, conducting and analysis phase of the researches. In addition to methods which help to prevent missing data, there are some treatment modalities that can decrease problems of missing data. A simple method is to surrogate missing data by a fixed value such as the mean (in normally distributed data) or median (in skewed data). This approach might artificially reduce the variance and affect the

strength of relationships with other variables since all missing data are replaced by a single value. Recent developments in analysis of missing data, consist of EM algorithm and MICE, provided methods to deal with missing data adequately.^[6]

Current available methods for managing missing data, assume missing takes place at random and ignore the possibility that unmeasured variables might interact with interventions to change outcomes.^[9] Our result showed that greater efforts to retain participants identified as being at-risk for missing data collection based on their socio-economic and demographic factors may be an essential step to improve the validity of research studies. As predictors of missing data in each study could differ from the others, it's important to evaluate predictors of missing data in other studies especially large national surveys.

In current study classification tree analysis was used in addition to logistic regression model. Classification trees present a clear, logical model that can be understood easily by people who are not mathematically inclined. The representation of the classification tree method is closer to medical reasoning and also high-order interaction of the risk-factors can easily be shown.^[10]

The result of classification tree analysis showed that questionnaires related to children in urban area whose dmft is under 5 and haven't visited dentist are predicted to have missing data more than other groups. This finding is similar to findings of logistic regression method.

In conclusion, area of living, number of children in a family, dmft index, sibling order, brushing teeth and history of visiting dentist are significant predictors for missing data when asking about cariogenic diet's of children from their parents. By knowing predictors of missing data, we can decline the chance of missing in data collection phase and manage the problem of non-random missing data in analysis phase.

Acknowledgements

This project was supported financially by the Kerman Research Center, Kerman University of Medical Sciences, Kerman, Iran.

References

1. Caufield PW, Griffen AL. Dental caries: An infectious and transmissible disease. *Pediatr Clin North Am* 2000;47:1001-19.
2. Tinanoff N, Palmer CA. Dietary determinants of dental caries and dietary recommendations for preschool children. *J Public Health Dent* 2000 Summer;60:197-206.
3. Streiner DL. The case of the missing data: Methods of dealing with dropouts and other research vagaries. *Can J Psychiatry* 2002;47:68-75.
4. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.
5. Peat J, Barton B. *Medical statistics: A guide to data analysis and critical appraisal*. 1st ed. John Wiley & Sons 2005;12-15.
6. Baneshi MR, Talei AR. Does the missing data imputation method affect the composition and performance of prognostic models? *Iran Red Crescent Med J* 2012;14:31-6.
7. Samadzadeh H, Hessari H. *Oral Health Situation of Iranian Children 1997*. Tehran, Ministry of Health and Medical Education, Under-Secretary of Health, Oral Health Bureau; 1999.
8. WHO. *Oral Health Surveys: Basic Methods*. 4th ed. Geneva: World Health Organization 1997;4-9.
9. Jerant A, Chapman BP, Duberstein P, Franks P. Is personality a key predictor of missing study data? An analysis from a randomized controlled trial. *Ann Fam Med* 2009;7:148-56.
10. Nagy K, Reiczigel J, Harnos A, Schrott A, Kabai P. Tree-based methods as an alternative to logistic regression in revealing risk factors of crib-biting in horses. *J Equine Vet Sci* 2010;30:21-6.

How to cite this article: Shahravan A, Hessari H, Baneshi MR, Rad M, Haghdoost AA. Predictors of missing data when asking parents about their children's diet based on "Oral Health Situation of Iranian Children" Survey. *J Indian Soc Pedod Prev Dent* 2015;33:100-3.

Source of Support: Nil, Conflict of Interest: None declared.