

Bayesian Network Modeling in Discovering Risk Factors of Dental Caries in Three-Year-Old Children

Łaguna W.*^{1, A-F}, Bagińska J.^{2, B, C, E, F}, Oniśko A.^{1, A, C-F}

1. Białystok University of Technology, Faculty of Computer Science, Poland
2. Medical University of Białystok, Department of Dentistry Propaedeutics, Poland

A- Conception and study design; B - Collection of data; C - Data analysis; D - Writing the paper; E- Review article; F - Approval of the final version of the article; G - Other (please specify)

ABSTRACT

Purpose - The aim of this study was to use probabilistic graphical models to determine dental caries risk factors in three-year-old children. The analysis was conducted on the basis of the questionnaire data and resulted in building probabilistic graphical models to investigate dependencies among the features gathered in the surveys on dental caries.

Materials and Methods - The data available in this analysis came from dental examinations conducted in children and from a questionnaire survey of their parents or guardians. The data represented 255 children aged between 36 and 48 months. Self-administered questionnaires contained 34 questions of socioeconomic and medical nature such as nutritional habits, wealth, or the level of education. The data included also the results of oral examination by a dentist. We applied the Bayesian network modeling to construct a model by learning

it from the collected data. The process of Bayesian network model building was assisted by a dental expert.

Results - The model allows to identify probabilistic relationships among the variables and to indicate the most significant risk factors of dental caries in three-year-old children. The Bayesian network model analysis illustrates that cleaning teeth and falling asleep with a bottle are the most significant risk factors of dental caries development in three-year-old children, whereas socioeconomic factors have no significant impact on the condition of teeth.

Conclusions - Our analysis results suggest that dietary and oral hygiene habits have the most significant impact on the occurrence of dental caries in three-year-olds.

Keywords - dental caries, Bayesian network, learning from data, risk assessment

DOI

*Corresponding author

Wojciech Łaguna

Białystok University of Technology, Faculty of Computer Science, Poland

e-mail: wojciech.laguna@gmail.com

Received: 19.03.2019

Accepted: 17.04.2019

Progress in Health Sciences

Vol. 9(1) 2019 pp 118-125

© Medical University of Białystok, Poland

INTRODUCTION

Dental caries is a worldwide problem. According to Global Burden of Disease, untreated caries in deciduous teeth is the 10th most prevalent condition which affects about 621 million children [1].

Dental caries has been recently included in a group of non-communicable diseases because it is caused by behaviorally based risk factors. Several general and local aspects promoting the development of carious lesions were identified. Among general risk factors, the most frequently indicated were gender, age, ethnicity, geographic location, and socioeconomic background. Other variables were the pattern of oral hygiene, consumption of sugar, frequency of dental check-ups, and use of fluorides.

Several studies were conducted to establish associations between the prevalence and experience of dental caries and a particular risk factor. Most of these studies adopted classical statistical methods.

Bayesian networks [2] are graphical models that are based on sound rules of the probability theory. They offer a framework for explicit modeling of probabilistic relationships which can be elicited from domain experts or learned directly from the data by means of existing algorithms. Learning Bayesian network algorithms allow to discover relationships between modeled variables which are represented by means of an acyclic directed graph. The graphical structure of Bayesian networks offers an additional insight into the modeled domain while reasoning with these models provides an explanation of their outcomes. The Bayesian network modeling proved to be a powerful tool for modeling complex uncertain knowledge. It was successfully applied in solving diagnostic or prognostic problems in medicine [3].

It was also proved that the Bayesian network modeling could be useful in epidemiological research [4].

However, the Bayesian network modeling was applied only in a limited number of studies on dental caries [5-10].

This paper presents the capabilities of Bayesian network model analysis in discovering risk factors of dental caries in three-year-old children. The analysis was conducted on the basis of the questionnaire data and resulted in the development of a probabilistic graphical model used to investigate

dependencies among the features gathered in the surveys on dental caries.

MATERIALS AND METHODS

Materials

Our analysis was based on the retrospective data collected between 2002 and 2003 from 255 children (108 girls and 147 boys) aged between 36 and 48 months (mean age of 43 months) in randomly selected kindergartens in Bialystok and in small towns and villages in the Podlasie region, Poland. That data was previously analyzed by one of the authors (JB) [11].

The collected data contained a dental examination and a self-administered questionnaire consisting of two parts. The first part (30 questions) was developed at the Department of Epidemiology, Medical University of Lodz, Poland, for the purpose of the Polish National Oral Health Survey; the second part (4 questions) was developed by the authors. The dental status was assessed with the dmft (the dmft index stands for decayed, missing, and filled teeth) index with a plane dental mirror and a periodontal probe according to the WHO guidelines [12].

Each tooth was described as caries-free, decayed (d), extracted due to caries (m) or with dental filling (f).

The questionnaire was filled out by 255 parents or guardians. The questionnaire contained questions of socioeconomic, demographic, and medical nature.

The collected data allowed us to analyze the impact of various socioeconomic factors such as habits, wealth, or the level of education on the condition of teeth in three-year-old children.

Based on the dental examination results, we divided the studied sample into three groups depending on the severity of dental caries:

- all teeth were healthy (healthy teeth)
- there were one up to three teeth with dental caries (moderate dental caries)
- there were more than four teeth with dental caries (significant dental caries).

The questionnaire data showed that 18.4% of examined children had healthy teeth, 35.3% had moderate dental caries, and 46.3% were classified as children with significant dental caries.

The characteristics of our data with respect to gender and dental caries status were shown in Table 1.

Table 1. The intensity of dental caries in the studied group

	Quantity	Healthy teeth	Moderate dental caries	Advanced dental caries
		Qty (%)	Qty (%)	Qty (%)
Girls	108	21 (19.4%)	44 (40.7%)	43 (39.8%)
Boys	147	26 (17.7%)	46 (31.3%)	75 (51.0%)
Total	255	47 (18.4%)	90 (35.3%)	118 (46.3%)

Methods

To analyze the data, we have applied the Bayesian network modeling that is a multivariate approach incorporating several variables in one framework. The Bayesian network consists of two parts: (1) a graphical structure modeling variables by means of nodes and probabilistic relationships among these variables represented by directed arcs; (2) conditional probability distributions that quantify relationships among neighboring variables. To explain the approach, we have created a simplified Bayesian network that models *Body Mass*

Index (BMI), Income, and Smoking as risk factors of *Diabetes* and *Number of teeth* as an effect of *Diabetes*. An example Bayesian network model was presented in Fig. 1. The arcs between the nodes in Fig. 1 represent probabilistic relationships, for example, the arc between *BMI* and *Diabetes* indicates that abnormal *BMI* can lead to *Diabetes*. This relationship was also quantified by a conditional probability distribution $P(Diabetes/BMI, Income)$ and was presented in Fig. 1.



Figure 1: Example of Bayesian network model

Once a Bayesian network model has been built, we can perform a reasoning, i.e. given observed variables in the model we can calculate a posterior probability distribution that can be further interpreted as a quantitative risk value. For example, we could calculate the posterior probability of *Diabetes* given observed values of the nodes *BMI*, and *Income*:

- $P(\text{Diabetes} = \text{yes} \mid \text{BMI} = \text{normal}, \text{Income} = \text{high}) = 0.08,$
- $P(\text{Diabetes} = \text{yes} \mid \text{BMI} = \text{overweight_or_obese}, \text{Income} = \text{low}) = 0.32.$

The Bayesian network modeling approach allows making in-depth analyses e.g. on the strength of influence or the diagnostic value.

The strength of influence is calculated on the basis of differences between the a posterior marginal probability of a child node, as a parent node changes [13].

The higher value, the stronger dependency between two variables was observed. In our example model, we have identified the following pairs of nodes with the highest strength of influence:

- *Diabetes* and *Number of teeth*
- *BMI* and *Diabetes*.

The diagnostic value allows analyzing the impact of observable variables on a target variable. Target variables are defined as variables with a diagnostic or explanatory interest.

In our example Bayesian network, we set the *Diabetes* as a target variable.

The assessment of diagnostic value is based on the cross-entropy between each of the observed variables and the target variable.

The higher diagnostic value we observed the bigger impact the variable had on the target variable. Fig. 2 shows a list of observable variables ranked by their diagnostic value.

The left-hand side of Fig. 2 presents two possible states of *Diabetes*.

The right-hand side of Fig. 2 shows the model variables sorted by their diagnostic value descending. The example shows that the *BMI* has the biggest impact on *Diabetes*.

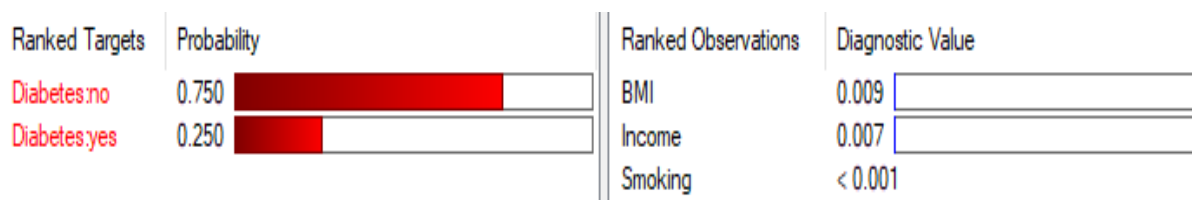


Figure 2. Observable variables ranked by their diagnostic value

A graphical structure and numerical parameters of Bayesian networks can be learned from data, elicited from expert knowledge or both. Bayesian network models learned from data are the examples of an unsupervised machine learning method.

This approach allows dealing with large data sets with thousands of variables. In this article, we described a model that was founded on data with some expert knowledge introduced.

Our Bayesian network models were created and tested using SMILE, an inference engine, and GeNIe, a development environment for reasoning in graphical probabilistic models, both developed at the Decision Systems Laboratory, University of Pittsburgh, and available at <http://www.bayesfusion.com>.

RESULTS

Learning the Bayesian network model

The collected data and the expert knowledge allowed us to build a Bayesian network model.

The data contained 255 observations and 33 variables. Our model was built based on 33 variables: 31 variables were of socioeconomic and demographic nature while two variables were of a medical nature and represented dental caries (dental caries and dental filling).

The discrete type of studied data allowed us to use a PC algorithm [14] to learn a Bayesian network structure. The resulting model includes 33 nodes, 39 arcs, and 2,037 numerical parameters. The PC algorithm identified relationships between the variables as presented in Fig. 3.

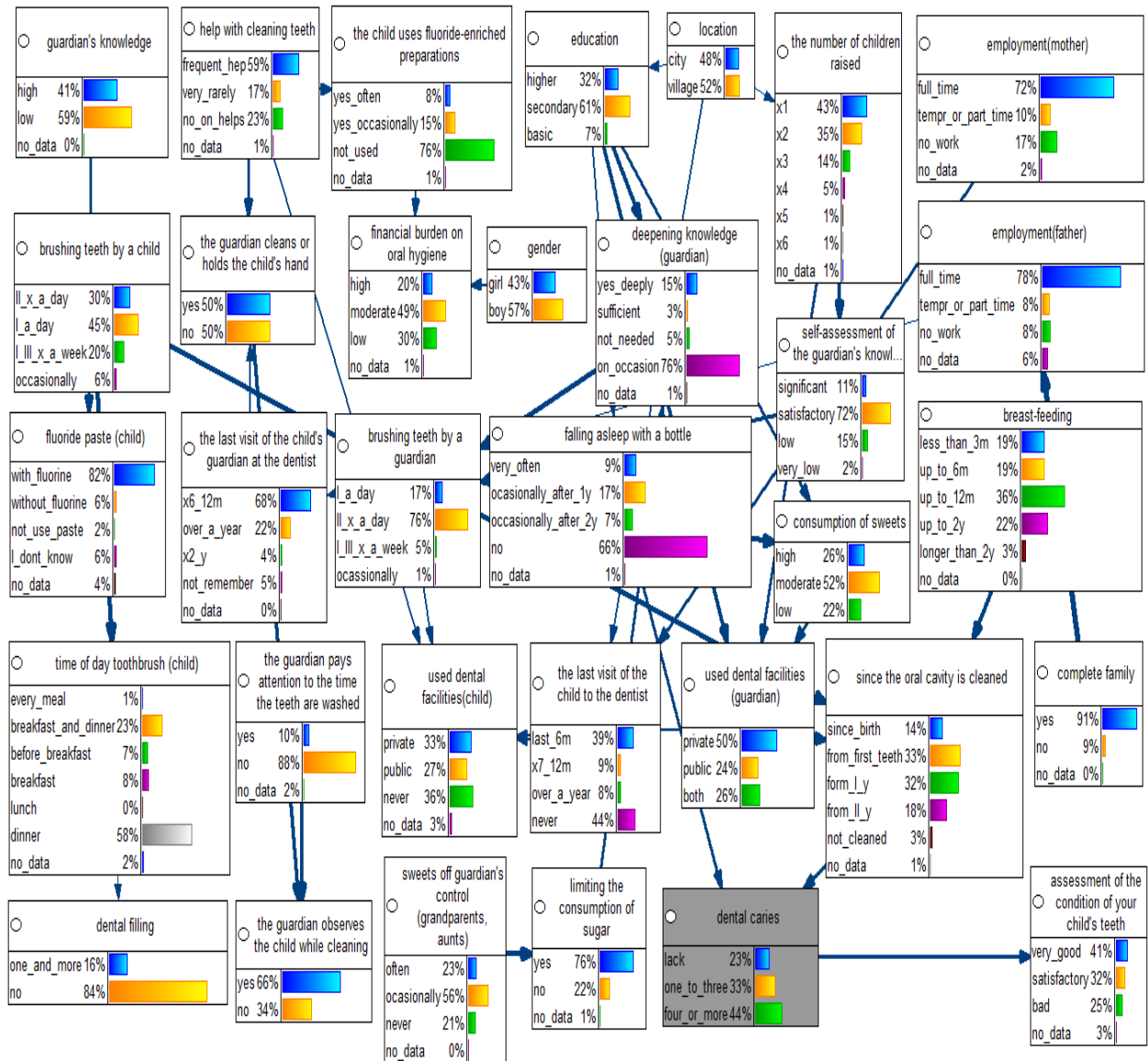


Figure 3. Bayesian network model for dental caries

The Bayesian network modeling allows combining objective data with expert knowledge. While building the model, we have focused on time tiers to avoid a situation where a future event depends on the past one. For example, dental caries cannot influence breast-feeding, whereas dental caries may possibly be caused by breast-feeding.

In our model we have identified the following pairs of nodes with the highest strength of influence:

- brushing teeth by a child and time of day toothbrush (child),
- education and deepening knowledge (guardian),
- complete family and employment (father)

- the guardian cleans or holds the child's hand and the guardian observes the child while cleaning,
- falling asleep with a bottle and used dental facilities (guardian).

Fig. 4 shows a list of variables ranked by their diagnostic value.

The left-hand side of Fig. 4 presents the levels of dental caries.

The right-hand side of Fig. 4 shows the model variables sorted by their diagnostic value descending. The variables with the highest diagnostic value in our model are:

- falling asleep with a bottle,
- since the oral cavity is cleaned,
- brushing teeth by a child.

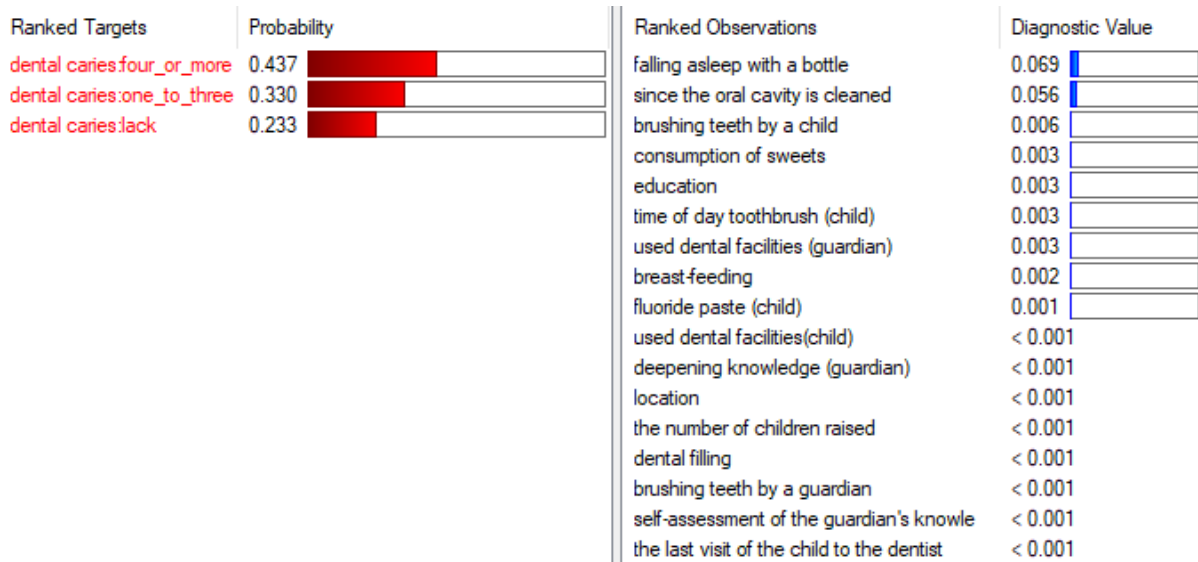


Figure 4. Observable variables ranked by their diagnostic value

Bayesian network models allow to observe how a change in an observable variable influences the probability distribution of the variable in question. Fig. 5 and Fig. 6 show a part of the model presented in Fig. 3. In our analysis, we have focused on six variables with the highest diagnostic value. Fig. 5 shows our model with two observed variables, i.e., high consumption of sweets and secondary education level of a guardian. In this case, the

posterior probability of lack of dental caries is equal to 19%. Fig. 6 shows our model with the observed low consumption of sweets and high education of the guardian. In this scenario, the posterior probability for the state *lack of dental caries* is equal to 26%.

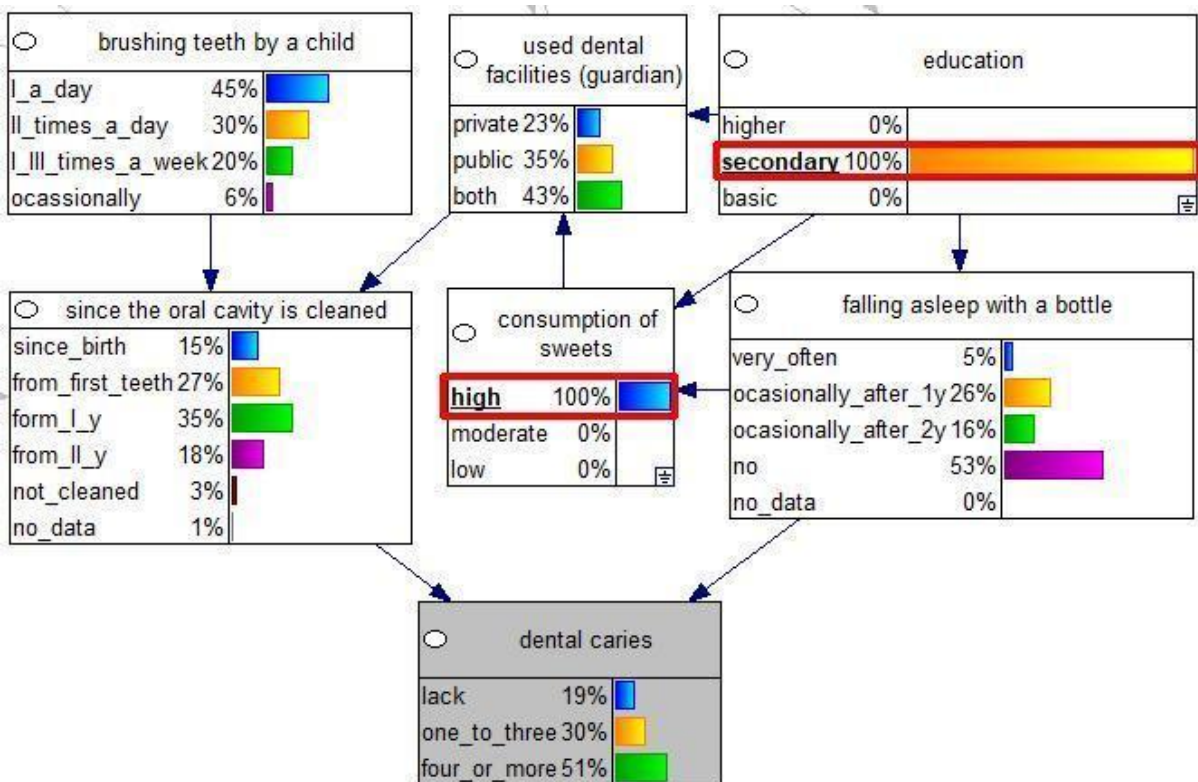


Figure 5. Example of reasoning: A high level of sweets consumption and secondary level of education have been observed

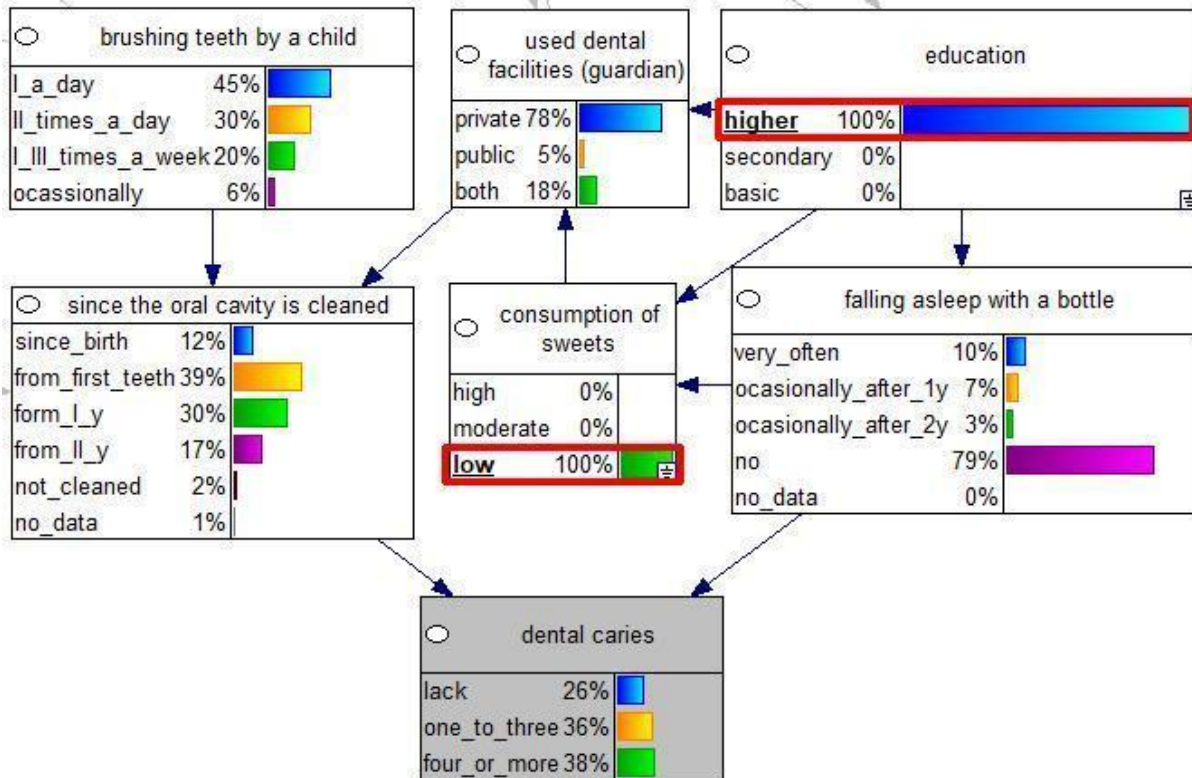


Figure 6. Example of reasoning: A low level of sweets consumption and higher education observed

We would like to note that the above observations propagate to other nodes in the model. Based on the subset of the model presented in Fig. 5 and Fig. 6, it can be observed that the posterior probability of the variable *Falling asleep with a bottle* for the state *no* has changed from 53% to 79%.

DISCUSSION

Our study showed that three-year-old children presented a very high level of dental caries. Such condition is typical for Polish children. A recent epidemiological survey revealed that only 20% of Polish children aged between 5 and 12 were caries-free [15] while for the European countries the World Health Organization (WHO) aimed at no more than 20% of six-year-olds suffering from dental caries by 2020.

The possibility of reasoning under uncertainty makes the Bayesian network modeling a powerful analytic tool for a multifactorial disease like dental caries. By learning the model from the data, we have discovered relationships between the variables. Such analysis also reduces the bias associated with small and non-randomized samples [6].

Bagińska, while analyzing the same data [11], identified a relationship between the gender and the condition of teeth, whereas our Bayesian network analysis indicated that the dependency

between the two variables was weak. According to Bagińska [11], there were three main risk factors of the caries prevalence: age since the oral cavity was cleaned, falling asleep with a bottle and consumption of sweets. Our Bayesian network model indicated that falling asleep with a bottle and age since the oral cavity is cleaned had a significant impact on dental caries. Furthermore, according to our model, the frequency of sugary food consumption at the age of three years seemed to be less important for the development of dental caries in children of this age. However, improper eating habits at the age of three may determine the future progression of the disease.

The Bayesian network analysis clearly indicated which children were at a great risk of developing dental caries. Based on our results, a simple questionnaire could be developed and used during routine pediatric check-ups to identify the most exposed children and to provide them adequate prevention strategies. Our model also showed how changes in daily routine can influence the probability of the disease. A modification of only one aspect (e.g. guardian education) decreased the chances of feeding a child during the sleep by several percentages, thereby reduced a risk of caries development. Such information is of a great educational value. Less educated parents of preschool children probably have to be more motivated to avoid risky behaviours.

To our knowledge, there are only a few studies that applied Bayesian network modeling to evaluate dental caries risk factors and dental status. Kemoli and Chepkwony [10] found this type of modeling suitable to discriminate specific factors of early childhood caries in Kenyan three- to six-year-old children. Similarly to us, they concluded that dietary habits were not significant in predicting the levels of the dmft index and that parents' education and employment status influenced child's oral condition. However, contrary to our results, they found boys being more prone to caries than girls. Wen et al. [5] used a Bayesian network analysis to establish the influence of household variables on the prevalence of dental caries in siblings in a high-risk population. They confirmed that this approach was more accurate than a generalized linear mixed model. Bhatia et al. [7] proposed using Bayesian networks to build a decision-making tool for dentists that would assist them in choosing the best treatment plan for dental caries.

CONCLUSIONS

The aim of this study was to use probabilistic graphical models to determine the risk factors of dental caries in three-year-old children. We have applied the real-world data to learn a Bayesian network model. The process of Bayesian network model building was additionally assisted by a dental expert. The results of our analysis suggest that the dietary and hygiene habits in the first three years of life have the most significant impact on the occurrence of dental caries in three-year-old children.

REFERENCES

1. Kassebaum N.J., Bernabé E., Dahiya M., Bhandari B., Murray C.J.L., Marcenes W.. Global Burden of Untreated Caries: A Systematic Review and Metaregression. *JDR*. 2015 May;94(5):650–58.
2. Pearl J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.
3. Lucas P.FJ., Gaag L., Abu-Hanna A. Bayesian networks in biomedicine and health-care. *Artif Intel Med*. 2004 Mar; 30:201–14.
4. Dunson DB: Commentary: Practical advantages of Bayesian analysis of epidemiologic data. *Am J Epidemiol*. 2001 June;153:12(15):1222–6.
5. Wen A, Weyant RJ, McNeil DW, Crout RJ, Neiswanger K, Marazita ML, Foxman B. Bayesian Analysis of the Association between Family-Level Factors and Siblings' Dental Caries. *JDR Clin Trans Res*. 2017 Jul;2(3):278-86.
6. Matranga D, Campus G, Castiglia P, Strohmenger L, Solinas G. Italian deprivation index and dental caries in 12-year-old children: a multilevel Bayesian analysis. *Caries Res*. 2014;48(6):584-93.
7. Bhatia A., Singh R., Using Bayesian Network as Decision making system tool for deciding Treatment plan for Dental caries, *JAIR*. 2013 July;2(2): 93-6.
8. Bandyopadhyay D., Reich B.J., Slate E.H. Bayesian modeling of multivariate spatial binary data with applications to dental caries. *Stat Med*. 2009 Dec 10;28(28):3492-508.
9. Komárek A, Lesaffre E, Härkänen T, Declerck D, Virtanen JI. A Bayesian analysis of multivariate doubly-interval-censored dental data. *Biostatistics*. 2005 Jan;6(1):145-55.
10. Kemoli AM, Chepkwony F. Applying Bayesian Model to Predict Socio-demographic and Occlusal Determinants of Early Childhood Caries (ECC) *Pesq Bras Odontoped Clin Integr*. 2017 Feb; 7(1):e3452.
11. Bagińska J.. Assessment of dental status in 3-year-old children from the Podlaskie Province. PhD thesis, Medical Academy of Bialystok, 2004.
12. World Health Organization, Oral Health Surveys, Basic Methods, 4th Edition. Geneva, 1997.
13. Koiter J.R. Visualizing Inference in Bayesian Networks. Master's thesis, Delft University of Technology, 2006.
14. Spirtes P., Glymour C. N., Scheines R. Causation, Prediction, and Search. MITpress, Cambige 2000.
15. Tomaszewski M, Matthews-Brzozowska T. Privatization trends in the sector of dental services in Poland, in respondents' opinion - preliminary report. *Art of Dentistry*. 2017;2(64):76-83. (Polish)