

Bayesian Network Model to Evaluate the Effectiveness of Continuous Positive Airway Pressure Treatment of Sleep Apnea

Olli-Pekka Ryyänänen, MD, Dr. Med.Sci.^{1,2}, Timo Leppänen, PhD^{3,4}, Pekka Kekolahti, Lic. Sci (Technology)⁵, Esa Mervaala, MD, PhD^{4,6}, Juha Töyräs, PhD^{3,4}

¹Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland; ²General Practice & Primary Health Care Unit, Kuopio University Hospital, Kuopio, Finland; ³Department of Applied Physics, University of Eastern Finland, Kuopio, Finland; ⁴Department of Clinical Neurophysiology, Kuopio University Hospital, Kuopio, Finland; ⁵Department of Communications and Networking, School of Electrical Engineering, Aalto University, Espoo, Finland; ⁶Institute of Clinical Medicine, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

Objectives: The association between obstructive sleep apnea (OSA) and mortality or serious cardiovascular events over a long period of time is not clearly understood. The aim of this observational study was to estimate the clinical effectiveness of continuous positive airway pressure (CPAP) treatment on an outcome variable combining mortality, acute myocardial infarction (AMI), and cerebrovascular insult (CVI) during a follow-up period of 15.5 years (186 ± 58 months). **Methods:** The data set consisted of 978 patients with an apnea-hypopnea index (AHI) ≥5.0. One-third had used CPAP treatment. For the first time, a data-driven causal Bayesian network (DDBN) and a hypothesis-driven causal Bayesian network (HDBN) were used to investigate the effectiveness of CPAP. **Results:** In the DDBN, coronary heart disease (CHD), congestive heart failure (CHF), and diuretic use were directly associated with the outcome variable. Sleep apnea parameters and CPAP treatment had no direct association with the outcome variable. In the HDBN, CPAP treatment showed an average improvement of 5.3 percentage points in the outcome. The greatest improvement was seen in patients aged ≤55 years. The effect of CPAP treatment was weaker in older patients (>55 years) and in patients with CHD. In CHF patients, CPAP treatment was associated with an increased risk of mortality, AMI, or CVI. **Conclusions:** The effectiveness of CPAP is modest in younger patients. Long-term effectiveness is limited in older patients and in patients with heart disease (CHD or CHF).

Keywords: Sleep Apnea Syndromes, Continuous Positive Airway Pressure, Bayesian Analysis, Patient-Specific Modeling, Outcome Assessment (Health Care)

Submitted: April 6, 2018

Revised: July 31, 2018

Accepted: September 21, 2018

Corresponding Author

Olli-Pekka Ryyänänen, MD, Dr. Med.Sci.

Department of Public Health and Clinical Nutrition, University of Eastern Finland, P.O. Box 1627, Kuopio 70211, Finland. Tel: +358-40-5141-741, E-mail: olli-pekka.ryyananen@uef.fi (<https://orcid.org/0000-0002-9253-7491>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. Introduction

Obstructive sleep apnea (OSA) is a common nocturnal breathing disorder affecting about 8% of the Finnish adult population [1]. Several studies have reported an association between OSA and increased mortality [2,3]. OSA has been shown to increase the risk of stroke or death from any cause [4].

Continuous positive airway pressure (CPAP) is a standard treatment for OSA [5]. CPAP treatment, for example, is shown to improve results in Epworth sleepiness scale questionnaires, quality of life, and subjective sleepiness [6].

The aim of this study was to assess the clinical effectiveness of CPAP treatment on an outcome variable combining all-cause mortality, acute non-fatal myocardial infarction (AMI), and non-fatal cerebrovascular insult (CVI) during a follow-up period (186 ± 58 months). This analysis was done using a default hypothesis that CPAP has an effect on the mentioned combined outcome. This combined outcome was chosen because all-cause mortality, stroke, and coronary heart disease are the most important clinical consequences of OSA [7].

In this study, a Bayesian network model was chosen as a tool for analysis. The Bayesian network approach affords certain advantages over standard frequentist methods in analyzing data collected in real practice. For example, Bayesian network analysis provides a transparent representation of relationships between system variables using different sources of data. It can handle complicated data sets with missing data, outliers, and nonlinear relationships, and the results of the analysis can be presented in a visual form that is easy to interpret [8-11]. The visual form uses directed acyclic graph (DAG), from which direct and indirect effects, common causes and effects can be discovered and mathematically expressed [12].

II. Methods

1. Novel Sleep Apnea Parameters

Diagnosis of OSA is based on daytime symptoms (e.g., daytime sleepiness) and an apnea-hypopnea index (AHI) or an oxygen desaturation index (ODI) [13]. We previously introduced novel desaturation severity (DesSev) and obstruction severity (ObsSev) parameters that account for the severity aspect of individual apnea, hypopnea, and desaturation events [14,15]. The definitions of the novel parameters are presented in Table 1.

2. Patients

The database used in this study consisted of 2,037 consecutive patients referred for night polygraphy in the Department of Clinical Neurophysiology at Kuopio University Hospital (a large referral hospital in Eastern Finland) between 1992 and 2003. The original data set consisted of 119 variables, e.g., variables from polygraph recordings, treatments, and medications. There were 984 subjects omitted from the study from the study due to having an AHI lower than 5.0. In addition, 51 subjects were removed due to several missing values, and 24 subjects were removed due to having oral device treatment, leaving 978 patients for the analysis.

All the recordings were registered using a custom-made ambulatory device, *Unisalkku* [2,14-16], and they were re-analyzed using standard respiratory rules developed by the American Academy of Sleep Medicine (AASM) [13], as in our previous studies [14-19].

In the present study, the follow-up time was defined as the time between the polygraph recording and death, AMI, or CVI; for the rest of the patients, it was the time between the polygraph recording and June 2014. Causes of death were acquired from Statistics Finland (Helsinki, Finland) in June 2014, and information about diseases, morbidities, and treatments was collected from the patients' medical records at Kuopio University Hospital. The subpopulations of the data set have been used previously [14-19]. More detailed infor-

Table 1. Definitions of the novel parameters

Parameter	Definition
Obstruction severity (s%)	$\frac{\sum_{n=1}^L (ApDur_n \times DesArea_n) + \sum_{m=1}^L (HypDur_m \times DesArea_m)}{Indextime}$
Desaturation severity (%)	$\frac{\sum_{n=1}^L DesArea_n}{Indextime}$

Individual apnea and hypopnea durations are denoted as *ApDurs* and *HypDurs*, respectively, and the individual desaturation area as *DesArea* (5%). *Indextime* denotes the total analyzed time of the polygraph recording, and *L* is the number of events in question.

mation about data collection and measurements is available in previous papers.

3. Bayesian Network Analysis

The statistical analysis was performed with Bayesian networks by using the BayesiaLab 5.3.3 tool [20]. A Bayesian network can be described as a DAG. It determines the factorization of a joint probability distribution over the variables (nodes of the DAG), where the factorization is defined as directed arcs of the DAG. A Bayesian network structure (i.e., a DAG) is constructed either manually or with machine learning based on observational data, for example, by a domain expert. We introduce a third alternative for structural learning—enabled by the tool we used—called expert-assisted machine learning, where the expert sets restrictions for the structural learning algorithm. Of the two main structural learning alternatives, constraint-based search and score-based learning [21], we applied the latter method. The search algorithm was Taboo [22], and the scoring method was two-stage minimum description length (MDL) [23].

A trade-off exists in MDL between the model’s complexity and the model’s fit to the data. The optimum model is one in which MDL (Model|Data) is at its minimum; in other words, simple model structures are preferred. The tool we used also offered the possibility to weigh the complexity part with a structural coefficient (SC) for situations in which the default value (SC = 1) does not produce credible results from the structural learning according to the experts’ prior knowledge

or research data. The structural coefficient is discussed more in Kekolahti et al. [24].

The objective in expert-assisted machine learning is to produce a Bayesian network in which arc directions correspond to causal assumptions of the data-generating model. In other words, when an arc exists from variable A to B, variable A is the cause of variable B, but if no arc exists, no direct causal relationship exists between them. Expert-assisted machine learning was used in the study in two ways, which are summarized below.

(1) A causal DAG is consistent with the research data. This structure is called a data-driven causal Bayesian network (DDBN). The restrictions set for the learning are the following. First, temporal indexes (relative temporal order between variables in the research data) are defined for variables that, based on the learning algorithm, can construct a structure in which the time-wise arc direction is from the older to the newer variable. Thus, situations in which a newer variable points to an older variable are blocked. Second, the number of variables can be limited in the model if they do not form any kind of dependency with other variables or if the variables are not relevant to the study. Third, the learning algorithm is informed that the learned arc direction between two variables is prohibited if the direction proposed by the Taboo algorithm does not make logical sense. In this case, two other alternatives, namely, the arc is missing and an opposite arc direction, are still allowed. This phase also contains discretization of the numerical values into meaningful intervals

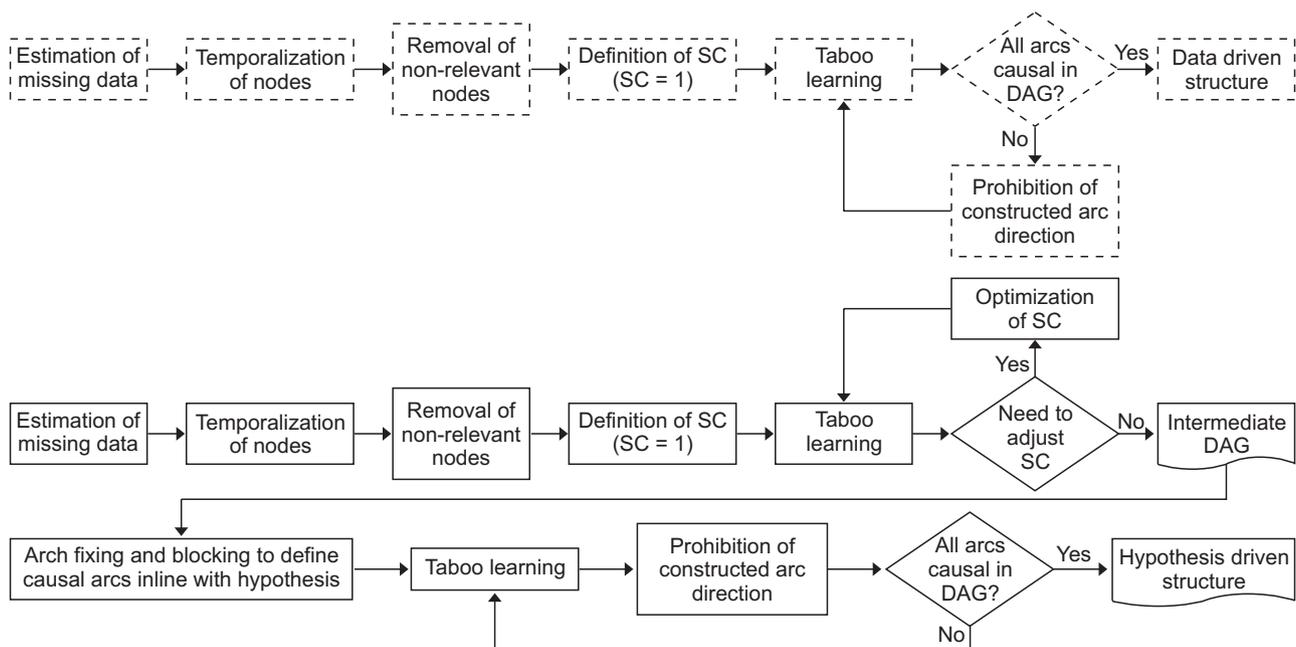


Figure 1. Process used to construct a hypothesis-driven (solid line) and a data-driven structure (dotted line). SC: structural coefficient, DAG: directed acyclic graph.

[25].

(2) A causal DAG is consistent with the hypothesis regarding the research question. We call this structure, simply, a hypothesis-driven causal Bayesian network (HDBN). The restrictions set for the learning are as follows. First, temporal indexes are defined for the variables. Second, variables can be excluded from the model if they do not form any kinds of dependency with other variables or if the variables are not relevant to the study. For example, the Markov blanket can be used for this phase. Third, the SC is adjusted if its default value does not produce credible structures according to the hypothesis, and the numerical values are discretized into meaningful intervals. Fourth, based on the hypothesis and prior knowledge, an arc is drawn manually and fixed between two variables to indicate their causal relation if the learning does not produce it automatically. Fifth, the learning algorithm is informed that the learned arc direction between two variables is prohibited if the direction does not make logical sense. In this case, two other alternatives, namely, the arc is missing and the arc direction is opposite, are still allowed.

Figure 1 describes the two expert-assisted learning processes, DDBN and HDBN, used in the study. Once the structural learning has been completed, parameter learning focuses on how the variables quantitatively relate to each other. For each variable in a DAG, conditional probability tables are estimated with the maximum likelihood method from the frequencies observed in the research data. This information is used to define the causal strength between two variables as information gain, i.e., as Kullback–Leibler divergence (D_{KL}). It provides a natural method for this study to compare distributions of two connected variables [26]. That is to say, we estimate the strength of a specific arc as D_{KL} in the context of the entire DAG. What if this arc were removed but all the others remained? Furthermore, direct effect (DE) is calculated between each variable and the outcome variable to compare the causal strength of the variables on the outcome variable. DE is based on Jouffé's proprietary likelihood matching algorithm [12], and it estimates the causal dependency between two variables by measuring the impact of a conditional mean of each state of variable A on the mean of variable B (outcome variable) with Kullback's minimum cross-entropy method MinxEnt [27] and by keeping the values of all other variables fixed. DE is especially suitable for situations where the dependency between two variables is linear.

The research data contained 3.02% missing data (total data before excluding variables), whose type was missing at ran-

dom (MAR). To maintain the number of samples, samples with missing data were kept, and the missing data were estimated by using a structural equation model (EM) algorithm [28].

The number of variables was reduced from 119 to 19 using an augmented Markov blanket algorithm. In this preliminary analysis, the SC value was set to 0.6 to find all potentially affecting variables. Variables connected to the variable *Outcome total* were included in the analysis. Sleep apnea parameters and CPAP treatment were selected by using a local SC value of 0.4 for them.

The discretization of the numerical variables was performed manually by using two alternative methods: (1) a decision tree algorithm, setting the variable *Outcome total* as the target, or (2) clinically commonly used thresholds (when using a decision tree algorithm was not possible). The discretized values as well as the total data set are presented in Table 2.

A temporal index (TI) was assigned to each variable to indicate the relative temporal order between variables, as seen in Table 2. To do this, the variables were divided into eight time categories according to knowledge about the variables' appearance. Thus, the variable *age* had a TI = 1 (oldest known measured value), and *Outcome total* had a TI = 8 (last measurement at the end of the follow-up period), for instance.

Arcs between the nodes indicate causality fulfilling the temporality criterion (newer variable cannot point to older variable as a function of time). However, arcs between variables having the same TI show no causality. For example, arcs between sleep apnea parameters like *DesSev* → *AHI* do not indicate causality because both variables were measured at the same time and they have the same TI value. Expert opinion was used to determine causality in a case with an obvious wrong direction of the arc. As an example, an arc direction of *CHD* → *Diabetes* was manually forbidden, but the opposite direction and no arc were allowed.

In the next step, the modeling process was changed from a DDBN to an HDBN. According to the default hypothesis, CPAP was considered to have a DE on *Outcome total*, even though this hypothesis was not supported by the DDBN. The model was simplified by limiting the number of variables to include only the most prominent ones (nine variables). The variable *Diuretic* was dropped because it was considered to be a marker, not a causal factor for *Outcome total*.

In the HDBN approach, an arc was manually added from variable CPAP to *Outcome total*. In the inference phase (i.e., when the constructed model was used), the variable CPAP

Table 2. Variables in the data set in temporal index order

Variable name	Explanation	Variable type	Value distribution	Discretization of numerical and interval scale variables	Temporal index	Number of missing values (%)
Age (yr)	Patient's age at baseline	Numerical background variable	21–84 (min–max) 50.5 ± 9.7 (mean ± SD)	≤55 1 = 650 (67.2%) >55 2 = 318 (32.8%)	1	0 (0)
Gender	Patient's gender	Nominal background variable	Male 1 = 801 (81.9%) Female 2 = 177 (18.1%)	-	1	0 (0)
Recruitment time	Date of patient recruitment	Interval scale background variable	Time range: 01/22/1992–04/22/2003 Median, 09/08/1997	1 = 1992–1996 n = 403 (41.2%) 2 = 1997–1999 n = 273 (27.9%) 3 = 2000–2003 n = 302 (30.9%)	2	0 (0)
BMI (kg/m ²)	Body mass index at baseline	Numerical background variable	19.9–74.0 (min–max) 32.0 ± 6.9 (mean ± SD)	≤30 1 = 459 (47.4%) >30 2 = 509 (52.6%)	2	10 (1)
Smoking	Tobacco smoking at baseline	Nominal background variable	No 1 = 428 (43.8%) Quit 2 = 273 (27.9%) Yes 3 = 277 (28.3%)	-	2	0 (0)
Diabetes	Diagnosed diabetes at baseline	Binary background variable	No 0 = 688 (71.8%) Yes 1 = 270 (28.26%)	-	3	20 (2)
BB	Diagnosed hypertension at baseline	Binary background variable	No 0 = 386 (40.3%) Yes 1 = 572 (59.7%)	-	3	20 (2)
Cancer	Diagnosed cancer at baseline	Binary background variable	No 0 = 855 (89.2%) Yes 1 = 103 (10.8%)	-	3	20 (2)
CHF	Diagnosed congestive heart failure at baseline	Binary background variable	No 0 = 858 (89.6%) Yes 1 = 100 (10.4%)	-	4	20 (2)
CHD	Diagnosed coronary heart disease at baseline	Binary background variable	No 0 = 802 (83.9%) Yes 1 = 154 (16.1%)	-	4	20 (2)
PCI	Previous cardiac infarction	Binary background variable	No 0 = 905 (94.5%) Yes 1 = 53 (5.5%)	-	4	20 (2)
Somnolence	Self-reported daytime somnolence at baseline	Binary background variable	No 0 = 247 (25.3%) Yes 1 = 731 (74.7%)	-	4	20 (2)
Diuretic	Use of diuretic medication at baseline	Binary background variable	No 0 = 668 (70.1%) Yes 1 = 285 (29.9%)	-	5	25 (3)

Table 2. Continued

Variable name	Explanation	Variable type	Value distribution	Discretization of numerical and interval scale variables	Temporal index	Number of missing values (%)
AHI	Apnea-hypopnea index at baseline	Numerical background variable	5.0–148.4 (min-max) 24.0 ± 21.7 (mean ± SD)	5.0–15.0 1 = 472 (48.3%) 15.1–30.0 2 = 251 (25.7%) ≥30.1 3 = 255 (26.0%)	6	0 (0)
ODI	Oxygen desaturation index at baseline	Numerical background variable	0–149.0 (min-max) 22.5 ± 21.5 (mean ± SD)	5.0–15.0 1 = 513 (53.0%) 15.1–30.0 2 = 221 (22.8%) ≥30.1 3 = 234 (24.2%)	6	0 (0)
ObsSev	Obstruction severity at baseline	Numerical background variable	0–594.7 (min-max) 35.9 ± 63.8 (mean ± SD)	0–30.0 1 = 725 (75.0%) 30.1–80.0 2 = 142 (14.6%) ≥80.1 3 = 101 (10.4%)	6	37 (4)
DesSev	Desaturation severity at baseline	Numerical background variable	0.08–16.9 (min-max) 1.0 ± 1.8 (mean ± SD)	0–1.0 1 = 748 (77.2%) 1.1–3.0 2 = 135 (14.0%) ≥3.1 3 = 85 (8.8%)	6	1 (0)
CPAP	Continuous positive airway pressure treatment	Binary intervention variable	No 0 = 635 (64.9%) Yes 1 = 343 (35.1%)	-	7	0 (0)
Outcome total	Combined variable of mortality, AMI, and CVI in follow-up period	Binary outcome variable	Dead, AMI, or CVI 1 = 252 (25.8%) Alive, no AMI, or CVI 0 = 726 (74.2%)	-	8	0 (0)

Baseline indicates the time when the patients were recruited into the study.

BMI: body mass index, CHF: congestive heart failure, CHD: coronary heart disease, AHI: apnea-hypopnea index, ODI: oxygen desaturation index, ObsSev: obstruction severity, DesSev: desaturation severity, CPAP: continuous positive airway pressure, AMI: acute myocardial infarction, CVI: cerebrovascular insult.

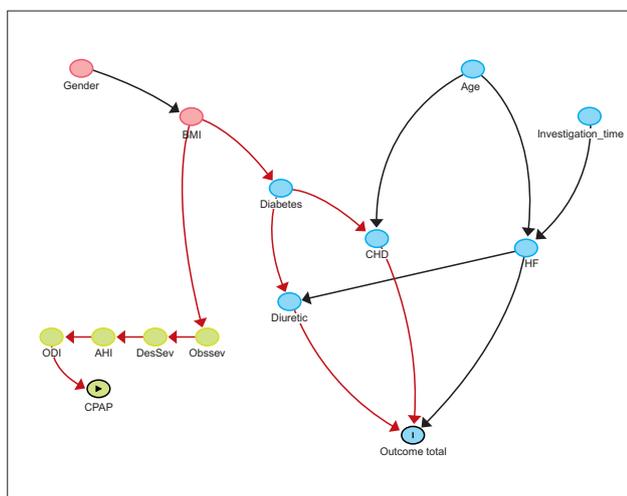


Figure 2. Data-driven model of factors associated with target variable *Outcome total*. Nodes are presented in temporal index order, with parent nodes at the top and child nodes at the bottom. Other variables were dropped from the analysis. Blue node with vertical bar indicates outcome variable; green node with triangle, exposure variable; green nodes, ancestors of exposure variable; red nodes, common ancestors of both exposure and outcome; blue nodes, ancestors of outcome; green arrow, causal path; red arrow, biasing path; and black arrows, other connections.

was set to be an intervention. In this way, real causal dependencies between CPAP and the outcome variable could be identified when this model was purged of unwanted associational backdoor paths between them [12].

Figures 2–4 were drawn with DAGitty software [29].

The Research Ethics Committee of the Hospital District of Northern Savo, Kuopio, Finland approved the protocol, and all the subjects gave written informed consent (No. 127/2004 and 14/2013).

III. Results

The variables with values, distributions, discretizations, temporal indices, and number of missing data are presented in Table 2.

The mean length of the follow-up period was 186 months (standard deviation 58 months, variation 0–276 months). During the follow-up period, altogether 185 patients died (18.9%), of which 154 were men and 31 were women. In addition, 55 men (8.5%) and 12 women (8.2%) had AMI, CVI, or both during the follow-up period.

A total of 252 patients died or had AMI, CVI, or both during the follow-up period. The 209 men (26.12%) and 43 women (24.3%) comprised 25.8% of all patients. Of the

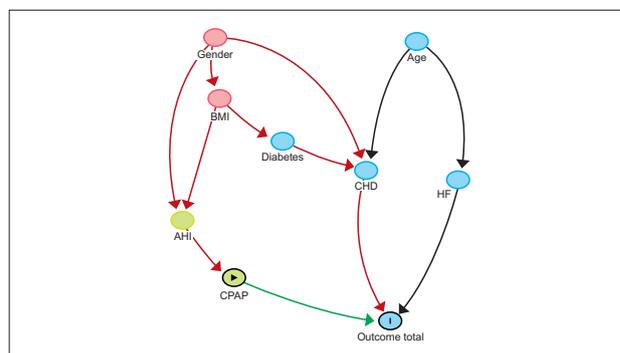


Figure 3. Simplified hypothesis-driven model with target variable *Outcome total*. Nodes are presented in temporal index order, with parent nodes at the top and child nodes at the bottom. According to the hypothesis, an arc from CPAP to *Outcome total* is added and fixed to the model. Blue node with vertical bar indicates outcome variable; green node with triangle, exposure variable; red nodes, common ancestors of both exposure and outcome; blue nodes, ancestors of outcome; green arrow, causal path; and black arrows, other connections.

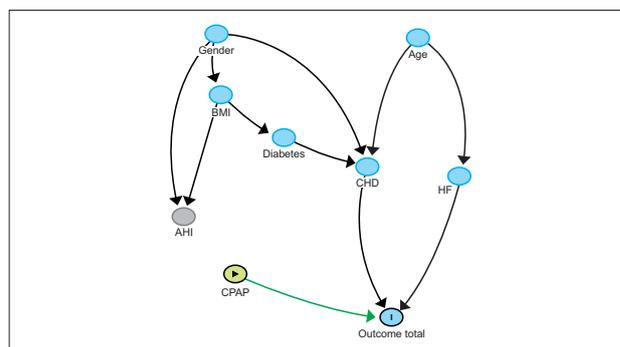


Figure 4. Simplified hypothesis-driven model with target variable *Outcome total*. According to the hypothesis, CPAP was set to an intervention mode and separated from associative paths. Blue node with vertical bar indicates outcome variable; yellow node with triangle, exposure variable; yellow nodes, ancestors of exposure variable; blue nodes, ancestors of outcome; green arrow, causal path; and black arrows, other connections.

patients, 343 (35.1%) had used CPAP treatment for at least 6 months or had continued CPAP at the end of the follow-up.

The DDBN model made by the Taboo algorithm (SC = 1) for the outcome variable *Outcome total* is presented in Figure 2. In the DDBN model, variables *CHD*, *Diuretic*, and *CHF* were causally associated with the outcome. No causal association between sleep apnea parameters or CPAP and the outcome variable was seen. Instead, there was a path between CPAP and *Outcome total* consisting of associational dependencies. There was a weak association between *AHI* and *Outcome total* due to common causes *BMI* and *Gender*.

Table 3. Relationship analysis for the DDBN presented in Figure 2

Parent	Child	Kullback–Leibler divergence	Pearson correlation
AHI	ODI	1.18	0.96
DesSev	ObsSev	0.59	0.87
DesSev	AHI	0.49	0.66
CHF	Diuretic	0.11	0.40
BMI	DesSev	0.08	0.28
ODI	CPAP	0.08	0.31
BMI	Diabetes	0.06	0.29
Diabetes	Diuretic	0.06	0.26
Recruitment time	IT	0.04	-0.22
Age	CHF	0.04	0.22
Gender	DesSev	0.03	-0.13
CHD	Outcome total	0.03	0.26
CHF	CHD	0.03	0.25
Diuretic	Outcome total	0.03	0.29
Age	CHD	0.03	0.22
Gender	CHD	0.02	-0.13
CHF	Outcome total	0.02	0.32
Gender	BMI	0.01	0.14

DDBN: data-driven causal Bayesian network, AHI: apnea-hypopnea index, ODI: oxygen desaturation index, DesSev: desaturation severity, ObsSev: obstruction severity, CHF: congestive heart failure, BMI: body mass index, CPAP: continuous positive airway pressure, CHD: coronary heart disease.

The variable *Recruitment time* was included in the DDBN model (Figure 2) because the patient recruitment period was long (11 years), and *Recruitment time* was considered a potential source of bias. There was an association between *Recruitment time* and the variable *CHF*, indicating that congestive heart failure was a more common finding in patients before the year 2000 than after. However, there was no DE between *Recruitment time* and *Outcome total*.

The relationship analysis of the DDBN model with Kullback–Leibler divergence and Pearson correlation is presented in Table 3. Sleep apnea parameters were strongly associated with each other (for example, *AHI*→*ODI* had the strongest association).

Analysis of DE on the target outcome variable *Outcome total* is presented in Table 4. Variables *CHF*, *CHD*, and *Diuretic* have strong direct effects on *Outcome total*. Sleep apnea parameters and *CPAP* have only a minimal DE on the target. In other words, based on the DDBN approach, there is no

Table 4. Direct effects on the target *Outcome total* in the DDBN presented in Figure 2

Node	Standardized direct effect	Contribution (%)
CHF	0.20	33.8
CHD	0.19	33.0
Diuretic	0.19	32.6
Diabetes	0.02	0.3
Age	-0.00	0.1
Recruitment time	-0.00	0.0
ObsSev	0.00	0.0
BMI	-0.00	0.0
Gender	0.00	0.0
CPAP	0.00	0.0
DesSev	-0.00	0.0
ODI	0.00	0.0
AHI	-0.00	0.0

DDBN: data-driven causal Bayesian network, CHF: congestive heart failure, CHD: coronary heart disease, ObsSev: obstruction severity, BMI: body mass index, CPAP: continuous positive airway pressure, ODI: oxygen desaturation index, AHI: apnea-hypopnea index.

Table 5. Relationships between variables in the HDBN presented in Figure 3

Parent	Child	Kullback–Leibler divergence	Pearson correlation
AHI	CPAP	0.08	0.31
BMI	AHI	0.07	0.28
BMI	Diabetes	0.06	0.29
Heart failure	Outcome total	0.05	0.32
Age	CHD	0.04	0.22
CHD	Outcome total	0.03	0.28
Age	Heart failure	0.03	0.25
Diabetes	CHD	0.03	0.18
Gender	CHD	0.02	-0.13
Gender	AHI	0.02	-0.12
CHD	Heart failure	0.02	0.24
Gender	BMI	0.01	0.14
CPAP	Outcome total	0.01	-0.03

HDBN: hypothesis-driven causal Bayesian network, AHI: apnea-hypopnea index, CPAP: continuous positive airway pressure, BMI: body mass index; AMI: acute myocardial infarction, CHD: coronary heart disease.

Table 6. Direct effects on the target *Outcome total* in the HDBN presented in Figure 3

Node	Standardized direct effect	Contribution (%)
Heart failure	0.28	47.2
CHD	0.23	38.6
CPAP	-0.03	5.7
BMI	0.01	2.0
Diabetes	0.01	1.7
Age	0.01	1.7
AHI	0.01	1.6
Gender	0.01	1.5

HDBN: hypothesis-driven causal Bayesian network, CHD: coronary heart disease, CPAP: continuous positive airway pressure, BMI: body mass index, AMI: acute myocardial infarction.

causal relationship between *CPAP* and the outcome variable.

The HDBN model is presented in Figure 3. The relationship analysis of this model is presented in Table 5, and direct effects on the target are presented in Table 6. In this model several paths from *CPAP* to the target were found; only one path is causal, i.e., the direct link from *CPAP* to *Outcome total*. All other paths from *CPAP* to *Outcome total* are associated with *BMI* or *Gender* as a common cause.

The HDBN model with the variable *CPAP* set as an intervention is presented in Figure 4. When *CPAP* is an intervention, this intervention variable is separated from all non-causal associations. This model was fixed independently for each value of the variables, and the results are presented in Table 7. In general, CPAP treatment showed a 5.3 percentage points improvement in *Outcome total* in comparison with no treatment. The most improvement was seen in patients aged 55 years or less (8.4% improvement with CPAP in comparison with no treatment). In patients with CHF, CPAP treatment showed a 10.2% increase in risk of mortality, AMI, or CVI (HDBN models number 16–17 in Table 7).

IV. Discussion

This analysis is, as far as we know, the first study in which an expert-assisted Taboo learning process with MDL scoring and causal Bayesian networks have been used to estimate clinical effectiveness. No causal query can be answered from data alone, without causal information that lies outside the data. Therefore, expert knowledge is required to complement the analysis [12]. This knowledge was exploited in the study in multiple ways, e.g., defining the known temporality between the variables, blocking non-relevant links from a

causal point of view, and adding causal links based on the default hypothesis. But are the discovered dependences really causal in the sense in which it is defined in [12] as *do-calculus*-equation? The implemented causal analysis follows the guidelines in [30]. Therefore, we can claim that, within the observed variables, the dependences are causal. However, due to weak dependences between multiple variables, causal dependences similarly are weak. Therefore, this has led to some differences between data-driven and hypothesis-driven networks when MDL scoring has been used.

This analysis used patient data obtained from a large referral hospital. We consider the data, which included information about diagnosis of sleep apnea as well as deaths and serious complications, to be very reliable and almost free from information bias.

To avoid modeling biases, several alternative versions were used for discretization and temporal indices, and expert knowledge was used to set arcs and variables for the final analysis. An analysis for mortality alone was also done. The differences between the models were minor.

In a study by Kendzerska et al. [31], the following factors were prognostic factors for cardiovascular disease in sleep apnea patients: time spent with oxygen saturation, sleep time, awakenings, periodic leg movements, heart rate, and daytime sleepiness. In our study, the same factors were not associated with the combined outcome. In our study, all-cause mortality was 18.9%, which is line with the results of Marshall et al., [32] who found 20-year all-cause mortality to be 19.4%.

In the DDBN, no direct association between sleep apnea parameters or CPAP treatment and the outcome variable was found. A weak non-causal association between *AHI* and *Outcome total* can explain the results reported by Rich et al. [3]. BMI was clearly a common cause for both CPAP treatment (and for all the variables in the path from BMI to CPAP) and *Outcome total*.

In this study, the follow-up time was long, at an average of 15.5 years. The weak effect of sleep apnea parameters and CPAP treatment on the outcome variable might be explained by the long follow-up period. According to Meinow et al. [33], health-related indicators are unstable, and their effect is strongest in a 1–2-year follow-up. In a longer follow-up, all health-related factors become weaker predictors of mortality.

In the HDBN, a long-term beneficial effect of CPAP treatment was found. This effect was generally a 5.3 percentage points improvement in the risk of death, AMI, or CVI. This result suggests that the dependency between *CPAP* and *Outcome total* consists of a causal dependency and spurious

Table 7. Fixation table of the HDBN with CPAP set to the intervention

Model number	Fixed values	Number of cases involved	Values of <i>Outcome total</i> Alive, no AMI or CVI = 0 Dead, AMI or CVI = 1
1	No fixation	978	0 = 74.2% 1 = 25.8%
2	CPAP 0 = 100%	635	0 = 72.4% 1 = 27.6%
3	CPAP 1 = 100%	343	0 = 77.7% 1 = 22.3%
4	Age ≤ 55 yr = 100% CPAP 0 = 100%	427	0 = 74.3% 1 = 25.7%
5	Age ≤ 55 yr = 100% CPAP 1 = 100%	230	0 = 80.4% 1 = 19.6%
6	Age > 55 yr = 100% CPAP 0 = 100%	208	0 = 68.5% 1 = 31.5%
7	Age > 55 yr = 100% CPAP 1 = 100%	113	0 = 72.0% 1 = 28.0%
8	AHI ≤ 15 = 100% CPAP 0 = 100%	375	0 = 72.6% 1 = 27.4%
9	AHI ≤ 15 = 100% CPAP 1 = 100%	97	0 = 77.9% 1 = 22.1%
10	AHI 15.1–30.0 = 100% CPAP 0 = 100%	163	0 = 72.3% 1 = 27.7%
11	AHI 15.1–30.0 = 100% CPAP 1 = 100%	52	0 = 77.6% 1 = 22.4%
12	AHI > 30 = 100% CPAP 0 = 100%	166	0 = 71.9% 1 = 28.1%
13	AHI > 30 = 100% CPAP 1 = 100%	52	0 = 77.1% 1 = 22.9%
14	CHF 0 = 100% CPAP 0 = 100%	570	0 = 76.5% 1 = 23.5%
15	CHF 0 = 100% CPAP 1 = 100%	308	0 = 83.5% 1 = 16.5%
16	CHF 1 = 100% CPAP 0 = 100%	65	0 = 36.6% 1 = 63.4%
17	CHF 1 = 100% CPAP 1 = 100%	35	0 = 26.4% 1 = 73.6%
18	CHD 0 = 100% CPAP 0 = 100%	536	0 = 77.6% 1 = 22.4%
19	CHD 0 = 100% CPAP 1 = 100%	288	0 = 81.8% 1 = 18.2%
20	CHD 1 = 100% CPAP 0 = 100%	99	0 = 46.5% 1 = 53.5%

Table 7. Continued

Model number	Fixed values	Number of cases involved	Values of <i>Outcome total</i> Alive, no AMI or CVI = 0 Dead, AMI or CVI = 1
21	CHD 1 = 100% CPAP 1 = 100%	54	0 = 53.0% 1 = 47.1%

The unfixed model (model number 1) presents direct distributions of the material. In models 2 to 21, each model has been fixed to a certain value (100% indicates the fixed value). The table shows how the results of the target variable *Outcome total* change with different fixations. Only selected fixations are shown.

HDBN: hypothesis-driven causal Bayesian network, CPAP: continuous positive airway pressure, AMI: acute myocardial infarction, CVI: cerebrovascular insult, AHI: apnea-hypopnea index, CHF: congestive heart failure, CHD: coronary heart disease.

associational dependencies enabled by *BMI* and *Gender* as common causes. This result can be compared with the study by Jennum et al. [34] who found that CPAP therapy is associated with reduced all-cause mortality in males, but not significantly in females.

Besides the long follow-up, our results differ from analyses done using conventional methods in two other ways. First, our study aimed to estimate clinical effectiveness, which differs from efficacy measurement in randomized trials. Secondly, most conventional methods in sleep apnea research using observational data are unable to distinguish direct effects from associational effects.

In addition, in patients with CHF, an increased risk of death, AMI, or CVI was seen when CPAP was used. This result is opposite to the findings of previous studies [35,36], which indicated a beneficial effect of CPAP on CHD and CHF. This result suggests that an unknown factor exists that mediates the association between CPAP and CHF. The result can be compared to those of some studies that have shown a detrimental effect of CPAP in CHF patients [37-39]. The divergence between DDBN and HDBN may be due to differences between subgroups.

We consider that the methodology used in this study gives a realistic view of treatment effectiveness. Bayesian methods also have potential value in analyzing similar problems in other contexts. The prognosis of OSA and the effectiveness of CPAP can be estimated on an individual level using prognostic factors in patients' demographic factors, comorbidity, and the results of sleep polygraphs. The effectiveness of CPAP is seen in patients without other diseases, but in more severely ill patients, the prognosis is determined by the underlying diseases. CPAP is an effective treatment that loses its effectiveness in patients with serious cardiovascular disease.

Conflict of Interest

OPR is a shareholder in Wisane Ltd., a company producing analyses in health care. The other authors have no conflicts of interest.

References

1. Kronholm E, Sallinen M, Suutama T, Sulkava R, Era P, Partonen T. Self-reported sleep duration and cognitive functioning in the general population. *J Sleep Res* 2009;18(4):436-46.
2. Muraja-Murro A, Eskola K, Kolari T, Tiihonen P, Hukkanen T, Tuomilehto H, et al. Mortality in middle-aged men with obstructive sleep apnea in Finland. *Sleep Breath* 2013;17(3):1047-53.
3. Rich J, Raviv A, Raviv N, Brietzke SE. All-cause mortality and obstructive sleep apnea severity revisited. *Otolaryngol Head Neck Surg* 2012;147(3):583-7.
4. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med* 2005;353(19):2034-41.
5. Sullivan CE, Issa FG, Berthon-Jones M, Eves L. Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares. *Lancet* 1981;1(8225):862-5.
6. Giles TL, Lasserson TJ, Smith BH, White J, Wright J, Cates CJ. Continuous positive airways pressure for obstructive sleep apnoea in adults. *Cochrane Database Syst Rev* 2006;(1):CD001106.
7. Wang X, Ouyang Y, Wang Z, Zhao G, Liu L, Bi Y. Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: a meta-analysis of prospective cohort studies. *Int J Cardiol* 2013;169(3):207-14.

8. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technol Assess* 2000;4(38):1-130.
9. Sheingold SH. Can Bayesian methods make data and analyses more relevant to decision makers? A perspective from Medicare. *Int J Technol Assess Health Care* 2001;17(1):114-22.
10. Allmark P. Bayes and health care research. *Med Health Care Philos* 2004;7(3):321-32.
11. Woertman W, van der Wilt GJ. Estimating the effectiveness of HPV vaccination in the open population: a Bayesian approach. *Value Health* 2013;16(4):604-9.
12. Pearl J. *Causality*. 2nd ed. Cambridge: Cambridge University Press; 2009.
13. American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep* 1999;22(5):667-89.
14. Kulkas A, Tiihonen P, Eskola K, Julkunen P, Mervaala E, Toyras J. Novel parameters for evaluating severity of sleep disordered breathing and for supporting diagnosis of sleep apnea-hypopnea syndrome. *J Med Eng Technol* 2013;37(2):135-43.
15. Kulkas A, Tiihonen P, Julkunen P, Mervaala E, Toyras J. Novel parameters indicate significant differences in severity of obstructive sleep apnea with patients having similar apnea-hypopnea index. *Med Biol Eng Comput* 2013;51(6):697-708.
16. Leppanen T, Toyras J, Muraja-Murro A, Kupari S, Tiihonen P, Mervaala E, Kulkas A. length of individual apnea events is increased by supine position and modulated by severity of obstructive sleep apnea. *Sleep Disord* 2016;2016:9645347.
17. Muraja-Murro A, Kulkas A, Hiltunen M, Kupari S, Hukkanen T, Tiihonen P, et al. Adjustment of apnea-hypopnea index with severity of obstruction events enhances detection of sleep apnea patients with the highest risk of severe health consequences. *Sleep Breath* 2014;18(3):641-7.
18. Leppanen T, Sarkka M, Kulkas A, Muraja-Murro A, Kupari S, Anttonen M, et al. RemLogic plug-in enables clinical application of apnea-hypopnea index adjusted for severity of individual obstruction events. *J Med Eng Technol* 2016;40(3):119-26.
19. Muraja-Murro A, Kulkas A, Hiltunen M, Kupari S, Hukkanen T, Tiihonen P, et al. The severity of individual obstruction events is related to increased mortality rate in severe obstructive sleep apnea. *J Sleep Res* 2013;22(6):663-9.
20. BayesiaLab [Internet]. Franklin (TN): Bayesia USA; c2018 [cited at 2017 Jul 14]. Available from: <http://www.bayesialab.com/>.
21. Kjaerulff UB, Madsen AL. *Bayesian networks and influence diagrams*. New York (NY): Springer-Verlag; 2008.
22. Bai X. *Tabu search enhanced Markov blanket classifier for high dimensional data sets*. Pittsburg (PA): Carnegie Mellon University; 2005.
23. Grunwald PD. *The minimum description length principle*. Cambridge (MA): MIT Press; 2007.
24. Kekolahti P, Karikoski J, Riikonen A. The effect of an individual's age on the perceived importance and usage intensity of communications services: a Bayesian network analysis. *Inf Syst Front* 2015;17(6):1313-33.
25. Kekolahti P, Kilkki K, Hammainen, H, Riikonen A. Features as predictors of phone popularity: an analysis of trends and structural breaks. *Telematics Inform* 2016;33(4):973-89.
26. Kasza J, Solomon P. Comparing score-based methods for estimating Bayesian networks using the Kullback-Leibler divergence. *Commun Stat Theory Methods* 2015;44(1):135-52.
27. Shamilov A, Asan Z, Giriftinoglu C. Estimation by MinxEnt principle. *Proceedings of the 9th WSEAS International Conference on Applied Mathematics (WSEAS)*; 2006 May 27-29; Istanbul, Turkey. p. 436-40.
28. Friedman N. The Bayesian structural EM algorithm. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*; 1998 Jul 24-26; Madison, WI. p. 129-138.
29. Textor J, Hardt J, Knuppel S. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology* 2011;22(5):745.
30. Pearl J, Mackenzie D. The ladder of causation. In: *The book of why: the new science of cause and effect*. New York (NY): Basic Books; 2018. p. 23-52.
31. Kendzerska T, Gershon AS, Hawker G, Leung RS, Tomlinson G. Obstructive sleep apnea and risk of cardiovascular events and all-cause mortality: a decade-long historical cohort study. *PLoS Med* 2014;11(2):e1001599.
32. Marshall NS, Wong KK, Cullen SR, Knuiman MW, Grunstein RR. Sleep apnea and 20-year follow-up for all-cause mortality, stroke, and cancer incidence and mortality in the Busselton Health Study cohort. *J Clin Sleep Med* 2014;10(4):355-62.
33. Meinow B, Kareholt I, Parker MG, Thorslund M. The effect of the duration of follow-up in mortality analysis:

- the temporal pattern of different predictors. *J Gerontol B Psychol Sci Soc Sci* 2004;59(3):S181-9.
34. Jennum P, Tonnesen P, Ibsen R, Kjellberg J. All-cause mortality from obstructive sleep apnea in male and female patients with and without continuous positive airway pressure treatment: a registry study with 10 years of follow-up. *Nat Sci Sleep* 2015;7:43-50.
 35. Kasai T, Narui K, Dohi T, Yanagisawa N, Ishiwata S, Ohno M, et al. Prognosis of patients with heart failure and obstructive sleep apnea treated with continuous positive airway pressure. *Chest* 2008;133(3):690-6.
 36. Gottlieb DJ, Yenokyan G, Newman AB, O'Connor GT, Punjabi NM, Quan SF, et al. Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure: the sleep heart health study. *Circulation* 2010;122(4):352-60.
 37. Cowie MR, Woehrle H, Wegscheider K, Angermann C, d'Ortho MP, Erdmann E, et al. Adaptive servo-ventilation for central sleep apnea in systolic heart failure. *N Engl J Med* 2015;373(12):1095-105.
 38. Yogasundaram H, Oudit GY. Increased mortality associated with adaptive servo-ventilation therapy in heart failure patients with central sleep apnea in the halted SERVE-HF Trial. *Can J Cardiol* 2015;31(9):1202-3.
 39. Woehrle H, Cowie MR, Eulenburg C, Suling A, Angermann C, d'Ortho MP, et al. Adaptive servo ventilation for central sleep apnoea in heart failure: SERVE-HF on-treatment analysis. *Eur Respir J* 2017;50(2):1601692.