

Analysis of risk factors of hip fracture with causal Bayesian networks

Alex Aussem¹, Pascal Caillet², Zara Klemm¹, Maxime Gasse¹, Anne-Marie Schott², and Michel Ducher³

¹ Universite de Lyon, CNRS

Universite Lyon 1, LIRIS, UMR5205, F-69622, France

² Hospices Civils de Lyon, Pole Information Medicale Evaluation Recherche, Lyon, F-69003, France ; Universite de Lyon, Universite Lyon 1, Lyon, F-69000, France ; INSERM U1033, Lyon, F-69003, France

³ Hospices Civils de Lyon, Groupement Hospitalier de Geriatrie, Francheville, F-69340, France

Abstract. We explore a practical approach to learn a *plausible* causal Bayesian network from a combination of non-experimental data and qualitative assumptions that are deemed likely by health experts. The method is based on the incorporation of prior expert knowledge in the form of partial pairwise ordering constraints between variables into a recent constraint-based Bayesian network structure learning algorithm. The learning process ends up with a partially oriented graph. The remaining undirected edges are then oriented according to the expert understanding. We show that the causal graph not only provides a statistical profile of the population under study but also offers a simple guideline principle to identify accessible sets of confounding variables for each causal relation under interest. To illustrate the potential of the proposed approach, we estimate the strength of the causal effect of psychotropic drugs, gait speed, body mass index and bone mineral density on the risk of hip fracture from a prospective cohort study EPIDOS sample, which included more than 7500 elderly osteoporotic women followed-up during 4 years. Our findings suggest that an intervention programme aimed at preventing physical deterioration and maintaining bone mass density should tend to reduce the risk of hip fracture among elderly.

1 Introduction

Osteoporotic fractures, including hip fractures, are a global health concern associated with significant morbidity and mortality as well as a major economic burden [1, 10]. This disease is caused by a modification of the bone's structure which translates into an increased risk of hip fractures [12]. Hip fractures commonly result in permanent disability, institutionalization or death, and are one of the most damaging fractures among elderly people [10]. Thus, identifying the causal risk factors for the development of timely interventions, such as pharmacotherapy, to limit bone structure degradation in the elderly osteoporotic population is an important challenge.

In this paper, we investigate the feasibility and usefulness of causal bayesian networks (BNs) in this setting. BNs learning algorithms search for statistical (not necessarily causal though) relationships between the disease and all potential risk factors simultaneously. This approach shows an increasing popularity in the medical domain [19, 18, 4, 8, 7], partly due to the advantages of a graphical representation that facilitates the communication between domain experts and knowledgeable engineers. BNs exhibit the variable statistical relationships qualitatively by means of a directed acyclic graph (DAG) and thus are relatively easy to interpret by non-statisticians. The ubiquity of DAG models in statistical applications stems primarily from their causal interpretation. However, BNs are only models of statistical association and it is a common misinterpretation to assume that arcs in a BN model denote causality, they denote only statistical dependency. In fact, two BNs that are observationally equivalent cannot be distinguished without resorting to manipulative experimentation or temporal information. Expert knowledge and biological understanding is clearly essential, since it is more than a statistical data analysis exercise to provide a causal representation of the data generating process.

The emphasis in this study is placed on integrating medical domain knowledge and statistical data analysis to produce a plausible causal DAG in which the edges are interpreted as causalities by the domain expert. The method is based on the incorporation of prior expert knowledge in the form of pairwise constraints into a BN structure learning algorithm called H2PC that appeared recently in the machine learning literature [15]. The method ends up with a partially oriented graph. The remaining undirected edges are then confronted to the knowledge of the domain expert and directed according to his causal interpretation. We then use the causal graph to infer the presence of confounding factors and we estimate the causal odds ratios to discern the causal information content of each risk factor based on the do-calculus developed by Pearl [21] from observational data alone. Finally, our findings are compared with the ones obtained by traditional logistic regression, published recently in the epidemiologic literature.

2 Background

2.1 Bayesian networks

Formally, a BN is a tuple $\langle \mathbb{G}, P \rangle$, where $\mathbb{G} = \langle \mathbf{U}, \mathbf{E} \rangle$ is a directed acyclic graph (DAG) with nodes representing the variables in the domain \mathbf{U} , and edges representing direct probabilistic dependencies between them. P denotes the joint probability distribution on \mathbf{U} . The BN structure encodes a set of conditional independence assumptions: that each node X_i is conditionally independent of all of its non-descendants in \mathbb{G} given its parents. These independence assumptions, in turn, imply many other conditional independence statements, which can be extracted from the network using a simple graphical criterion called d-separation [20]. We say that $\langle \mathbb{G}, P \rangle$ satisfies the *faithfulness condition* if the d-separations in \mathbb{G} identify *all and only* the conditional independencies in P . Two graphs are

said *equivalent* if they encode the same set of conditional independencies via the d-separation criterion. These DAGs are said observationally equivalent - that is, every probability distribution that is compatible with one of the DAGs is also compatible with the other. Two DAGs are equivalent *iff* they have the same underlying undirected graph and the same set of v-structures (i.e., uncoupled head-to-head meetings $X \rightarrow Y \leftarrow Z$) [20]. Furthermore using the concept of a Markov boundary, we can identify the minimum set of variables that shield off the target variable (hip fracture in our case) from the influence of the remaining variables. Furthermore, if P is faithful to a DAG (as we will assume in the sequel), then the Markov boundary of a variable X : it consists of the parents, the children and the spouses of X in \mathbb{G} .

2.2 Learning the structure

The problem of finding the DAG that encodes the conditional independencies present in the data attracted a great deal of interest over the last years [24, 25, 2, 15, 23, 27]. Ideally the DAG should coincide with the dependence structure of the global distribution, or it should at least identify a distribution as close as possible to the correct one in the probability space. This step, called structure learning, is similar in approaches and terminology to model selection procedures for classical statistical models. Basically, constraint-based (CB) learning methods systematically check the data for conditional independence relationships and use them as constraints to construct a partially oriented graph representative of a BN equivalence class, whilst search-and-score (SS) methods make use of a goodness-of-fit score function for evaluating graphical structures with regard to the data set. Hybrid methods attempt to get the best of both worlds: they learn a skeleton with a CB approach and constrain on the DAGs considered during the SS phase. There are many excellent treatments of BNs which survey the learning methods (see [17] for instance). In this study, we use a novel hybrid algorithm for BN structure learning, called Hybrid HPC (H2PC) [15]. H2PC was shown experimentally to outperform several state-of-the-art algorithms on several benchmarks with various data sizes, in terms of goodness of fit to new data and in terms of the quality of the network structure itself, which is close to the true dependence structure of the data. We point the reader to [24, 15] for further details.

Of course, H2PC, like any other algorithm, is sensitive to the particular data set at hand [14]. So, we cannot simply accept our chosen structure as a true representation of the underlying distribution. Averaging over the sampled structures that are generated by a sampling process produces models that are more robust, have greater confidence and place less reliance on a single data set. Several approaches exist: generating samples of the BN structure from its marginal posterior distribution using Monte Carlo Markov Chain (MCMC) [16], or using bootstrapping methods for computing a statistical confidence features within a BN [26]. In this study, we make use of the bootstrapping method to generate a more robust network structure. Confidence in a particular edge is measured as a percentage of the number of times this edge actually appears in

the set of reconstructed graphs. If an edge has a confidence above the threshold, it is included in the consensus network. Thus, only the dependencies that have enough support will be captured and represented in the final consensus DAG.

2.3 Case study

In this study, we used the EPIDOS cohort described in [11] which consists of 7598 women aged 75 years or older. The mean age was 80.5 years (std=3.8). These women were recruited between January, 1992, and January, 1996 in five French cities: Amiens, Lyon, Montpellier, Paris, and Toulouse and followed up by mailed questionnaires every 4 months during 4 years. Women who were not able to walk independently and those who had a hip fracture or bilateral hip replacement were excluded. Femoral-neck BMD by dual-photon X-ray absorptiometry and potential fall-related risk factors were assessed, which included self-reported physical capacity, neuromuscular function, mobility, visual function, history of previous falls and use of medication. During an average of 3.8 years of follow-up, 293 women suffered a hip fracture. After this 4 year period, only the vital status was regularly assessed until 2010 by checking the French national registry of death (INSEE).

Based on literature, especially to the FRAX tool [5] and on expert knowledge, we used a subset of 15 variables (out of 70) to describe the study population (see Table 1): age, body mass index at inclusion, current or past use of corticoids during 3 months or more, t-score at femoral neck, number of falls during the 6 months before inclusion, weekly intake of alcohol, tobacco smoking status, history of hip fracture since 55 of age, parental history of hip fracture, gait speed, Five Times Sit to Stand Test results which is a proxy of the motor performances of the patients, number of recorded chronic diseases (among diabetes, depression, glaucoma, cataract, angora, Parkinson disease and hypertension), current or past use of vitamin D in the past year, current psychotropic drug use and the outcome variable, i.e. hip fracture. Hip fracture over the 4 years was investigated every 4 month by mailed questionnaire and ascertained by X-rays radiography by an expert rheumatologist. Data were discretized when needed according to the expert knowledge [11].

3 Experiments

3.1 Causal graph discovery with expert priors

An often discussed topic when working with BNs is how the expert knowledge can be incorporated in the model before or during the learning process. Edge orientation could partially (under certain assumptions) be achieved from observational data alone but such judgments are more reliable when they are anchored onto fundamental blocks of the domain expert knowledge [21]. In our view, an expert cannot safely construct a fully specified DAG even if he feels confident about variables relationships. Expert elicitation is expensive, time-consuming and relies on experts having full knowledge. On the other hand, automated learning

| Name | Description | Values |
|---------------|---|--|
| Fracture | Hip-fracture during the 4-years follow-up | binary |
| Age | Age at study inclusion | < 80 , $80 \leq 85$, $85 \leq 90$, > 90 |
| Chron_disease | Number of chronic diseases | binary: < 2 , ≥ 2 |
| Psycho | Use of sedatives or anxiolytics at inclusion | binary |
| Vit_D | Use of vitamin D at inclusion or history of vitamin D one year before inclusion | binary |
| Glucoc | Use of glucocorticoids at inclusion or history of glucocorticoids one year before inclusion | binary |
| Alcohol | Daily intake of alcohol in g | binary: $1 \leq 20$, > 20 |
| Tobacco | Tobacco smoking | none, former, actual |
| Gait_speed | Gait speed at inclusion in m/s | < 0.60 , $0.6 \leq 0.85$, $0.85 \leq 1$, > 1 |
| Test_5 | Five chair test (time to sit down and stand up 5 times) in s | $1 \leq 9$, $9 \leq 16$, $16 \leq 23$, > 23 , incapable |
| BMI | Body mass index at inclusion | low, normal, obesity |
| BMD | T-score of BMD of the neck at inclusion | binary: normal or ≤ 1 , ≤ 2.5 |
| Falls | Number of falls during 6 months before inclusion | binary: ≤ 2 , > 2 |
| Earl_Frac | History of fracture from age of 55 to inclusion | binary |
| Par_Frac | History of hip-fracture in the parents | binary |

Table 1. Variables included in the study.

from observational data is restricted to identification of statistical dependency. We therefore asked our expert for constraints to limit the possible edge orientations in the graph structure as proposed in [4]. The expert was provided with a cross table shown in Table 2. The result is a set of priors for pairwise relations. These assumptions are partly based on the notion that causality only occurs in the forward time direction (e.g. Earl_Frac $<$ Fracture) and also on the fact that intuition deems many cause-effect relationships impossible (e.g. Tobacco $<$ Age is deemed impossible). The use of partial ordering that constrains arcs in a causal BN is standard in BN learning (e.g., Tetrad IV and K2) and was adopted in other medical studies [4]. So $X < Y$ becomes the structural constraint that X cannot be a descendant of Y . Note however that this soft constraint in no way implies a direct or indirect causal path between X and Y though. The constraints given to the learning algorithm are listed in Table 2. Satisfying these constraints does not establish, of course, the validity of the causal model postulated. Our hope is that the resulting DAG will best fit the independence statements present in the data and the constraints expressed by the expert.

The resulting consensus DAG shown in Figure 1. It is the average over 200 graphs learned from bootstrapped data. All edges that appeared in more than

| Cause | Age | Bmi | Bmd | Gsp | 5TS | Efr | Pfr | Crdr | Vtd | Glu | Psy | Alc | Tob | Fal | Fra |
|---------------|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|
| Age | | < | < | < | < | < | < | < | < | < | < | < | < | < | < |
| BMI | | | < | < | | | | | | | | | | | < |
| BMD | | | | | | < | | | | | | | | | < |
| Gait_speed | | | < | | | | | | | | | | | | < |
| Test_5 | | | < | < | | | | | | | | | | | < |
| Earl_Frac | | | | | | | | | | | | | | | < |
| Par_Frac | < | < | < | < | < | < | | < | < | < | < | < | < | < | < |
| Chron_disease | | | < | < | < | | | | | | < | | | | < |
| Vit_D | | | | | | | | | | | | | | | < |
| Gluco | | | | | | | | | < | | | | | | < |
| Psycho | | | | < | | | | | | | | | | | < |
| Alcohol | | | < | < | < | | < | | | < | | | | | < |
| Tobacco | | | < | < | < | | < | | | < | | | | | < |
| Falls | | | < | | | | | | | | | | | < | < |
| Fracture | | | | | | | | | | | | | | | < |

Table 2. Pairwise ordering constraints obtained by the expert elicitation process. For X in Row, and Y in Column, $X < Y$ indicates that X cannot be a descendant of Y .

25% of the graphs were included in the final causal graph. As may be seen, the DAG provides a coherent picture of the population under study. It reflects the investigator's subjective and qualitative knowledge of causal influences in the domain. The substantive assumptions embodied in this causal DAG are negative causal assertions which are conveyed through the links missing from the graph. For example, the missing arrow between Chronic disease and Fracture signifies the investigator's understanding that Chronic disease does not affect the risk of hip fracture directly; its entire influence on hip fracture is mediated by BMI, Test_5 and Psycho. Alcohol intake and tobacco smoking are associated but none of them is associated with hip fracture. Likewise, parental history of hip fracture is a singleton node in the graph. These three variables were thus discarded from our analysis. We may also deduce, using the d-separation criterion, that BMD and Chron Disease are independent given Age and BMI; that BMI and Gait Speed are independent given Chron Disease. The graph reveals that gait speed is consequence of age, consumption of psychotropics and Test_5 and that the number of falls 6 months before inclusion is a direct consequence of gait speed. Age and BMI were found to influence directly the BMD and Psychotropic drug use were found to influence directly the gait speed and the Test_5 test results, which in turn influenced history of previous falls (Falls). It is noticeable that neither history of hip fracture nor previous fall (Falls) were directly associated with hip fracture in this model, despite the fact that the hypothesis of Falls being a cause of fracture but not the opposite was allowed in the expert constraints. We also observe Vit_D is descendant of BMD, BMI and Earl_Frac. This is somewhat surprising and calls for an explanation. The reason is that vitamin D was prescribed before inclusion in the study in view of the BMD and BMI

levels of the women, as well as her fracture history. This is clearly a limitation intrinsic to our static analysis (incorporation of time is left to future work).

Finally, the only two variables that were directly linked to fracture were gait speed and bone mineral density. Like embedded feature selection methods, BN learning techniques have an important computational advantage as they combine model construction with feature selection. Rephrased in the language of probability, {Gait Speed, BMD} is the Markov boundary of Hip Fracture. In other words, these two variables shield off hip fracture from the influence of the remaining ones. The area under the ROC curve (AUC) of our causal BN, estimated by 10-fold cross-validation, was 0.71 [0.68 – 0.73], not statistically different from that obtained with the logistic regression. This is interesting insofar as {BMD, Gait Speed} are the only two predictive variables effectively used for predicting the occurrence of hip fracture by our model.

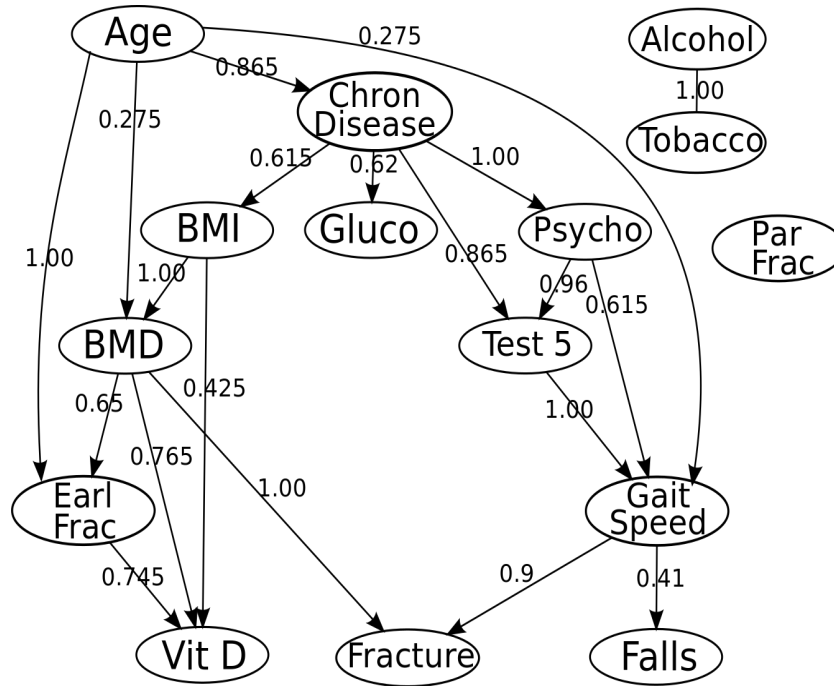


Fig. 1. The consensus causal DAG obtained by bootstrapped H2PC using the pairwise constraints along with the edge confidence values.

| Variable | Causal BN | Logistic regression | |
|---|------------------|---------------------|---------|
| | COR (95%CI) | OR (95%CI) | p-value |
| BMD (T-score) | | | |
| $\geq -2.5std$ | Reference | Reference | |
| $\leq -2.5std$ | 4.15 (3.65-4.72) | 3.54 (2.29-5.46) | < 0.05 |
| Gait speed ($m.s^{-1}$) | | | |
| < 0.6 | Reference | Reference | |
| $0.6 \leq 0.85$ | 0.55 (0.48-0.63) | 0.65 (0.48-0.87) | < 0.05 |
| $0.85 \leq 1$ | 0.29 (0.25-0.33) | 0.36 (0.25-0.52) | < 0.05 |
| ≥ 1 | 0.17 (0.14-0.19) | 0.21 (0.13-0.33) | < 0.05 |
| Psychotropics | | | |
| No | Reference | Reference | |
| Yes | 1.56 (1.38-1.77) | 1.32 (1.02-1.69) | < 0.05 |
| BMI | | | |
| 15-25 | Reference | Reference | |
| < 15 | 1.85 (1.55-2.22) | 1.67 (0.98-2.87) | 0.06 |
| > 25 | 0.68 (0.6-0.78) | 0.64 (0.42-0.98) | < 0.05 |

Table 3. Causal Odds ratios learned from the causal BN by adjusting for confounders compared to statistical Odds ratios obtained from the standard logistic regression.

3.2 Exposure effect estimation and confounder selection

In this study, we consider the causal odds ratios (abbreviated $COR(X, Y)$). When X and Y are not confounded, the causal effect $COR(X, Y)$ is of course the standard odd ratio $OR(X, Y)$. However, we may observe from our causal DAG that other factors associated with both Psycho and Hip Fracture may confound the targeted causal effect. We have to unblock the extraneous flow of influence between Psycho and Hip Fracture, which appear under the rubric of spurious correlation. It is called spurious because it is not part of what we seek to estimate - the causal effect of Psycho and Hip Fracture in the target population. Patients who receive different treatments (e.g. sedative or anxiolytics) tend to differ in health characteristics (i.e. Chron Disease), so biasing naive estimates of treatment effects. The common method of reducing confounding bias in the analysis of causal effects is to adjust for a set of confounders. Choosing which variables to use for adjustment in studies with many measured covariates is an important step for ensuring the validity of effect estimates. Given the causal DAG, a sufficient set for estimating the causal effect of X on Y is any set of non-descendants of X that d-separate X from Y after removing all arrows emanating from X (see [21] pp. 80). This criterion, called *back-door*, provides a mathematical definition of confounding and helps researchers identify accessible sets of variables worthy of measurement. If a set of variables \mathbf{Z} satisfies the back-door criterion relative to (X, Y) , then the causal effect of X on Y , denoted as $P(y|do(x))$, is given by the formula

$$P(y|do(x)) = \sum_{\mathbf{z}} P(y|x, \mathbf{z})P(\mathbf{z})$$

So, we used our causal graph to identify confounders when calculating certain exposures on the outcome hip-fracture. In table 4, different valid sets of confounders are listed for each cause-effect relation under interest. The selection of confounders is predicated, of course, on Fig. 1 being the correct data-generating model. We used the set with the smaller sampling variability or the outcome-related set as advocated by Judea Pearl [22] because modeling the outcome mechanism is a much safer strategy for estimating causal effects in observational studies. For instance, adjusting for variables Gait speed is therefore more appropriate than {Chron Disease, Age} for evaluating the effect of BMD and BMI on hip fracture. Likewise, Gait Speed satisfies the back door criterion relative to the causal effects of BMD and BMI on hip fracture. So causal BNs offer a simple guideline principle to identify accessible sets of confounding variables for each causal relation.

The causal Odds ratios learned from the causal BN by adjusting for confounders are reported in Table 3. As may be observed, current use of sedative or anxiolytics was associated with an increase of hip fractures risk (COR = 1.56 [1.38-1.77]) after adjusting for Chron Disease. Having a BMI higher than 25 appears to have a protective effect (COR = 0.68 [0.60-0.78]), contrasting with the trend observed for women having a BMI below 15 (COR = 1.85 [1.55-2.22]). Concerning gait speed, we observed that the higher the measured gait speed, the less the patient was prone to sustain a fracture, suggesting an important effect of gait over the fracture risk. As expected, a low bone mass density was positively associated with sustaining a hip fracture in the next four years (COR = 4.15 [3.65-4.72]).

Not discussed here, we also used a logistic regression model with a conservative backward selection approach (retained threshold of $p < 0.25$ for prior inclusion, and $p > 0.10$ for exclusion of variables in the model, which are common thresholds in epidemiology). The logistic regression analysis found a statistically significant association of fracture with several variables, including sedative or anxiolytics use, Age, BMD, BMI, gait speed, personal history of fracture and history of more than 2 falls in the previous semester. The statistical Odds ratios obtained from the standard logistic regression were adjusted upon these variables. In other words, we did not use the causal graph for identifying the variables that must be controlled to obtain unconfounded effect estimates. We simply used the prevailing practice in epidemiology. The resulting Odds ratios are also shown in Table 3 for comparison. As may be observed, the ORs obtained by logistic regression are in good agreement with those obtained using the causal DAG. The CORs values are somewhat larger for low BMD values and large BMI.

3.3 Discussion and related work

Our findings are in nice agreement with previous studies as we shall discuss now. A recent study involving causal effect estimation of bazedoxifene acetate on

| Exposure | Admissible sets of confounders |
|------------|---|
| BMD | {Chron Disease, Age} ; {Gait Speed}; {Age, BMI} |
| Gait_speed | {Chron Disease, Age} ; { BMD }; {Age, BMI} |
| Psycho | {Chron Disease} |
| BMI | {Chron Disease} |

Table 4. Variables confounding the effect of BMD, gait speed, psychotropic drugs and BMI on fracture.

fracture by use of structural equation modeling found age and body mass index to be causally linked to BMD that, in turn, had an effect on fractures [3]. Our results suggest that gait-speed is causally associated with fracture and mediate the effect of other variables, including age. These results are supported by another study performed in a sample of 469 older adults in 1996 that found that an improvement in usual gait speed predicted a better survival over 36 months [13]. These observations confirm that an optimal prevention of hip fracture has to be thought as a multi-component intervention, at least involving preservation of structural properties of bone and improvement of gait in the meantime. Another recent study involving the synthesis of the results of 160 studies available from the literature in a single BN and the evaluation of its predictive performances on a sample of 288 institutionalized elderly patients found that psychotropic drugs was also a predictive factor for fall [6], which is in turn strongly and causally associated to hip-fracture. Furthermore, the initial analysis of EPIDOS data using Cox models and accounting for time found also a great influence of gait speed and BMD in the occurrence of hip fracture [11] and some other study involving more traditional analysis accounting for time showed similar results (see [9] and references therein). Incorporation of time, competitive risk and hierarchical structures embedded in the data in BN modeling is actually an important field of research in bioinformatics and suggest avenues for future research. This is left to future work.

4 Conclusion

This paper focused on identification of the relationships between the occurrence of hip fracture and its potential risk factors using BNs, with the emphasis on integrating medical domain knowledge and statistical data analysis and testing their usefulness in causal representation of the hip fracture epidemiology. From an epidemiological perspective, the present study confirms, among other interesting findings, that use of sedatives or anxiolytics, bone mineral density, body mass index are direct or indirect causal risk factors for hip fracture. It was reassuring that this BN analysis uncovered previously established relationships between the above risk factors and hip fracture. Our findings tend to confirm that intervention programs targeting gait improvement and limiting the decrease in bone mineral density should be relevant.

This study had several limitations. First, the participants were volunteers who lived independently at home, and are probably healthier than average for their age, which can limit the generalization of our results to all older women. Our results may not be applicable to less mobile and less healthy women, such as nursing-home residents. However, mobile women who live independently at home are more likely to participate in preventive programs than their less healthy peers. The fact that women were preferentially selected to the cohort, depending on the values of some variables in the our model, is likely to introduce spurious associations, i.e. selection bias, not to mention the problem entailed by the unmeasured confounders that may, under certain circumstances, amplify the bias [22]. So, our conclusions need to be regarded with caution.

Second, our model is quite simplistic, of course, as it is based on certain assumptions that all experts might not agree to - for instance, that there is no direct link between past use of corticoids and fracture. The model would need to be refined then, and we might end up with a graph containing twenty variables or more. Nonetheless, there is no need to worry if another experts tells us we should take this or that factor into account. On the contrary, the graph welcomes such new ideas, because it is so easy to add new observed or unobserved factors into the model. The rules of do-calculus permit a clinician to merely glance at the graph and decide if he can compute the effect of one variable on another.

Overall, the advantage of the BN method is not that it will identify the true causes, but rather that it will perform initial data exploration to unearth new knowledge in a semi-automated and rapid fashion as well as give hints about causal mechanism implied in the studied phenomenon.

Acknowledgments

The authors thank Marco Scutari for sharing his *bnlearn* package in *R*. The experiments reported here were performed on computers funded by a French Institute for Complex Systems (IXXI) grant.

References

1. G.S. Collins and K. Michalsson. Fracture risk assessment: State of the art, methodologically unsound, or poorly reported? *Curr Osteoporos Rep*, 2012.
2. A. Aussem et al. Analysis of lifestyle and metabolic predictors of visceral obesity with bayesian networks. *BMC Bioinformatics*, 11:487, 2010.
3. J. Detilleux et al. A bayesian path analysis to estimate causal effects of bazedoxifene acetate on incidence of vertebral fractures, either directly or through non-linear changes in bone mass density. *Stat Methods Med Res*, 2002.
4. J. Flores et al. Incorporating expert knowledge when learning bayesian network structure: A medical case study. *Artif. Intell. Med.*, 53(3):181–204, 2011.
5. J.A. Kanis et al. Development and use of frax in osteoporosis. *Osteoporos*, 2010.
6. L. Lalonde et al. Bayesian networks: a new method for the modeling of bibliographic knowledge: Application to fall risk assessment in geriatric patients. *Med Biol Eng Comput.*, 2013.

7. M. Lappenschaar et al. Multilevel bayesian networks for the analysis of hierarchical health care data. *Artificial Intelligence in Medicine*, 57(3):171–183, 2013.
8. M. Velikova et al. Exploiting causal functional relationships in bayesian network modelling for personalised healthcare. *Int. J. Approx. Reasoning*, 55(1):59–73, 2014.
9. M.H. Edwards et al. Clinical risk factors, bone density and fall history in the prediction of incident fracture among men and women. *Bone*, 2013.
10. O. Strom et al. Osteoporosis: burden, health care provision and opportunities in the EU. *Archives of Osteoporosis*, pages 1–97, June 2011.
11. P. Dargent-Molina et al. Fall-related factors and risk of hip fracture: the epidos prospective study. *Lancet*, 348(9021):145–9, 1996.
12. P.A. Downey et al. Bone biology and the clinical implications for osteoporosis. *Phys Ther.*, 86(1):77–91, 2006.
13. S.E. Hardy et al. Improvement in usual gait speed predicts better survival in older adults. *J Am Geriatr Soc*, 2007.
14. N. Friedman and D. Koller. Being bayesian about network structure. a bayesian approach to structure discovery in bayesian networks. *Machine Learning*, 50(1-2):95–125, 2003.
15. M. Gasse, A. Aussem, and H. Elghazel. Comparison of hybrid algorithms for bayesian network structure learning. In *Proceeding ECML-PKDD'12*, Lecture Notes in Computer Science, pages 58–73. Springer-Verlag Berlin Heidelberg, 2012.
16. D. Heckerman, D. Geiger, and D.M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.
17. D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
18. F.I. Lewis and M.P. Ward. Revealing the complexity of health determinants in resource-poor settings. *American Journal of Epidemiology*, 2012.
19. F.I. Lewis and M.P. Ward. Improving epidemiologic data analyses through multivariate regression modelling. *Emerging Themes in Epidemiology*, 10(4), 2013.
20. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, USA, 1988.
21. J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, England, 2000.
22. J. Pearl. Invited commentary: understanding bias amplification. *American journal of epidemiology*, 174(11):1223–1227, October 2011.
23. J. Pena. Finding consensus bayesian network structures. *Journal of Artificial Intelligence Research*, 42:661–687, 2012.
24. S. Rodrigues de Moraes and A. Aussem. An efficient learning algorithm for local bayesian network structure discovery. In *Proceedings ECML-PKDD'10*, volume 6323 of *Lecture Notes in Computer Science*, pages 164–169. Springer-Verlag Berlin Heidelberg, 2010.
25. M. Scutari. Learning bayesian networks with the bnlearn R package. *Journal of Statistical Software*, 35(3):1–22, 2010.
26. E. Steele and A. Tucker. Consensus and Meta-analysis regulatory networks for combining multiple microarray gene expression datasets. *Journal of Biomedical Informatics*, 41(6):914–926, 2008.
27. I. Tsamardinos, L.E. Brown, and C.F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.