

## Handling Missing Entries in Monitoring a Woman's Monthly Cycle and Controlling Fertility

Anna Łupińska-Dubicka<sup>1</sup>

<sup>1</sup> Faculty of Computer Science, Bialystok University of Technology, Poland

**Abstract.** Even a small percentage of missing data can cause serious problems with analysis, reducing the statistical power of a study and leading to wrong conclusions being drawn. In the case of monitoring a woman's monthly cycle, missing entries can appear even in a woman experienced in fertility awareness methods. Due to the fact that in a system of controlling a woman's fertility, it is the most important to predict the day of ovulation and, ultimately, to determine the fertile window as much precisely as possible, much attention should be paid to the quality of the used data. This paper presents the results of handling missing observations as far as predicting the time during the cycle when a woman can become pregnant is concerned. Data taken from a multinational European study of daily fecundability was used to learn the quantitative part of the variety of a higher-order dynamic Bayesian network modeling a woman's monthly cycle. The main goal of this paper is to examine whether omitting observations has an influence on the model's reliability. The accuracy of comparison was examined based on two measures: the average percentage length of the infertile time during the monthly cycle and average percentage of days inside the fertile window classified as infertile.

### Introduction

Missing data (or missing values) are a common problem in statistical analysis and most practical databases contain missing values of some of their attributes. This can have a significant effect on the conclusions that can be drawn from the data. There are several reasons why data may be missing. Sometimes it results from malfunctioning equipment, sometimes the value of the attribute is not known, or the data was not entered correctly. However, regardless of the reason for the missing values, the fact that a measurement is missing is a complication for any algorithm that analyzes the data.

This paper presents the results of handling missing values in the problem of modeling a dynamic process. The vehicle for experiments was the problem of monitoring a woman's monthly cycle. In the case of monitoring

a woman's monthly cycle, the main goal is to predict the day of ovulation and, ultimately, to determine the fertile window. Days inside the fertile window that are classified as infertile are false negatives. If the model is used to avoid pregnancy, it is crucial to reduce the false negative rate to zero. Days that are marked as fertile and are outside the fertile window are false positives. The smaller the false positive rate, the closer the predicted day of ovulation is to the real day of ovulation, which can be helpful for couples seeking pregnancy. The data being used throughout this paper was drawn from a multinational European study of daily fecundability, which enrolled women from centers providing services on fertility awareness and natural family planning. According to calculations, there were 63,152 missing values (21.65% of all entries) in the data set. There was not a single record without missing values.

This experiment involved learning the conditional probability distribution of woman's monthly cycle models from the multinational European fecundability study data set, using different methods for dealing with missing data. A variety of higher-order dynamic Bayesian network models were used. The chosen approaches to handling missing data have an impact only on the quantitative part of the model. The qualitative part, i.e. the graphical structure, remained unchanged in each of the tested approaches. The accuracy of comparison was tested using the following measures: the average percentage length of the infertile time during the monthly cycle and the average percentage of false negatives (i.e. the days inside the fertile window that were classified as infertile).

The remainder of the paper is structured as follows: the following section presents the background of the analyzed problem, including basic information about a woman's monthly cycle with a short description of the applied data set, the issue of missing data and techniques applied to dealing with it, as well as a short introduction to Bayesian networks and their extension, i.e. dynamic Bayesian networks. This section also introduces the created dynamic Bayesian network models of a woman's monthly cycle. The succeeding section presents the results of an experimental comparison. Finally, the last section concludes the paper.

## **Materials and Methods**

The data of 881 participants of an Italian study of daily fecundability were analyzed. Between the years 1992 and 1996, 881 women from seven European centers (Milan, Verona, Lugano, Duesseldorf, Paris, Lon-

don and Brussels) recorded a total of over seven thousand monthly cycles. The women participating in the study satisfied the following five entry criteria: (1) experienced in use of Fertility Awareness Method (FAM), (2) married or in a stable relationship, (3) between 18th and 40th birthday at admission, (4) had at least one menses after cessation of breastfeeding or after delivery, (5) not taking hormonal medication or drugs affecting fertility. In addition, neither partner could be permanently infertile and both had to be free from any illness that could affect fertility. For each woman, several dynamic Bayesian network models were created. The intention was to simulate the use of the DBN model by women who want to become pregnant or want to avoid pregnancy. The conducted experiment involved learning the conditional probability distribution of DBNs, using different methods for dealing with missing data.

***Physiology of a Woman's Monthly Cycle.*** A woman's monthly cycle is driven by a highly complex interaction between hormones produced by three organs of the body: the hypothalamus, the pituitary gland, and the ovaries. There are four main hormones involved in the menstrual cycle process: estrogen, progesterone, follicle stimulating hormone (FSH), and luteinizing hormone (LH).

A woman's monthly cycle can be divided into four phases: (1) menstruation, (2) the follicular phase, (3) ovulation, and (4) the luteal phase (Figure 1). The length of each phase may vary from woman to woman and from cycle to cycle. Menstruation begins with the first day of bleeding. During the follicular phase (or the proliferative phase), the follicles in the ovary mature. The main hormone controlling this stage is estrogen. Just before ovulation, the level of estrogen is high enough to cause an increased release of luteinizing hormone and, as a result, the egg is released from the ovary. The luteal phase (or the secretory phase) is the latter phase of the menstrual cycle. The main hormone associated with this stage is progesterone, whose level is significantly higher during the luteal phase than during the other phases of the cycle.

In addition to measurable blood hormone levels, there are several easily accessible indicators of phases of the cycle: raised the basal body temperature (BBT) after ovulation, presence of cervical mucus, and changes in the position and consistency of the cervix. BBT is defined as the body temperature measured immediately after awakening and before any physical activity has been undertaken. To increase the reliability of this indicator, temperature should be measured every day at the same time. BBT follows a cyclical biphasic pattern, shifting near ovulation from the low to the high

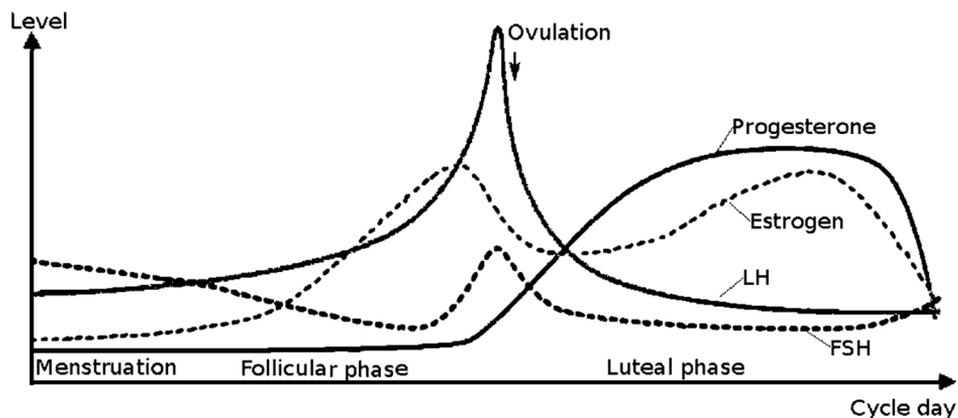


Figure 1. Levels of hormones during the phases of a woman's monthly cycle (Weschler, 2006)

phase. Metabolism is slower in the pre-ovulatory phase of the cycle, which results in a slightly lower body temperature. Following ovulation, as a result of an increased level of progesterone in the body, women typically experience an increase in the BBT of at least  $0.2^{\circ}\text{C}$ . BBT remains higher until menstruation occurs or, if a woman becomes pregnant, until the end of the pregnancy. Sometimes BBT can rise due to causes other than ovulation. This atypical rise is treated as disturbance and can be caused by a change in conditions around the time of measurement, such as later measurement time, lack of sleep, high stress levels, travel, or illness.

In the analyzed data set, the menstrual cycle was defined as the interval in days between the first day of menstrual bleeding in two neighboring cycles, where day 1 was the first day of fresh red bleeding, excluding any preceding days with spotting. The day of ovulation was identified in each cycle from records of basal body temperature and mucus symptoms. Daily mucus observations were classified into four classes; ranging from a score of 1 (no discharge and dry) to 4 (transparent, stretchy, slippery). The cervical mucus peak day was defined as the last day with best quality mucus, within a woman's specific cycle. If there were different mucus observations on a single day, the most fertile characteristic of the observed mucus determined the classification. To determine BBT shift, the three over six rule was used, i.e. the first time in the menstrual cycle when three consecutive temperatures were registered, all of which were above the average temperature of the last six preceding days.

In the analysis, only 3,432 (out of 7,017) cycles from 236 (out of 881) women were included. All women who recorded fewer than seven cycles were

excluded, because a woman needs at least six cycles to become familiar with the chosen fertility awareness method. In addition, cycles with no uniquely identified mucus peak or BBT shift days were excluded, as the proposed models used these values to determine the beginning of post-ovulatory infertility.

**Missing Data.** According to Little and Rubin (2002), one can distinguish three classes of a possible mechanism that account for missing data. Each of these mechanisms has unique characteristics both in terms of the reasons for the missing data, and the implications of the specific type of missingness. A brief description of each class of missing data is presented below.

The first class is Missing Completely at Random (MCAR). MCAR means that the missing data mechanism is unrelated to the values of any variables, whether missing or observed. For example, data that is missing because survey participants accidentally skipped questions are likely to be MCAR. If the observed values are essentially a random sample of the full data set, a complete case analysis gives the same results as the full data set would. Unfortunately, most missing data is not MCAR.

The opposite class is Non-Ignorable (NI). NI means that the missingness of data is not random and the missing data mechanism is related to the missing values. It commonly occurs when people do not want to reveal something very personal or unpopular about themselves. For example, if the research concerns mental health and people who have been diagnosed as depressed are less likely than others to report their mental status, the data is not missing at random. Clearly, the mean mental status score for the available data would not be an unbiased estimate of the mean that would have been obtained with complete data.

In between these two extremes there is a third class: Missing at Random (MAR). MAR requires that the cause of the missing data is unrelated to the missing values, but may be related to the observed values of other variables. For example, well-educated people are less likely to reveal their income than those with lower education.

A key distinction is whether the mechanism is ignorable (i.e., MCAR or MAR) or non-ignorable. Various approaches for handling ignorable missing data have been developed. Non-ignorable missing data is more challenging and requires a different approach. In general, methods dealing with ignorable missing data can be divided into the following categories (Little & Rubin, 2002): (a) procedures based on completely recorded units, (b) imputation-based procedures, and (c) model-based procedures.

Procedures based on completely recorded units simply omit cases with missing data and analyze the remaining ones. In those cases with a small fraction of records containing missing values, this method works well. However, when many records contain missing values, it becomes unreliable. Applying this method would result in an almost empty or even completely empty data set. In the case of MCAR data, such behavior affects the overstatement of statistical error. When used on non-MCAR data, these methods may affect unrepresentativeness of the data in relation to the whole data set. If the missing values belong to the NI class, removing rows may affect the pursuit of biased estimation.

Imputation-based procedures fill in the missing values and the resultant completed data is analyzed by standard methods. The commonly used procedures for imputation include hot deck imputation and mean/mode/median imputation. Mean imputation consists of replacing the missing data for a given feature (attribute) by the mean of all known values of that attribute in the class to which the instance with the missing attribute belongs. However, the mean can be affected by the presence of outliers and thus, to assure robustness, it seems natural to use the median instead. In this case, the missing data for a given feature is replaced by the median of all known values of that attribute in the class that the instance with the missing value belongs to. In the case of a missing value in a categorical feature, mode imputation can be used instead of either mean or median imputation. In hot deck imputation, recorded units in the sample are used to substitute values.

Model-based procedures define a model for the observed data and basing inferences on the likelihood or posterior distribution under that model, with parameters estimated by procedures such as maximum likelihood. These methods include, among others, iterative methods such as EM algorithm (Lauritzen, 1995) or Gibbs sampling (Yi & Li, 2011), and methods based on probability intervals, for example, the deterministic “bound and collapse” method (Ramoni & Sebastiani, 1999).

As was calculated, there were 63,152 missing values (21.65% of all entries) in the data set. Figure 2 presents the number of cases in the data set as a function of the number of missing values per woman. It can be seen that there was not a single record without missing values. The author’s hypothesis is that missing data in the multinational European fecundability study data set do not belong to non-ignorable class. They can be both missing completely at random (women simply forgot to chart observation on a particular day) or missing at random (women failed to note the observation due to some simplification, for example during bleeding the woman did not

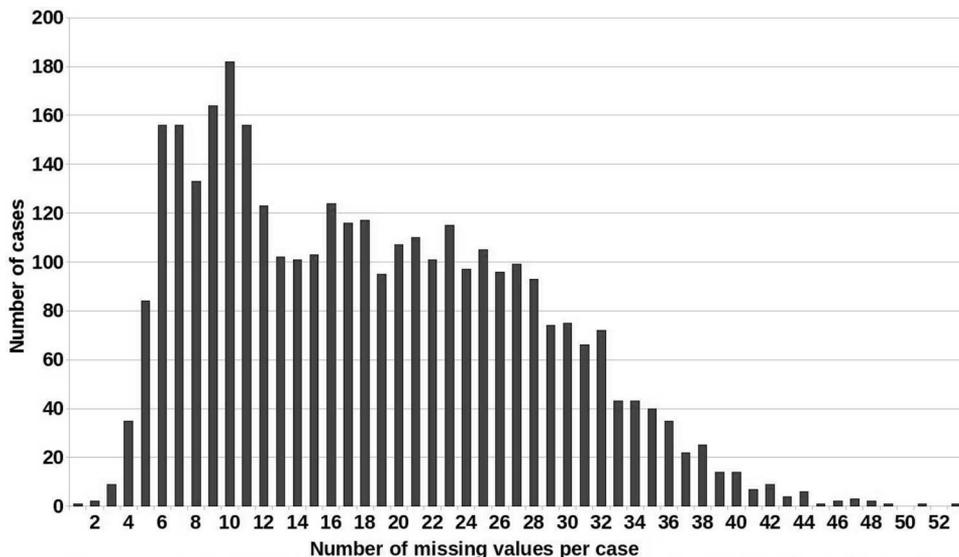


Figure 2. Number of cases as a function of the number of missing values per case in the multinational European fecundability study data set (Colombo & Masarotto, 2000)

chart her basal body temperature or, after the BBT shift, she did not take mucus observation). However, bearing in mind the type of study the data was obtained from (women experienced in the use of the Fertility Awareness Methods), it could be hypothesized that most of the missing data are missing at random cases.

**Dynamic Bayesian Networks.** Bayesian networks (BNs) (Cooper, 1990; Pearl, 1986), also called belief networks or causal networks, belong to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. Formally, a BN  $\mathcal{B}$  is a pair  $\langle G; \Theta \rangle$ , where  $G$  is an acyclic directed graph in which nodes represent random variables  $X_1, \dots, X_n$  and edges represent direct dependencies between pairs of variables.  $\Theta$  represents the set of parameters that describes the probability distribution for each node  $X_i$  in  $G$ , conditional on its parents in  $G$ , i.e.  $P(X_i|Pa(X_i))$ , where  $Pa(X_i)$  stands for the parents of node  $X_i$ . The structure of the graph is often given as a causal interpretation, convenient from the point of view of knowledge engineering and user interfaces.

BNs allow to compute probability distributions over subsets of their variables conditional on other subsets of observed variables (Neapolitan, 2003).

While Bayesian networks are a powerful tool for representing uncertainty, they do not provide a direct mechanism for representing temporal dependencies. Most events occurring in everyday life are not detected based on a particular point in time. They can be described through the multiple states of observations that all together lead up to a final event. The ability to model temporal aspects of the domain effectively plays a crucial role in modeling the World.

Dynamic Bayesian networks (DBNs) (Dean & Kanazawa, 1989; Mihaĳlovic & Petkovic, 2001) are a temporal extension of Bayesian networks for modeling dynamic systems. While a Bayesian network shows the cumulative probability distribution over a set of random variables independent of time, the dynamic Bayesian networks can be seen as a multidimensional representation of a random process. DBNs allow to interpret the present, reconstruct the past and predict the future. Phenomena are time bound, and the specific location in time is governed by the “earlier and later” relationship. Time is treated as a discrete variable, mostly due to the computational complexity of the inference algorithm. It should be noted that term *dynamic* means that the system’s development is modelled over time as opposed to the model’s structure and parameters changing over time, even though the latter is theoretically possible (Friedman et al., 1998; Robinson & Hartemink, 2008).

A DBN is a directed, acyclic graphical model of a stochastic process (Ghahramani, 1998). It consists of time-steps and each time-step contains its own variables. In the most common approach it is usually assumed that the network meets the Markov property, i.e. the values of future states of the process are determined only by its current state, regardless of the past. In other words, the future states of the process are conditionally independent of the past states. Such a network is called a first-order network. Usually, the DBN is defined as a pair of BNs ( $\mathcal{B}^1; \mathcal{B}^\rightarrow$ ), where  $\mathcal{B}^1$  represents a priori probability distribution  $P(Z^1)$  of the model. Typically,  $Z^t = (U^t; X^t; Y^t)$ , where  $U$  represents the input variables,  $X$  represents the hidden variables, and  $Y$  represents the output variables of the model.  $\mathcal{B}^\rightarrow$  is a two time-slice BN (2TBN), which defines transition distribution  $P(Z^t|Z^{t-1})$  as follows (Murphy, 2002):

$$P(Z^t|Z^{t-1}) = \prod_{i=1}^n P(Z_i^t|Pa(Z_i^t))$$

where  $Z_i^t$  is the  $i^{\text{th}}$  node at time  $t$ .  $Pa(Z_i^t)$  are the parents of  $Z_i^t$  from the

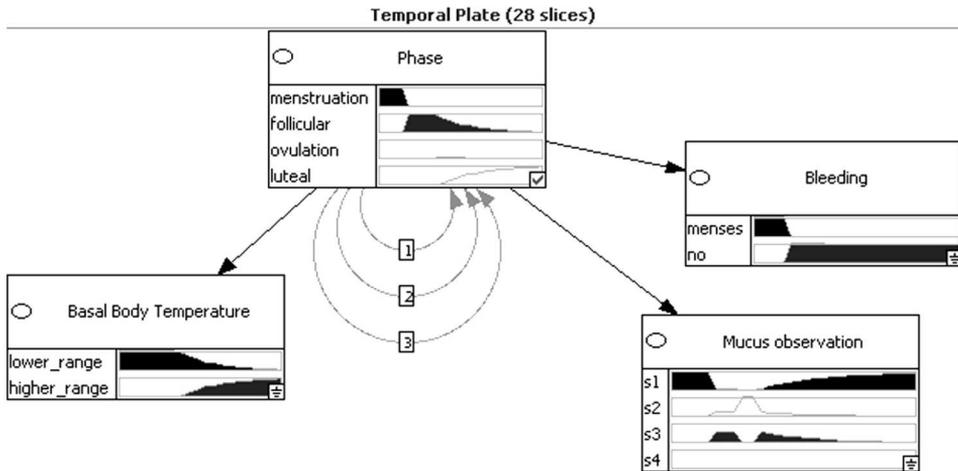
same or previous time-slice. The joint probability distribution for a sequence of length  $T$  can be obtained by *unrolling* the  $\mathcal{B}^\rightarrow$  network:

$$P(Z^{1:T}) = \prod_{t=1}^T \prod_{i=1}^n P(Z_i^t | Pa(Z_i^t))$$

It is worth mentioning that the most practical uses of DBNs involve temporal influences of the first order, i.e., influences between neighboring time steps. This choice is a convenient approximation influenced by the existence of efficient algorithms for first-order models and limitations of the available tools. Many real world systems, however, have memory that spans beyond their current state. Although DBNs are widely used nowadays, they focus generally on first- and second-order DBNs and there are only few application of higher-order models in real world systems (Daly et al., 2009; Łupińska-Dubicka & Drużdżel, 2011; McNaught & Zagorecki, 2009; Onisko et al., 2009; Perrin et al., 2003).

Furthermore, while any DBN model should contain at least one first-order influence (if that were not the case, some slices would be disconnected from the model), a model of order  $k$  does not need to include influences of all orders between 1 and  $k - 1$  (Łupińska-Dubicka & Drużdżel, 2015).

***Application of Dynamic Bayesian Networks for Monitoring Woman's Monthly Cycle.*** The computer models of a woman's monthly cycle based on a modeling technique known as dynamic Bayesian networks, described in details in (Łupińska-Dubicka & Drużdżel, 2011, Łupińska-Dubicka & Drużdżel, 2015), used in the experiments, combine information retrieved from BBT charting with observations of cervical mucus secretions. They contain a variable Phase with four states: menstruation, follicular, ovulation, and luteal, standing for the phases of the menstrual cycle. Three discrete observation variables were included in the data set: Basal Body Temperature, Bleeding, and Mucus observation, which are readily available to any woman. BBT has two possible values: lower range and higher range, representing the temperature before and after the BBT shift, respectively. Bleeding describes whether the woman had menses on a particular day or not. Mucus observation can have one of the four states ( $s_1$  through  $s_4$ ), described in detail in (Dunson et al., 2001). Time was modeled explicitly as  $n$  time steps, where  $n$  is the number of days of the longest monthly cycle of the particular woman. The model is  $k$ -order, i.e. it contains temporal influences between 1 and  $k$ . In the example model presented in Figure 3,  $k = 3$ , which means that the model contains temporal influences from 1 to 3.



**Figure 3.** An example of a third-order DBN model of a woman’s monthly cycle. Its temporal influences range from 1 to 3

For each woman, nine DBNs of temporal orders ranging from 1 to 9 were created. Additionally, a model with a structure that can change after each cycle was proposed for each woman, called “*SEL*” model. Its structure was changed by adding or removing temporal arcs, bearing in mind that the first-order arc is essential and cannot be removed. The procedure of changing the structure of a model can be described as follows: after each cycle the minimal and the most frequent day of ovulation in a woman’s last 12 cycles were calculated. By dividing these values by two, the order of temporal arcs to appear in the model was obtained. Moreover, those orders that were no longer necessary were removed from the model.

In order to obtain more meaningful results and to compensate for the absence of much data, the initial structure of a model and its parameters were determined based on domain knowledge and adjusted to a particular woman using data for her first six cycles. While a woman’s body can also change over time together with the characteristics of the cycle, the model’s parameters after each cycle were updated using the woman’s last 12 cycles at most.

**Experimental Comparison of the Approaches.** In the case of monitoring a woman’s monthly cycle, the main goal is to predict the right time of ovulation and based on it to determine the fertile window. If the main goal of a model is to avoid pregnancy, it is critical to reduce the number of days inside the fertile window that were incorrectly classified as infertile (false negatives). On the other hand, in the case when the model is used

to help couples who seek pregnancy, the number of days outside the fertile window and classified as fertile (false positives) should be minimized.

The number of fertile days during a menstrual cycle is difficult to specify, as it depends on the life span of the ovum and sperm, which varies from person to person and from cycle to cycle. It is generally believed that an ovum can be fertilized only within the first 24 hours after ovulation (Royston, 1982). Many authors (Kippley & Kippley, 1996; Potter, 1961; Rötzer, 1968; Szymański, 2004; Wilcox et al., 1995; World Health Organization, 1983) agree that the start of the fertile interval is strictly connected with changes in vaginal discharge and, in particular, estrogenic-type cervical mucus secretions. However, they differ in their estimates of the length of the fertile window. In this experiment, the definition of fertile window was based on Wilcox et al. (1995), who define it as the period between the day of ovulation minus five days and the day of ovulation plus one day. Apart from the number of false negatives, the percentage lengths of the pre-ovulatory and the post-ovulatory phases, and the percentage length of the fertile window were chosen as the comparison criterion. The numbers of fertile and infertile days in all cycles were determined and divided by the total length of the cycle, for each woman and for each cycle. Effectively, the percentage of all days that were classified as infertile and the percentage of all days that were classified as fertile were obtained. These two numbers (which add up to 100%) are a good indication of the precision of the method. If the method avoids false negatives perfectly, the larger the percentage of infertile days and the smaller the percentage of fertile days, the more precise the method is and the better approximation of the ovulation day it provides. At each time step (i.e. every day of the cycle), the proposed DBN models computed the most probable day of ovulation. If a time interval between the current day and the day with the highest probability of ovulation equaled at least six days, the current day was marked as infertile. In the other cases, the current day was the beginning of the fertile period. To find the beginning of the post-ovulatory phase, the model used the BBT shift. The third day after the BBT shift was considered as infertile.

The main idea behind the conducted experiment was to examine whether omitting observations has an influence on the reliability of the DBN model, by learning the conditional probability distribution, using different methods for dealing with missing data. In the case of each of the methods, the structure of the model was fixed and the tested approaches had impact only on the models' numerical parameters. In this comparison, two approaches, briefly introduced previously, were taken into account: *mode im-*

*putation* (due to the fact that all variables in the data set are discrete) and *hot-deck imputation*. The method based on completely recorded units could not be used due to the fact that discarding records in the case of the used data set would lead to an empty data set (see Figure 2). In both approaches, the missing values were estimated for each woman separately. In the case of *mode imputation*, the dominant values that appeared on a given day in all women's cycles were calculated. In the case of *hot-deck imputation*, the most similar case was defined as the case with the smallest distance between particular variables on particular days of the cycle. In the case when the closest record also had a missing value on a particular day, the next one in order of similarity was chosen. An additional method for completing missing data was also proposed, named "*normal*"-*value imputation*. In this case, values neighboring the missing one were taken into consideration and compared, and those that minimized the model's false negatives were chosen. In the case of *Basal body temperature*, the more important variable value was *lower range*, as this attribute value assumed that ovulation had not yet occurred. In the case of *Mucus*, the more important variable value was the more fertile type of mucus, for example value  $s_3$  is more important than value  $s_2$ . The values calculated by the model learned from the data set with missing values were also included as a baseline in terms of performance.

Figure 4 shows the result of comparison of the average percentage of infertile days during a monthly cycle for different models and different approaches to dealing with missing data. For all the implemented methods, the relationship between the order of the model and the number of infertile days during a cycle is similar: the higher order, the longer the infertile period. It can also be noticed that the models learned from the data set not containing missing values resulted in shorter numbers of infertile days. This is particularly apparent in case of models with lower orders, for example in case of the first-order model this difference equals roughly 10%. While models learned using *hot-deck imputation* seem to perform best, especially in the case of higher order models, the overall performance of the implemented methods is similar.

Figure 5 presents the result of a comparison between the average percentages of false negatives. It can again be observed that the overall performance of the tested approaches is similar. The models learned using "*normal*"-*value imputation* performed worse; however, for eight- and ninth-order models, they indicated a slightly smaller number of false negatives compared to other methods. Nevertheless, these variations are generally around a few hundredths of a percent.

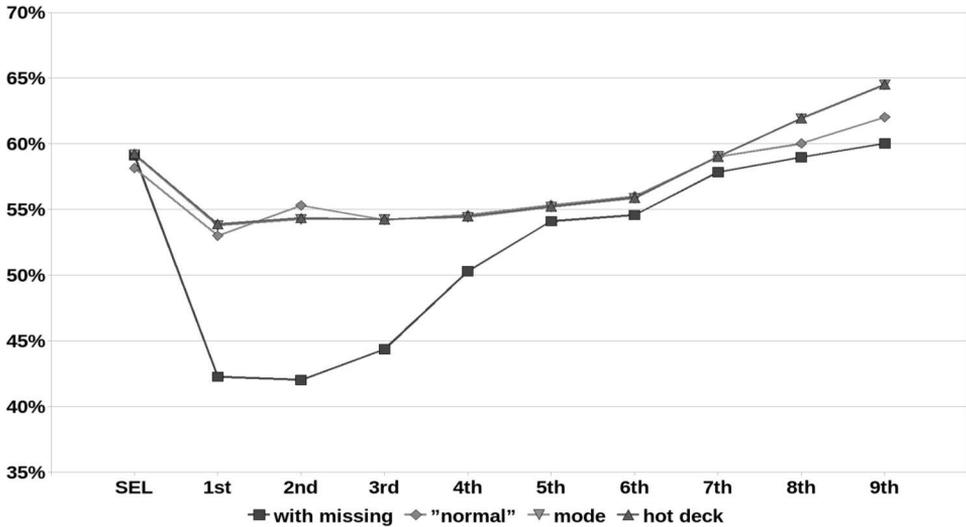


Figure 4. Average percentage of days classified as infertile during the monthly cycle for each of the compared DBN models for the selected approaches to handling missing data. The x-axis (horizontal) in the figure represents a particular model created for each woman (nine DBNs of temporal orders ranging from 1 to 9 and a model with a structure that can change after each cycle was proposed, called “SEL” model)

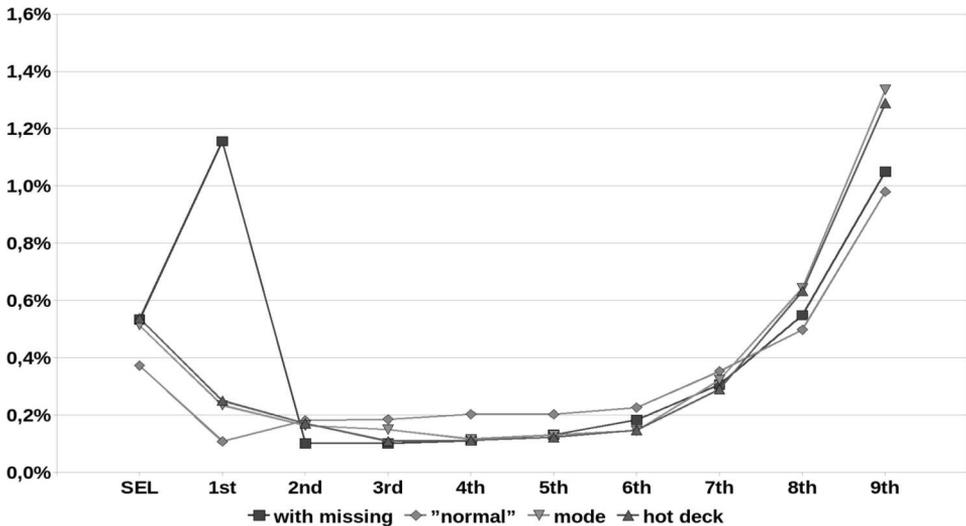


Figure 5. The average percentage of false negatives during the monthly cycle for each of the compared DBN models for selected approaches to handling missing data. The x-axis (horizontal) in the figure represents a particular model created for each woman (nine DBNs of temporal orders ranging from 1 to 9 and a model with a structure that can change after each cycle was proposed, called “SEL” model)

## Conclusions

In this paper, the diagnostic accuracy of different models for monitoring a woman's monthly cycle of several methods dealing with incomplete data was tested. In each case, all models had the same graphical structure. The results show that the choice of a method dealing with missing data has some influence on the predictive performance and error measured by means of false negatives; however, the accuracy for most of the tested methods was similar, with slightly better performances of *hot-deck* and *mode* imputation in the case of calculating the length of the fertile window and "normal" imputation in the case of the number of average percentages of false negatives, in comparison with other approaches. It is interesting to note that while there were some performance differences between the applied methods, they were minimal. One would expect that given the large percentage of missing data in the database, even a small improvement in dealing with the missing values should lead to a noticeable improvement. This, however, was not the case. A very probable explanation is the fact that these results confirm the hypothesis that missing data in the multinational European fecundability study data set belong to the missing at random class and can well be compensated for by the EM algorithm during the learning process of the model parameters and no additional method for filling the missing values in monthly cycle observations is required.

## REFERENCES

- Colombo, B., & Masarotto, G. (2000). Daily Fecundability: First Results from a New Data Base. *Demographic Research*, 3:5. doi: 10.4054/DemRes.2000.3.5
- Cooper, G. F. (1990). The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks. *Artificial Intelligence*, 42(2-3), 393-405.
- Daly, R., Edwards, K. D., O'Neill, J. S., Aitken, S., Millar, A. J., & Girolami, M. (2009). Using Higher-Order Dynamic Bayesian Networks to Model Periodic Data from the Circadian Clock of Arabidopsis Thaliana. In V. Kadiramanathan, G. Sanguinetti, M. Girolami, M. Niranjan & J. Noirel (Eds.), *Pattern Recognition in Bioinformatics* (pp. 67-78). PRIB 2009. Lecture Notes in Computer Science: Vol. 5780. Springer, Berlin, Heidelberg.
- Dean, T., & Kanazawa, K. (1989). A Model for Reasoning About Persistence and Causation. *Computational Intelligence*, 5(2), 142-150.
- Dunson, D. B., Sinai, I., & Colombo, B. (2001). The Relationship between Cervical Secretions and the Daily Probabilities of Pregnancy Effectiveness of the TwoDay Algorithm. *Human Reproduction*, 16(11), 2278-2282.

- Friedman, N., Murphy, K., & Russell, S. (1998). Learning the Structure of Dynamic Probabilistic Networks. In G. F. Cooper & S. Moral (Eds.), *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 139–147). Morgan Kaufmann Publishers.
- Ghahramani, Z. (1998). Learning Dynamic Bayesian Networks. In C. L. Giles & M. Gori (Eds.), *Adaptive Processing of Sequences and Data Structures* (pp. 168–197). Lecture Notes in Computer Science: Vol. 1387. Springer, Berlin, Heidelberg.
- Kippley, J., & Kippley, S. (1996). *The art of natural family planning* (4th edition). Cincinnati: The Couple to Couple League.
- Lauritzen, S. L. (1995). The EM Algorithm for Graphical Association Models with Missing Data. *Computational Statistics and Data Analysis*, 19(2), 191–201.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (second edition). Chichester: Wiley.
- Lupińska-Dubicka, A., & Druzdziel, M. J. (2011). Modeling dynamic systems with memory: What is the right time-order? In *The 8th Bayesian Modelling Applications Workshop* (pp. 75–82). Barcelona, Spain.
- Lupińska-Dubicka, A., & Druzdziel, M. J. (2015). Modeling Dynamic Processes with Memory by Higher Order Temporal Models. In A. Hommersom & P. Lucas (Eds.), *Foundations of Biomedical Knowledge Representation* (pp. 219–232). Lecture Notes in Artificial Intelligence: Vol. 9521. Springer International Publishing.
- McNaught, K. R., & Zagorecki, A. (2009). Using Dynamic Bayesian Networks for Prognostic Modelling to Inform Maintenance Decision Making. In *Proceedings of Industrial Engineering and Engineering Management, IEEE 2009* (pp. 1155–1159). Hong Kong, China.
- Mihajlovic, V., & Petkovic, M. (2001, October). *Dynamic Bayesian Networks: A State of the Art* (Technical Report No. TR-CTIT-01-34). Enschede: Centre for Telematics and Information Technology, University of Twente.
- Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning* (PhD Dissertation). University of California Berkeley, Computer Science Division.
- Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Prentice Hall.
- Onisko, A., Druzdziel, M. J., & Austin, R. M. (2009). Application of Dynamic Bayesian Networks to Cervical Cancer Screening. In *Proceedings on XI International Conference on Artificial Intelligence, AI-24'2009* (pp. 5–14). Siedlce, Poland: Publishing House of the University of Podlasie.
- World Health Organization (1983). A Prospective Multicentre Trial of the Ovulation Method of Natural Family Planning. III. Characteristics of the Menstrual Cycle and of the Fertile Phase. *Fertility and Sterility*, 40(6), 773–778.
- Pearl, J. (1986). Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, 29(3), 241–288.

- Perrin, B.-E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., & Buc, D. F. (2003). Gene Network Inference Using Dynamic Bayesian Networks. *Bioinformatics*, *19*(Suppl. 2), ii138–ii148.
- Potter, R. (1961). Length of the Fertile Period. *Milbank Quarterly*, *39*, 132–162.
- Ramoni, M., & Sebastiani, P. (1999). Learning Conditional Probabilities from Incomplete Data: An Experimental Comparison. In *Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics* (pp. 260–265). Morgan Kaufmann, San Mateo, CA.
- Robinson, J. W., & Hartemink, A. J. (2008). Non-stationary dynamic Bayesian networks. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21*, NIPS 2008 (pp. 1369–1376). Curran Associates, Inc.
- Rötzer, J. (1968). Supplemented Basal Body Temperature and Regulation of Conception. *Archives of Gynecology and Obstetrics*, *206*(2), 195–214.
- Royston, J. P. (1982). Basal Body Temperature, Ovulation and the Risk of Conception, with Special Reference to the Lifetimes of Sperm and Egg. *Biometrics*, *38*(2), 397–406.
- Szymański, Z. (2004). *Płodność i Planowanie Rodziny*. Wydawnictwo Pomorskiej Akademii Medycznej.
- Weschler, T. (2006). *Taking Charge of Your Fertility: The Definitive Guide to Natural Birth Control, Pregnancy Achievement, and Reproductive Health*. Collins.
- Wilcox, A. J., Weinberg, C. R., & Baird, D. D. (1995). Timing of Sexual Intercourse in Relation to Ovulation. Effects on the Probability of Conception, Survival of the Pregnancy, and Sex of the Baby. *New England Journal of Medicine*, *333*(23), 1517–1521.
- Yi, W., & Li, Z. (2011). Processing of Missing Values Using Gibbs Sampling. In *Proceedings of the 2011 Third International Conference on Measuring Technology and Mechatronics Automation*, Vol. 2 (pp. 927–930). Shangshai, China.