Ranking the effect of chronodisruption-based biomarkers in reproductive health

Ana G. Rúa[1] · Noelia Rico[2] · Ana Alonso[1] · Elena Díaz[1] · Irene Díaz[2]

Abstract

Chronodisruption alters circadian rhythms, which has negative consequences on different pathologies and mental disorders. This work studies whether factors related to chronodisruption of circadian rhythms motivated by shift works influence on reproductive health or not. In particular, this influence is studied on four particular aspects related to reproductive health: reproductive health disease, first pregnancy attempt, problems during pregnancy and gestation period. Some explainable machine learning models based on trees have been employed. These methods provided information about the importance of each predictor. The most important variables provided by each method were aggregated using a ranking aggregation function in order to reach a consensus ranking of variables that made possible to understand whether the chronodisruption factors had an effect on each of the aspects studied. The data have been obtained from 697 health professionals. Information about classical biomarkers, sleep quality indices and also other new variables related to eating jet lag, sleep hygiene and how the sleep is affected by shift works were considered as input data. Experiments have shown how some of these novel biomarkers are ranked in the top positions of the issues studied in relation to reproductive health. In particular, the light level and the use of electronic devices, which are features related to chronodisruption, are highlighted as biomarkers.

Ana G.Rúa and Noelia Rico have contributed equally to this work.

✉ Irene Díaz
sirene@uniovi.es

Ana G. Rúa
anagrua@hotmail.com

Noelia Rico
noeliarico@uniovi.es

Ana Alonso
alonsoana@uniovi.es

Elena Díaz
elenadr@uniovi.es

1 Department of Functional Biology, University of Oviedo, Julián Clavería, 33006 Oviedo, Asturias, Spain

2 Department of Computing, University of Oviedo, Jesús Arias de Velasco, 33005 Oviedo, Asturias, Spain

1 Introduction

Biological clocks are natural timing devices that generate biological oscillations, which affects the regulation of circadian rhythms. These oscillations are usually synchronized with the environment, but can be disturbed by external signals. If these signals are not in concordance with the internal clock, then the system experiences a misalignment called *desynchronization* [1]. Modern lifestyle habits such as shift work, night work, changes in eating schedules or poor sleep quality can cause desynchronization between the biological clock and physiological rhythms at different levels. This is a risk factor for the initiation and development of various pathological processes. In fact, the relationship between shift work and the incidence of metabolic disorders, pathologies related to reproduction and also with alterations in mental health has been demonstrated [2, 3].

The effects of rotating shifts on pregnancy complications are associated with an increased risk of preterm

delivery, gestational hypertension and even with lower weight of the offspring for their gestational age [2]. Nurses who work during the night report a greater number of irregular menstrual cycles than those who work during the day [4]. Social jet lag, i.e., discrepancy between biological and social rhythms, is a marker of chronodisruption associated with severe menstrual symptoms, suggesting the association between the circadian system and the reproductive function in females [5]. In this way, altered sleeping at night was associated with reduced fecundability. An inverse association was observed between shorter sleep duration and fecundability [6].

On the other hand, poor sleep quality in pregnant women during the second trimester of pregnancy is associated with stress, anxiety and depressive symptoms suggesting that sleep quality could be used as a modifiable factor to improve the psychological health among pregnant women, including the risk of adverse birth outcomes. Therefore, improving sleep could in turn improve psychological health and at the same time promote the overall well-being of pregnant women [7–9].

Machine learning is an increasingly used tool in health related to reproductive or mental health among others [10, 11]. For example, machine learning models can predict postpartum hemorrhage using data available at the time of admission for labor [12]. In addition, machine learning was also applied to predict preterm delivery [13], identifying as key predictor the light at night exposure, i.e., exposure to bright light at night due to the use of electronic devices such as computers, mobile phones, television, tablets or artificial light itself. Actually, feature selection is a long-standing problem in machine learning and other related fields (see, for example, [14]).

It is well known that there does not exist a standard machine learning method associated to an specific problem. For this reason, the common approach is to test several methods and select the one considered the best in some sense. However, due to many different factors, sometimes it is difficult to select the best method and in this case it is possible to combine the information provided by the different machine learning methods in some sense. In this context, it is possible to adapt and apply methods from the *social choice theory* field (see, for example, [15]) that provide a consensus conclusion from the output provided by each classifier.

Taking into account that the birth rate progressively declines in recent years [16] and the anxiety and psychological disorders that both attempted pregnancy and disturbances during pregnancy causes in future pregnant mothers [17], it seems interesting to study the usefulness of predictive models to determine the influence of chronodisruption on reproductive health and on disturbances during pregnancy.

Although there are some works where the effect of chronodisruption and eating jet lag over pregnancy is studied, to the best of our knowledge, there is no evidence about the use of machine learning as a tool to identify some chronodisruption markers as indicators of the reproductive health for shift worker women. In addition, the possible influence of sleep quality on the time of pregnancy, i.e., the time needed to conceive once the woman starts to try to become pregnant is still under study. Thus, the goal of this work is to identify possible biomarkers related to chronodisruption factors to properly study reproductive health using a machine learning approach.

The remaining of the paper is organized as follows. In Sect. 2, the data available for this studied are described as well as the variables to study. Sect. 3 describes the methods employed to identify biomarkers. Sect. 4 describes the experiments, and finally in Sect. 5, the discussion and some conclusions are drawn.

2 Description of the data

A descriptive and cross-study was performed with 697 women health professionals (nurses, nursing assistants, midwives, physiotherapists, occupational therapists, laboratory technicians, etc.), who work in different hospitals or in different primary care centers of Oviedo, Asturias, Spain. An overall of nine public hospitals was considered for this study. All of them offer integral patient care, three are reference hospital complexes with a number of beds between 434 for the smallest and 1,039 for the largest, and another six are regional hospitals with fewer beds (between 84 and 201). The study was approved by the Clinical Research Ethics Committee of the Principality of Asturias. All the participants have been informed about the objective of the study and have signed an informed consent. The variables obtained are described in the following subsections.

2.1 Classical markers

Although the main objective of this work is to study the impact of variables related to the sleep habits in reproductive health, some classical markers are also included to consider as much information as possible. These are related to basic information (age, body mass index or BMI, marital status). Also "smoker" was considered having in mind that tobacco smoking is one factor associated with complications in pregnancy. The possible values these variables can take are the following. Note that age is collected as a continuous variable and further categorized.

- Age (< 30, $[30 − 40)$, $[40 − 50)$, $[50 − 60)$ and ≥ 60).

- BMI (< 18.5- [18.5, 25), [25, 30) and > 30).
- Marital status (single, married, widower or divorced).
- Smoker (Yes/No).

In addition, as shift work has been previously proved to be an important variable in order to predict reproductive health [13], the following markers are also considered:

- Shift work during pregnancy (Yes/No).
- Current shift work (Yes/No).
- Work position (emergency, consultation, hospitalization, ICU, etc).

2.2 Sleep quality

The Pittsburgh Sleep Quality Index (PSQI) [18] is a self-rated questionnaire which assesses sleep quality and disturbances over a 1-month time interval. Seven different scores between 0 and 3 points are obtained from nineteen individual items, and these scores are: subjective sleep quality (*pQuality*), sleep latency (*pLatency*), sleep duration (*pDuration*), habitual sleep efficiency (*pEfficiency*), sleep disturbances (*pDisturbances*), use of sleeping medication (*pMedication*), and daytime dysfunction (*pDysfunction*). The score for any item indicates that there is no difficulty when is equal to 0, while a score of 3 indicates a severe difficulty in that item. The sum of these seven scores yields one global score (*pittsburg*) ranging from 0 (indicating the absence of difficulties) to 21 (indicating severe difficulties in all the areas studied).

2.3 Sleep hygiene

Having a good sleep hygiene means to have daily routines related to the sleeping time and also a comfortable bedroom environment that promote good quality of the sleep, (i.e., consistent and uninterrupted sleep) [19]. This can be achieved by having a bedroom free of disruptions and also building healthy relaxing habits before going to sleep. Some of these factors related to the sleep hygiene are measured through the following variables:

- Exposure to light at night (Yes/No).
- Level of light during sleeping (Integer number in the interval [1, 6]). Six possibilities are offered, ranging from level 1 (I sleep with a mask or I cannot see my hands), to level 6 (I can read comfortably).
- Use of electronic devices before sleeping (No/Less than 1 h/Between 1 and 2 h/Around 3 h/More than 4h).
- Average hours of sleeping during 1) workdays with morning shift (*hsm*) 2) workdays with evening shift (*hst*) 3) workdays with night shift (*hsn*) 4) workdays (*hsdt*) 5) free days (*hsdd*).The variable set to 0 when the

characteristics of the individual do not fit any of the established models.

- Social jet lag (*jls*), defined as the difference between midpoint of sleep in working days and free days.
- Social jet lag night shift (*jlsn*) defined as the difference between the midpoint of sleep in working days with night shift and free days. The value is set to 0 if the person does not work in night shift.
- Social jet lag daylight shift (*jlst*) that is defined as the difference between the midpoint of sleep in working days with daylight shift and free days. The value is set to 0 if the person does not work in daylight shift. Social Jet lag was calculated in all cases following the methodology previously described by other authors [19].
- The person sleeps alone or accompanied (*sleepWith*), which can take any of the values: No / Yes, in the same bed / Yes, in the same bedroom but not the same bed / Yes, in another bedroom).

2.4 Eating jet lag

The variables introduced in this section are referred to the delay on the eating times produced by the different work shifts. Eating Jet Lag is a marker of the regularity of daily routines.

- Eating midpoint in 1) morning shift (*pmtm*) 2) evening shift (*pmtt*) 3) night shift (*pmtn*) 4) free days (*pmd*). It is computed for the different work shifts as

$$\text{Eating midpoint} = \frac{tlm - tfm}{2} + tfm,$$

where *tlm* is the timing of the last meal and *tfm* the timing of the first meal [20].
- Eating jet lag when working in 1) morning shift (*jlatm*) 2) evening shift (*jlatt*) 3) night shift (*jlatn*). It is computed as the difference (in absolute value) between eating midpoint on free days and eating midpoint on work days (for the different shift works).
- Delay in breakfast time when working in 1) morning shift (*rhdtm*) 2) evening shift (*rhdtt*) and 3) night shift (*rhdtn*).
- Delay in lunch time when working in 1) morning shift (*rhctm*) 2) evening shift (*rhctt*) and 3) night shift (*thctn*)
- Delay in dinner time when working in 1) morning shift (*rhcetm*) 2) evening shift (*rhcett*) and 3) night shift (*rncetn*) The variability or delay in the timing of breakfast, lunch or dinner depending on the work shift was calculated [19] as the difference in hours (in absolute value) between the time of breakfast (resp. lunch, dinner) on free days and the time of breakfast (resp. lunch, dinner) on work days.

2.5 Measuring reproductive health

The hypothesis of this work is that some of the variables introduced in sects. 2.1–2.4 could influence over different issues related to reproductive health. In particular, four different aspects related to the reproductive health are studied. These aspects will be studied separately in different problems. The variables of interest are:

- **Reproductive health disease**, which is *Yes* if the woman has presented any disease related to reproductive health and *No* otherwise.The main diseases were polycystic ovary, endometriosis, myoma, uterine malformations, tubal obstruction and early ovarian failure.
- **Attempt time**, referring to the time that the woman needs to achieve successful pregnancy for the first child.
- **Problems during pregnancy**, which takes the value *No* if the woman did not have any problem during her pregnancy and *Yes* if she had any problem, such as preeclampsia (gestational hypertension with albuminuria), hypertension (gestational hypertension), gestational diabetes or any other.
- **Gestation period**, which is *short* if the length of the pregnancy is under 37 weeks and *normal* otherwise.

3 Methods

As previously stated, the aim of this study is to identify the biomarkers (from the ones detailed in sects. 2.1–2.4) that influence on each of the four variables presented in Sect. 2.5. For each one of the problems, these biomarkers are identified using a procedure that can be divided in three different steps:

- STEP 1. Training different classification models.
- STEP 2. Get the most important variables associated to each classification model.
- STEP 3. Determine the importance of each biomarker.

The above introduced procedure is applied to solve four different problems, considering each time a dataset formed by all the possible predictor variables and one of the four interest variables (Reproductive health disease, Attempt time, Problems during pregnancy, Gestation period) as the output that must be predicted. Note that the training set depends on the variable under study (due to missing values on the variable under study). Appendix A shows the details related to each dataset.

3.1 Machine learning methods

Classification models aim at predicting the value of a variable class [21] based on previously observed data. Therefore, a key purpose of any learning algorithm is to build models with good generalization capability [22]. Among the many different families of classification algorithms, decision trees are very popular in biological contexts due to their high interpretability. This kind of algorithms obtain tree representations, where each node tests the input variables until the leaf nodes are reached, which are labelled with the classes to classify the objects.

Every algorithm for building classification trees must define, at least, the splitting criteria, the stopping conditions and the pruning method (if any). This allows to define different algorithms based on trees with the same underlying idea. Among them, CART trees are binary decision trees that use the Gini index to select the nodes at each point of the model [23]. Taking also as main step this tree representation, the *random forest* algorithm [24] is a classifier consisting of an ensemble of decision trees, where each tree is constructed by applying a given number of times an algorithm on a random sub-sample of examples and variables of the training set. The prediction of the random forest is obtained by a majority vote over the predictions of the individual trees.

Due also to its high level of interpretability, models based on trees are often used for feature selection in machine learning as the above defined strategy naturally ranks by how well each feature improves the purity of the node. Nodes with the greatest decrease in impurity happen at the start of the trees, while nodes with the least decrease in impurity occur at the end of trees [25], being therefore the ones that happen at the beginning more important in order to determine the output.

Implementations of these algorithms can be easily found in any modern programming language. In this work, the R programming language has been used, more concretely the methods `rpart` and `rpart2` that implement CART trees and also the `randomForest` method. The difference between the two tree-based algorithms selected consist in the constraints of the model that each algorithm enforces. More concretely, the main difference between these two implementations of the CART trees are the parameters used and how they affect in the training process for

building the optimal tree. On the one hand `rpart` uses the complexity parameter, which is the value of the minimum improvement in the model needed at each node for creating a new split for pruning the tree and avoided overfitting. On the other hand, `rpart2` uses the max tree depth that the tree is allowed to grow. The implementantion of `randomForest` has also parameters regarding the number of trees and variables randomly sampled.

3.1.1 Validation process and resampling techniques

For all the problems, the data sets defined have been divided in training set and testing set in a proportion 80%-20%, respectively, using the training set for parameter setting of the algorithms and building the models, and the testing set for the validation of the models in unseen data. Cross-validation is a common model validation technique to train machine learning models on a limited data sample. The dataset is divided into k groups. Then k models are trained leaving one of the folds for evaluation. The performance of the model is the average of the results obtained in the testing folds associated with each model. On the other hand, the bootstrap method [26] estimates quantities about a population by averaging estimates from multiple small data samples. Then, samples are constructed by drawing observations from a large data sample one at a time and returning them to the data sample after they have been chosen. Models using these two different validation techniques have been trained for the studied problems in this work.

In addition, as the input data associated to each of the binary problems studied are quite unbalanced, four sampling techniques are tested, seeking for improving the results of the obtained models. Basic techniques *upsampling* (resp. *downsampling*) increases (resp. decreases) the number of samples of the minority class to fight unbalancing. More advanced techniques have been also tested: ROSE [27] is a bootstrap-based technique that generates synthetic examples from a conditional density estimate of the two classes; SMOTE [28] is another technique that over-samples the minority class and under-samples the majority class. More concretely, SMOTE first selects a minority class instance at random and finds its k nearest minority class neighbors. Then, a new synthetic instance is created by choosing randomly one of the k nearest neighbors and generating a convex combination in a the hyper plane that connects the object with this neighbor.

3.1.2 Evaluation of the models

In order to evaluate the performance of the models, metrics computed using the *confusion matrix* have been used [24]. Considering the binary target variable with a positive and a negative class, *TP* (resp. *TN*) represents the number of positive (resp. negative) examples correctly classified, *FP* is the number of negative examples classified as positive, and *FN* is the number of positive examples classified as negative [24]. Using this definition of confusion matrix, the *Sensitivity* and *Specificity* are computed as follows:

$$Sensitivity = \frac{TP}{TP + FN}, \quad Specificity = \frac{TN}{TN + FP}.$$

For the problems defined in this work (see Sect. 2.5), the minority class has always been considered the positive class as it is the object of interest. Thus, the positive class for the *reproductive health disease* problem is *Yes* (i.e., having a problem). For the *attempt time* problem, *more than a year* is the positive class, which is less common and therefore is interesting to study the factors that influence this value. *Having a problem during pregnancy* is the event that the model must detect so is the positive class. Lastly, *having a gestation period lower than expected* is also the positive class, as is interesting to understand why this happens.

3.2 Consensus ranking

Methods that aim at drawing conclusions from the preferences expressed by several voters over a set of candidates are deeply studied in the field of *social choice theory* (see for example [15]). When the voter preferences are expressed as a ranking, these can be aggregated using ranking rules in order to obtain a winning candidate. Many of these ranking rules allow to obtain also the complete winning ranking.

A set of rankings over the voters is known as profile of rankings. The Kemeny method [29] is a distance-based method that selects as winner of the election the candidate in the first position of the ranking that is the closest to the profile of rankings. To do so, it considers the distance from all the possible rankings that can be created by permuting the set of candidates of the profile and the permutation that is closest to the profile is elected as winner. This distance measure between two rankings is based on the pairwise comparison of the candidates such that:

- One point is annotated every time that two candidates are in different order in the two rankings.
- Half of a point is annotated every time that two candidates are tied in one of the rankings and ordered in the other.

To compute the final distance each permutation of the set of candidates is compared to all the rankings of the profile, updating accordingly to the previous procedure after each comparison. This method ensures that if a candidate wins all the other candidates by majority in the pairwise

comparison, then this candidate is always chosen as winner [30].

Notice however that, by the definition of this distance, there could be more than one ranking with the same distance to the profile, and all of them would be the ranking winner of the election. If all those rankings have the same winning candidate, then the winning candidate can be chosen anyway as candidate winner, but how to choose the winner otherwise remains unclear and it is a field of current study [31]. Moreover, this method has an additional drawback to be applied in practice due to its complexity, as determining the Kemeny ranking is an NP-hard problem. However, recently published optimization techniques allows to apply this method to a greater number of candidates [32], which is enough to apply to the data considered in this work.

Besides this ranking rule, a easier method to determine the winner of the election is the one proposed by Borda,

each candidate as many points as the number of candidates that are ranked in a worse position than it in a certain ranking. Summing the individual points that the candidate has in each of the rankings of the profile of rankings the final score is obtained. The candidates are then ranked in decreasing order according to their score, making the candidate with the most points the winner. Kemeny and Borda Count methods are combined in this work to take advantage of both methods.

3.3 Biomarker identification

The methodology proposed for the identification of biomarkers considers all the methods previously introduced in this section. It is detailed in Algorithm 1.

Algorithm 1 Methodology used for biomarker identification

---

for each algorithm i in {`rpart`, `rpart2`, `randomForest`} do
    for each metric evaluation j in {`Sensitivity`, `BalancedAccuracy`} do
        for each resampling method k in {normal, `up`, `down`, `SMOTE`, `ROSE`} do
            for each validation method ℓ in {cross validation, bootstraping} do
                Train and evaluate models on the subsamples of the training set
                using `caret::train()` in R to determine the best parameter setting;
            end
            Choose the best parameter configuration of the model trained
             with algorithm i resampling method k and validation method ℓ
             according to j;
            Build a model using the best parameter configuration using i
             in all the training set;
            Compute the results of the model on the testing set;
        end
    end
    Choose the best model obtained for the data using the algorithm i according
    to their results on the testing set;
end
At this point, three different models are selected (i.e. the best model
obtained using each algorithm i);
Create an empty profile of rankings;
for each model do
    Determine the ranking of most important variables;
    Add this ranking to the profile of rankings;
end
Apply Kemeny to the profile of rankings to determine the most important variable;
if Kemeny gives only one solution
    Establish the winning ranking as the ranking of most important biomarkers;
else *(Kemeny gives more than one solution)*
    Apply Borda Count to establish the winning ranking, allowing tied variables;
end

---

nowadays known as *Borda Count* [33]. This method tries to elect broadly acceptable candidates, rather than those preferred by a majority. [34]. Borda Count method gives to

First of all, different classification models are trained using trees and random forest algorithms, getting the best

configuration for each one according to the parameter optimization.

Unfortunately, how to define which are these *best* models is not straightforward. Actually this task is even more complicated when working with unbalanced datasets, as in this case (see Fig. 1). Therefore during the training process both Sensitivity and Balanced Accuracy are considered. It is considered that the best model in testing is the one that maximizes the Sensitivity but always ensures a minimum threshold on the Specificity.

Once the models are selected, the next step is to determine the most important variables using these models. CART trees measure variable importance by the reduction in the sum of the loss function associated to each variable in each node is sum. Random forest obtains the importance of each variable averaging the minimal depth values for all trees in the random forest. Variable ranking importance has been computed using the function `varImp` of the `caret` package in R.

The rankings of most important variables according to each model are added as part of a profile of rankings. These rankings are then aggregated using Kemeny ranking rule. If Kemeny produces more than one ranking, Borda count is applied to obtain the final consensus order of most important biomarkers.

4 Experiments and results

In this section, the biomarkers identified for each of the individual problems studied in this work are presented. All the experiments have been performed in R 4.0.3 using RStudio v1.4.1103 with a computer with 3.4 GHz Intel Core i5 with 4 kernels and 8 Gb RAM.

Following the methodology proposed in Algorithm 1, for each problem, the best model obtained with each algorithm is kept and the rankings of importance of the variables are obtained. Although some of the models can give a ranking up to twenty most important variables, in this work only the six most important variables are considered for the ranking aggregation. This threshold is established in order to obtain an accurate set of important variables.

4.1 First pregnancy attempt

In order to illustrate the methodology followed to obtain the results presented in this work, this problem is set as example and it is more detailed. Therefore, exceptionally for this problem the number of important variables considered has been increased to eight instead of six, so more aspects of the methodology can be clarified.

Table 1 presents the three different rankings obtained with each of the selected models. Although many of the variables appear in more than one of the rankings, notice how not all are common, and thus the need of establishing a consensus ranking for determining the variable that influences the output the most. Consider the ranking obtained with the algorithm `rpart`, which is:

$$pmd \succ workType \succ pEfficiency \succ age \succ pittsburgh \succ hsm$$

This ranking does not contain all the variables that are present in the three rankings. However, in order to aggregate the rankings, the Kemeny method must evaluate all the possible permutations of the set of candidates. Therefore, all rankings must consider the union of all the variables that appear in the three rankings. Thus, it is necessary to modify this ranking adding the variables that were not in the original ranking as tied in the last position. That means that according to that method all the added variables are equal and worse than the ones originally given in the ranking:

$$pmd \sim workType \sim pEfficiency \sim age \sim pittsburgh \sim hsm \sim smoker \sim$$
$$\sim hsdt \sim pDuration \sim hst \sim jlatm \sim lightLevel \sim pmtt \sim$$
$$\sim rhdtm \sim pSleepiness \sim pregnantShift$$

In the same way, the ranking given by `rpart2` adding these variables is:

$$pmd \succ pittsburgh \succ rhdtm \succ hsm \succ pmtt \succ pEfficiency \succ hst \succ jlatm \succ$$
$$\succ pDuration \sim age \sim smoker \sim hsdt \sim lightLevel \sim$$
$$\sim pSleepiness \sim pregnantShift \sim workType$$

And the ranking given by `randomForest` is now:

$$pSleepiness \succ hsdt \succ pDuration \succ pregnantShift \succ jlatm \succ$$
$$\succ workType \succ lightLevel \succ pmd \succ age \sim pEfficiency \sim smoker \sim hsm \sim$$
$$\sim hst \sim pittsburgh \sim pmtt \sim rhdtm$$

These three rankings form the profile of rankings. After applying the Kemeny method to this profile of rankings, more than one solution is obtained. However, the winner is always the variable *pmd*, showing that this biomarker is the most important according to the aggregated opinion of these three models. Now, in order to determine a single ranking as consensus ranking, the Borda Count ranking rule is applied. The points obtained by each variable are:

- age: 19
- smoker: 16
- hsdt: 25.5
- hsm: 25.5
- hst: 16
- jlatm: 22.5
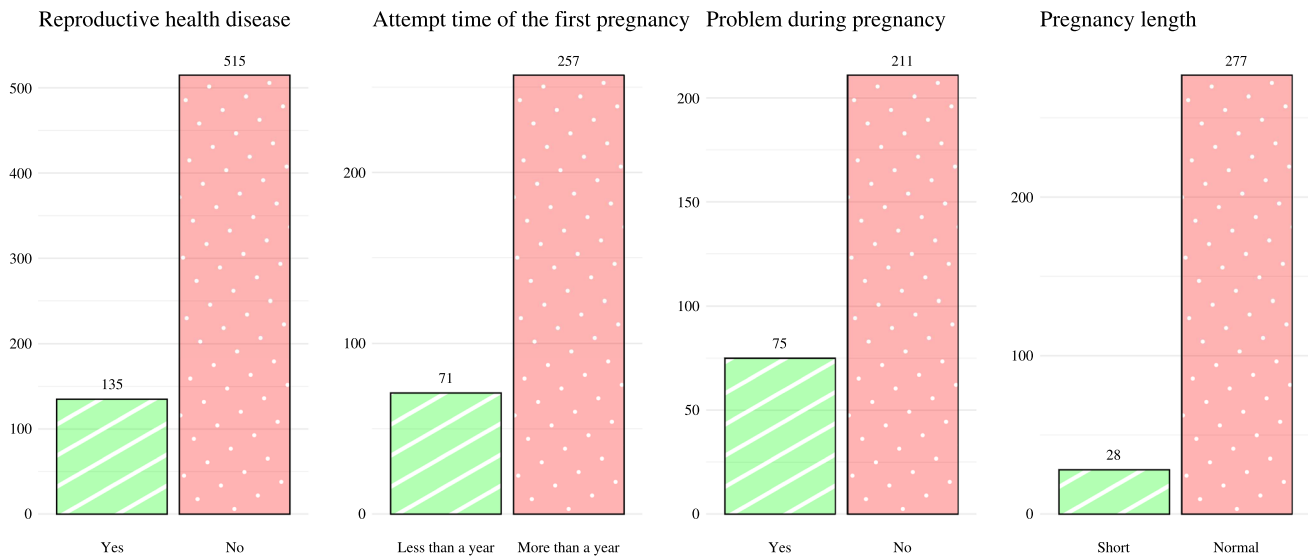- lightLevel: 16
- pDuration: 20

Fig. 1 Distribution of the interest variables that are the main class of study in each of the four problems

- pEfficiency: 26.5
- pittsburgh : 28.5
- pmd: 38
- pmtt: 18
- pSleepiness: 22
- pregnantShift: 19
- rhdtm: 20
- workType: 27.5

After sorting all the variables decreasingly using their score, the consensus ranking showing the importance of the

biomarkers according to the criterion of the three methods is:

$$pmd \succ pittsburgh \succ workType \succ pEfficiency \succ hsdt \sim hsm \succ jlatm \succ$$
$$\succ pSleepiness \succ pDuration \sim rhdtm \succ age \sim pregnantShift \succ$$
$$\succ pmtt \succ smoker \sim hst \sim lightLevel$$

Notice how the middle point in the daylight (*pmd*), a chronodisruption biomarker, is the most important variable according to the models. Also variables related to the pittsburgh test as pEfficiency, pSleepiness and pDuration appear in this ranking. Furthermore, many of the introduced variables related to the amount of time of sleep in work days and more concretely with the turn in which the person work appear in this ranking (*hsdt*, *hsm*, *hst*). From the classical biomarkers only *age* and *smoker* appear, which indicates that in presence of the proposed biomarkers other classical measures lose importance.

4.2 Gestation period

The results obtained for predicting gestation period are shown in Table 2.

The Kemeny method gives also for this problem more than one solution. However, in this case, is not possible to determine the winning variable only using the Kemeny ranking as it returns more than one solution and in some of them *pmtm* is in the first position but also *pQuality* is in some others.

The ranking of biomarkers obtained after applying the Borda Count method is shown below:

Table 1 Results of the models (right) and ranking of most important variables according to each of them (left) for the *attemp time of the first child* problem

	rpart	rpart2	randomForest
1st	pmd	pmd	pSleepiness
2nd	workType	pittsburgh	hsdt
3rd	pEfficiency	rhdtm	pDuration
4th	age	hsm	pregnantShift
5th	pittsburgh	pmtt	jlatm
6th	hsm	pEfficiency	workType
7th	smoker	hst	lightLevel
8th	hsdt	jlatm	pmd

	rpart	rpart2	randomForest
train Sens	0.49	0.50	0.77
train Spec	0.88	0.53	0.55
test Sens	0.57	0.64	0.57
test Spec	0.80	0.59	0.59

$$pmtm \succ jls \succ pQuality \succ hsdd \succ device \succ$$
$$\succ sleepWith \sim pittsburgh \sim rhdtn \succ hsm \succ rhctn$$

Notice how social jet lag (*jls*) is a variable that appears in all the rankings and consequently is ranked in the second position of the Borda ranking. The importance of this variable reassures the idea that chronodisruption has a negative impact on reproductive issues. Again, the hours of sleep are important. Moreover, the use of devices is also listed in this final ranking of important biomarkers, remarking what has been shown in the literature before in related problems with the pregnancy attempt time [13].

4.3 Problems during pregnancy

When reproductive health is studied in terms of the possible problems during pregnancy, the results obtained are shown in Table 3.

The results after applying Borda Count are:

$$jlatm \succ pLatency \succ hsdt \sim hsm \succ sleepWith \sim hsdd \succ pQuality \sim$$
$$\sim jls \sim rhcett \succ pmd \succ pmtm \succ pSleepiness \succ rhctn \sim rhdtm$$

Again, all the biomarkers related to the hours of sleep are present in high positions of the rankings, as happened with he previous problems.

4.4 Reproductive health disease

Reproductive health is studied in terms of reproductive health disease issues, the results are those shown in Table 4.

In this case, the age is the first variable in all the rankings, so it is clear that the Kemeny method returns solutions always with this variable in the first position. To obtain one single solution, Borda Count produces the following ranking:

$$age \succ pmtm \succ lightLevel \succ jltn \succ jls \succ pDuration \sim$$
$$\sim hsm \sim pittsburgh \succ device \sim hsdt \sim pmtn \succ hsn \sim pPerturbation$$

Again the light level and the use of electronic devices as well as other of the introduced variables related to chronodysruption are present as most important variables in the consensus ranking.

5 Discussion

In this work, a methodology to identify biomarkers based on consensus ranking and machine learning is presented. A ranking of relevant variables is obtained from the the different classifiers employed. These rankings are then aggregated to provide a more robust set of important variables. In particular, this methodology has been shown as an effective tool to seek for the variables that influence chronodisruption on reproductive health. In addition, our results corroborate and expand those previously obtained [13], which showed that machine learning is a promising tool to predict the influence of chronodisruption in preterm births and identified the use of electronic devices at night and consequently the exposure to high levels of lightness at night as a risk factor for preterm labor.

With regard to the biomarkers, this study shows that the chronodisruption produced by shift works in humans affects their daily routines and have a great impact on different aspects related to the reproductive health. Factors associated with *eating jet lag* or *sleep hygiene* are specially relevant for first pregnancy attempt time, pregnancy disturbances or duration, that are three of the four issues related to reproductive health studied in this work. The results indicate that the main biomarker is *eating jet lag* and its components since, respectively, the most important factor in each case is *pmd*, *jlatm* and *pmtm*. Thus, night shift work previously related to various pathologies [35] is, according to our model, the main factor to take into account for pregnancy disturbances, being also an important factor for the time until reaching the first pregnancy, although in this case behind other indicator factors of chronodisruption such as sleep timing and quality.

Note that *eating delay* caused by the different work shifts reinforce the idea that shift work is a risk factor for pregnancy, since three of the four aspects studied in relation to reproductive health are affected by this factor. In case of duration of the pregnancy, *eating delays* are those produced during the night shift, while for the timing of

Table 2 Results of the models (right) and ranking of most important variables according to each of them (left) for the *length of the pregnancy*

	rpart	rpart2	randomForest
1st	device	pQuality	pmtm
2nd	sleepWith	pittsburgh	rhdtn
3rd	jls	jls	pQuality
4th	pmtm	pmtm	hsdd
5th	hsm	hsdd	rhctn
6th	hsdd	hsm	jls

	rpart	rpart2	randomForest
Train Sens	0.56	0.74	0.71
Train Spec	0.77	0.45	0.57
Test Sens	0.60	1.00	0.60
Test Spec	0.86	0.58	0.77

Table 3 Results of the models (right) and ranking of most important variables according to each of them (left) for the *problem during pregnancy*

	rpart	rpart2	randomForest
1st	sleepWith	hsdd	jlatm
2nd	pQuality	rhcett	jls
3rd	pLatency	hsm	pmd
4th	jlatm	hsdt	pmtm
5th	pSleepiness	pLatency	hsdt
6th	hsm	rhctn	rhdtm

	rpart	rpart2	randomForest
Train Sens	0.37	0.69	0.78
Train Spec	0.66	0.37	0.58
Test Sens	0.60	0.53	0.67
Test Spec	0.62	0.52	0.55

pregnant or pregnancy alterations like diabetes, preeclampsia or hypertension, the most important *eating delays* are those produced during the morning shift. To our knowledge, this is the first evidence that eating jet lag is a risk factor for the evolution and term of pregnancy.

On the other hand, other authors have found that *social jet lag* and *sleep quality* were predictors for stress symptoms among pregnant women [36]. In this way, elevated levels of depression and anxiety were associated with

Table 4 Results of the models (right) and ranking of most important variables according to each of them (left) for the *reproductive health disease*

	rpart	rpart2	randomForest
1st	age	age	age
2nd	jltn	lightLevel	lightLevel
3rd	jls	pmtm	pmtm
4th	pDuration	hsm	pittsburgh
5th	hsdt	device	pmtn
6th	pmtm	pPerturbation	hsn

	rpart	rpart2	randomForest
Train Sens	0.64	0.65	0.79
Train Spec	0.61	0.65	0.89
Test Sens	0.52	0.52	0.44
Test Spec	0.60	0.64	0.79

obstetric complications, preterm labor and pain relief under labor, also has implications for fetal and neonatal well-being and behavior [37]. Based on our results and the well-known relationship between mental health and the development of pregnancy, we infer that a simple intervention on sleeping and eating habits, trying to take regular breakfast, lunch and dinner times every day of the week, as well as improving our quality of sleep, it is possible to intervene to improve both mental health during pregnancy and the development of the pregnancy itself.

Finally, we have found that for reproductive health disease the first influencing factor is age. Besides, several chronodisruption indicators are also predictive factors (*jls*, quality of sleep or sleep hours). However, in this case the variables related to delays in feeding do not seem to be predictors of reproductive health disease. These results are in concordance with the risks related to night shift work previously found in relation to some diseases such as endometriosis [38] or breast and ovarian cancer [39, 40]. Obviously, this is associated with reduced emotional health [41].

To sum up, we have proposed a methodology that allows to identify factors that affect reproductive health. This methodology combines the results of several machine learning models into just one using consensus rules to avoid biases. This fact is important as the dataset is not too large. In addition, we have identified some new factors related to chronodisruption for the first time. These are the main strengths of our work. As future work, we plan to overcome some of the weaknesses of the proposal. In particular, we will try to gather a larger sample and then to improve the performance of the baseline models.

Appendix A: Description of the datasets

This appendix shows a graphical description of the dataset used to solve each problem.

A.1 Reproductive health disease

The dataset of the *reproductive health disease* contains the 650 samples of the original dataset that have no missing values. Among these, 135 have a problem and therefore they take the value of the positive class *Yes* (striped green in the figures) and the remaining 515 take the value *No* (dotted red in the figures) (Figs. 2, 3, 4, 5).
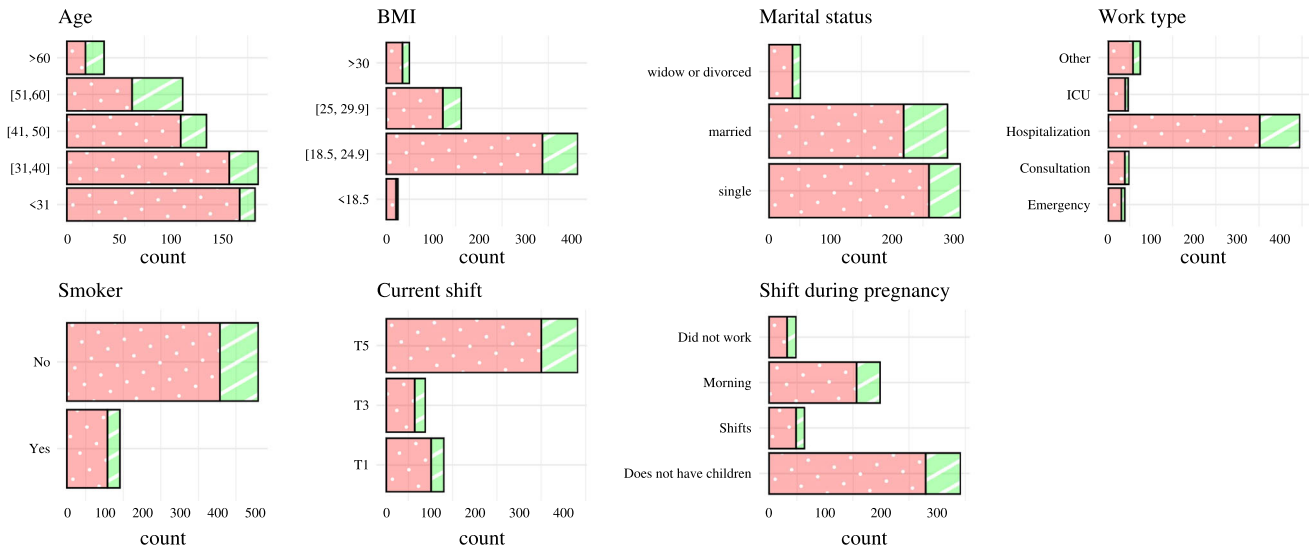
Fig. 2 Distribution of the classical markers according to the class (striped green = Yes, dotted red = No) for the dataset used in the reproductive health disease problem
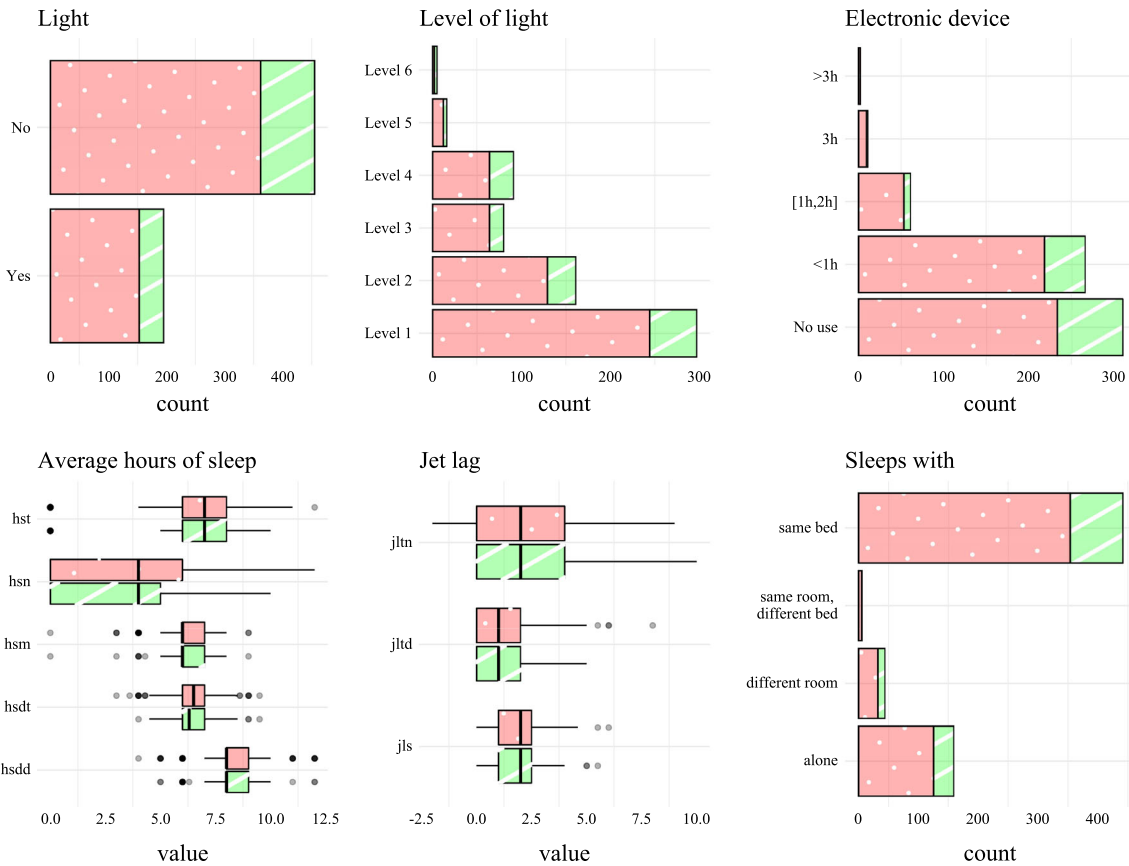

Fig. 3 Distribution of the variables related to the sleep hygiene according to the class (striped green = Yes, dotted red = No) for the dataset used in the reproductive health disease problem
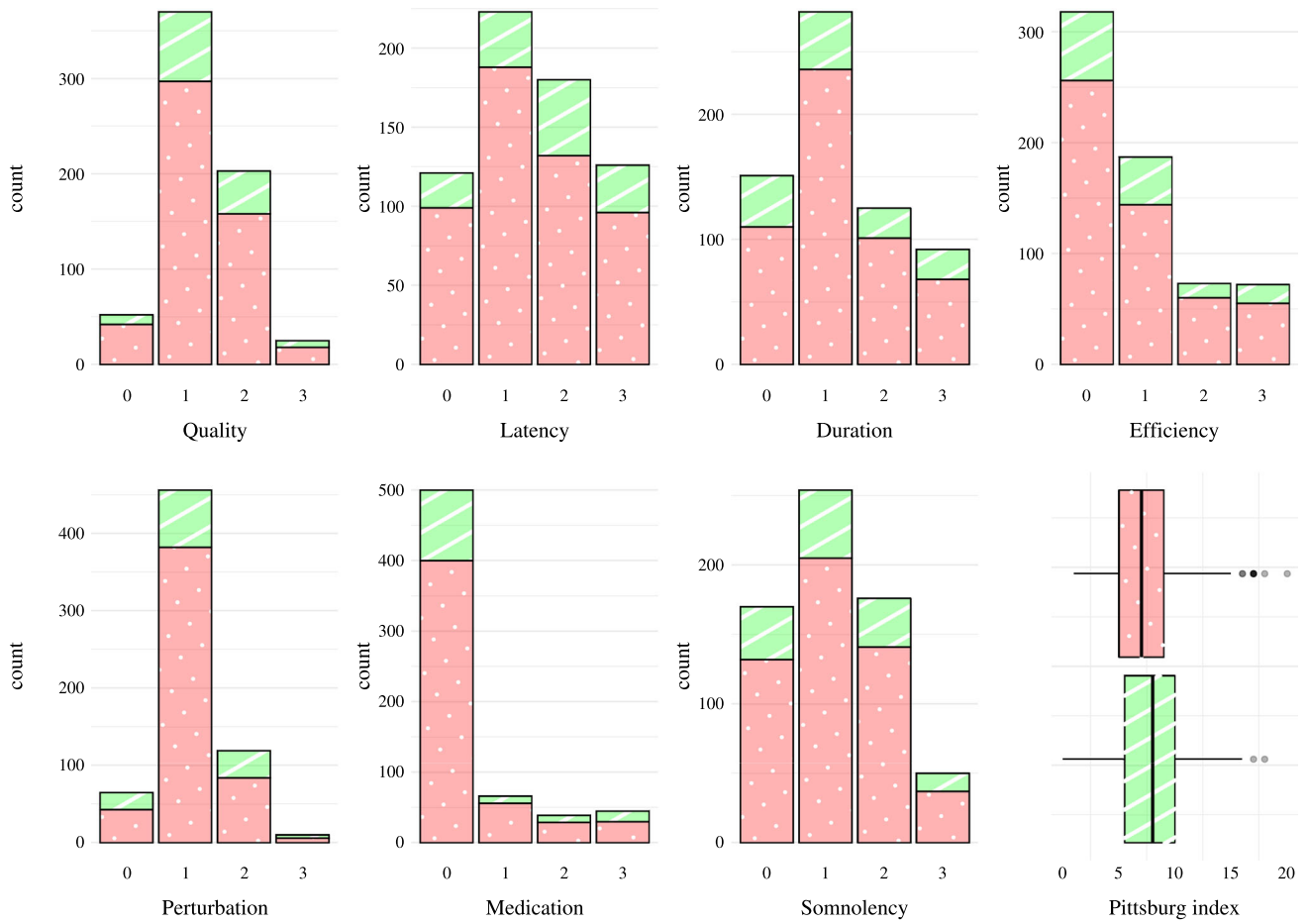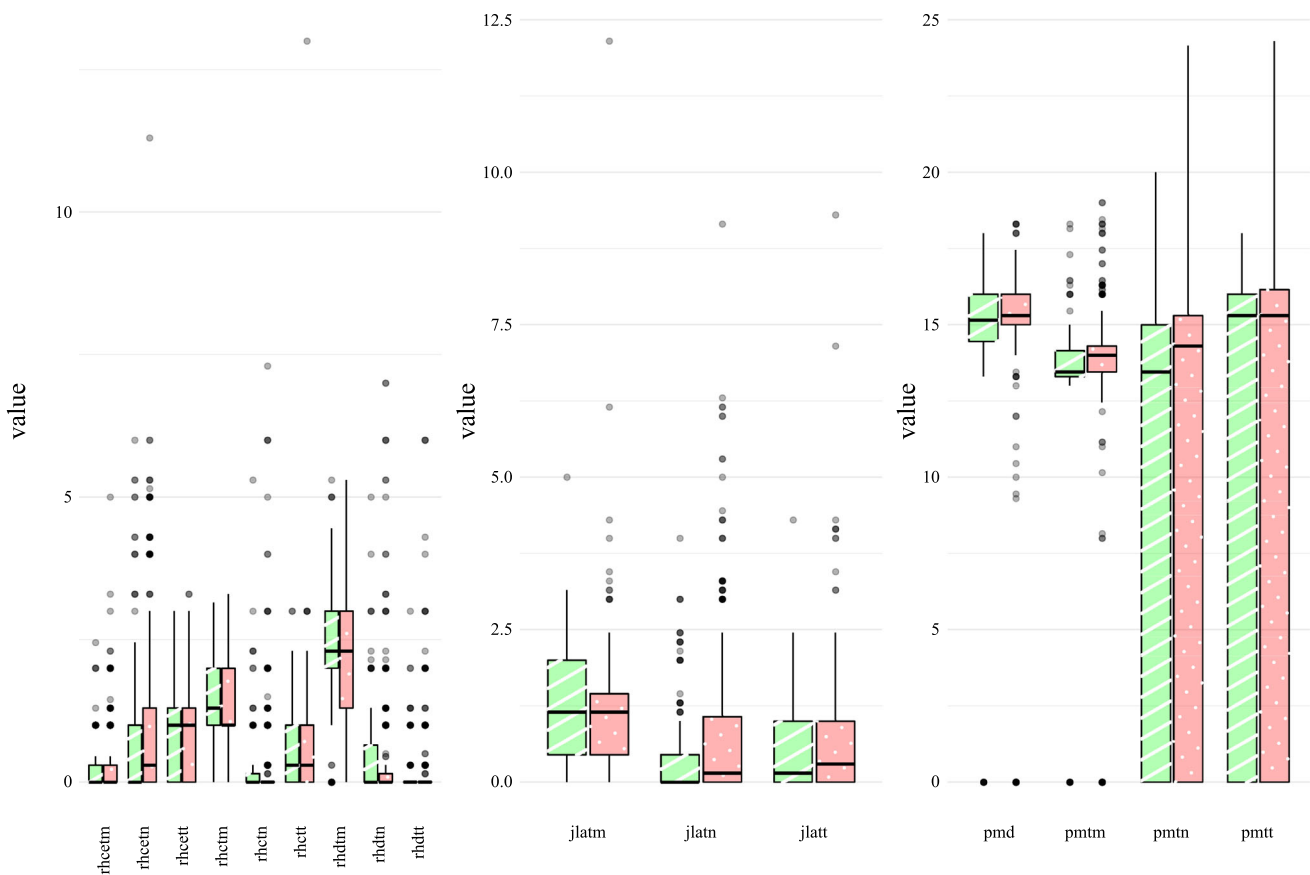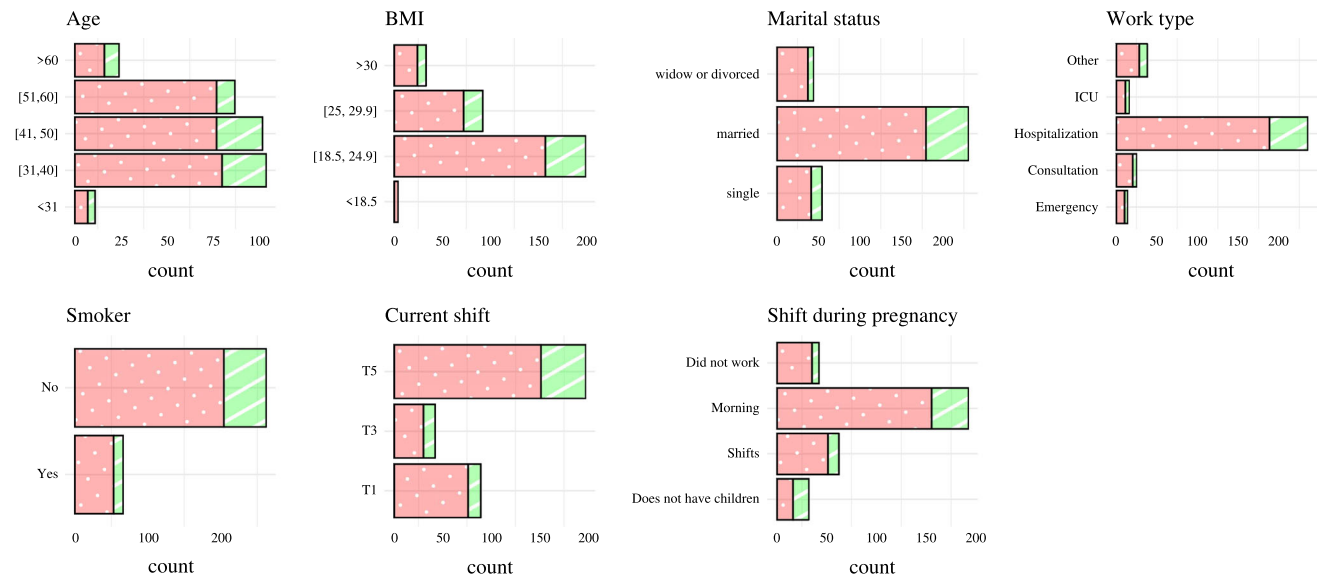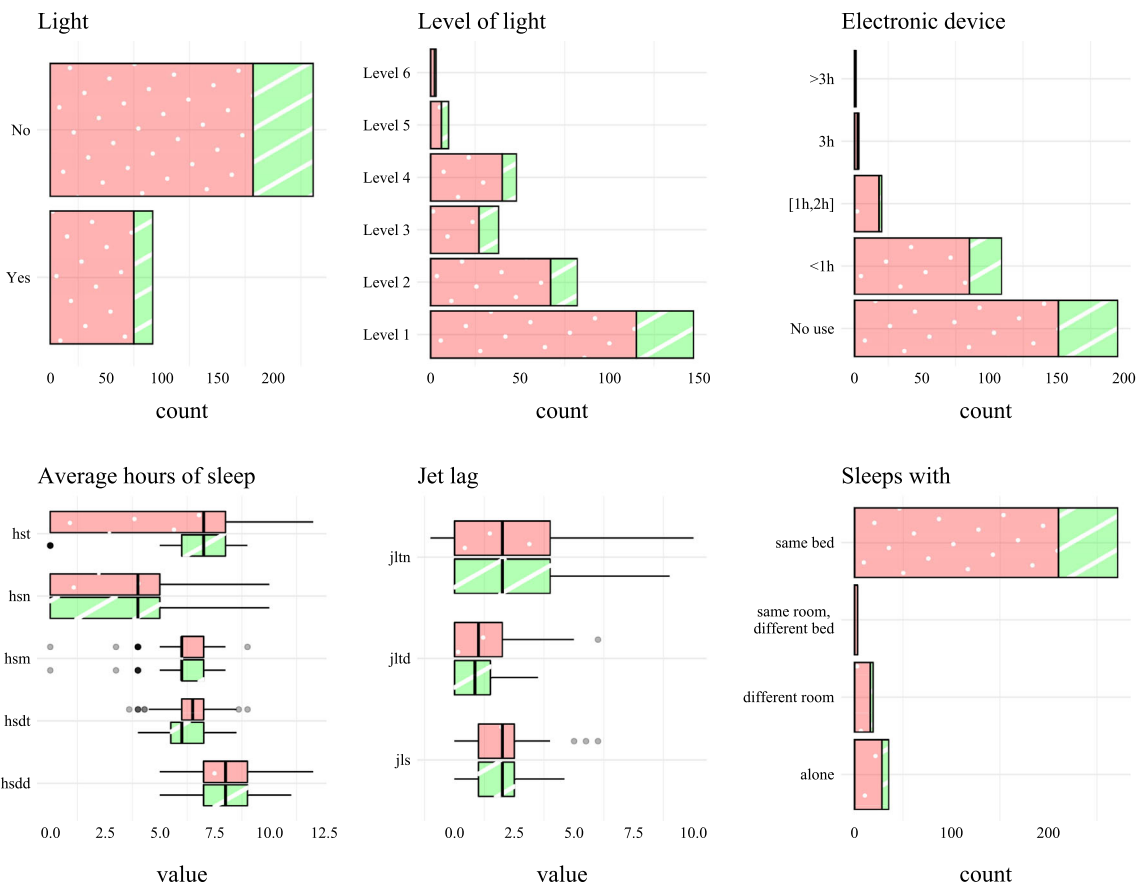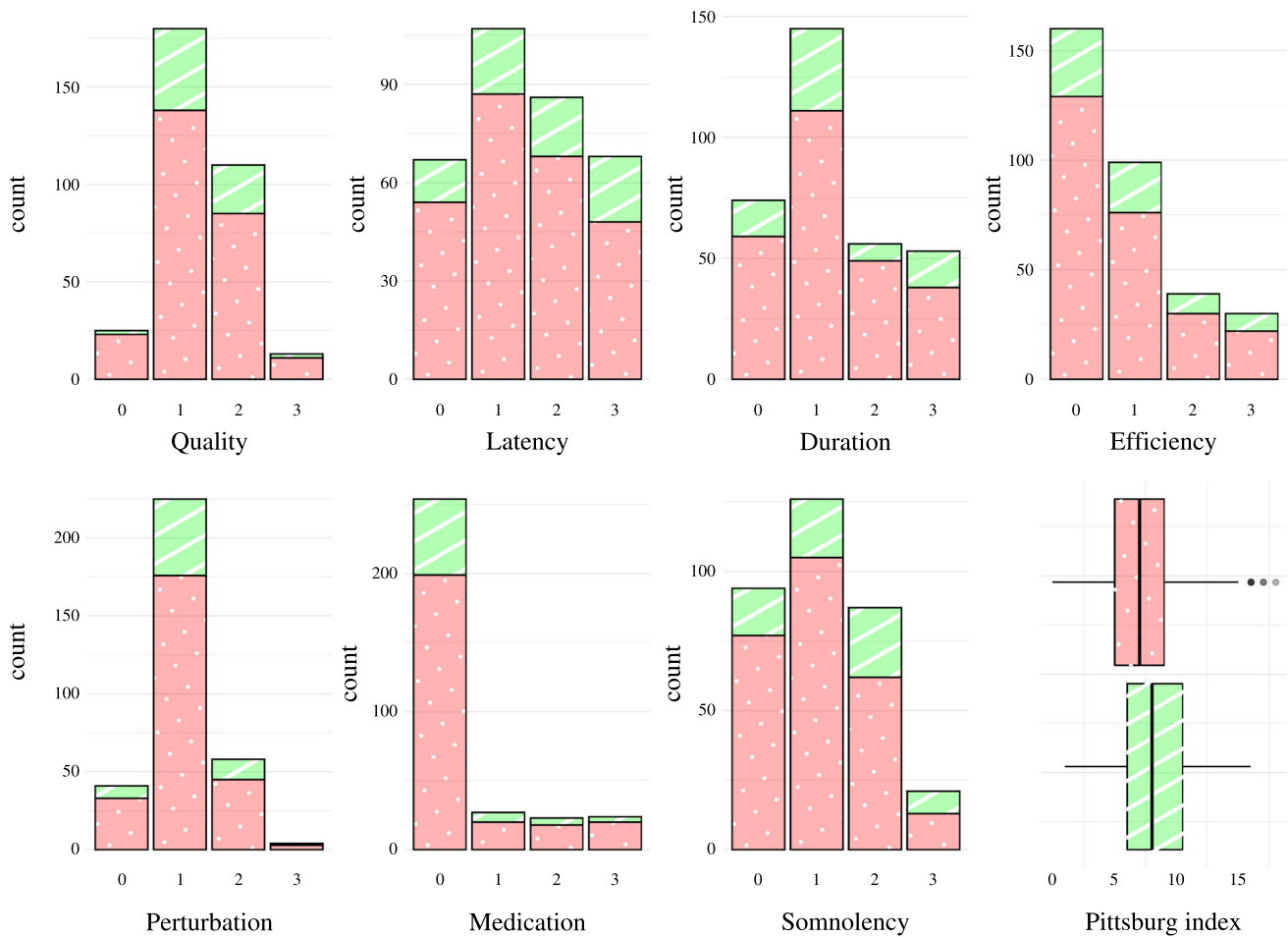
Fig. 4 Distribution of the Pittsburgh test variables according to the class (striped green = Yes, dotted red = No) of the dataset used in the reproductive health disease problem

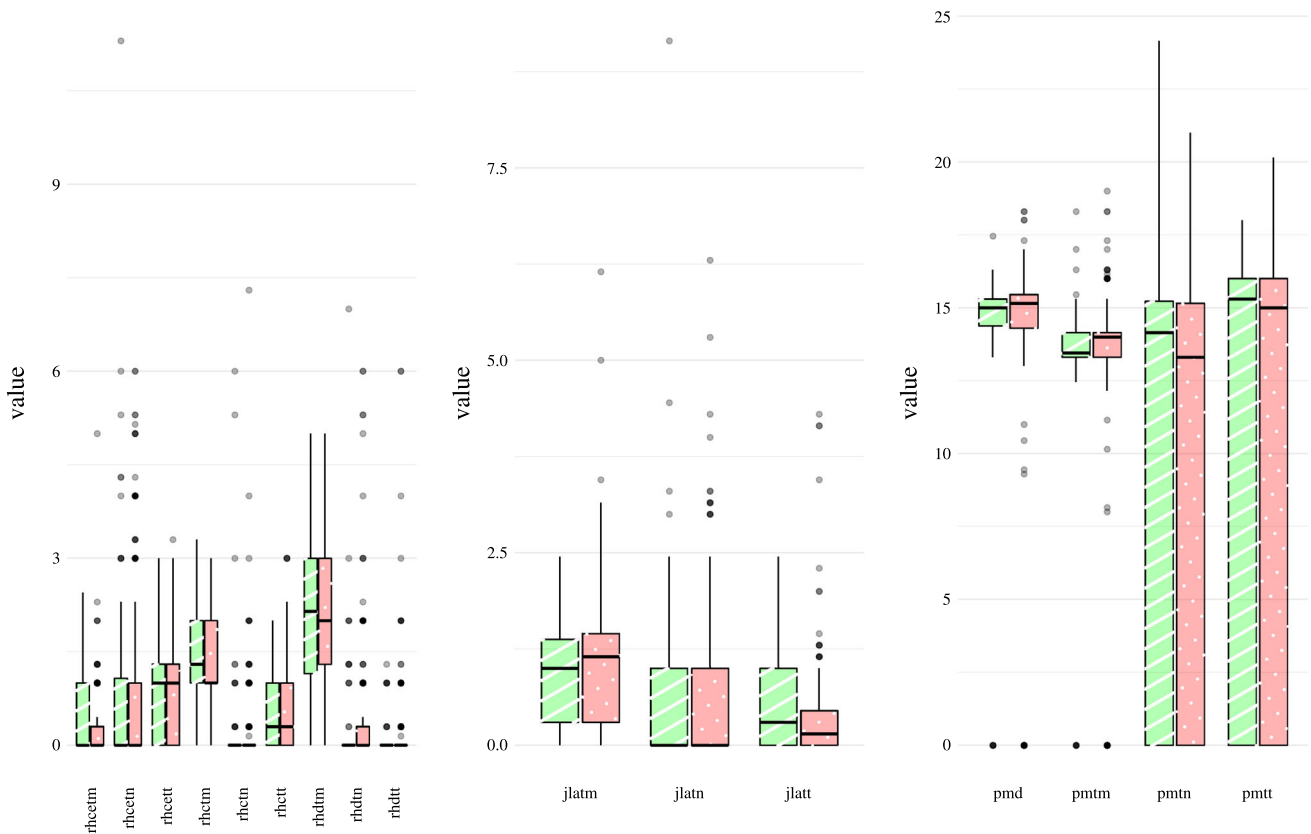**Fig. 5** Distribution of the variables related to the jet lag in the diet colored according to the class (striped green = Yes, dotted red = No) for the dataset used in the reproductive health disease problem

A.2 Attempt time

The dataset of the *first pregnancy attempt* contains the 328 samples of the original dataset that have no missing values and have a value different to *did not try* as answer to this on the original form. Among these, 71 have tried for more than one year which is the positive class and is represented by striped green in the figures. The remaining 257 take tried for one year or less (dotted red in the figures). The values in Fig. 6 with value *Does not have children* for the *shift during pregnancy* plot corresponds with not ended pregnancies and miss-carriages (Figs. 7, 8, 9).
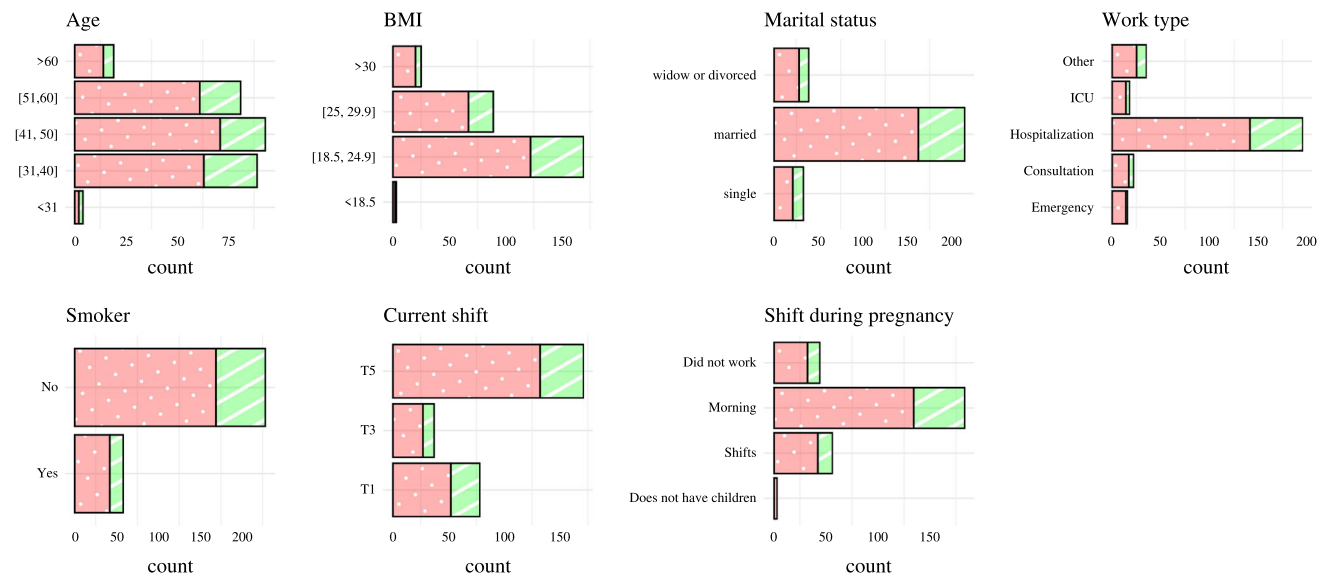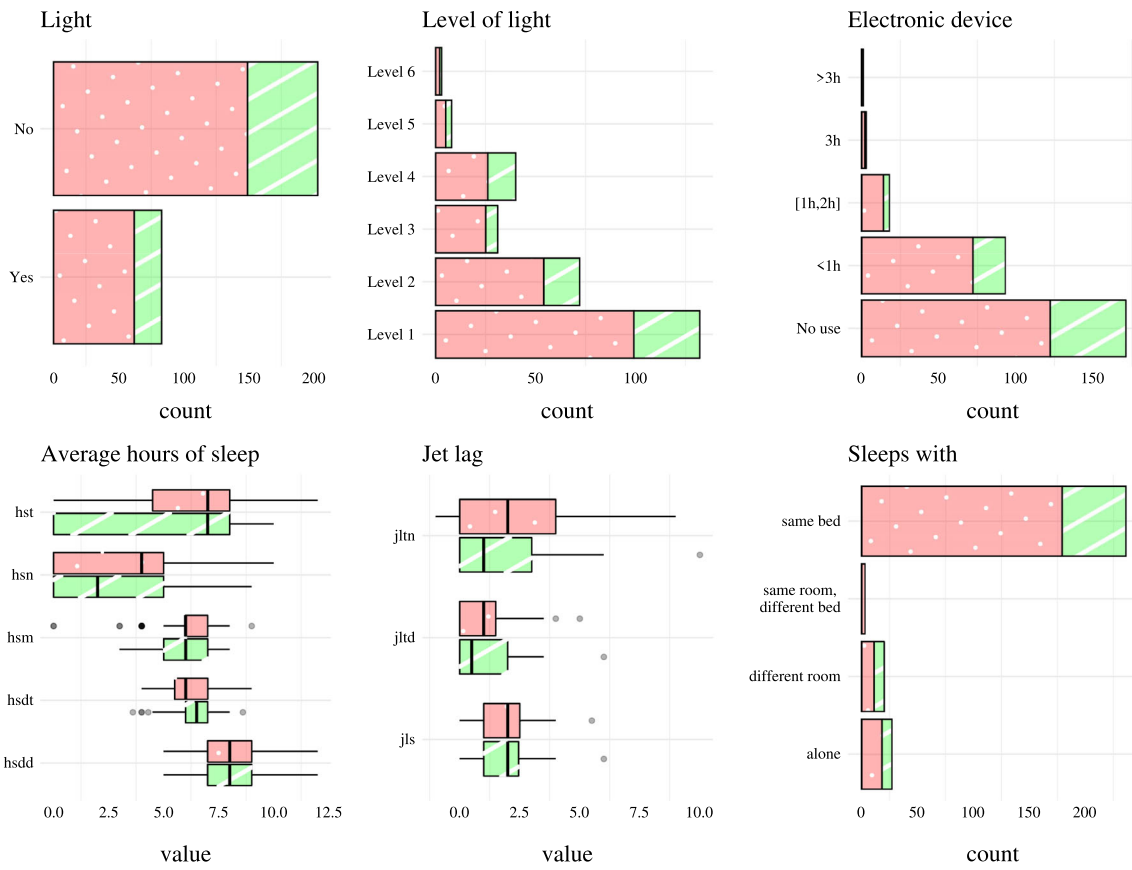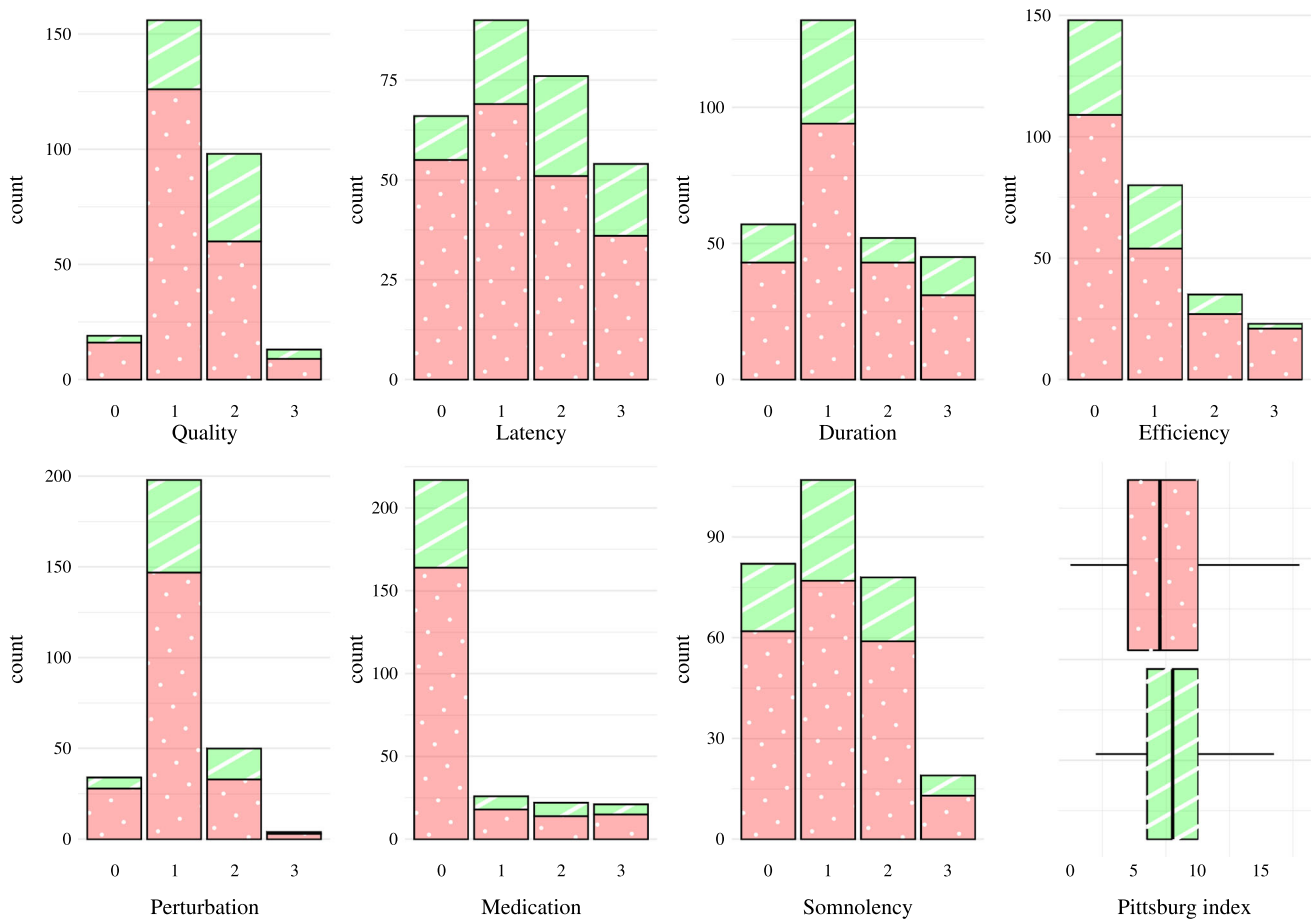
Fig. 6 Distribution of the classical markers according to the class (striped green = short, dotted red = normal) for the dataset used in the attempt time problem

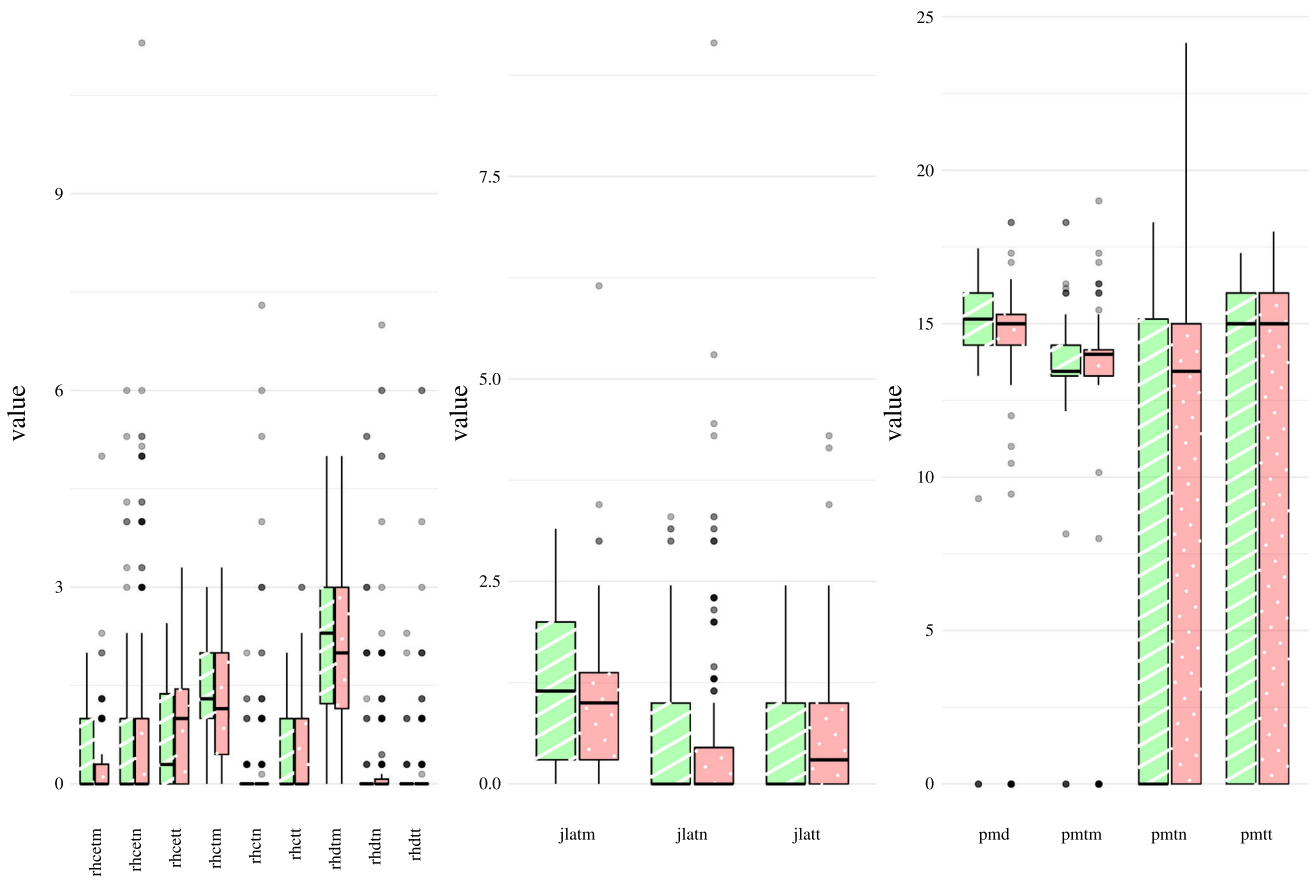**Fig. 7** Distribution of the variables related to the sleep hygiene according to the class (striped green = short, dotted red = normal) for the dataset used in the attempt time problem

Fig. 8 Distribution of the pittsburgh test variables according to the class (striped green = short, dotted red = normal) for the dataset used in the attempt time problem

A.3 Problem during pregnancy

The dataset of the *problem during pregnancy* contains the 286 samples of the original dataset that have no missing values and an answer different to *no children* in this column. Among these, 75 had a problem and therefore they take the value of the positive class *Yes* (striped green in the figures) and the remaining 211 take the value *No* (dotted red in the figures) (Figs. 10, 11, 12, 13).

A.4 Gestation period

The dataset of the *gestation period* problem contains the 305 samples of the original dataset that have no missing values and have a value different to *never been pregnant* as answer to this variable in the original form. Among these, 28 had a pregnant lower than 37 weeks, which is the positive class as it is the value of interest to study (it is represented by striped green in the figures). The remaining

Fig. 9 Distribution of the variables related to the jet lag in the diet colored according to the class (striped green = short, dotted red = normal) for the attempt time problem
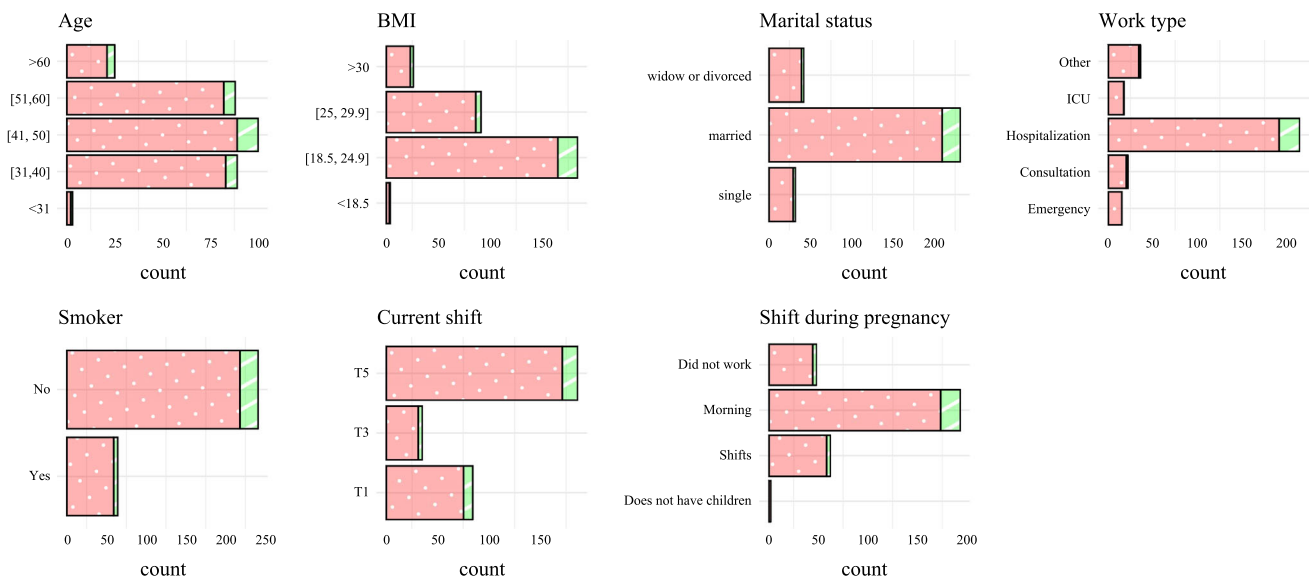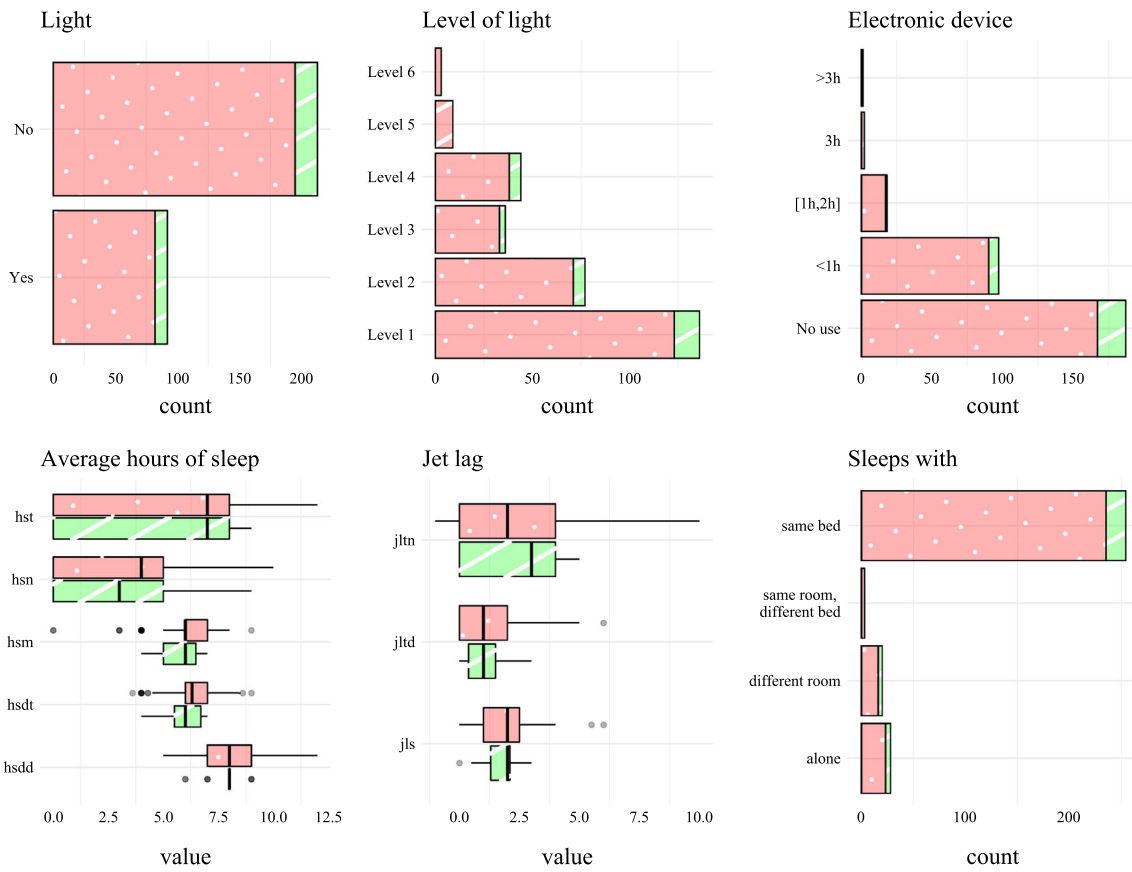

Fig. 10 Distribution of the classical markers according to the class (striped green = Yes, dotted red = No) for the dataset used in the prediction of the problem during pregnancy

Fig. 11 Distribution of the variables related to the sleep hygiene according to the class (striped green = Yes, dotted red = No) for the dataset used in the reproductive health disease problem
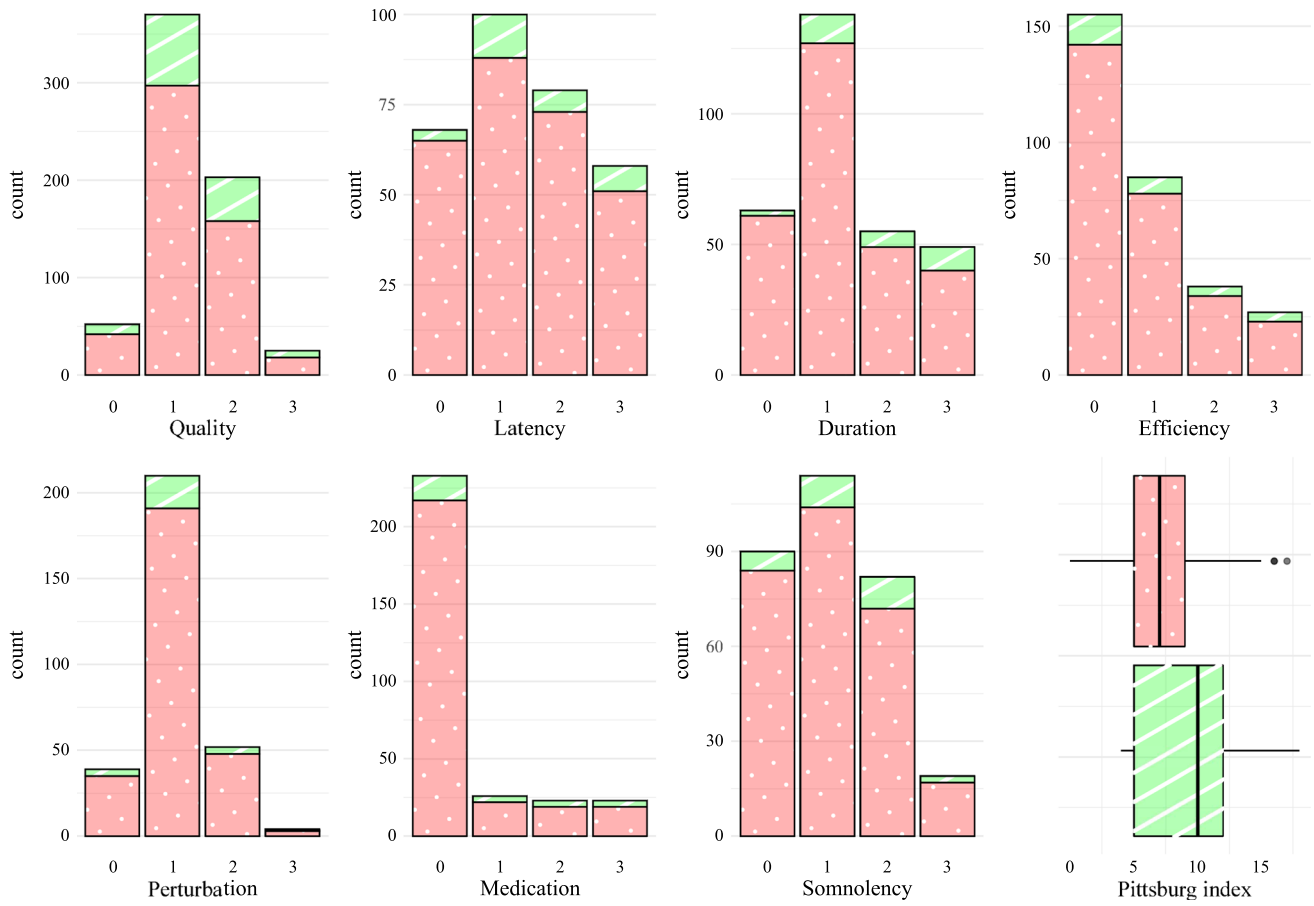
Fig. 12 Distribution of the pittsburgh test variables according to the class (striped green = short, dotted red = normal) for the dataset used in the prediction of problem during pregnancy
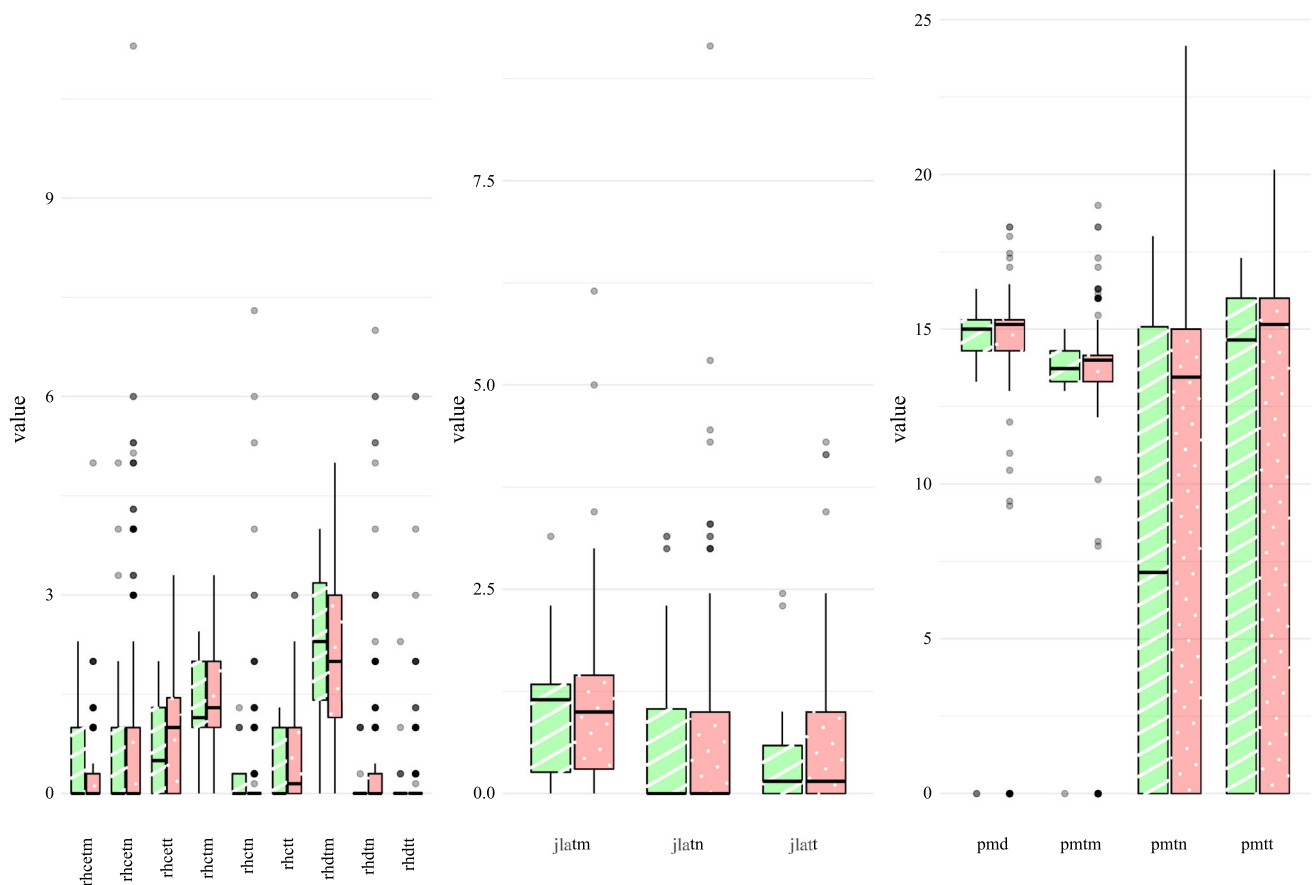
Fig. 13 Distribution of the variables related to the jet lag in the diet colored according to the class (striped green = short, dotted red = normal) for the dataset used in the prediction of problem during pregnancy

Fig. 14 Distribution of the classical markers according to the class (striped green = Yes, dotted red = No) for the dataset used in the prediction of the length of the pregnancy

Fig. 15 Distribution of the variables related to the sleep hygiene according to the class (striped green = Yes, dotted red = No) for the dataset used in the reproductive health disease problem

Fig. 16 Distribution of the pittsburgh test variables according to the class (striped green = short, dotted red = normal) for the dataset used in the prediction of the length of the pregnancy

Fig. 17 Distribution of the variables related to the jet lag in the diet colored according to the class (striped green = short, dotted red = normal) for the dataset used in the prediction of the length of the pregnancy

277 have a pregnancy of 37 weeks or longer. The values in Fig. 14 with value *Does not have children* for the *shift during pregnancy* plot corresponds with not ended pregnancies (Figs. 15, 16, 17).

Conflict of interest The authors declare no conflict of interest.

Informed consent The study was approved by the Clinical Research Ethics Committee of the Principality of Asturias. All the participants have been informed about the objective of the study and have signed an informed consent.

References

1. Touitou Y, Coste O, Dispersyn G, Pain L (2010) Disruption of the Circadian System by Environmental Factors: Effects of Hypoxia. Advanced drug delivery reviews, Magnetic Fields and General Anesthetics Agents. https://doi.org/10.1016/j.addr.2010.06.005
2. Cai C, Vandermeer B, Khurana R, Nerenberg K, Featherstone R, Sebastianski M, Davenport MH (2019) The impact of occupational shift work and working hours during pregnancy on health outcomes: a systematic review and meta-analysis. Am J Obstet Gynecol. https://doi.org/10.1016/j.ajog.2019.06.051
3. Rivera AS, Akanbi M, O'Dwyer LC, McHugh M (2020) Shift work and long work hours and their association with chronic health conditions: a systematic review of systematic reviews with meta-analyses. PLoS ONE. https://doi.org/10.1371/journal.pone.0231037
4. Wang Y, Gu F, Deng M, Guo L, Lu C, Zhou C, Chen S, Xu Y (2016) Rotating shift work and menstrual characteristics in a cohort of chinese nurses. BMC Women's Health. https://doi.org/10.1186/s12905-016-0301-y

5. Komada Y, Ikeda Y, Sato M, Kami A, Masuda C, Shibata S (2019) Social jetlag and menstrual symptoms among female university students. Chronobiol Int. https://doi.org/10.1080/07420528.2018.1533561

6. Willis SK, Hatch EE, Wesselink KJA, Rothman K, Mikkelsen EM, Wise LA (2019) Female sleep patterns, shift work, and fecundability in a north american preconception cohort study. Fertil Steril 111:1201–1210. https://doi.org/10.1016/j.fertnstert.2019.01.037

7. Massarotti C, Gentile G, Ferreccio C, Scaruffi P, Remorgida V, Anserini P (2019) Impact of infertility and infertility treatments on quality of life and levels of anxiety and depression in women undergoing in vitro fertilization. Gynecol Endocrinol. https://doi.org/10.1080/09513590.2018.1540575

8. Sutter-Dallay AL, Giaconne-Marcesche V, Glatigny-Dallay E, Verdoux H (2004) Women 1with anxiety disorders during pregnancy are at increased risk of intense postnatal depressive symptoms: a prospective survey of the matquid cohort. Eur Psychiatry. https://doi.org/10.1016/j.eurpsy.2004.09.025

9. Mancuso RA, Schetter CD, Rini CM, Roesch SC, Hobel CJ (2004) Maternal prenatal anxiety and corticotropin-releasing hormone associated with timing of delivery. Psychosom Med. https://doi.org/10.1097/01.psy.0000138284.70670.d5

10. Curchoe CL, Bormann CL (2019) Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE (2018). J Assist Reprod Genet. https://doi.org/10.1007/s10815-019-01408-x

11. Shatte ABR, Hutchinson DM, Teague SJ (2019) Machine learning in mental health: a scoping review of methods and applications. Psychol Med. https://doi.org/10.1017/S0033291719000151

12. Venkatesh KK, Strauss RA, Grotegut CA, Heine RP, Chescheir NC, Stringer JSA, Stamilio DM, Menard KM, Jelovsek JE (2020) Machine learning and statistical models to predict postpartum hemorrhage. Obstet Gynecol. https://doi.org/10.1097/AOG.0000000000003759

13. Díaz E, Fernández-Plaza C, Abad I, Alonso A, González C, Díaz I (2021) Machine learning as a tool to study the influence of chronodisruption in preterm births. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-021-02906-6

14. Montañés E, Combarro E.F, Díaz I, Ranilla J (2005) Towards automatic and optimal filtering levels for feature selection in text categorization. In: Famili, A.F., Kok, J.N., Sánchez, J.M.P., Siebes, A., Feelders, A.J. (eds.) Advances in Intelligent Data Analysis VI. In: 6th international symposium on intelligent data analysis, IDA 2005, Madrid, Spain, September 8-10, 2005, Proceedings. Lecture Notes in Computer Science, vol. 3646, pp. 239–248. Springer, Heidelberg. https://doi.org/10.1007/11552253_22

15. Pérez-Fernández R, Rademaker M, Alonso P, Díaz I, Montes S, Baets BD (2016) Representations of votes facilitating monotonicity-based ranking rules: from votrix to votex. Int J Approx Reasoning. https://doi.org/10.1016/j.ijar.2016.01.007

16. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK (2021) National Vital Statistics Reports, Vol. 70(2) Births: Final Data for 2019

17. Gdanska P, Drozdowicz-Jastrzebska E, Grzechocinska B, Radziwon-Zaleska M, Wegrzyn P, Wielgos M (2017) Anxiety and depression in women undergoing infertility treatment. Ginekol Pol. https://doi.org/10.5603/GP.a2017.0019

18. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A (2016) The pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev. https://doi.org/10.1016/j.smrv.2015.01.009

19. Wittmann M, Dinich J, Merrow M, Roenneberg T (2006) Social jetlag: misalignment of biological and social time. Chronobiol Int. https://doi.org/10.1080/07420520500545979

20. Roenneberg T, Wirz-Justice A, Merrow M (2003) Life between clocks: daily temporal patterns of human chronotypes. J Biol Rhythm. https://doi.org/10.1177/0748730402239679

21. Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York. https://doi.org/10.1007/978-1-4614-6849-3

22. Mitchell T. Machine Learning Textbook

23. Breiman L (2001) Statistical modeling: the two cultures. Stat Sci. https://doi.org/10.1214/ss/1009213726

24. Shalev-Shwartz S, Ben-David S (2013) Understanding machine learning: from theory to algorithms. Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

25. Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and regression trees. Routledge, Milton Park. https://doi.org/10.1201/9781315139470

26. Johnson RW (2001) An introduction to the bootstrap. Teaching Stat. https://doi.org/10.1111/1467-9639.00050

27. Lunardon N, Menardi G, Torelli N (2014) Rose: A package for binary imbalanced learning. R Journal https://doi.org/10.32614/rj-2014-008

28. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. J Artifi Intell Res. https://doi.org/10.1613/jair.953

29. Shapiro S, Hellman G (1993) Mathematics without numbers. JSTOR. https://doi.org/10.2307/2215793

30. De Condorcet N (2016) Essai sur l'application de l'analyse' la probabilite des decisions rendues a la pluralite des voix. Cambridge University Press, Cambridge, United Kingdom

31. Muravyov SV, Baranov PF, Emelyanova EY (2019) How to transform all multiple solutions of the kemeny ranking problem into a single solution. J Phys Conf Ser. https://doi.org/10.1088/1742-6596/1379/1/012053

32. Rico N, Vela CR, Pérez-Fernández R, Díaz I (2021) Reducing the computational time for the kemeny method by exploiting condorcet properties. Mathematics. https://doi.org/10.3390/math9121380

33. de Borda J.-C (1781) Mémoire sur les élections au scrutin. Histoire de l'Académie Royale des Sciences

34. Ali A, Meilǎ M (2012) Experiments with kemeny ranking: what works when? Math Soc Sci. https://doi.org/10.1016/j.mathsocsci.2011.08.008

35. Cho YM, Ryu SH, Lee BR, Kim KH, Lee E, Choi J (2015) Effects of artificial light at night on human health: a literature review of observational and experimental studies applied to exposure assessment. Chronobiol Int. https://doi.org/10.3109/07420528.2015.1073158

36. Teoh AN, Kaur S, Shukri NHM, Shafie SR, Bustami NA, Takahashi M, Lim PJ, Shibata S (2021) Psychological state during pregnancy is associated with sleep quality: preliminary findings from my-care cohort study. Chronobiol Int. https://doi.org/10.1080/07420528.2021.1902338

37. Alder J, Fink N, Bitzer J, Hosli I, Holzgreve W (2007) Depression and anxiety during pregnancy: a risk factor for obstetric, fetal and neonatal outcome? a critical review of the literature. J Maternal-Fetal Neonatal Med 20:189–209

38. Marino JL, Holt VL, Chen C, Davis S (2008) Shift work, hclock t3111c polymorphism, and endometriosis risk. Epidemiology. https://doi.org/10.1097/EDE.0b013e31816b7378

39. Stock D, Knight JA, Raboud J, Cotterchio M, Strohmaier S, Willett W, Eliassen AH, Rosner B, Hankinson SE,

Schernhammer E (2019) Rotating night shift work and meno-pausal age. Human Reproduct. https://doi.org/10.1093/humrep/dey390

40. Haus EL, Smolensky MH (2013) Shift work and cancer risk: potential mechanistic roles of circadian disruption, light at night, and sleep deprivation. Sleep Med Rev. https://doi.org/10.1016/j.smrv.2012.08.003

41. McKinlay SM, Brambilla DJ, Posner JG (1992) The normal menopause transition. Am J Human Biol. https://doi.org/10.1002/ajhb.1310040107